

---

# WISTA

## Wirtschaft und Statistik

---

Florian Dumpert

**Maschinelles Lernen im Statistischen Bundesamt**

Oliver Reitz

**Erfassung von Ernteerträgen mit Satellitenbildern und Maschinellern Lernen – das Projekt FernEE 2.0**

Matthias Keller | Thomas Körner

**Haushalte, Familien und Lebensformen mit Einwanderungsgeschichte – Auswirkungen auf die Erwerbsbeteiligung von Eltern**

Alejandro Donado | Benedikt  
Kuckelkorn | Yannik Weigelt |  
Daniel Seeger

**Daten als eigenständiges Vermögensgut in den Volkswirtschaftlichen Gesamtrechnungen – ein Werkstattbericht**

Marlon Suárez | Lieselotte Hofmann

**Die Nutzendenumfrage zur Kulturstatistik – Konzept und Ergebnisse**

---

# 4 | 2024

#### ABKÜRZUNGEN

D	Durchschnitt (bei nicht addierfähigen Größen)
Vj	Vierteljahr
Hj	Halbjahr
a. n. g.	anderweitig nicht genannt
o. a. S.	ohne ausgeprägten Schwerpunkt
Mill.	Million
Mrd.	Milliarde

#### ZEICHENERKLÄRUNG

–	nichts vorhanden
0	weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts
.	Zahlenwert unbekannt oder geheim zu halten
. . .	Angabe fällt später an
X	Tabellenfach gesperrt, weil Aussage nicht sinnvoll
I oder —	grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt
/	keine Angaben, da Zahlenwert nicht sicher genug
( )	Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist
	Abweichungen in den Summen ergeben sich durch Runden der Zahlen.
	Tiefer gehende Internet-Verlinkungen sind hinterlegt.

# INHALT

3	Editorial
4	Kennzahlen
8	Aktuelle Informationsangebote
10	Kurznachrichten
17	Florian Dumpert <b>Maschinelles Lernen im Statistischen Bundesamt</b> <i>The use of machine learning at the Federal Statistical Office</i>
29	Oliver Reitz <b>Erfassung von Ernteerträgen mit Satellitenbildern und Maschinellern Lernen – das Projekt FernEE 2.0</b> <i>The FernEE 2.0 project – Using satellite imagery and machine learning to capture crop yields</i>
39	Matthias Keller, Thomas Körner <b>Haushalte, Familien und Lebensformen mit Einwanderungsgeschichte – Auswirkungen auf die Erwerbsbeteiligung von Eltern</b> <i>Immigration history in the context of households, families and living arrangements – Implications for the labour force participation of parents</i>

## INHALT

- 58 Alejandro Donado, Benedikt Kuckelkorn, Yannik Weigelt, Daniel Seeger  
**Daten als eigenständiges Vermögensgut in den Volkswirtschaftlichen Gesamtrechnungen – ein Werkstattbericht**  
*Recording data as a separate asset in national accounts – A workshop report*

- 71 Marlon Suárez, Lieselotte Hofmann  
**Die Nutzendenumfrage zur Kulturstatistik – Konzept und Ergebnisse**  
*Survey of users of cultural statistics – Approach taken and survey results*

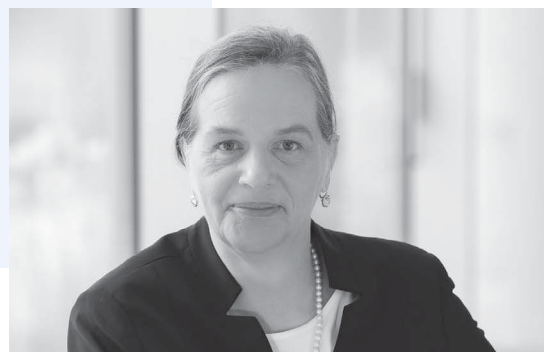
---

# EDITORIAL

---

Dr. Ruth Brand

---



## LIEBE LESERIN, LIEBER LESER,

mit den Ergebnissen des jährlich repräsentativ bei rund 1 % der Bevölkerung durchgeführten Mikrozensus stehen aussagekräftige statistische Daten zur Struktur sowie zur wirtschaftlichen und sozialen Lage der Bevölkerung zur Verfügung, mit Informationen zu Familie und Lebenspartnerschaft, Arbeitsmarkt und Erwerbstätigkeit, Beruf und Ausbildung. Das Statistische Bundesamt hat mit dem Berichtsjahr 2023 im Mikrozensus das [Konzept der Einwanderungsgeschichte](#) eingeführt und setzt es nun auch auf der Ebene von Haushalten und Lebensformen um. Lesen Sie in einem Beitrag in dieser WISTA-Ausgabe, wie sich Haushalte mit Einwanderungsgeschichte von denen ohne Einwanderungsgeschichte unterscheiden und welche Auswirkungen dies auf die Erwerbsbeteiligung von Eltern hat.

Verfahren Maschinellen Lernens und Künstlicher Intelligenz halten Einzug in alle Lebensbereiche und auch in die amtliche Statistik, wie ein Aufsatz über die Historie und aktuelle Entwicklungen zum Einsatz Maschinellen Lernens im Statistischen Bundesamt zeigt. In einer zunehmend digitalisierten Welt nimmt die Bedeutung von Daten als eigenständiges Vermögensgut erheblich zu. Wie aber lässt sich ihr Wert bemessen und in das Bruttoinlandsprodukt integrieren? Das Statistische Bundesamt untersucht mit einem von den Vereinten Nationen empfohlenen und auf dem Kostensummenansatz basierenden Berechnungsmodell, wie Investitionen in Daten und darauf anfallende Abschreibungen erfasst werden können.

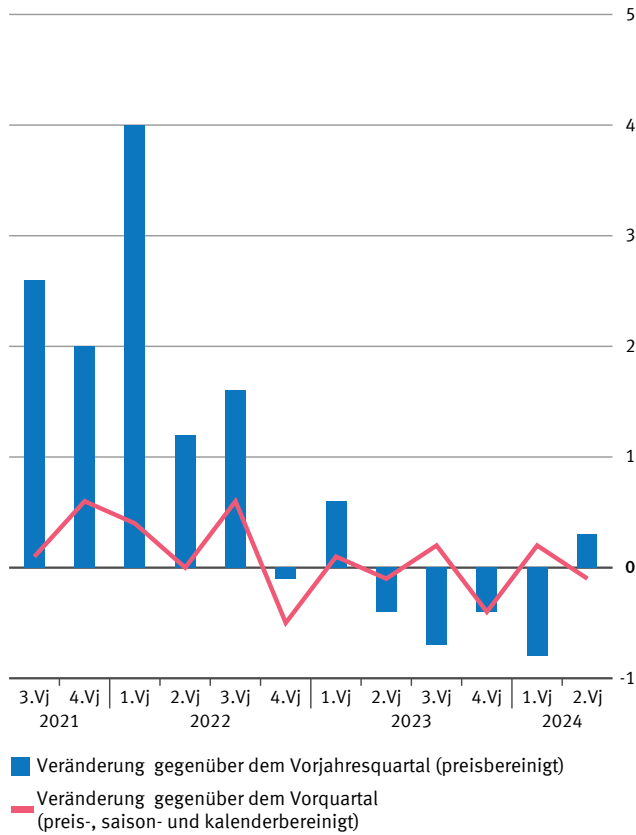
In weiteren Beiträgen informieren wir Sie über ein Projekt zur Erfassung von Ernteerträgen mit Satellitenbildern und Maschinellern Lernen sowie zur aktuellen dritten Projektphase zur Etablierung einer fundierten, bundesweiten Kulturstatistik.

Ich wünsche Ihnen eine aufschlussreiche Lektüre dieser WISTA-Ausgabe.

*Ruth Brand*

Präsidentin des Statistischen Bundesamtes

**Bruttoinlandsprodukt**  
in %

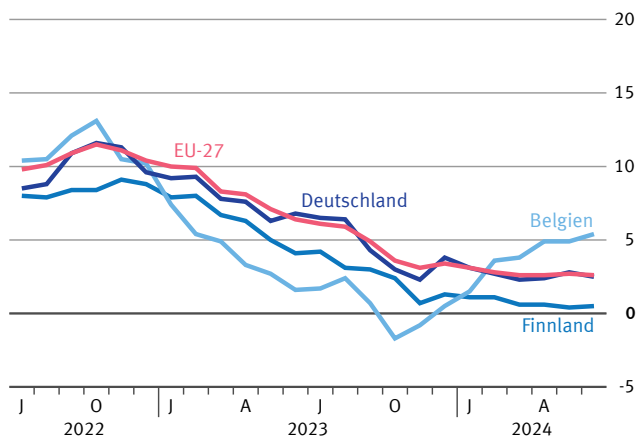


**Verbraucherpreisindex**  
2020 = 100

2023	2024
Januar 114,3	Januar 117,6
Februar 115,2	Februar 118,1
März 116,1	März 118,6
April 116,6	April 119,2
Mai 116,5	Mai 119,3
Juni 116,8	Juni 119,4
Juli 117,1	Juli 119,8
August 117,5	
September 117,8	
Oktober 117,8	
November 117,3	
Dezember 117,4	

**119,8** ↑ **2,3 %**  
Veränderung zum Vorjahresmonat

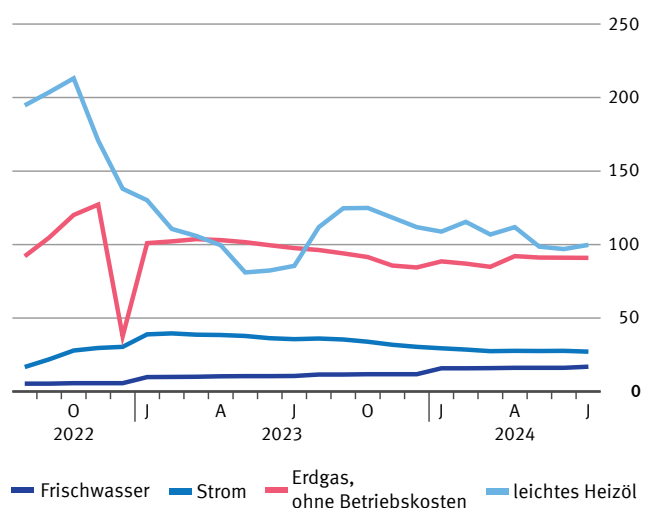
**Harmonisierter Verbraucherpreisindex insgesamt**  
Veränderung gegenüber dem Vorjahresmonat in %



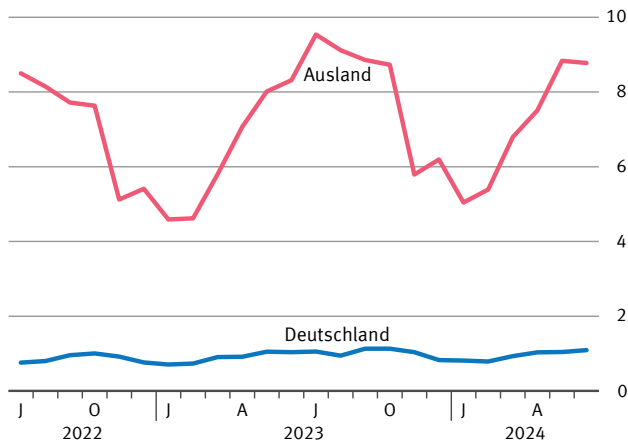
Dargestellt sind neben Deutschland und der Europäischen Union insgesamt (EU-27) die Länder mit der höchsten und der niedrigsten Veränderungsrate innerhalb der EU.

Stand: 09.08.2024

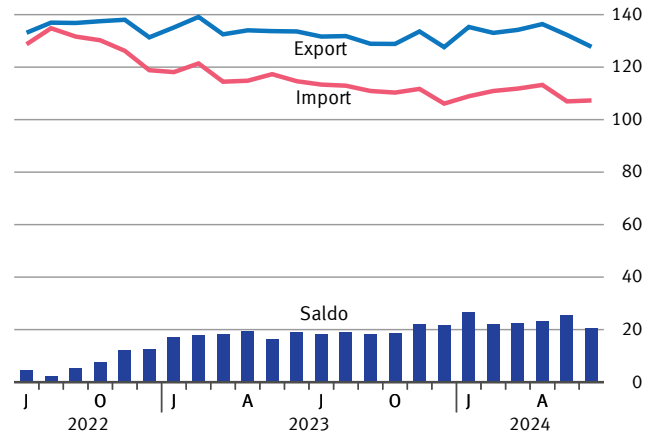
**Entwicklung der Verbraucherpreise für Energie und Wasser**  
Preisabstand in % gegenüber dem Jahr 2020



**Einsteigende Flugpassagiere auf deutschen Flughäfen nach dem Streckenziel in Mill.**

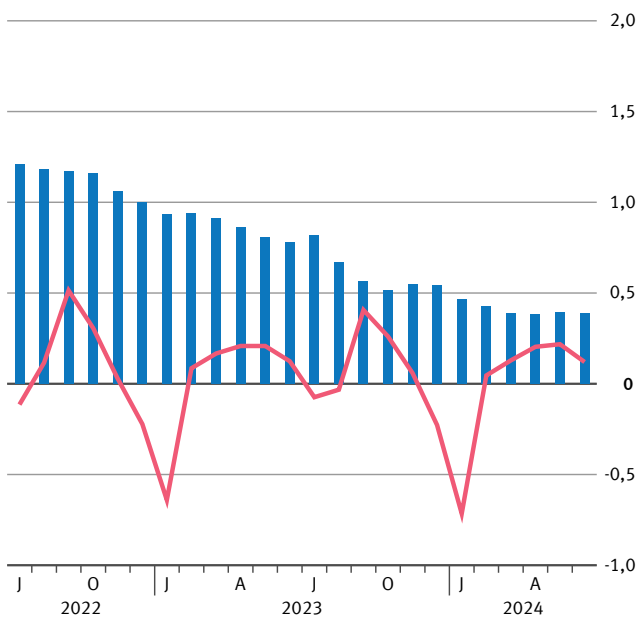


**Außenhandel in Mrd. EUR**



Kalender- und saisonbereinigte Werte nach dem Verfahren X13 JDemetra+. – Vorläufiges Ergebnis.

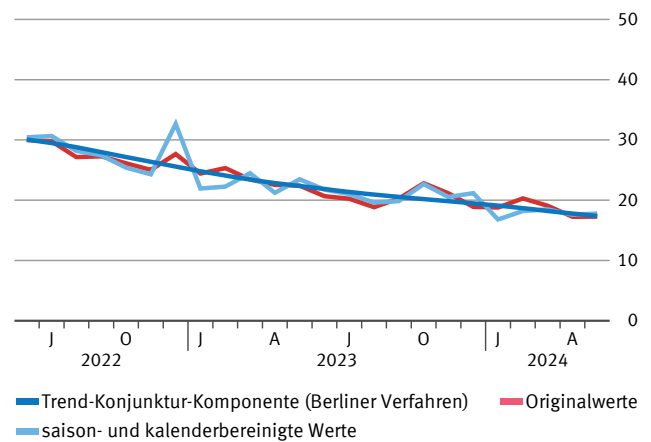
**Erwerbstätige mit Wohnort in Deutschland Veränderungen in %**



■ Veränderung gegenüber dem Vorjahresmonat  
 ■ Veränderung gegenüber dem Vormonat

Stand: 09.08.2024

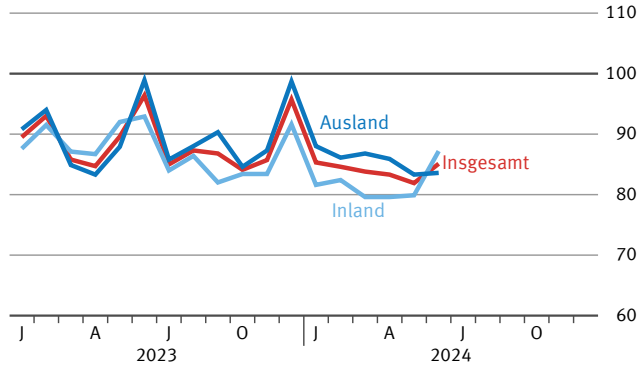
**Genehmigte Wohnungen Wohn- und Nichtwohngebäude, in 1 000**



■ Trend-Konjunktur-Komponente (Berliner Verfahren) ■ Originalwerte  
 ■ saison- und kalenderbereinigte Werte

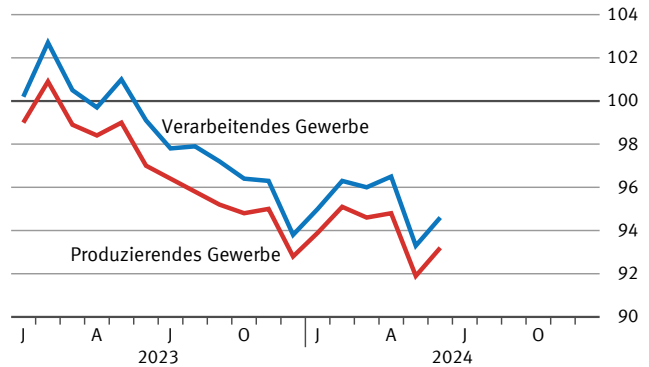
## Kennzahlen und Indikatoren

**Auftragseingang im Verarbeitenden Gewerbe**  
Volumenindex 2021 = 100



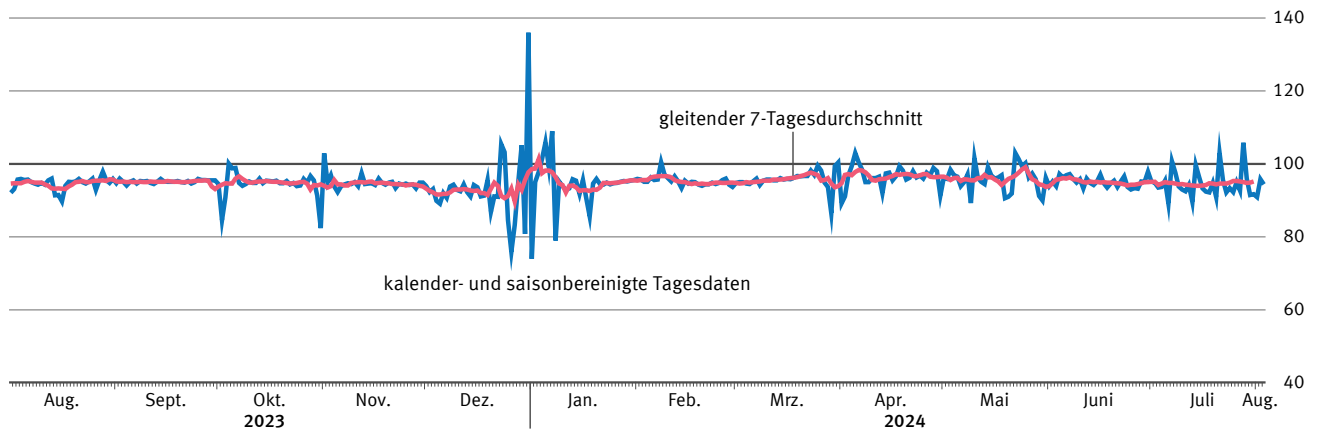
Kalender- und saisonbereinigter Wert nach dem Verfahren X13 JDemetra+. – Vorläufiges Ergebnis.

**Produktion im Produzierenden und Verarbeitenden Gewerbe**  
Index 2021 = 100



Kalender- und saisonbereinigte Werte nach dem Verfahren X13 JDemetra+. – Vorläufiges Ergebnis.

**Lkw-Maut-Fahrleistungsindex**  
2021 = 100



Quellen: Bundesamt für Logistik und Mobilität, Deutsche Bundesbank, Statistisches Bundesamt

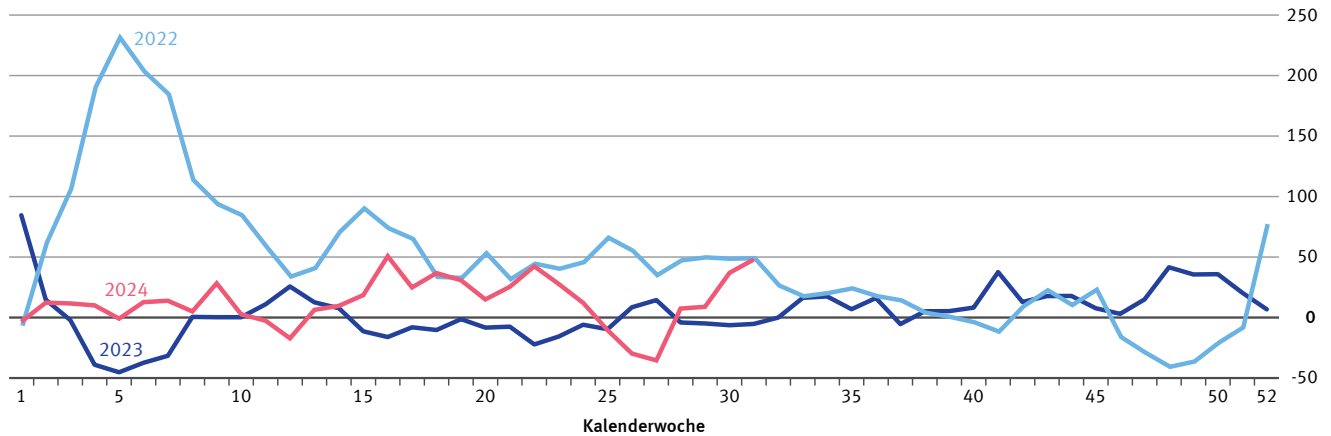
Stand: 09.08.2024



## Kennzahlen und Indikatoren

### Neue Kreditverträge nach Kalenderwochen

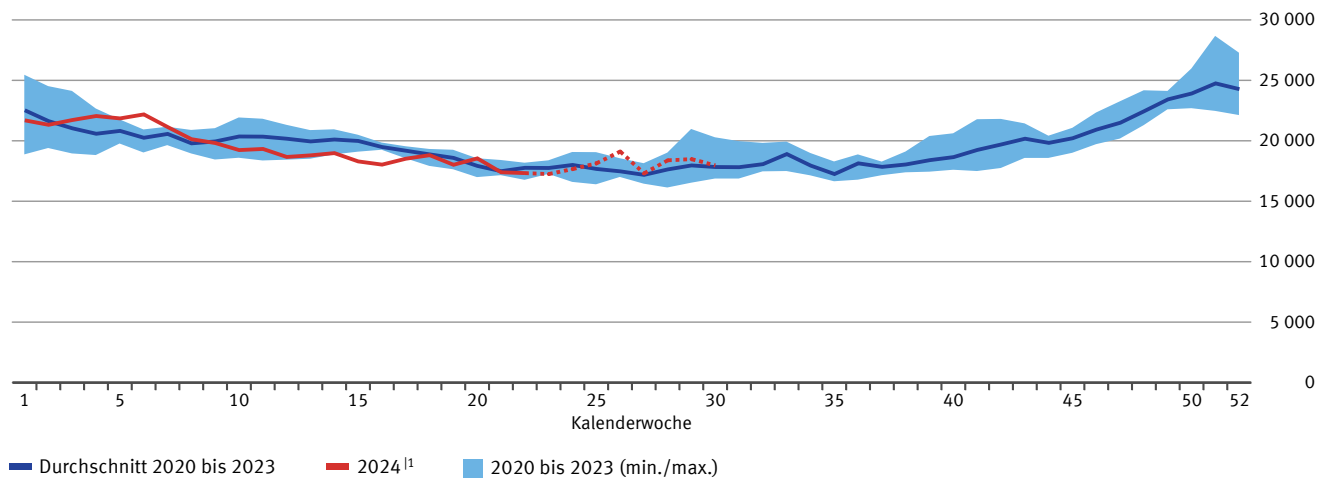
Veränderung gegenüber der entsprechenden Vorjahreswoche in %



Anfang 2022 zeigte sich mit Veränderungsraten von 100 % und mehr ein starker Anstieg im Vergleich zum Vorjahr; dabei handelt es sich um Sondereffekte, die seitens des Datenlieferanten nicht bereinigt werden konnten.

Quelle: SCHUFA Holding AG; Berechnung: Statistisches Bundesamt

### Wöchentliche Sterbefallzahlen in Deutschland



Gestrichelte Werte enthalten Schätzanteil.

1 Sonderauswertung der vorläufigen Sterbefallzahlen.

Stand: 09.08.2024



### Ukraine

Der Angriff Russlands auf die Ukraine und die damit verbundenen Sanktionen haben starke Auswirkungen auf Wirtschaft und Bevölkerung sowie den Energie-sektor. Auf einer Sonderseite zum Thema stellt das Statistische Bundesamt relevante Daten zur Verfügung. Über die Seite gelangt man auch zum [zentralen Hilfs-portal](#) der Bundesregierung für Geflüchtete aus der Ukraine.

➤ [www.destatis.de/Im-Fokus/Ukraine](http://www.destatis.de/Im-Fokus/Ukraine)



### Dashboard Deutschland

Das vom Statistischen Bundesamt entwickelte Datenportal bietet hochaktuelle und hochfrequente Zahlen, Daten und Fakten zu den Themen Arbeitsmarkt, Bauen und Wohnen, Energie, Finanzen, Konjunktur und Wirtschaft sowie Ukraine. Es trägt damit zu einem faktenbasierten demokratischen Diskurs der Öffentlichkeit und zur evidenzbasierten Entscheidungsfindung durch Politik und Verwaltung bei. Der integrierte Pulsmesser Wirtschaft bietet Einblicke in das aktuelle wirtschaftliche Geschehen, intuitives und einfaches Vergleichen von Daten sowie das Erkennen von konjunkturellen Entwicklungen und Zusammenhängen mithilfe täglicher, wöchentlicher, monatlicher und vierteljährlicher Indikatoren.

➤ [www.dashboard-deutschland.de](http://www.dashboard-deutschland.de)



### EXSTAT – Experimentelle Statistiken

In der Rubrik „EXSTAT – Experimentelle Statistiken“ veröffentlicht das Statistische Bundesamt regelmäßig neue, innovative Projektergebnisse. Sie entstehen auf der Grundlage neuer Datenquellen und Methoden. Im Reifegrad und in der Qualität unterscheiden sie sich von amtlichen Statistiken, insbesondere in Bezug auf Harmonisierung, Erfassungsbereich und Methodik. Dennoch sind es Ergebnisse der Statistischen Ämter des Bundes und der Länder, die interessante, neue Perspektiven auf verschiedene Themenfelder der Statistik bieten.

➤ [www.destatis.de/exstat](http://www.destatis.de/exstat)



### Ergebnisse des Zensus 2022

Wie viele Menschen leben in Deutschland? Wie wohnen und arbeiten sie? Diese Fragen beantworten die Ergebnisse des Zensus 2022. Im Wesentlichen geht es dabei um zwei Ziele: die Ermittlung aktueller Bevölkerungszahlen für Deutschland sowie Informationen zum Wohnraum mit der Gebäude- und Wohnungszählung.

➤ [www.zensus2022.de](http://www.zensus2022.de)



### Klima

Der Klimawandel ist eine der größten Herausforderungen der heutigen Zeit, alle Bereiche der Gesellschaft sind betroffen. Wie beeinflusst unsere Lebens- und Wirtschaftsweise das Klima? Wie wirkt sich die Umstellung hin zu mehr Klimaschutz gesamtgesellschaftlich aus? Was bedeutet sie für unseren Alltag – vom Weg zur Arbeit bis zum aktuellen Strompreis? Wo zeigen sich die Folgen des Klimawandels? Daten und Fakten zum Thema Klima, Klimawandel und Klimaschutz sind gebündelt unter

➤ [www.destatis.de/klima](http://www.destatis.de/klima)



### Fachkräfte

Fachkräftemangel und Arbeitskräftebedarf sind zunehmend wichtige Faktoren für die wirtschaftliche Entwicklung in Deutschland. Daten und Fakten dazu bündelt das Statistische Bundesamt auf einer eigenen Sonderseite. Das Angebot umfasst die Bereiche Demografie, Erwerbstätigkeit, Bildung und Zuwanderung – und wird sukzessive erweitert.

➤ [www.destatis.de/fachkraefte](http://www.destatis.de/fachkraefte)

# KURZNACHRICHTEN

## IN EIGENER SACHE

### IAOS Young Statisticians Prize 2024



Beim Young Statisticians Prize 2024 der International Association for Official Statistics (IAOS) haben Adrian Urban und Simon Rommelspacher aus dem Statistischen Bundesamt mit ihrem Paper „RUMS – how to compare structures of enterprise groups?“ den dritten Platz erreicht. Ausgezeichnet wurden fachliche Beiträge, die sich mit aktuellen methodischen und strategischen Fragen im Bereich der amtlichen Statistik auf regionaler, nationaler oder internationaler Ebene beschäftigen und mögliche Lösungswege aufzeigen.

Das Paper beschreibt ein Ähnlichkeitsmaß, das die beiden Autoren für den Vergleich von Daten zu Unternehmensgruppen entwickelt haben. RUMS ermöglicht den Vergleich der komplexen Strukturen von Unternehmensgruppen zu zwei verschiedenen Zeitpunkten oder aus zwei unterschiedlichen Registern und bewertet deren Ähnlichkeit. Es wird sowohl für die nationale Pflege der Unternehmensgruppensdaten als auch zum Beliefern und Nutzen des EuroGroups-Registers eingesetzt.

Das Paper beschreibt ein Ähnlichkeitsmaß, das die beiden Autoren für den Vergleich von Daten zu Unternehmensgruppen entwickelt haben. RUMS ermöglicht den Vergleich der komplexen Strukturen von Unternehmensgruppen zu zwei verschiedenen Zeitpunkten oder aus zwei unterschiedlichen Registern und bewertet deren Ähnlichkeit. Es wird sowohl für die nationale Pflege der Unternehmensgruppensdaten als auch zum Beliefern und Nutzen des EuroGroups-Registers eingesetzt.

### Verleihung des Innovationspreises 2024

Der Statistische Beirat berät das Statistische Bundesamt in Grundsatzfragen und vertritt dabei die Belange der Nutzenden, Befragten und Produzierenden der Bundesstatistik. Im Rahmen seiner 71. Jahrestagung am 27. Juni 2024 hat der Statistische Beirat Beschäftigte des Statistischen Bundesamtes für innovative Arbeiten ausgezeichnet, die zur Qualitätssteigerung in der Statistikproduktion führen, innovative fachliche Lösungen bereitstellen oder die nutzerorientierten Serviceleistungen verbessern.

Der erste Platz ging an Miriam Engel, Alexander Irmer, Nelli Krüger, Elke Nagel, Max Julius Schöne, Christiane Theisen und Joël Van Hoorde für die Entwicklung einer mobilen App als neues Erhebungsinstrument in den freiwilligen Haushaltserhebungen. Das Projekt überzeugte die Jury durch seine große Relevanz – durch den Einsatz der App reduzieren sich die Aufwände für die Befragten und für das Statistische Bundesamt substanziell. Aktuell wird die App bei den [Laufenden Wirtschaftsrechnungen \(LWR\)](#) eingesetzt.

Für seine Arbeit am „Tool zur Digitalisierung und computergestützten Auswertung kameraler Haushaltspläne“ in den Umweltökonomischen Gesamtrechnungen (UGR) konnte Johannes Knecht den zweiten Preis entgegennehmen. Das Tool schließt bestehende Datenlücken und verbessert die Datenqualität in Erhebungen zu einem gesellschaftspolitisch hoch relevanten Bereich. Dabei wurde die Verbesserung für Befragte und Produzierende belastungsarm umgesetzt.

Die Laudationes auf die prämierten Arbeiten hielten die Mitglieder der Jury des Statistischen Beirats, Frau

Solveigh Jäger (Bundesverband der Deutschen Industrie e.V.), Prof. Stefan Bender (Rat für Sozial- und Wirtschaftsdaten) und Prof. Dr. Joachim Wilde (Hochschulkonferenz).

## AUS ALLER WELT

---

### 21. Sitzung des CSSP

Am 17. und 18. Juni 2024 fand die 21. Sitzung des Committee on Statistics and Statistical Policy (CSSP) der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) statt. Die OECD-Statistik hat wesentlichen Einfluss auf Themen, die nahezu zeitgleich auch in den jeweiligen statistischen Gremien der Europäischen Union (EU) bis hin zum Europäischen Parlament übergeordnet beraten werden. Das CSSP behandelt neben strategischen Themen der Weiterentwicklung auch fachspezifische Fragen der amtlichen Statistik.

Der langjährige Chefstatistiker der OECD, Paul Schreyer, ist seit Juli 2024 im Ruhestand. Sein Nachfolger ist Steve MacFeely, vorher Direktor für Daten und Analytik bei der Weltgesundheitsorganisation.

Die Themen Imputationen und Modellierungen in der amtlichen Statistik sind wichtige Teilaspekte der OECD-Qualitätsrichtlinien. Verbindliche Kriterien für den Einsatz von Imputationen gibt es jedoch bisher ausschließlich auf nationaler Ebene. Auch der Umgang mit privat gehaltenen Daten sowie Cybersicherheit wird künftig in den Qualitätsrichtlinien thematisiert.

Zur Diskussion standen auch Maßnahmen, um die Rücklaufquoten bei Erhebungen zu erhöhen. Die Teilnahmequote bei Haushaltserhebungen ist bei einkommensschwachen sowie bildungsfernen Bevölkerungsgruppen geringer. Dem entgegenwirken würden eine Auskunftspflicht und verschiedene weitere Optionen, wie ein reduzierter Merkmalskatalog sowie der mögliche Einsatz sogenannter Smart-Features (beispielsweise Nutzung von GPS-Daten oder das Fotografieren von Kassenzetteln, die automatisiert weiterverarbeitet werden).

Weiterhin stellte die OECD eine erste Definition von Künstlicher Intelligenz (KI) in der amtlichen Statistik sowie fünf Prinzipien und fünf Empfehlungen zu ihrem

Einsatz vor. Wichtige Prinzipien sind die Wahrung der Grundrechte der Menschen sowie Transparenz hinsichtlich der KI-Nutzung und deren Auswirkungen auf statistische Daten.

Das Projekt „WISE“ (Well-being, Inclusion, Sustainability and Equal Opportunity – Wohlbefinden, Eingliederung, Nachhaltigkeit und Chancengleichheit) untersucht eine Vielzahl gesellschaftlicher Aspekte. Unter anderem gehört dazu die quantitative Darstellung der sozialen Verbundenheit in der Gesellschaft. Eine Befragung der amtlichen Datenproduzenten ergab unter anderem, dass das Thema Einsamkeit in nur etwa einem Drittel der relevanten Erhebungen enthalten ist. Künftige Arbeiten sollen nach Einschätzung der OECD deshalb Aspekte der sozialen Bindungen besser erfassen, um die Haushaltserhebungen dementsprechend ausweiten zu können. Außerdem stellte das WISE-Projekt einen Ansatz vor, mit dem die Kosten- und Lastenverteilungen, die sich durch den Klimawandel ergeben, statistisch gemessen werden können. Im Zuge eines OECD-Projekts wird eine Datenplattform aufgebaut mit dem Ziel, auf deren Grundlage die Folgen des Klimawandels selbst auf regionaler Ebene berechnen zu können.

### Themenseite zur Weltbevölkerung

Die Weltbevölkerung hat sich nach Berechnungen der Vereinten Nationen in den vergangenen 50 Jahren verdoppelt und inzwischen die Schwelle von 8 Milliarden Menschen überschritten. Mehr als die Hälfte aller Menschen lebt in Asien, weniger als jeder Zehnte in Europa. Das Statistische Bundesamt hat eine Themenseite zusammengestellt, auf der Daten zum Bevölkerungsanstieg in den vergangenen Jahrzehnten sowie zur künftigen Entwicklung in den bevölkerungsreichsten Staaten der Erde verfügbar sind.

➤ [www.destatis.de](https://www.destatis.de)

### AUS EUROPA

#### 72. Plenarsitzung der CES

Die Konferenz der Europäischen Statistiker (CES) tagte am 20. und 21. Juni 2024 in Genf. Im Mittelpunkt stand das Seminar zu Künstlicher Intelligenz (KI) und Machine Learning (ML). Es fasste die seit dem Jahr 2019 stattfindenden Initiativen zu Künstlicher Intelligenz in der amtlichen Statistik zusammen und nahm Bezug auf das UNECE CES White Paper „Large Language Models (LLM) for Official Statistics“. Das Whitepaper wurde im Dezember 2023 veröffentlicht und bietet einen umfassenden Überblick über LLM-Aspekte, die aus der Perspektive der amtlichen Statistik relevant sind.

Im gemeinsamen Seminar mit den kartographischen Ämtern zur Nutzung von Geodaten wurde ein Papier zu strategischen Leitlinien im Bereich Georeferenzierung vorgestellt. In der abschließenden Diskussion wurden gemeinsame Arbeitsbereiche für Statistik- und Geodaten identifiziert. Zu diesen Themen zählen unter anderem die Verknüpfung von Daten aus verschiedenen Quellen (Linked Data) und der Einsatz von Künstlicher Intelligenz bei der Integration von Geodaten und statistischen Daten.

Das Thema der subjektiv wahrgenommenen Armut ist sehr emotionsgeladen, sodass insbesondere der Statistik die Aufgabe zufällt, den Sachstand neutral darzustellen. Der CES wurde von der Task Force zur subjektiven Armut ein Papier vorgelegt, das konzeptionelle Hintergründe, Methodiken der Datenerhebung sowie Schlussfolgerungen und Empfehlungen enthält. Die europäischen Haushaltserhebungen bieten mit ihren regulären Frageprogrammen und Ad-hoc-Jahresmodulen bereits Ansätze, das Wohlergehen der Menschen zu messen.

Die Überarbeitung des internationalen Regelwerks der Volkswirtschaftlichen Gesamtrechnungen (System of National Accounts – SNA 2008) schreitet voran. Zwölf der angenommenen Empfehlungen haben konzeptionelle Änderungen zur Folge, vier davon zugleich direkte Auswirkungen auf das Bruttoinlandsprodukt (BIP). Deutschland hat, wie andere Staaten, Bedenken hinsichtlich der Auswirkungen und Machbarkeit ausgewählter Empfehlungen. Ein zu hoher Anteil von Berechnungen auf Basis von Modellen birgt das Risiko, keine aussagekräftigen

und robusten Ergebnisse zu erhalten. Rechtsverbindlich für Deutschland ist das Europäische System Volkswirtschaftlicher Gesamtrechnungen (ESVG), welches im Anschluss an das SNA überarbeitet wird.

#### Europäischer Statistikwettbewerb 2024



An der siebten Ausgabe des Europäischen Statistikwettbewerbs (European Statistics Competition – ESC) nahmen mehr als 22 000 Schülerinnen

und Schüler aus 20 EU-Mitgliedstaaten teil. In der Altersklasse der Mittelstufe errang das Team MATHLETICS der Sportschule im Olympiapark – Poelchau-Schule aus Berlin einen hervorragenden dritten Platz.

Für das Finale drehte jedes Team ein Video zu dem Thema „Früher war alles besser! Stimmt das oder nicht?“ auf Basis europäischer Daten. Alle Videos stehen online zur Verfügung:

➤ [www.esc2024.eu](http://www.esc2024.eu)

### AUS DEM INLAND

#### Veröffentlichung der Ergebnisse des Zensus 2022 gestartet

Mit einer Pressekonferenz haben das Statistische Bundesamt, das Statistische Landesamt Nordrhein-Westfalen (IT.NRW) und das Bayerische Landesamt für Statistik am 25. Juni 2024 die Veröffentlichungsphase der Ergebnisse des Zensus 2022 begonnen.

Die Ergebnisse umfassen Informationen zu den Bevölkerungszahlen, zur Bevölkerungsstruktur in Bund, Ländern und Kommunen sowie zu Bildung und Erwerbstätigkeit. Außerdem liefert der Zensus 2022 Daten zu Wohnungen und Gebäuden in Deutschland, darunter neue Merkmale, die im Zensus 2011 noch nicht enthalten waren, wie Nettokaltmieten oder Energieträger der Heizung.

Zunächst wurden auf der Zensus-Webseite Regionaltabellen auf Gemeindeebene und der [Zensus-Atlas](#) auf Basis von Gitterzellen veröffentlicht. In einem weiteren Schritt wurden ab Mitte Juli 2024 Daten in der [Zensusdatenbank](#) zur Verfügung gestellt, anfangs zu den Themenblöcken Bevölkerung (Demografie) und zu Gebäuden und Wohnungen. Die Datenbank und Informationsangebote wie das Zensus-Ergebnisportal werden bis Ende September 2024 vervollständigt und veröffentlicht.

➤ [www.zensus2022.de](http://www.zensus2022.de)

## 71. Jahrestagung des Statistischen Beirats

Der Statistische Beirat hat in seiner 71. Tagung am 27. Juni 2024 über Wege diskutiert, die Nutzung amtlicher Daten in der Forschung zu fördern. Dabei wurden mehrere Punkte erörtert: die Vergabe eines Forschungsauftrags an das Statistische Bundesamt, die gesetzliche Verankerung der Forschungsdatenzentren (FDZ) der Statistischen Ämter des Bundes und der Länder sowie die Bereitstellung der Daten und die Finanzierungsmöglichkeiten der Forschungsdatenzentren. Zusätzlich diskutierte der Statistische Beirat darüber, eine Mikrodatentreuhandstelle für öffentliche Daten einzurichten und die Möglichkeiten der Datenzusammenführung zu erweitern. Zu der Bitte der kommunalen Statistikstellen nach einem verbesserten Datenzugang für Einzelangaben und eine dauerhafte Speicherung der Adresse oder Geokoordinate zu ermöglichen sprach der Beirat Empfehlungen aus.

## VERANSTALTUNGEN

---

### Demografiegespräch „Wählerschaft im Wandel“

Das Statistische Bundesamt bietet in Zusammenarbeit mit dem Bundesinstitut für Bevölkerungsforschung die „Berliner Demografiegespräche“ an. Die Reihe richtet sich insbesondere an die Zielgruppe Politik und Verwaltung. Kurze Vorträge mit anschließender Diskussion beleuchten aktuelle demografische Themen aus Sicht

der amtlichen Statistik und der Forschung. So wird ein knapper und informativer Einblick geboten und zu Gesprächen über aktuelle Ergebnisse angeregt.

Am 29. Mai 2024 fand die Veranstaltung „Wählerschaft im Wandel“ statt. Kurz vor der Europawahl stieß das Thema – mit Fokus auf Zuwanderung und Änderungen im Staatsangehörigkeitsrecht – auf besonderes Interesse. Weiterhin ging es um eine Reform des Europawahlgesetzes. Sie bewirkt, dass 2024 auch Jugendliche im Alter von 16 und 17 Jahren wahlberechtigt waren.

➤ [www.destatis.de](http://www.destatis.de)

## 1. Stuttgarter VGR-Kolloquium

Gemeinsam mit dem Arbeitskreis Volkswirtschaftliche Gesamtrechnungen der Länder richtete das Statistische Bundesamt am 6. und 7. Juni 2024 das diesjährige VGR-Kolloquium erstmals in Stuttgart aus. Die traditionsreiche Veranstaltungsreihe, die in den letzten Jahren in Berlin und zuvor in Rostock stattfand, widmet sich neben theoretischen und konzeptionellen Grundlagen der Volkswirtschaftlichen Gesamtrechnungen (VGR) auch verwandten aktuellen Themen der amtlichen Statistik. Die Treffen ermöglichen einen fachlichen Austausch der verschiedenen Akteure aus amtlicher Statistik, Wissenschaft, Wirtschaft, Ministerien und Instituten zu den Themen der VGR und leisten somit einen Beitrag zu Kooperation und Koordination.

Schwerpunkte in diesem Jahr waren unter den Titeln „BIP und Nachhaltigkeit“ sowie „Was soll die VGR messen?“ in erster Linie konzeptionelle Grundlagen der VGR und aktuelle Aktivitäten in diesem Bereich. Die Vorträge aus dem Statistischen Bundesamt befassten sich vor allem mit der Revision der VGR-Rahmenwerke SNA und ESGV sowie den Verbindungen der Ökosystemgesamtrechnungen damit. Die Teilnehmenden am abendlichen Panel diskutierten die Themen „Datenanalyse per KI“ und „Experimentelle Konjunkturdaten“. Tenor aller Teilnehmenden war dabei, dass KI vielfältige Möglichkeiten in der Statistik bietet, die es noch zu erschließen gilt. Einigkeit bestand auch darin, dass die Bürokratiebelastung durch Statistik sehr gering ist und kein Potenzial für weitere Reduzierungen gesehen wird. Mit „Statistik trifft Wirtschaft“ war eine gemeinsame Veranstaltung mit der Industrie- und Handelskammer Region Stuttgart am ersten Abend überschrieben.



### 15. Wissenschaftliche Tagung des ADM, der ASI und des Statistischen Bundesamtes

Am 20. und 21. Juni 2024 fand die 15. Wissenschaftliche Tagung des Statistischen Bundesamtes gemeinsam mit dem Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM) und der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI) statt. Sie widmete sich dem Thema „Datenerhebung, Datenqualität und Datenethik in Zeiten von künstlicher Intelligenz“.

Die Veranstaltung bot den Teilnehmenden aus amtlicher Statistik, Wissenschaft und Privatwirtschaft eine Plattform für den interdisziplinären Austausch und die Diskussion über die aktuellen Einsatzmöglichkeiten, Herausforderungen und Risiken von künstlicher Intelligenz. So kann diese beispielsweise die Codierungen von großen Datenmengen übernehmen und damit teils (händische) Routinetätigkeiten, teils anspruchsvolle Tätigkeiten ersetzen.

Jedoch erfordert der Umgang mit sensiblen Daten ein hohes Maß an Verantwortung in Bezug auf Datenschutz und -sicherheit. Wichtige rechtliche Aspekte im Zusammenhang mit der Erstellung und Verwendung von KI-Modellen sind weiterhin Urheberrechte, Geheimhaltung, Arbeitsrecht und Diskriminierungsgefahr.

Die vollständige Tagungsdokumentation steht zur Verfügung unter

➤ [www.destatis.de](http://www.destatis.de)

### STATISTIK VISUALISIERT

---

#### Zensus-Atlas

Wie viele Menschen leben in Deutschland? Wie wohnen und arbeiten sie? Der Zensus-Atlas umfasst als interaktive kartografische Anwendung erste Ergebnisse des Zensus 2022 zu Bevölkerung, Gebäuden und Wohnungen und bietet kleinräumige Daten auf Gitterzellen-Basis an – auch über die Grenzen der Gemeinden hinaus.

Die zugrunde liegenden Daten aus dem Zensus 2022 stehen – hochauflösend georeferenziert – in 10-km-, 1-km- und 100-m-Gitterzellen als Zip-Dateien zum Download zur Verfügung. Damit können die Ergebnisse des Zensus auch unabhängig von administrativen Regionaleinheiten verarbeitet werden. Die Geheimhaltung der Ergebnisse aus dem Zensus 2022 ist auf allen Gitterzelebeneen sichergestellt.

➤ [atlas.zensus2022.de](https://atlas.zensus2022.de)

#### Unfallatlas mit neuen Funktionen

Wo passieren an meinem Wohnort die meisten Verkehrsunfälle? Wo ist man mit dem Fahrrad besonders gefährdet? Das und mehr zeigt der interaktive Unfallatlas der Statistischen Ämter des Bundes und der Länder. Damit ist es nun möglich, auch die Entwicklung an einzelnen Unfallstellen im Laufe der Jahre zu betrachten. Und: In der aktuellen Ausgabe werden erstmals Fahrradunfälle auf Rad- und Feldwegen dargestellt.

➤ [unfallatlas.statistikportal.de](https://unfallatlas.statistikportal.de)

#### Tourismusatlas aktualisiert

Aus welchen Herkunftsländern reisen die meisten Touristinnen und Touristen nach Deutschland? Wohin reisen Gäste aus Japan am liebsten? Die Webanwendung visualisiert ausgewählte Jahresergebnisse der bundesweiten Monatserhebung im Tourismus der deutschen amtlichen Statistik. Sie bietet verschiedene Auswertungsmöglichkeiten ab dem Berichtsjahr 2018 an und ist nun um das Berichtsjahr 2023 erweitert worden. Im Tourismusatlas können kleinräumige Darstellungen abgerufen werden, die in Form von Rasterkarten präsentiert werden. Unter dem Button „Analyse und Interpretation“ sind beispielsweise saisonale Muster der touristischen Nachfrage erkennbar, die aufgrund der Corona-Maßnahmen in den vergangenen Jahren überlagert wurden und sich nun im Jahr 2023 wieder stabil zeigen.

➤ [gis-hsl.hessen.de](https://gis-hsl.hessen.de)



## NEUERSCHEINUNGEN

---

### Bildung in Deutschland 2024

„Bildung in Deutschland“ erscheint alle zwei Jahre als umfassende und empirisch fundierte Bestandsaufnahme des deutschen Bildungswesens: von der frühkindlichen Bildung über die allgemeinbildende Schule und die non-formalen Lernwelten im Schulalter, die berufliche Ausbildung und Hochschulbildung bis hin zur Weiterbildung im Erwachsenenalter.

Der zehnte Bildungsbericht führt die Berichterstattung mit bereits bekannten Indikatoren zum deutschen Bildungswesen fort und präsentiert gleichzeitig neue Indikatoren. Ein Schwerpunkt Kapitel geht vertiefend auf berufliche Bildung ein und gibt damit einen Überblick über Governance, berufliche Orientierung, Weiterqualifizierung, Qualitätssicherung und weitere Aspekte beruflicher Bildung.

➤ [www.wbv.de](http://www.wbv.de)

### Neue Veröffentlichungen der OECD

#### OECD-Bericht zu Künstlicher Intelligenz in Deutschland

Deutschland war 2018 eines der ersten Länder, die eine nationale Strategie für Künstliche Intelligenz (KI) beschlossen haben. Ziel der Strategie ist es, Wachstum und Wettbewerbsfähigkeit zu fördern und eine verantwortungsvolle und vertrauenswürdige Entwicklung Künstlicher Intelligenz sicherzustellen.

Der „OECD-Bericht zu Künstlicher Intelligenz in Deutschland“ umfasst ein internationales Benchmarking des deutschen Ökosystems der Künstlichen Intelligenz und diskutiert Fortschritte bei der Umsetzung der nationalen KI-Strategie. Der Bericht erörtert Deutschlands Stärken, Schwächen, Chancen und Herausforderungen im KI-Bereich und gibt Empfehlungen zur Gestaltung der KI-Politik in Deutschland in den kommenden Jahren ab. Darüber hinaus befasst er sich mit der KI-Infrastruktur und enthält Sonderkapitel zu den drei Sektoren Öffentlicher Sektor, Nachhaltigkeit sowie Einsatz im Gesundheitswesen.

➤ [manage.oecd-berlin.de](http://manage.oecd-berlin.de)

#### Society at a Glance 2024 – OECD Social Indicators

Die OECD-Länder verzeichnen seit längerem einen Rückgang ihrer Geburtenraten. In den Nullerjahren kam es zu einer vorübergehenden Stabilisierung, seit der globalen Finanzkrise von 2007/2008 setzt sich der Abwärtstrend aber fort. Im OECD-Durchschnitt betrug die zusammengefasste Geburtenziffer 2022 nur noch 1,5 Kinder je Frau und blieb damit deutlich hinter dem Bestandserhaltungsniveau von 2,1 Kindern je Frau zurück.

„Society at a Glance“ gibt anhand von Indikatoren einen Überblick über soziale Trends und politische Entwicklungen in den OECD-Ländern. Verglichen werden etwa Scheidungsraten, Beschäftigungsquoten, die Verfügbarkeit von Kinderbetreuung, Einkommensungleichheit, Sozialausgaben, Kindersterblichkeit sowie Lebens- und Arbeitszufriedenheit. Schwerpunkt der diesjährigen Ausgabe sind Fertilitätstrends.

➤ [manage.oecd-berlin.de](http://manage.oecd-berlin.de)

#### OECD Employment Outlook 2024:

##### The Net-Zero Transition and the Labour Market

Die Arbeitsmärkte haben sich im vergangenen Jahr weiter gut entwickelt. Viele OECD-Länder verzeichneten ein historisch hohes Beschäftigungsniveau und eine niedrige Arbeitslosigkeit. In den meisten Ländern haben sich die Erwerbstätigenquoten von Frauen im Vergleich zu vor der Pandemie stärker verbessert als die von Männern.

Der „OECD Employment Outlook“ untersucht die jüngsten Arbeitsmarktentwicklungen und -aussichten in den OECD-Mitgliedstaaten. In dieser Ausgabe werden auch die Auswirkungen des Übergangs zu Netto-Null-Emissionen bis 2050 auf den Arbeitsmarkt und die Arbeitsplätze von Millionen von Arbeitnehmerinnen und Arbeitnehmern bewertet. Während sich die Gesamtbeschäftigung nicht wesentlich ändern wird, werden Arbeitsplätze in schrumpfenden emissionsintensiven Industrien verloren gehen und gleichzeitig andere in expandierenden emissionsarmen Aktivitäten entstehen. Viele Tätigkeiten werden sich grundlegend verändern.

➤ [manage.oecd-berlin.de](http://manage.oecd-berlin.de)

Zu dieser Veröffentlichung in englischer Sprache gibt es auch eine Ländernotiz für Deutschland auf Deutsch:

➤ [www.oecd.org](http://www.oecd.org)



# MASCHINELLES LERNEN IM STATISTISCHEN BUNDESAMT

Ein Überblick über die Historie seit 2015  
und aktuelle Entwicklungen

Florian Dumpert

📌 **Schlüsselwörter:** Digitalisierung – Automatisierung – methodische Weiterentwicklung – Zusammenarbeit mit der Wissenschaft – Technologie

## ZUSAMMENFASSUNG

Maschinelles Lernen ist ein wichtiger Bestandteil der amtlichen Statistik. Der Artikel stellt erstmals umfassend die Historie sowie aktuelle Entwicklungen zum Maschinellen Lernen in der amtlichen Statistik in Deutschland dar. Er zeigt drei charakteristische Phasen auf, von den Anfängen über die quantitative und qualitative Erweiterung hin zur Reifung und Etablierung, und benennt anstehende Herausforderungen. Dabei liegt der Fokus auf der Bundesstatistik und dem Statistischen Bundesamt, ergänzt durch Verweise auf den Statistischen Verbund und internationale Entwicklungen.

📌 **Keywords:** digitalisation – automation – methodological development – collaboration with academia – technology

## ABSTRACT

*Machine learning is an essential element of official statistics. This article is the first to present a comprehensive account of the history and current developments of machine learning in official statistics in Germany. It looks at three characteristic phases, detailing the beginnings of machine learning, its quantitative and qualitative expansion and its subsequent maturation and establishment, and identifies upcoming challenges. The focus of the article is on federal statistics and the Federal Statistical Office, with references to the German Official Statistics Network and international developments.*



**Dr. Florian Dumpert**

leitet das Referat „Künstliche Intelligenz, Big Data“ des Statistischen Bundesamtes, das sich beispielsweise mit Verfahren des Maschinellen Lernens und der Imputation befasst. Der Diplom-Mathematiker beschäftigt sich unter anderem mit methodischen Fragestellungen beim Einsatz dieser Verfahren und verantwortet die Weiterentwicklung der Bundesstatistik mittels dieser Verfahren.

## 1

### Einleitung

Dieser Aufsatz beleuchtet das Maschinelle Lernen (ML) in der deutschen amtlichen Statistik und berücksichtigt dabei besonders die Perspektive des Statistischen Bundesamtes. Die Einsatzzwecke von Maschinellern Lernen in der amtlichen Statistik sind breit gestreut: Sie reichen von für sich stehenden Klassifikationen und Regressionen über ML-basiertes Zusammenführen von Datensätzen bis hin zum Einsatz von Maschinellern Lernen, um fehlerhafte und fehlende Werte zu ersetzen. Die Historie des Einsatzes von Maschinellern Lernen im Statistischen Bundesamt beginnt ungefähr 2015 – eine kurze Zeitspanne im Vergleich zur Historie des Statistischen Bundesamtes selbst, das 2023 seinen 75. Geburtstag feierte (Statistisches Bundesamt, 2023a). Das ist auf den ersten Blick nicht besonders überraschend, handelt es sich bei Maschinellern Lernen oder (häufig synonym verwendet) bei Künstlicher Intelligenz (KI) doch um eine „State of the art“-Technologie. Ob sich die Anwendung von Verfahren des Maschinellen Lernens tatsächlich noch als vergleichsweise „jung“ bezeichnen lässt, hängt vor allem von der Charakterisierung des Begriffs des Maschinellen Lernens ab.<sup>1</sup> Das Statistische Bundesamt nutzt die folgende, auch graduell erfüllbare Charakterisierung:

Bei Maschinellern Lernen handelt es sich

- › häufig (nicht immer!) um nichtparametrische statistische Methoden,
- › um Muster und Zusammenhänge zu erkennen,
- › mit einem Schwerpunkt – zumindest beim sogenannten überwachten Lernen – auf der Prädiktion (und weniger auf der Erklärung),
- › die genutzt werden, um konkrete Fragestellungen zu beantworten, ohne die Lösung explizit vorgegeben zu bekommen,
- › somit im Allgemeinen daten- und nicht modellgetrieben sind

---

1 Wobei Charakterisierung hier nicht im Sinne einer exakten „genau dann, wenn“-Beziehung zu verstehen ist.

- › und sich dadurch auszeichnen, dass der Raum der Lösungen (gemeint ist der Hypothesenraum) häufig so groß ist, dass er (annähernd) alle Muster und Zusammenhänge enthält.<sup>2</sup>

Die Charakterisierung erfolgt somit nicht in der Art, dass die Liste möglicher Verfahren (lineare Regression, logistische Regression, Klassifikations- und Regressionsbaum, Support Vector Machine, Neuronales Netz und weitere Verfahren) nach „klassischer Statistik“<sup>3</sup> und Maschinellern Lernen sortiert würde. So kann beispielsweise die lineare Regression – je nach Anwendungsfall und wie sie trainiert (das heißt angepasst) wird – sowohl klassisch als auch im Sinne des Maschinellen Lernens betrachtet werden. Je eher die oben genannten Aspekte erfüllt sind, desto eher wird von Maschinellern Lernen gesprochen.<sup>4</sup>

Die folgenden Kapitel 2, 3 und 4 beschreiben den Einsatz und die Weiterentwicklung von Maschinellern Lernen im Statistischen Bundesamt in den Jahren 2015 bis 2017, 2018 bis 2020 sowie 2021 bis 2023. Kapitel 5 geht auf aktuelle Entwicklungen ein, Kapitel 6 skizziert abschließend Herausforderungen und wie die Zukunft von Maschinellern Lernen in der amtlichen Statistik aussehen könnte.

## 2

### Die Anfänge: 2015 bis 2017

Die Anfangsjahre des Maschinellen Lernens im Statistischen Bundesamt waren geprägt von einem Herantasten und ersten Versuchen des Einsatzes. Vorerfahrungen bestanden kaum, nur wenige Beschäftigte hatten bereits Kontakt mit diesen Verfahren. Das verwundert nicht, denn selbst die meisten neu eingestellten Beschäftigten mit Hochschulabschluss hatten im Rahmen ihrer universitären Curricula zu diesem Zeitpunkt kaum oder keinen Kontakt zu Maschinellern Lernen. So hing es

---

2 Der letzte Punkt besagt, dass prinzipiell – wenngleich im Einzelfall gegebenenfalls nicht de facto – alle (messbaren) Muster und Zusammenhänge beliebig genau abgebildet werden können, es also keine grundsätzlich nicht darstellbaren Muster und Zusammenhänge gibt.

3 Dieser Begriff wird in diesem Aufsatz nicht näher erläutert, sondern dient als gedanklicher Gegenpol zum Maschinellen Lernen.

4 Eine Quantifizierung oder gar die Angabe eines Schwellenwertes ist müßig.

zunächst von aus Eigeninteresse erworbenen Kenntnissen zu Maschinellern Lernen sowie von der Möglichkeit zur Zusammenarbeit mit Externen ab, wo (das heißt in welcher Fachstatistik) ML-Projekte stattfinden konnten. Diese dienten jedoch nicht ausschließlich dem reinen Erkunden neuer Möglichkeiten. Bereits zu Beginn war die Zielsetzung deutlich: Mittels Maschinellern Lernen sollten Arbeitsschritte im statistischen Produktionsprozess<sup>5</sup> effizienter gestaltet oder Auswertungsmöglichkeiten erweitert werden, um damit die Qualität der Statistiken zu verbessern. Erste Projekte umfassten beispielsweise die Identifikation im statistikrechtlichen Sinne nicht relevanter Einheiten in der Handwerksstatistik (Feuerhake/Dumpert, 2016) oder die Erkennung von Unternehmen, die dem sogenannten Dritten Sektor angehören (Dumpert und andere, 2016). Diese und weitere Beispiele werden auch in Dumpert/Beck (2017) diskutiert.

Charakteristisch für die zunehmende Befassung mit dem Maschinellern Lernen war einerseits ein stetiges Hinzulernen hinsichtlich der Herausforderungen, die der Einsatz von Maschinellern Lernen mit sich brachte. Diese Herausforderungen waren sowohl allgemeiner Art (zum Beispiel der Umgang mit „imbalanced data“-Situationen) als auch der spezifischen Situation im Statistischen Bundesamt geschuldet. Beispielsweise war die Nutzung der Programmiersprache R zu dieser Zeit im Statistischen Bundesamt noch nicht etabliert, sodass Wege gefunden werden mussten, die entwickelten ML-Lösungen überhaupt auf statistischen Einzeldaten einsetzen zu können. Andererseits gab es zu dieser Zeit auch noch keine einheitliche Vorstellung davon, wie der Einsatz von Maschinellern Lernen die amtliche Statistik jenseits der genannten Einzelfälle bereichern könnte. Darüber hinaus mussten ML-bezogene Kooperationen mit der Wissenschaft und der diesbezügliche Austausch mit anderen (statistischen) Ämtern erst aufgebaut werden. Von Anfang an wurde großer Wert darauf gelegt, den Fortschritt und auch die untersuchten und verwendeten Verfahren durch Publikationen und Vorträge sowohl innerhalb der amtlichen Statistik als auch gegenüber der Fachöffentlichkeit transparent zu machen. Dieses Vorgehen erfüllte bereits die Forderung in den aktuellen

Debatten zur Regulierung von Künstlicher Intelligenz, den Einsatz solcher Verfahren bei der Erstellung eines Produkts bekanntzugeben.<sup>6</sup>

Somit ist festzuhalten, dass die Umsetzung der ersten Projekte, der „low hanging fruits“, stark von einzelnen Personen und deren Engagement (sowohl auf strategischer als auch auf fachlicher Ebene) abhängig war. Wäre es dabei geblieben, wäre die Zukunft des Maschinellern Lernens in der Bundesstatistik schnell infrage gestellt worden.

---

### 3

## Qualitative und quantitative Erweiterung: 2018 bis 2020

---

Die erfolgreichen Projekte der Anfangsjahre eröffneten die Möglichkeit, das Thema höherrangig zu behandeln und eine generelle Aussage über den Einsatz von Maschinellern Lernen in der amtlichen Statistik zu treffen. Das Maschinelle Lernen wurde den Digitalisierungsthemen des Statistischen Bundesamtes (und später auch des Statistischen Verbunds<sup>7</sup>) zugeordnet und Teil der Digitalen Agenda: „Eines von vier Leuchtturmprojekten unserer Digitalen Agenda ist die Durchführung eines ‚Proofs of Concept Machine Learning‘.“ (Statistisches Bundesamt, 2019; hier: Seite 15). Der im Jahr 2018 durchgeführte Proof of Concept umfasste folgende Teilbereiche:

- › Klärung der Begrifflichkeit,
- › Konzept zum Informationsaustausch im Statistischen Bundesamt,
- › Umfrage bei Statistikinstitutionen zum Einsatz von Maschinellern Lernen,
- › Abfrage im Statistischen Bundesamt zum Einsatz von Maschinellern Lernen,
- › Gedanken zur notwendigen Infrastruktur,
- › Handlungsempfehlungen.

---

<sup>5</sup> Für eine Erläuterung des allgemeinen statistischen Produktionsprozesses siehe Wirtschaftskommission für Europa der Vereinten Nationen – UNECE ([statswiki.unece.org](https://statswiki.unece.org)) sowie Blumöhr und andere (2017), hier: Kapitel 3.

<sup>6</sup> Beispiele sind selbstverpflichtende Leitlinien für den KI-Einsatz (Denkfabrik Digitale Arbeitsgesellschaft im Bundesministerium für Arbeit und Soziales, 2022), die Stellungnahme des Deutschen Ethikrats (Deutscher Ethikrat, 2023) sowie die europäische KI-Verordnung.

<sup>7</sup> Den Statistischen Verbund bilden die Statistischen Ämter des Bundes und der Länder.

Der Informationsaustausch sollte das vorhandene Wissen zu Maschinellern Lernen in Form von Notizen und Dokumenten an einer zentralen Stelle bündeln und für alle Beschäftigten auffindbar und nutzbar machen. Ein zentrales System, das dieses leisten konnte, gab es zu diesem Zeitpunkt noch nicht, es musste auf eine Plattform aus der Softwareentwicklung zurückgegriffen werden. Zudem dienten Veranstaltungsformate für verschiedene Zielgruppen dem Informationsaustausch im Statistischen Bundesamt. Sie stellten die Chancen und die zu bedenkenden Aspekte von Maschinellern Lernen, die Beispielprojekte aus den Jahren 2015 bis 2017 sowie die Ergebnisse der Umfrage bei anderen Statistikinstitutionen vor und zur Diskussion. Themen der Umfrage unter allen europäischen Statistikämtern, vielen deutschen und einigen weiteren ausländischen Statistikproduzenten waren im Wesentlichen der Inhalt des jeweiligen Projekts (Zielsetzung), der aktuelle Status (Idee, Entwicklung, Test, Produktion), das oder die eingesetzte(n) Verfahren und die eingesetzte Software. Ergebnis des Proofs of Concept 2018 war, dass einige Ämter und Institutionen Maschinellern Lernen bereits erproben, dass hierfür häufig baumbasierte Verfahren (wie Random Forest) eingesetzt wurden, dass aber viele Ansätze noch in der Entwicklungsphase, mithin nicht im Produktivbetrieb waren. Häufig wurden Aufgaben aus dem Bereich der Klassifikation (im Unterschied zur Regression) mittels Maschinellern Lernen bearbeitet.

Bei der Infrastruktur wurde der Fokus auf die Notwendigkeit der Bereitstellung von R gelegt. Alternativen (zum Beispiel MATLAB) und Ergänzungen (zum Beispiel Python) wurden ebenfalls diskutiert. Damit verbunden wurde deutlich, dass Arbeitsplatz-PCs alleine nicht ausreichen, um Maschinellern Lernen im Statistischen Bundesamt zu betreiben. Auch die Notwendigkeit der Nutzung von Grafikprozessoren für den Einsatz von Deep Learning hatte der Proof of Concept bereits ergeben. Der Proof of Concept schloss mit Handlungsempfehlungen für die Leitung des Statistischen Bundesamtes: ein zentrales Referat für das Thema Maschinellern Lernen einrichten, die notwendige Infrastruktur bereitstellen, die Statistischen Ämter der Länder einbeziehen, Innen- und Außenkommunikation praktizieren sowie mit der Wissenschaft zusammenarbeiten (Beck und andere, 2018a; Beck und andere, 2018b).

Zum 1. August 2018 wurde das Referat „Maschinellern Lernen und Imputationsverfahren“ im Statistischen Bun-

desamt eingerichtet und damit erstmals ein zentrales Kompetenzzentrum zu diesen Themen organisatorisch verankert.<sup>8</sup> Ebenfalls 2018 wurde der erste Platz des Innovationspreises des Statistischen Beirats<sup>9</sup> für die bisherigen Arbeiten im Bereich Maschinellern Lernen vergeben. Mit der Einrichtung des Referats und der Würdigung der bisherigen Arbeiten durch den Statistischen Beirat einher gingen ein personeller Aufwuchs und damit eine quantitative Steigerung der untersuchten Fragestellungen (unter anderem Schmidt, 2020). Neue Anwendungsfälle wurden in gemeinsamen Workshops von Fachreferaten und zentralem Kompetenzzentrum erarbeitet. Damit verbunden zeigte sich erstmals, dass eine Koordinierung der Arbeiten notwendig ist. Darüber hinaus wurde deutlich, dass es offene methodische Fragestellungen bezüglich der Anwendung von Maschinellern Lernen in der amtlichen Statistik gab (und bis heute gibt) und dass die nationale wie internationale Vernetzung ausgebaut werden sollte.

Auf nationaler Ebene gründete sich im Jahr 2018 eine Gruppe namens „KI-Labor“, die schließlich im 2020 eingerichteten Arbeitskreis Maschinellern Lernen (AK ML) aufging. Dieses Gremium besteht aus Vertreterinnen und Vertretern des Statistischen Bundesamtes und der Statistischen Ämter der Länder und versteht sich als „Treiber des Prozesses“ im Statistischen Verbund rund um das Thema Maschinellern Lernen. Es übernimmt dort die Steuerung der mit der Thematik verbundenen Aufgaben und Maßnahmen und vernetzt die beteiligten Akteure und Stakeholder.

Besonders wichtig für die internationale Vernetzung war das Machine Learning Project 2019/2020 der High-Level Group for the Modernisation of Official Statistics der Wirtschaftskommission der Vereinten Nationen für Europa (UNECE HLG-MOS). Dieses Projekt hat erstmals einen direkten Austausch auf internationaler Ebene zum Einsatz von Maschinellern Lernen in der amtlichen Statistik ermöglicht (United Nations, 2022). Details zu diesem Projekt und eine entsprechende Einordnung für die deutsche amtliche Statistik liefert Dumpert (2021).

---

8 Das Thema Imputation war zuvor organisatorisch mit der statistischen Geheimhaltung verbunden.

9 Der Statistische Beirat ist das nach § 4 Bundesstatistikgesetz berufene Gremium, welches das Statistische Bundesamt in statistischen Fachfragen berät und das die Belange der Nutzerinnen und Nutzer der Bundesstatistik vertritt.



### 4

## Reifung und Etablierung: 2021 bis 2023

In den Jahren 2021 bis 2023 hat sich der Umgang mit und der Einsatz von Maschinellern Lernen in der Bundesstatistik (und auch in einigen Statistischen Landesämtern) zunehmend etabliert. In vielen Fällen wird nun Maschinelles Lernen, einschließlich Natural Language Processing<sup>10</sup>, erprobt oder im statistischen Produktionsprozess eingesetzt (siehe dazu zum Beispiel Levagin und andere, 2022; Dumpert/Beck, 2023; Moritz und andere, 2024; Weißmann/Herbst, 2024; Limberg, 2024) oder beispielsweise seine Nützlichkeit bei der Erzeugung synthetischer Daten untersucht<sup>11</sup>. Gleichzeitig wurden ausstehende Aufgaben bearbeitet, zu nennen sind hier die methodische Weiterentwicklung sowie Fragen zu Qualitätsaspekten beim Einsatz Maschinellen Lernens in der amtlichen Statistik.

Im Bereich der Methodenfragen wurde das Thema Record Linkage aufgegriffen (Schnell, 2021). Darüber hinaus galt es, Fragen zum Hyperparameter-Tuning<sup>12</sup> zu klären. Hinsichtlich eher allgemeiner Fragen hierzu liefert Bischl und andere (2022) wichtige Einsichten. Für eher spezielle Fragestellungen (zum Beispiel die Visualisierung des Tuning-Prozesses) sei auf Bartz und andere (2023) verwiesen. Außerdem konnten 2022 die Arbeiten in einem Kooperationsprojekt mit der Ludwig-Maximilians-Universität München beginnen. Bearbeitet wurden Fragen zur Evaluation der Prädiktionsgüte von Maschinellen Lernverfahren in komplexen Situationen (Hornung und andere, 2023), zur Interpretierbarkeit von Ergebnissen (zum Beispiel Dandl und andere, 2023), zu Maschinellern Lernen bei besonderen Stichprobendesigns (Nalenz und andere, 2024) und zu Fairness (zum Beispiel Schenk/Kern, 2024). Die Arbeiten von Stock und anderen (2023) zu Federated Learning und

von Moritz und anderen (2024) zur Anwendbarkeit von Online Learning in der amtlichen Statistik erweiterten den Blick über die naheliegenden Einsätze hinaus.

Mit Blick auf Qualitätsaspekte beim Einsatz Maschinellen Lernens in der amtlichen Statistik wurde in Zusammenarbeit mit Kolleginnen und Kollegen aus Statistischen Landesämtern in einem mehrstufigen Prozess eine Liste von Dimensionen und Querschnittsaspekten erarbeitet (Saidani und andere, 2023). Diese leitet aus den generischen Ausführungen im Qualitätshandbuch der Statistischen Ämter des Bundes und der Länder (Statistische Ämter des Bundes und der Länder, 2021) die speziellen Anforderungen für Maschinelles Lernen ab. Dem vorausgegangen war ein Workshop mit Teilnehmerinnen und Teilnehmern aus Wissenschaft, amtlicher Statistik und Nutzenden (zum Beispiel Wirtschaftsforschungsinstituten) zu Qualitätsaspekten Maschinellen Lernens (Dumpert und andere, 2023). Darauf aufbauend werden aktuell Fragen nach konkreten Qualitätsrichtlinien und nach erweiterten Möglichkeiten des transparenten Umgangs mit Maschinellern Lernen in der amtlichen Statistik behandelt.

Ein ganz anderer Ansatz, Maschinelles Lernen als gewöhnliches Hilfsmittel zur (Teil-)Automatisierung von statistischen Produktionsprozessen oder zur Datenanalyse zu etablieren, waren die Hackathons in den Jahren 2021 und 2022. Sie standen unter den Aufgabenstellungen „Wie können wir zukünftig die Effektivität von nationaler Klimaschutzpolitik datenbasiert validieren?“ beziehungsweise „Welches innovative Produkt kann die amtliche Statistik entwickeln, um in Zukunft schneller relevante Daten in Krisen bereitzustellen?“. Der Hackathon 2021 wurde als interne Veranstaltung des Statistischen Bundesamtes durchgeführt, der [Hackathon 2022](#) war auch für Teilnehmerinnen und Teilnehmer aus statistischen Ämtern anderer Staaten offen. Anhand der Kriterien Relevanz, Innovationscharakter, Design, Automatisierung und Universalität hat eine Jury die Lösungen ausgewertet. Wesentliches Ziel der beiden Hackathons war es, Freiraum und Gelegenheit für eine intensive und interdisziplinäre Zusammenarbeit außerhalb der normalen Arbeitsroutinen zu schaffen. Gleichzeitig wurde auch das Vernetzen innerhalb des Statistischen Bundesamtes sowie mit Kolleginnen und Kollegen aus anderen statistischen Ämtern gefördert. Obwohl die Hackathons auch Wettbewerbscharakter hatten, waren es letztlich doch die übergreifenden Aspekte, die dieses Format för-

10 Natural Language Processing bezeichnet die Verarbeitung natürlicher Sprache durch Computer. Dabei wird natürliche Sprache (zum Beispiel Texte) in durch statistische Verfahren verarbeitbare Daten transformiert und mit geeigneten Verfahren, zum Beispiel des Maschinellen Lernens, weiterverarbeitet.

11 Dies geschieht beispielsweise im 2023 begonnenen [Forschungsprojekt „Anonymität bei integrierten und georeferenzierten Daten“ \(AnigeD\)](#).

12 Die Ermittlung geeigneter Detailspezifikationen der zu trainierenden ML-Verfahren wird Hyperparameter-Tuning genannt.

derlich für eine weitere Verbreitung von Maschinellern Lernen im Bewusstsein der Beschäftigten erscheinen ließen.<sup>13</sup>

Die Weiterentwicklung betrifft aber nicht nur methodische Fragen und Aspekte der Arbeitskultur im Statistischen Bundesamt, sondern in gleichem Maße die technologischen Möglichkeiten, die dem Statistischen Bundesamt zur Verfügung stehen. Während in den Anfangsjahren des Maschinellen Lernens noch auf Arbeitsplatzrechnern – also im besten Sinne lokal – gearbeitet werden musste, hat sich dies über die Zeit wesentlich verändert. Heute gibt es rechtliche und technische Möglichkeiten, Server- und (private) Cloudlösungen zu nutzen, was sinnvoll und notwendig für einen gewinnbringenden Einsatz von Maschinellern Lernen in der amtlichen Statistik ist.

13 Auch 2023 fand ein Hackathon im Statistischen Bundesamt statt, dieser jedoch ohne expliziten Bezug zu Künstlicher Intelligenz/ Maschinellern Lernen. Teams des Statistischen Bundesamtes waren darüber hinaus sehr erfolgreich beim European Big Data Hackathon 2023 ([Statistisches Bundesamt, 2023b, hier: Seite 14](#)).

## Übersicht 1

Arbeitspakete am AIML4OS-Grant zum Thema Künstliche Intelligenz und Maschinelles Lernen in der amtlichen Statistik

Nr.	Titel	Leitung	Beteiligung des Statistischen Bundesamtes
1	Project management and coordination	Irland	
2	Communication and community engagement	Italien	
3	ESS AI/ML lab: Technical infrastructure and organisational setup	Frankreich	
4	AI/ML state-of-play and ecosystem monitoring	Deutschland	✓
5	Standards, methodological and implementation frameworks	Niederlande und Deutschland	✓
6	Knowledge repository and training material	Polen	
7	Use Case: AI/ML on earth observation data, satellite imagery	Niederlande	
8	Use Case: Statistically valid and efficient editing and imputation in official statistics by AI/ML – with a special focus on editing	Deutschland	✓
9	Use Case: Statistically valid and efficient editing and imputation in official statistics by AI/ML – with a special focus on imputation	Spanien	✓
10	Use Case: From text to code – Experiences and potential of the use of AI/ML for classifying and coding	Deutschland	✓
11	Use Case: Applying ML for estimating firm-level supply chain networks	Niederlande	
12	Use Case: Large language models	Schweden	
13	Use Case: Generation of synthetic data in official statistics: techniques and applications	Italien und Österreich	✓

## 5

### Internationale Entwicklungen: 2024

Bereits im Frühjahr 2023 hat das Statistische Amt der Europäischen Union (Eurostat) die nationalen statistischen Institute aufgerufen, sich auf einen Grant zum Thema Künstliche Intelligenz und Maschinelles Lernen in der amtlichen Statistik (AIML4OS) zu bewerben. Das Statistische Bundesamt schloss sich seinerzeit einem Konsortium unter der Leitung Irlands an, das schließlich aus 15 beteiligten Staaten bestand und im März 2024 auch den finalen Zuschlag für die Förderung erhielt. Die Arbeiten im 4-Jahres-Grant begannen am 2. April 2024. Teilfinanziert durch den Grant wird bis 2028 an Querschnitts- und anwendungsspezifischen Themen gearbeitet. Einen Überblick über die Arbeitspakete gibt [Übersicht 1](#).

Im Frühjahr 2024 hat das Statistische Bundesamt die als wissenschaftliche Fachtagung konzipierte „Conference on Foundations and Advances of Machine Learning“ ausgerollt.



ning in Official Statistics“ veranstaltet.<sup>14</sup> Ziel war, den Austausch mit der Wissenschaft, anderen Behörden und anderen statistischen Ämtern national wie international zu stärken, Einblicke in die Herangehensweisen und Projekte anderer zu erhalten und hinsichtlich methodischer und technologischer Entwicklungen am Puls der Zeit zu sein. Inspiriert durch über 40 Vorträge haben rund 150 Teilnehmerinnen und Teilnehmer aus insgesamt 19 Staaten aktuelle Herausforderungen und Ansätze zu deren Lösungen präsentiert und diskutiert. Das Themenspektrum reichte dabei von

- › mathematisch-statistischen Fragestellungen (unter anderem zu Fehler und Unsicherheit, Einfluss des Stichprobendesigns und Resamplingverfahren) über
- › Technologie, Standards und Qualität (einschließlich Fairness und Reproduzierbarkeit sowie Fragestellungen rund um das Thema Prozessverbesserung) bis hin zu
- › konkreten Anwendungsfällen Maschinellen Lernens innerhalb und außerhalb der Statistikproduktion (unter anderem Datenintegration, Textklassifikation und Natural Language Processing, Codierung statistischer Einheiten auf Basis textueller Beschreibungen, Nutzung von Large Language Models, Datenvalidierung und Imputation).

## 6

### Herausforderungen und ein Blick in eine mögliche Zukunft

Dass Maschinelles Lernen in der amtlichen Statistik auch weiterhin eine wichtige Rolle spielen wird, steht außer Frage. Anforderungen aus dem Bundesstatistikgesetz<sup>15</sup> und dem Qualitätshandbuch der Statistischen Ämter des Bundes und der Länder<sup>16</sup> liefern Gründe dafür, insbesondere das daraus abgeleitete Bestreben, die

Produkte der amtlichen Statistik – auch wissenschaftlich – so verwertbar wie möglich für ihre Nutzerinnen und Nutzer zu erstellen (Qualität als „fitness for use“). Auch die Kommission Zukunft Statistik<sup>17</sup> empfiehlt, „lernende Algorithmen in der Entwicklung, Produktion und Evaluation eigener Produkte explorativ und transparent“ einzusetzen (Kommission Zukunft Statistik, 2024; hier: Seite 6). Darüber hinaus trifft der Fachkräftemangel auch den öffentlichen Dienst und die amtliche Statistik konkurriert mit anderen Aufgabenbereichen der öffentlichen Verwaltung um Finanzmittel. Bei der daher notwendigen (Teil-)Automatisierung von Produktionsschritten, gegebenenfalls einhergehend mit einer Digitalisierung und Standardisierung der Prozessschritte und der eingesetzten Werkzeuge, spielt Maschinelles Lernen notwendigerweise eine wichtige Rolle.

Bei allen Erfolgen, die hier bereits erzielt wurden, bleiben aber auch offene Fragen. Neben noch ungeklärten methodischen Fragestellungen und der konkreten Umsetzung des Qualitätsbegriffs und der Qualitätssicherung (siehe Kapitel 4) stellt die Integration der einzusetzenden ML-Lösungen in die IT-Systeme eine Herausforderung für die Zukunft dar. Die Einführung von MLOps<sup>18</sup>, das längst Industriestandard ist, steht für die Bundesstatistik noch aus, das Gleiche gilt für ein entsprechendes Management der mannigfaltigen Datenbestände der deutschen amtlichen Statistik. Aus diesen Datenbeständen speisen sich die für Maschinelles Lernen erforderlichen Trainings-, Validierungs- und Testdaten und in diese Bestände gehen durch ML-Verfahren bearbeitete Daten ein. Insbesondere mit Blick auf die Qualitätsdimension der Reproduzierbarkeit erscheinen diese beiden Punkte unabdingbar. Außerdem ist gerade bei den technologischen Möglichkeiten nie ein finaler Zustand erreicht. Durch die Weiterentwicklung auf methodischer und technologischer Seite ist die Infrastruktur der Informationstechnik immer wieder zu aktualisieren, zu verändern und anzupassen – sofern der ent-

14 Die Tagungsdokumentation mit Abstracts und Präsentationsfolien steht unter [www.destatis.de](http://www.destatis.de) zur Verfügung.

15 „Sie [die Bundesstatistik] gewinnt die Daten unter Verwendung wissenschaftlicher Erkenntnisse und unter Einsatz der jeweils sachgerechten Methoden und Informationstechniken.“ (§ 1 Satz 3 Bundesstatistikgesetz)


16 Statistische Ämter des Bundes und der Länder (2021). Hier insbesondere: Qualitätsgrundsätze G07 (solide Methodik), G08 (geeignete statistische Verfahren), G10.2 (Produktivitätspotenzial der Informationstechnologie).

17 [Die Kommission Zukunft Statistik](#) (KomZS) wurde vom Statistischen Bundesamt eingerichtet und mit der Erarbeitung von Empfehlungen für eine vorausschauende Programmplanung und eines Zielbilds der amtlichen Statistik für das Jahr 2030 beauftragt. Der Abschlussbericht der Kommission (Kommission Zukunft Statistik, 2024) wurde am 16. Januar 2024 der Leitung des Statistischen Bundesamtes überreicht.

18 Unter MLOps (machine learning operations) versteht man „Praktiken und Prozesse, die darauf abzielen, Modelle für maschinelles Lernen zuverlässig und effizient zu entwickeln, produktiv bereitzustellen, zu verwalten, zu überwachen und zu warten“ (Saidani und andere, 2023; hier: Seite 294).

sprechende Fortschritt für das Statistische Bundesamt nutzbar gemacht werden soll. Weitere erleichternde und hemmende Faktoren von Maschinellern Lernen in der amtlichen Statistik hat kürzlich Karanka (2023) herausgearbeitet.

Bereits heute haben die Kolleginnen und Kollegen in den Fach- und Querschnittsbereichen des Statistischen Bundesamtes die Möglichkeit, sich zum Thema Maschinelles Lernen fortzubilden. Ergänzend ist zu beobachten, dass ML-Themen mittlerweile auch Einzug in viele universitäre Curricula gefunden haben. Daher wird Maschinelles Lernen – zusätzlich begünstigt durch die fortschreitende einfache Nutzbarkeit der ML-Techniken im Standardumfeld des Computerarbeitsplatzes – perspektivisch ein gewöhnliches Werkzeug für viele Beschäftigte in der Statistikproduktion werden. Das aktuell bestehende zentrale Kompetenzzentrum könnte mittelfristig seine Ressourcen stärker zur Weiterentwicklung von Maschinellern Lernen in der amtlichen Statistik einsetzen. Das heißt, es könnte verstärkt methodische Fragestellungen bearbeiten, generische Komponenten entwickeln, die Einhaltung von Standards mit Bezug zu Maschinellern Lernen definieren und überwachen, die fachliche und technologische Integration von Maschinellern Lernen weiter vorantreiben und außerordentliche Fragestellungen lösen.

Obwohl derzeit Teil der öffentlichen Debatte, hat dieser Aufsatz das Thema generative KI nicht behandelt. Zwar gibt es erste Untersuchungen, inwiefern auch diese Klasse von Verfahren (hier vor allem Large Language Models) im Statistischen Bundesamt eingesetzt werden könnte, jedoch zeigten diese innerhalb der Kernprozesse der klassischen Statistikproduktion bislang keinen nennenswerten Vorteil. Es ist für die Zukunft jedoch angezeigt, generative KI für die amtliche Statistik weiter zu beobachten und zu bewerten. 

## LITERATURVERZEICHNIS

---

- Bartz, Eva/Bartz-Beielstein, Thomas/Zaefferer, Martin/Mersmann, Olaf. *Hyperparameter Tuning for Machine and Deep Learning with R*. Singapur 2023. DOI: [10.1007/978-981-19-5170-1](https://doi.org/10.1007/978-981-19-5170-1)
- Beck, Martin/Dumpert, Florian/Feuerhake, Jörg. *Machine Learning in Official Statistics*. 2018a. [Zugriff am 18. Juni 2024]. Verfügbar unter: [arxiv.org](https://arxiv.org)
- Beck, Martin/Dumpert, Florian/Feuerhake, Jörg. *Proof of concept machine learning – Abschlussbericht*. 2018b. [Zugriff am 18. Juni 2024]. Verfügbar unter: [www.destatis.de](https://www.destatis.de)
- Bischl, Bernd/Binder, Martin/Lang, Michel/Pielok, Tobias/Richter, Jakob/Coors, Stefan/Thomas, Janek/Ullmann, Theresa/Becker, Marc/Boulesteix, Anne-Laure/Deng, Difan/Lindauer Marius. *Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges*. In: Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. Jahrgang 13. Ausgabe 2/2023, Seite e1484 ff. DOI: [10.1002/widm.1484](https://doi.org/10.1002/widm.1484)
- Blumöhr, Torsten/Teichmann, Corina/Noack, Anke. *Standardisierung der Prozesse: 14 Jahre AG SteP*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2017, Seite 58 ff.
- Dandl, Susanne/Casalicchio, Giuseppe/Bischl, Bernd/Bothmann, Ludwig. *Interpretable Regional Descriptors: Hyperbox-Based Local Explanations*. In: Koutra, Danai/Plant, Claudia/Gomez Rodriguez, Manuel/Baralis, Elena/Bonchi, Francesco (Herausgeber). *Machine Learning and Knowledge Discovery in Databases: Research Track. ECML PKDD 2023*. Cham 2023, Seite 479 ff. DOI: [10.1007/978-3-031-43418-1\\_29](https://doi.org/10.1007/978-3-031-43418-1_29)
- Denkfabrik Digitale Arbeitsgesellschaft im Bundesministerium für Arbeit und Soziales. *Selbstverpflichtende Leitlinien für den KI-Einsatz in der behördlichen Praxis der Arbeits- und Sozialverwaltung*. 2022. [Zugriff am 18. Juni 2024]. Verfügbar unter: [www.bmas.de](https://www.bmas.de)
- Deutscher Ethikrat. *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz*. 2023. [Zugriff am 18. Juni 2024]. Verfügbar unter: [www.ethikrat.org](https://www.ethikrat.org)
- Dumpert, Florian. *Machine Learning in der amtlichen Statistik – Ergebnisse und Bewertung eines internationalen Projekts*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2021, Seite 53 ff.
- Dumpert, Florian/Beck, Martin. *Einsatz von Machine-Learning-Verfahren in amtlichen Unternehmensstatistiken*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Jahrgang 11. Ausgabe 2/2017, Seite 83 ff. DOI: [10.1007/s11943-017-0208-6](https://doi.org/10.1007/s11943-017-0208-6)
- Dumpert, Florian/Beck, Martin. *Verbesserung der Datengrundlage der Mindestlohnforschung mittels maschineller Lernverfahren*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Jahrgang 17. Ausgabe 1/2023, Seite 5 ff. DOI: [10.1007/s11943-023-00318-w](https://doi.org/10.1007/s11943-023-00318-w)

## LITERATURVERZEICHNIS

---

Dumpert, Florian/von Eschwege, Katja/Beck, Martin. [\*Einsatz von Support Vector Machines bei der Sektorzuordnung von Unternehmen\*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 1/2016, Seite 87 ff.

Dumpert, Florian/Wichert, Sebastian/Augustin, Thomas/Storfinger Nina. *Editorial Issue 3+4/2023*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Jahrgang 17. Ausgabe 3–4/2023, Seite 191 ff. DOI: [10.1007/s11943-023-00334-w](https://doi.org/10.1007/s11943-023-00334-w)

Feuerhake, Jörg/Dumpert, Florian. [\*Erkennung nicht relevanter Unternehmen in den Handwerksstatistiken\*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 2/2016, Seite 79 ff.

Hornung, Roman/Nalenz, Malte/Schneider, Lennart/Bender, Andreas/Bothmann, Ludwig/Bischl, Bernd/Augustin, Thomas/Boulesteix, Anne-Laure. *Evaluating machine learning models in non-standard settings: An overview and new findings*. 2023. [Zugriff am 19. Juni 2024]. Verfügbar unter: [arxiv.org/abs/2310.15108](https://arxiv.org/abs/2310.15108)

Karanka, Joni. *Facilitators and Blockers of ML Adoption in Official Statistics*. 2023. [Zugriff am 19. Juni 2024]. Verfügbar unter: [unece.org](https://unece.org)

Kommission Zukunft Statistik. *Bericht und Empfehlungen der Kommission Zukunft Statistik. Version 1.0*. 2024. [Zugriff am 19. Juni 2024]. Verfügbar unter: [www.destatis.de](https://www.destatis.de)

Levagin, Bogdan/Lange, Kerstin/Walprecht, Sylvana/Gerls, Fabian/Kühnhenrich, Daniel. [\*Vereinfachtes Verfahren zur interaktiven Schätzung des Erfüllungsaufwands mittels maschinellen Lernens\*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2022, Seite 53 ff.

Limberg, Heiko. [\*Potenziale von Clustering-Algorithmen für die Plausibilisierung im Außenhandel\*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 1/2024, Seite 54 ff.

Moritz, Steffen/Dumpert, Florian/Jung, Christian/Bartz-Beielstein, Thomas/Bartz, Eva. *Practical Applications of Online Machine Learning*. In: Bartz, Eva/Bartz-Beielstein, Thomas (Herausgeber). *Online Machine Learning*. Singapur 2024. Seite 71 ff. DOI: [10.1007/978-981-99-7007-0\\_7](https://doi.org/10.1007/978-981-99-7007-0_7)

Moritz, Steffen/Wynck, Frederik/Wiebels, Johannes. [\*Prognose der Abgabequote von Einkommensteuererklärungen bei Rentnerinnen und Rentnern\*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 2/2024, Seite 83 ff.

Nalenz, Malte/Rodemann, Julian/Augustin, Thomas. *Learning de-biased regression trees and forests from complex samples*. In: *Machine Learning*. Ausgabe 113. 2024, Seite 3379 ff. DOI: [10.1007/s10994-023-06439-1](https://doi.org/10.1007/s10994-023-06439-1)

Saidani, Younes/Dumpert, Florian/Borgs, Christian/Brand, Alexander/Nickl, Andreas/Rittmann, Alexandra/Rohde, Johannes/Salwiczek, Christian/Storfinger, Nina/Straub, Selina. *Qualitätsdimensionen maschinellen Lernens in der amtlichen Statistik*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Jahrgang 17. Ausgabe 3–4/2023, Seite 253 ff. DOI: [10.1007/s11943-023-00329-7](https://doi.org/10.1007/s11943-023-00329-7)

### LITERATURVERZEICHNIS

---

Schenk, Patrick Oliver/Kern, Christoph. *Connecting Algorithmic Fairness to Quality Dimensions in Machine Learning in Official Statistics and Survey Production*. 2024. [Zugriff am 19. Juni 2024]. Verfügbar unter: <https://arxiv.org/abs/2402.09328>

Schmidt, Elena. *Korrektur des Tätigkeitsschlüssels der Bundesagentur für Arbeit mit Hilfe maschineller Lernverfahren*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2020, Seite 37 ff.

Schnell, Rainer. *Expertise Maschinelles Lernen für Record Linkage – Endbericht*. 2021. Internes Dokument.

Statistische Ämter des Bundes und der Länder. *Qualitätshandbuch der Statistischen Ämter des Bundes und der Länder (Version 1.21)*. 2021. [Zugriff am 19. Juni 2024]. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

Statistisches Bundesamt. *Digitale Agenda des Statistischen Bundesamtes*. 2019. [Zugriff am 8. Juli 2024]. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

Statistisches Bundesamt. *WISTA Wirtschaft und Statistik. Ausgabe zu 75 Jahre Statistisches Bundesamt*. 2023a. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

Statistisches Bundesamt. *European Big Data Hackathon 2023: Erfolge für Teams des Statistischen Bundesamtes*. 2023b. In: WISTA Wirtschaft und Statistik. Ausgabe 2/2023. Kurznachrichten, Seite 14.

Stock, Joshua/Hauke, Oliver/Weißmann, Julius/Federrath Hannes. *The Applicability of Federated Learning to Official Statistics*. In: Quaresma, Paulo/Camacho, David/Yin, Hujun/Gonçalves, Teresa/Julian, Vicente, Tallón-Ballesteros, Antonio J. (Herausgeber). *Intelligent Data Engineering and Automated Learning – IDEAL 2023*. Cham 2023. Seite 70 ff. DOI: [10.1007/978-3-031-48232-8\\_8](https://doi.org/10.1007/978-3-031-48232-8_8)

United Nations (Vereinte Nationen). *Machine Learning for Official Statistics*. 2021. [Zugriff am 18. Juni 2024]. Verfügbar unter: [unece.org](http://unece.org)

Weißmann, Julius/Herbst, Tim. *Maschinelles Lernen im Basisregister für Unternehmen*. In: WISTA Wirtschaft und Statistik. Ausgabe 3/2024, Seite 67 ff.

## RECHTSGRUNDLAGEN

---

Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I Seite 2394), das zuletzt durch Artikel 14 des Gesetzes vom 8. Mai 2024 (BGBl. I Nr. 152) geändert worden ist.

Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828.

# ERFASSUNG VON ERNTEERTRÄGEN MIT SATELLITENBILDERN UND MASCHINELLEM LERNEN – DAS PROJEKT FernEE 2.0

Oliver Reitz

➤ **Schlüsselwörter:** Erntestatistik – Fernerkundung – Künstliche Intelligenz (KI) – Ackerbau – Sentinel-2

## ZUSAMMENFASSUNG

Die Ernte- und Betriebsberichterstattung stützt sich auf Schätzungen erfahrener Landwirtinnen und Landwirte. Es wird jedoch immer schwieriger, diese zu gewinnen und damit die Basis für eine hochwertige Erntestatistik für Feldfrüchte bereitzustellen. Um diese Basis zu verbreitern, nutzt ein Pilotprojekt des Hessischen Statistischen Landesamtes Satellitendaten und Maschinelles Lernen und modelliert damit Ernteerträge für vier wichtige Kulturen auf kleinräumiger Ebene. Von dort aus können die Erträge auf beliebige Verwaltungsebenen aggregiert werden. Dieses weitgehend automatisierte Verfahren wurde in den Jahren 2022 und 2023 in sieben Bundesländern getestet. Es wird ab 2024 bundesweit eingeführt, damit es künftig in die statistische Produktion integriert werden kann.

➤ **Keywords:** *crop production statistics – remote sensing – artificial intelligence (AI) – farming – Sentinel-2*

## ABSTRACT

*Crop production reporting is based on estimates of experienced farmers. It is becoming increasingly difficult, however, to recruit such experts and provide the basis for generating high-quality crop production statistics. To broaden this basis, a pilot project by the Land Statistical Office of Hessen uses satellite data and machine learning to model crop yields for four key crops at a small area scale. From there, the yields can be aggregated to any administrative level. This largely automated procedure was tested in seven German Länder in 2022 and 2023. Starting in 2024, it will be rolled out across Germany for future integration into statistical production.*

### Dr. Oliver Reitz

ist Geograph und Referent im Referat „Competence Center Geoinformation und SAS, Auswertungsunterstützung“ des Hessischen Statistischen Landesamtes. Seine Aufgabenschwerpunkte liegen in den Bereichen Fernerkundung, Geoinformation und Maschinelles Lernen.



## 1

### Einleitung

Unabhängige, flächendeckende und räumlich differenzierte Erntestatistiken sind eine wichtige Informationsbasis für Politik, Wirtschaft und die Öffentlichkeit. Sie bilden eine wesentliche Grundlage, um Versorgungsbilanzen zu erstellen und die Marktsituation auf nationaler und europäischer Ebene zu beurteilen. In der Erntestatistik basiert die Ermittlung von Ernteerträgen für Feldfrüchte auf zwei Erhebungen, der Ernte- und Betriebsberichterstattung (EBE) sowie der Besonderen Ernte- und Qualitätsermittlung (BEE). Die Besondere Ernte- und Qualitätsermittlung ist ein vergleichsweise aufwendiges Stichprobenverfahren zur Ermittlung von gewogenen Hektarerträgen auf Basis einer Zufallsauswahl an Feldern für die Hauptgetreidesorten, Winterraps und Kartoffeln. Die Ernte- und Betriebsberichterstattung basiert auf einer Vorausschätzung freiwillig meldender Berichterstatte(r)innen und Berichterstatte(r). (Breitenfeld, 2016). Aus methodischer Sicht betrachtet hängt die Güte der Schätzungen von den Erfahrungen und subjektiven Eindrücken der Meldenden ab. Sie ist damit mit mehr Unsicherheiten behaftet als das objektive Messverfahren der Besonderen Ernte- und Qualitätsermittlung. Beide Erhebungen liefern in Verbindung mit der Bodennutzungshaupterhebung zu einem frühen Zeitpunkt objektive und repräsentative Angaben über Menge und Qualität der Ernte ausgewählter wichtiger Fruchtarten für das gesamte Bundesgebiet und für die Länder.

Die BEE-Ergebnisse für die oben genannten Feldfrüchte werden ausschließlich auf Bundes- und Landesebene ausgewiesen; um die Erträge weiter zu regionalisieren, sind die Ergebnisse der Ernte- und Betriebsberichterstattung notwendig. Deren Verfahren ermöglicht eine Regionalisierung auf Kreisebene, jedoch aus Gründen der Geheimhaltung keine tiefere Unterteilung auf Gemeinde- oder Gitterzellenebene. Allerdings wurden selbst auf der Kreisebene in den letzten Jahren vermehrt fehlende Werte in der Erntestatistik verzeichnet. In Hessen betraf dies beispielsweise im Berichtsjahr 2023 durchschnittlich 38% aller Landkreise mit entsprechenden Anbauflächen (Hessisches Statistisches Landesamt, 2024). Ursache dafür ist, dass es zunehmend schwieriger wird, fachlich versierte Berichterstatte(r)nde zu gewinnen (Statistisches Bundesamt, 2024). In Hin-

blick auf deren Altersstruktur ist zudem davon auszugehen, dass sich dieses Problem weiter verschärfen wird.

Vor diesem Hintergrund ist die Nutzung von Satellitendaten in Kombination mit maschinellen Lernverfahren (ML) ein vielversprechender Ansatz, um eine flächendeckende Regionalisierung nicht nur zu gewährleisten, sondern Erntestatistiken auch auf noch kleinräumigeren Ebenen als bisher anbieten zu können. Dieses objektive Verfahren kann somit die subjektiven Schätzungen der Berichterstatte(r)nden ergänzen. Die Europäische Weltraumorganisation stellt durch das Copernicus-Erdbeobachtungsprogramm der Allgemeinheit kostenfrei und regelmäßig hochaufgelöste Satellitenbilder zur Verfügung. Dadurch bietet sich die Möglichkeit, den phänologischen Zustand<sup>1</sup> und die Entwicklung von Feldfrüchten effizient, flächendeckend und automatisiert zu erfassen, um so eine Datenbasis für maschinelle Lernverfahren zu schaffen. Eine Alternative zu solchen statistischen Modellen stellen prozessbasierte Ertragsmodelle dar (Arnold und andere, 2021). Diese sind jedoch von vordefinierten Annahmen der zugrunde liegenden komplexen pflanzenphysiologischen Prozesse abhängig (Leng/Hall, 2020). Daher sind sie im Vergleich zu maschinellen Lernverfahren tendenziell weniger flexibel und bedürfen in der Regel eines größeren Entwicklungsaufwandes.

Das Hessische Statistische Landesamt führte im Projekt FernEE (Fernerkundliche Erfassung von Ernteerträgen) ab 2018 Ertragsmodellierungen für vier wichtige Feldfrüchte – nämlich Wintergerste, Winterraps, Winterroggen und Winterweizen – für das Land Hessen durch (Ghosh und andere, 2022). Die Modellierung erfolgte retrospektiv, nachdem die Ernte eingefahren war. Hierfür wurden In-situ-Ertragsmessungen der Besonderen Ernte- und Qualitätsermittlung mit Satellitendaten und Schlaggeometrien aus dem Integrierten Verwaltungs- und Kontrollsystem (InVeKoS)<sup>2</sup> kombiniert, um statistische Ertragsmodelle zu trainieren. Ein Schlag ist hierbei eine zusammenhängende Fläche, die mit einer einheitlichen Kultur und von einem Betrieb bewirtschaftet wird.

1 Der phänologische Zustand bezieht sich auf äußerlich erkennbare Phasen im Jahreszyklus einer Pflanzenart.

2 System zur Verwaltung und Überwachung der Leistungen der Gemeinsamen Agrarpolitik der Europäischen Union. Antragstellende geben dort ihre landwirtschaftlichen Parzellen und angebauten Feldfrüchte an.



Ziel der in diesem Artikel vorgestellten nächsten Projektphase FernEE 2.0 ist, den vielversprechenden Ansatz für die Berichtsjahre 2022 und 2023 weiterzuentwickeln. Hierzu zählen vornehmlich die Ausweitung auf sieben am Projekt teilnehmende Länder, methodische Verbesserungen sowie die fortschreitende Automatisierung der Prozesskette. Die Ergebnisse sollen darüber hinaus als Entscheidungsgrundlage dienen, um eine Implementierung des fernerkundlichen Ansatzes in die operative Ertragsstatistik weiter voranzutreiben.

Kapitel 2 erläutert die verwendeten Daten und die Methodik des Verfahrens. Es folgt eine Darstellung der Validierungsergebnisse mit den gemessenen Erträgen auf Schlagebene sowie der Vergleich der amtlichen Statistik auf Kreisebene in Kapitel 3. In Kapitel 4 werden mögliche Fehlerquellen diskutiert und Kapitel 5 gibt einen Ausblick über anstehende Entwicklungen.

## 2

### Daten und Methoden

Die Methodik des Projektes basiert auf der Verknüpfung mehrerer unterschiedlicher Datenquellen durch eine Vielzahl von Prozessschritten. Eine Übersicht hierzu bietet [Grafik 1](#) auf Seite 32.

Die erste Säule der verwendeten Daten bilden georeferenzierte BEE-Volldrusch-Ertragsmessungen, die punktuell vorliegen und die Lernreferenz der statistischen Modelle darstellen. Ein Volldrusch bedeutet hierbei, dass ein gesamter Schlag und nicht bloß eine Probe gerodet wird. Die Daten haben die Statistischen Ämter der teilnehmenden Länder<sup>3</sup> für die vier oben genannten Winterfeldfrüchte bereitgestellt.

Die zweite Säule sind Daten aus InVeKoS. Diese Daten umfassen für jeden enthaltenen Schlag eines Landes neben der Lage und Geometrie auch die angebaute Feldfrucht.

Die dritte Säule bilden flächendeckende Raster aus fernerkundlichen, meteorologischen und bodenkundlichen Daten, die die erklärenden Variablen für die Modelle darstellen:

- › Aus dem Copernicus-Dataspace wurden Bilder der optischen Sentinel-2-Satelliten heruntergeladen. Diese bieten in verschiedenen Bändern spektrale Informationen vom sichtbaren Licht bis in den Infrarotbereich mit einer räumlichen Auflösung von bis zu zehn Metern und liefern von jedem Ort Deutschlands mindestens alle fünf Tage eine Aufnahme.
- › Monatliche meteorologische Raster für die Monate April bis Juli hat der Deutsche Wetterdienst bereitgestellt; sie umfassen solare Einstrahlung, Bodenfeuchte, Niederschlag und Temperatur.
- › Von der Webseite der Bundesanstalt für Geowissenschaften und Rohstoffe wurden zudem bundesweite Raster der nutzbaren Feldkapazität und der Luftkapazität heruntergeladen.
- › Weiterhin gingen die Koordinaten sowie die Geländehöhe der Schläge in die Modelle ein, die aus dem digitalen Geländemodell (DGM200) des Bundesamtes für Kartographie und Geodäsie abgeleitet wurde.

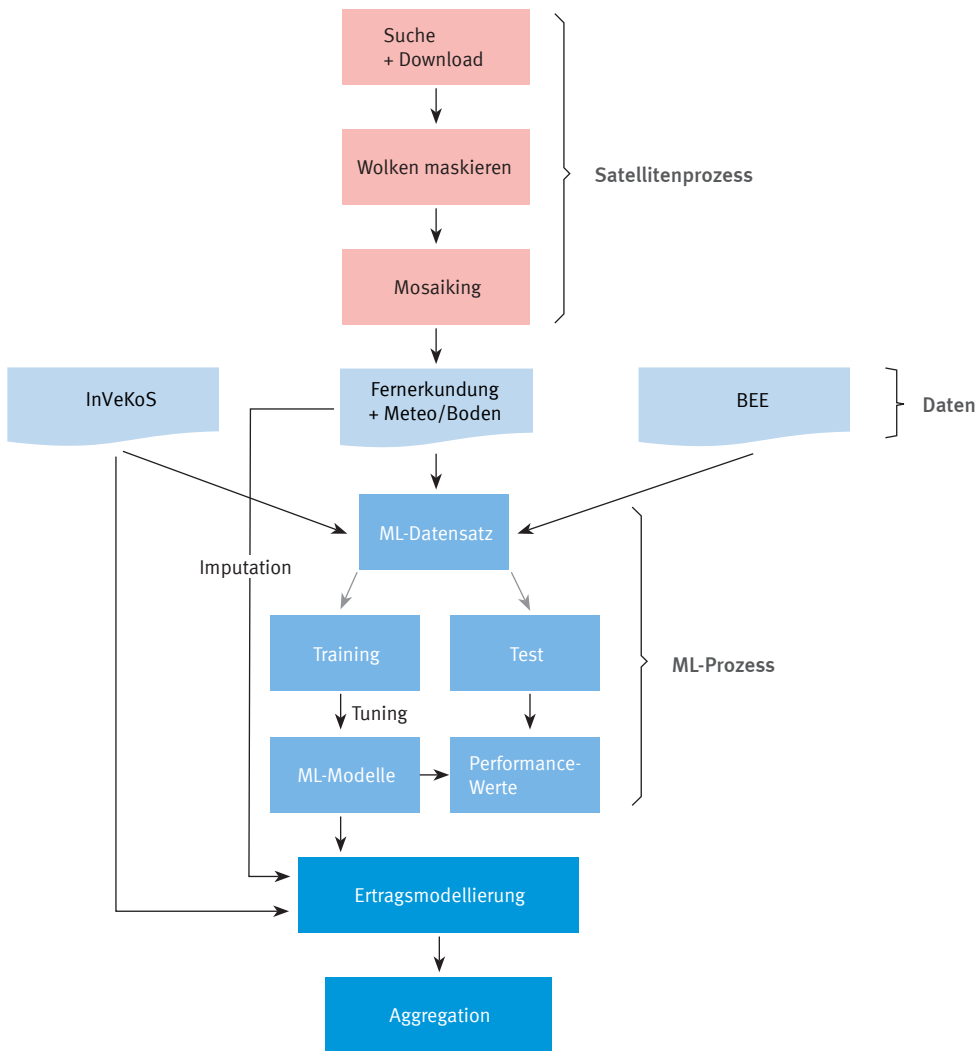
Bei der optischen Erdbeobachtung sind Wolken eine große Herausforderung. Um den Informationsverlust durch Wolkenbedeckung zu minimieren, wurden möglichst wolkenfreie Sentinel-2-Szenen zunächst innerhalb eines großen Zeitraums zwischen dem 1. Mai und dem 10. Juli der Jahre 2022 und 2023 gesucht. Dieser Zeitraum wurde gewählt, da er für Winterfrüchte in der Regel die entscheidende Wachstumsphase vor der Ernte umfasst. Innerhalb dieses großen Zeitraums wurden daraufhin zwei kleinere, zweiwöchige Zeiträume identifiziert, die möglichst viele der wolkenfreien Bilder enthalten.

Für diese zwei Zeiträume wurden danach die Bilder für die gesamte Fläche automatisiert heruntergeladen, verbleibende Wolken maskiert, das heißt Wolken durch fehlende Werte ersetzt, und die Bilder zu zwei Mosaiken je Land zusammengefügt. Die Mosaiken umfassen somit möglichst wolkenfreie Bilder von verschiedenen Tagen, die jedoch höchstens 14 Tage auseinanderliegen. Diese Begrenzung soll verhindern, dass die Sentinel-2-Daten allein aufgrund ihres Aufnahmedatums allzu unterschiedliche phänologische Zustände im gleichen Mosaik enthalten. Von jeder Sentinel-Szene wurden neun verschiedene Bilder verschiedener spektraler Kanäle genutzt, die Informationen vom sichtbaren

<sup>3</sup> Berlin-Brandenburg, Baden-Württemberg, Bayern, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Saarland.

**Grafik 1**

Vereinfachte Darstellung der Prozesskette im Projekt FernEE



InVeKoS: Integriertes Verwaltungs- und Kontrollsystem; BEE: Besondere Ernte- und Qualitätsermittlung; ML: Maschinelles Lernen.

Licht bis Infrarot abbilden. Aus diesen fernerkundlichen Daten wurden schließlich sieben verschiedene spektrale Vegetations- und Feuchteindizes<sup>4</sup> berechnet.

Um die Tabellen zum Training der maschinellen Lernverfahren zu erstellen, wurden die punktuellen BEE-Daten mit den flächigen InVeKoS-Daten zunächst räumlich verknüpft. Somit war es möglich, diejenigen Schlaggeome-

trien zu identifizieren, für die eine Ertragsmessung vorlag. Für jeden dieser Schläge wurde daraufhin der Mittelwert aller Pixel, die innerhalb seiner Grenzen lagen, je Raster berechnet. Um den Einfluss von Randbereichen der Felder zu minimieren, wurde eine zehn Meter breite Pufferzone am Rand der Schläge ausgeschlossen.

In der Tabelle sind damit jedem Schlag bodenkundliche, fernerkundliche und meteorologische Werte zugeordnet und zudem die Information, aus welchem Bundesland die Ertragsmessung kam. Im Modelltraining wurden ausschließlich Ertragsmessungen mit vollständigen, das heißt wolkenfreien, erklärenden Variablen genutzt.

<sup>4</sup> Normalized Difference Vegetation Index, Enhanced Vegetation Index, Normalized Difference Moisture Index, Normalized Difference Red Edge Index, Anthocyanin Reflectance Index, Chlorophyll Red-Edge, Modified Chlorophyll Absorption in Reflectance Index. Siehe [custom-scripts.sentinel-hub.com](https://custom-scripts.sentinel-hub.com)

Für jede der vier Feldfrüchte wurden eigene ML-Ertragsmodelle mithilfe des Caret-Toolkits in der Programmiersprache R trainiert (Kuhn, 2008). Dies geschah jeweils mit Daten aller sieben Länder, um eine möglichst große Trainingsbasis zu ermöglichen. Je Feldfrucht wurden verschiedene ML-Algorithmen<sup>5</sup> mit den gleichen Daten trainiert und zu einem Ensemble zusammengefasst. Dadurch können die jeweiligen Stärken der Algorithmen zur Geltung kommen und kann eine insgesamt robustere Modellierung durchgeführt werden. Die Modellparameter der einzelnen Algorithmen wurden mit einer internen, fünffachen Kreuzvalidierung optimiert (Tuning). Eine externe, ebenfalls fünffache Kreuzvalidierung wurde zur Evaluation der Modellgüte herangezogen. Hierbei wurden jeweils 20 % der Daten als Testdaten vom gesamten Trainingsprozess ausgeschlossen und am Ende von den finalen Modellen vorhergesagt. Dieser Prozess wurde daraufhin fünfmal wiederholt, bis alle Daten einmal vom Training ausgeschlossen wurden. Diese Validierung simuliert den tatsächlichen Anwendungsfall auf unbekannte Daten, wobei die Qualität der Modellierung im Vergleich zu den tatsächlich gemessenen Erträgen evaluiert werden kann.

Die Modellierung für das Berichtsjahr 2022 fand nur mit Daten aus dem Jahr 2022 statt. Für das Berichtsjahr 2023 wurden hingegen zwei Modellierungen verglichen. Erstens nur mit Daten aus 2023 (Single) und zweitens mit Trainingsdaten aus 2022 und 2023 (Multi), wobei auch hier die Testdaten immer nur aus 2023 kamen, um zwischen Single und Multi vergleichbare Ergebnisse zu produzieren. Mit dem Multi-Modell soll hierbei untersucht werden, ob durch die Hinzunahme eines zurückliegenden Jahres der positive Effekt einer größeren Trainingsbasis den potenziell nachteiligen Effekt einer schlechten Übertragbarkeit auf das aktuelle Jahr überwiegt.

Für die tatsächliche Anwendung der Modelle wurden zunächst alle erklärenden Variablen für alle InVeKoS-Schläge extrahiert. Aufgrund von Wolkenbedeckung fehlende Werte in den fernerkundlichen Variablen wurden daraufhin mit einem Random-Forest-Modell je Zeitraum aus den lückenlosen meteorologischen und bodenkundlichen Variablen aufgefüllt. Dann wurden die trainierten Ertragsmodelle auf diese Daten angewendet, sodass für jeden Schlag eine Ertragsmodellierung vorlag. Es war

nun möglich, diese Hektarerträge auf die Kreisebene und Landesebene zu aggregieren, indem sie nach Größe der Schläge gewichtet wurden. Schläge auf einer Grenze zwischen zwei Landkreisen wurden hierbei dem Kreis mit der größeren Überlappung zugeordnet.

## 3

### Ergebnisse

Die Ergebnisse der externen Kreuzvalidierung in [Tabelle 1](#) sind ein Indikator für die Qualität des Modells, auf bisher ungesehene Daten und Situationen zu reagieren. Ein niedriger sogenannter Generalisierungsfehler stellt sicher, dass die Modelle nicht nur die Trainingsdaten auswendig lernen, sondern den Ernteertrag auch in der Anwendung auf neue Schläge akkurat vorhersagen. Als Maß zur Evaluation der Modelle wurde der mittlere absolute Fehler (MAE) gewählt, da er direkt angibt, wie weit die Modellierungen im Mittel von den tatsächlichen Werten entfernt sind. Um die Fehler verschiedener Jahre und verschiedener Feldfrüchte miteinander vergleichen zu können, wurde zudem ein relati-

**Tabelle 1**

Ergebnisse der externen Kreuzvalidierung der Jahre 2022 und 2023<sup>1</sup>

Feldfrucht	Training	MAE (mittlerer absoluter Fehler)	nMAE (normalisierter [relativer] Fehler)
		dt/ha	%
2022			
Wintergerste	Single	9,66	12,80
Winterraps	Single	4,78	11,48
Winterroggen	Single	9,72	17,37
Winterweizen	Single	8,69	11,23
2023			
Wintergerste	Single	9,11	13,08
	Multi	9,05	12,99
Winterraps	Single	4,45	12,44
	Multi	4,40	12,31
Winterroggen	Single	9,45	20,31
	Multi	9,23	19,84
Winterweizen	Single	9,65	13,57
	Multi	9,43	13,26

<sup>5</sup> Neuronale Netze, Gradient Boosting, Random Forest, Gauß-Prozess, Support Vector Regression, Partial Least Square Regression.

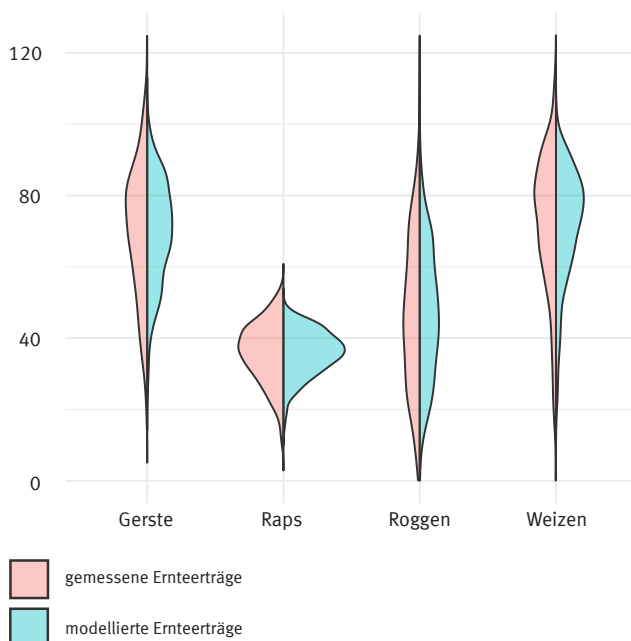
<sup>1</sup> Für 2023 mit Trainingsdaten nur aus dem Bezugsjahr 2023 (Single) und mit Trainingsdaten aus 2022 und 2023 (Multi).

ver Fehler berechnet. Für diesen normalisierten Fehler (nMAE) wurde der MAE durch den mittleren Ertrag aller gemessenen Erträge geteilt und mit 100 multipliziert, um den prozentualen Anteil des Fehlers am Ertrag anzugeben.

Sowohl im Jahr 2022 als auch im Jahr 2023 weisen die Modelle für Wintergerste, Winterraps und Winterweizen ähnliche relative Fehler auf (zwischen 11,23 und 13,57%), für Winterroggen waren diese jedoch in beiden Jahren deutlich höher (17,37 bis 20,31%). Im Vergleich der beiden Jahre haben die Modellierungen für 2022 zudem leicht niedrigere Fehlerwerte als für 2023. Eine vorangegangene satellitengestützte Ertragsschätzung der Berichtsjahre 2018 und 2019 wies zum Vergleich höhere relative Abweichungen zwischen 14 und 18 % für Winterweizen und 42 und 43 % für Winterraps auf (Arnold und andere, 2021). Weiterhin wurde für 2023 untersucht, inwiefern die Hinzunahme von Trainingsdaten eines vergangenen Jahres Auswirkungen auf die Modellqualität hat. Für alle vier Feldfrüchte wurde durch das Multi-Modell der Fehler leicht verringert (siehe Tabelle 1).

## Grafik 2

Verteilungsdichte der gemessenen und modellierten Ernteerträge für vier Winterfeldfrüchte im Jahr 2023  
Ertrag in Dezitonnen je Hektar



Anmerkung: Je breiter die Form, desto mehr Daten gibt es in diesem Wertebereich.

Darüber hinaus ist es von Interesse, welche Wertebereiche maßgeblich zu diesen Fehlern beigetragen haben. **Grafik 2** zeigt anhand der gestauchten Form der rechten Hälften, dass die modellierten Erträge eine geringere Spannweite als die tatsächlichen Erträge (linke Hälften) aufweisen. Für Ausreißer mit einem hohen Ertrag wurde also tendenziell ein zu niedriger Ertrag modelliert und für niedrige Ausreißer ein zu hoher, sodass es zu einer Stauchung um die ohnehin häufigen Werte kommt. Im Mittel und im Median wurden die Erträge durch die Modellierung hingegen sehr gut getroffen. Zu beachten ist zudem, dass sich diese Fehler bei Ausreißern auf der Schlagebene durch Aggregation auf die Kreisebene teilweise aufheben, sodass sich dieses Modellverhalten nicht eins zu eins auf höhere Ebenen überträgt.

**Tabelle 2** vergleicht zudem die finalen auf Kreisebene aggregierten Erträge mit den Kreiszahlen der amtlichen Statistik, sofern diese im Rahmen der Geheimhaltung verfügbar sind. Im Gegensatz zu den Fehlern in Tabelle 1 sind die Abweichungen hier jedoch keine Referenz zu den wahren Werten, sondern ein Vergleich mit einer anderen Schätzung. Sie gelten aufgrund einer unterschiedlichen Methodik eher als grober Richtwert, wie in Kapitel 4 näher ausgeführt wird. Generell sind die Abweichungen auf Kreisebene geringer als zu den gemessenen Erträgen auf Schlagebene. Auf Kreisebene sind die Abweichungen bei Roggen ebenfalls höher als bei den übrigen drei Feldfrüchten, für die zwischen den beiden Jahren keine deutlichen Unterschiede zu erkennen sind.

## Tabelle 2

Abweichung der auf Kreisebene aggregierten Modellerträge von den EBE-Kreiszahl der Regionaldatenbank<sup>1</sup>

Feldfrucht	Training	MAE (mittlerer absoluter Fehler)	nMAE (normalisierter [relativer] Fehler)
		dt/ha	%
2022			
Wintergerste	Single	5,19	7,46
Winterraps	Single	2,77	6,80
Winterroggen	Single	7,48	13,04
Winterweizen	Single	5,62	8,00
2023			
Wintergerste	Multi	4,88	7,42
Winterraps	Multi	2,68	7,27
Winterroggen	Multi	5,01	10,78
Winterweizen	Multi	5,15	8,02

<sup>1</sup> Die dargestellten Werte sind Mittelwerte über alle Kreise, für die entsprechende Statistiken vorhanden waren.

## 4

### Diskussion

Eine Herausforderung des fernerkundlichen Ansatzes besteht in dem von Jahr zu Jahr variablen Informationsverlust durch Bewölkung. Der hier beschrittene Lösungsweg der Mosaikierung innerhalb eines begrenzten Zeitraums hat folgende Vorteile: Er stellt möglichst viele originäre, nicht interpolierte Daten den ML-Modellen zur Verfügung und geht durch die vergleichsweise kleine Datenmenge sparsam mit begrenzten Speicherplatz- und Rechenkapazitäten um. Dieser Ansatz ist jedoch immer ein Kompromiss zwischen Informationsverlust durch Wolken und der Inklusion unterschiedlicher phänologischer Zustände.

Eine Alternative wäre, alle Bilder innerhalb eines langen Zeitraums, etwa von April bis Juli, unabhängig von ihrer Wolkenbedeckung zu verwenden und die vielen fehlenden Werte durch Wolken zeitlich zu interpolieren. Damit könnte zwar die phänologische Entwicklung der Feldfrüchte umfassender abgebildet werden, aber die Modelle würden auch mit deutlich mehr interpolierten Daten trainiert werden. Zudem würde dieser Ansatz deutlich mehr Speicherplatz- und Rechenkapazitäten beanspruchen, daher hat er sich als vorerst nicht durchführbar erwiesen.

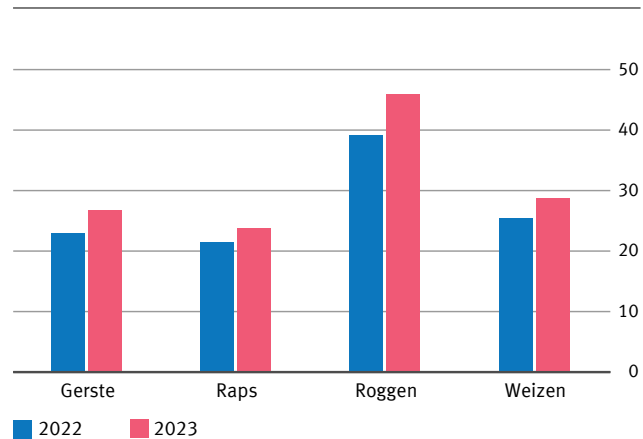
Verschiedene Gründe können zu den in Tabelle 1 dargelegten Modellabweichungen zu den tatsächlichen Ertragsmessungen der Besonderen Ernte- und Qualitätsermittlung beitragen. Die vergleichsweise geringe Zahl an BEE-Daten, die in das Modelltraining eingingen, hat wahrscheinlich dazu geführt, dass die Lernkurve der Modelle noch nicht ihr Optimum erreicht hat (Viering/Loog, 2023). Mehr Ertragsmessungen in den Trainingsdaten, vor allem des aktuellen Jahres, geben den Modellen diversere Beispiele, um die generellen statistischen Beziehungen zu den erklärenden Variablen lernen zu können und dadurch robustere Ergebnisse zu produzieren.

In der Modellierung traten hohe Fehler insbesondere bei sehr hohen und sehr niedrigen Erträgen auf. Die Spannweite der modellierten Erträge war dadurch geringer als diejenige der gemessenen Erträge. Dementsprechend gestaltet sich die Modellierung auf Schlagebene umso schwieriger, je höher die abzubildende Varianz in den

Ertragsdaten ist. Die Varianz war für Winterroggen und auch allgemein im Jahr 2023 vergleichsweise hoch, was die entsprechend höheren relativen Fehler erklären kann. [↗ Grafik 3](#)

**Grafik 3**

Variationskoeffizient je Winterfeldfrucht und Jahr in %



Anmerkung: Hohe Variationskoeffizienten sind ein Hinweis auf eine schwierigere Modellierung auf Schlagebene.

Weiterhin kann es andere wichtige erklärende Variablen geben, die bisher nicht in den Modellen berücksichtigt wurden. Zu nennen wäre hier die noch nicht flächendeckend vorliegende Information, ob ein Schlag ökologisch oder konventionell bewirtschaftet wird. In Einzelfällen kann auch der Ertrag schlecht mit den spektralen Satellitendaten korreliert sein. So sehen beispielsweise Schläge im Satellitenbild gleich oder sehr ähnlich aus, obwohl sie eigentlich einen unterschiedlichen Ertrag aufweisen. Darüber hinaus kann es nach dem letzten Satellitenbild noch zu lokalen Hagel- oder Sturmschäden an der Ernte gekommen sein. Solche Phänomene sind jedoch zu kleinräumig, um sie in den meteorologischen 1-km-Rasterdaten adäquat abbilden zu können.

Auch wenn sich die oben genannten Fehler bei der Aggregation größtenteils aufheben, können die auf Kreisebene aggregierten Ergebnisse von denen der amtlichen Statistik aus weiteren Gründen abweichen:

- › Zuerst wurde eine Vollerhebung aller verfügbaren InVeKoS-Schläge basierend auf den BEE-Volldruschmessungen modelliert. Die Schätzungen der Ernte- und Betriebsberichterstattung wurden daher in der Modellierung nicht berücksichtigt.

- › Die Aggregation auf höhere Ebenen erfolgte zudem nach Belegenheit des Schlages, die amtliche Erntestatistik folgt jedoch dem Betriebssitzprinzip. Technisch wäre eine Auswertung nach beiden Prinzipien zu realisieren, sofern Informationen zum Betriebssitz verfügbar sind.
- › Weiterhin geht aus den InVeKoS-Daten nicht hervor, ob die jeweiligen Flächen tatsächlich für den Korn-ertrag oder für die Ganzpflanzensilage genutzt wurden. Ob ein Schlag für den Korn-ertrag oder als Ganzpflanzensilage geerntet wird, ist eine individuelle Entscheidung der Landwirtinnen und Landwirte. Sie hängt neben dem erwarteten Ertrag auch von aktuellen Marktpreisen und Lagerbeständen ab und ist daher schwer in einem Modell abzubilden. Vor allem im Jahr 2023 kann dies aufgrund von verbreiteten Dauerniederschlägen zur Erntezeit zu Abweichungen von den tatsächlichen Korn-erträgen geführt haben.

Letztlich unterschied sich auch die Datenbasis der Besonderen Ernte- und Qualitätsermittlung zwischen FernEE 2.0 und der amtlichen Statistik. Aus vier verschiedenen Gründen wurde das Training der Modelle nicht mit allen theoretisch zur Verfügung stehenden BEE-Daten durchgeführt:

1. Es kamen aufgrund ihrer höheren Qualität ausschließlich Volldruschdaten zum Einsatz. In der Besonderen Ernte- und Qualitätsermittlung werden teilweise auch Probeschnitte durchgeführt.
2. Aus manchen Bundesländern wurden lediglich Daten verwendet, für deren Nutzung die Landwirtinnen und Landwirte ihr Einverständnis gegeben hatten.
3. Ertragsmessungen, deren Koordinaten nicht eindeutig einer InVeKoS-Geometrie der passenden Feldfrucht zugeordnet werden konnten (Koordinaten lagen zum Beispiel im Wald), konnten nicht weiterverwendet werden.
4. Für manche BEE-Daten waren aufgrund von Wolkenbedeckung keine entsprechenden fernerkundlichen Informationen verfügbar, weswegen sie vom Modelltraining ausgeschlossen wurden.

Trotz dieser Beschränkungen sind die Ergebnisse auf Kreisebene sehr vielversprechend. So konnten die relativen Abweichungen zu den Kreis- und Betriebsberichterstattungen im Vergleich zu den Ergebnissen in Arnold und andere (2021) zum Teil deutlich reduziert werden. Eine

exakte Übereinstimmung des fernerkundlichen Ansatzes mit den Kreis- und Betriebsberichterstattungen ist aus den oben genannten Gründen ohnehin nicht zu erwarten; vielmehr sind alle Methoden mit gewissen Unsicherheiten verbunden und weichen daher von den wahren Werten, also den tatsächlichen Erntemengen und Erträgen auf Kreisebene, ab.

---

## 5

### Fazit und Ausblick

---

Dieses Pilotprojekt bestätigt, dass der fernerkundliche Ansatz für die Erntestatistik nützlich sein kann. Die Ergebnisse sind dabei nicht nur sehr vielversprechend, sondern es wurden auch verschiedene Neuerungen implementiert, darunter die Integration von meteorologischen und bodenkundlichen Variablen sowie die Anwendung neuer ML-Algorithmen. Dies hat die Ergebnisse noch einmal deutlich verbessert.

Eine Herausforderung bestand darin, dass mit der Einbeziehung von sieben Bundesländern die Wahrscheinlichkeit einer flächendeckend geringen Bewölkung in einem gegebenen Zeitraum abnahm. Dem wurde mit einer Mosaikierung der Satellitenbilder begegnet.

Infolge dieser erfolgreichen Projektphase beschloss der Statistische Verbund<sup>16</sup> im April 2024, das Vorhaben für die Berichtsjahre 2024 bis 2026 fortzusetzen. Das Projekt wird vom Hessischen Statistischen Landesamt und dem Statistischen Bundesamt geleitet und wird drei maßgebliche Ziele umsetzen:


1. Das Verfahren zur retrospektiven Ertragserfassung wird auf die gesamte Bundesfläche ausgeweitet und weitere methodische Neuerungen werden eingeführt. Diese umfassen eine bereits in Erprobung befindliche komplexere Mosaikierung ebenso wie die Integration zusätzlicher Satellitenprodukte, wodurch der Informationsverlust durch Wolken weiter abnehmen wird.
2. Ein operativer Regelbetrieb wird erprobt und ein Verfahren erarbeitet, um den Ansatz dauerhaft in der amtlichen Statistik zu implementieren.

---

<sup>16</sup> Den Statistischen Verbund bilden die Statistischen Ämter des Bundes und der Länder.



3. Das Verfahren wird in zwei Machbarkeitsstudien auf weitere Feldfrüchte der Besonderen Ernte- und Qualitätsermittlung erweitert (Hafer, Triticale, Sommergerste) und eine frühzeitige Ertragsprognose im Sommer des laufenden Berichtsjahres mit maschinellen Lernverfahren getestet.

Die den Modellen zur Verfügung gestellte Datenbasis wird sich weiter verbessern, da es mehr Ertragsmessungen des aktuellen Berichtsjahres aus mehr Ländern wie auch insgesamt mehr Datenjahre geben wird. Dieses innovative Vorhaben zeigt somit, wie der Einsatz neuer Technologien einen wichtigen Beitrag zur Statistikproduktion im 21. Jahrhundert leisten kann. 

## LITERATURVERZEICHNIS

---

Arnold, Jasmin/Brandt, Patric/Gerighausen, Heike. *Erprobung der satellitengestützten Ertragsschätzung für die Agrarstatistik – Projekt SatAgrarStat*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2021, Seite 43 ff.

Breitenfeld, Jörg. *Wie erfolgt die statistische Ermittlung der Getreideernte?* In: Statistische Monatshefte Rheinland-Pfalz. Ausgabe 8/2016, Seite 760 ff. [Zugriff am 19. Juni 2024]. Verfügbar unter: [www.statistischebibliothek.de](http://www.statistischebibliothek.de)

Ghosh, Parmita/Mandal, Dipankar/Wilfling, Sarah/Hollberg, Jens/Bargiel, Damian/Bhattacharya, Avik. *Synergy of optical and synthetic aperture radar data for early-stage crop yield estimation: a case study over a state of Germany*. In: Geocarto International. Jahrgang 37. Ausgabe 25/2022, Seite 10743 ff. DOI: [10.1080/10106049.2022.2039306](https://doi.org/10.1080/10106049.2022.2039306)

Hessisches Statistisches Landesamt. *Statistische Berichte: Die Ernte ausgewählter Feldfrüchte in Hessen 2023*. 2024. [Zugriff am 19. Juni 2024]. Verfügbar unter: [statistik.hessen.de](http://statistik.hessen.de)

Kuhn, Max. *Building Predictive Models in R Using the caret Package*. In: Journal of Statistical Software. Band 28. Ausgabe 5/2008. DOI: [10.18637/jss.v028.i05](https://doi.org/10.18637/jss.v028.i05)

Leng, Guoyong/Hall, Jim W. *Predicting spatial and temporal variability in crop yields: an inter-comparison of machine learning, regression and process-based models*. In: Environmental Research Letters. Band 15. Ausgabe 4/2020. DOI: [10.1088/1748-9326/ab7b24](https://doi.org/10.1088/1748-9326/ab7b24)

Statistisches Bundesamt. *Qualitätsbericht – Ernte- und Betriebsberichterstattung (EBE): Feldfrüchte und Grünland*. 2024. [Zugriff am 28. Juni 2024]. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

Viering, Tom/Loog, Marco. *The Shape of Learning Curves: A Review*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Band 45. Ausgabe 6/2023, Seite 7799 ff. DOI: [10.1109/TPAMI.2022.3220744](https://doi.org/10.1109/TPAMI.2022.3220744)



# HAUSHALTE, FAMILIEN UND LEBENSFORMEN MIT EINWANDERUNGSGESCHICHTE – AUSWIRKUNGEN AUF DIE ERWERBSBETEILIGUNG VON ELTERN

Matthias Keller, Thomas Körner

➤ **Schlüsselwörter:** Gleichstellung – Diskriminierung – Arbeitsmarktbeteiligung von Müttern und Vätern – Mikrozensus – Arbeitsteilung

## ZUSAMMENFASSUNG

Das Statistische Bundesamt hat seit dem Jahr 2023 im Mikrozensus das Konzept der Einwanderungsgeschichte eingeführt und setzt es nun auch auf der Ebene von Haushalten und Lebensformen um. Der Beitrag stellt zunächst Strukturmerkmale von Haushalten und Lebensformen mit und ohne Einwanderungsgeschichte vergleichend dar. Dabei zeigt sich, dass Haushalte mit Einwanderungsgeschichte größer sind und Familien mit Einwanderungsgeschichte mehr Kinder haben. Davon ausgehend wird die Erwerbstätigkeit von Männern und Frauen sowie Vätern und Müttern mit und ohne Einwanderungsgeschichte betrachtet, ebenso die Arbeitsteilung von Paaren mit und ohne Einwanderungsgeschichte in Bezug auf bezahlte Arbeit.

➤ **Keywords:** *gender equality – discrimination – labour market participation of mothers and fathers – microcensus – division of work*

## ABSTRACT

*In 2023, the Federal Statistical Office introduced the concept of “people with immigration history” in the microcensus and is now also applying this approach at the level of households and living arrangements. This article first presents and compares structural characteristics of households and living arrangements in the context of immigration history (i.e. persons with and without an immigration history). The comparison reveals that households with an immigration history tend to be bigger and families with an immigration history are likely to have more children. On this basis, we analyse the employment rate of men and women and of fathers and mothers who have or do not have an immigration history. In addition, we examine the division of (paid) work between couples with and without an immigration history.*



**Matthias Keller**

ist Diplom-Volkswirt (FH) und im Referat „Bevölkerungsstatistische Auswertungen und Analysen aus dem Mikrozensus“ des Statistischen Bundesamtes tätig. Schwerpunkte seiner Arbeit sind Analysen und Sonderauswertungen sowie Publikationen zu privaten Haushalten und Lebensformen.



**Thomas Körner**

ist Soziologe und leitet im Statistischen Bundesamt das Referat „Bevölkerungsstatistische Auswertungen und Analysen aus dem Mikrozensus“. Seine Arbeitsschwerpunkte sind Analysen und Publikationen zu den Bereichen Haushalte und Lebensformen, Migration und Integration sowie Gleichstellung.

## 1

### Einleitung

Das seit 2005 im Mikrozensus verwendete Konzept „Migrationshintergrund“ stand in der Öffentlichkeit und auch bei der Wissenschaft seit einigen Jahren immer wieder in der Kritik. Vor diesem Hintergrund hat das Statistische Bundesamt die Empfehlungen der Fachkommission der Bundesregierung zu den Rahmenbedingungen der Integrationsfähigkeit (2021) aufgegriffen und im Jahr 2023 das Konzept der Eingewanderten und ihrer (direkten) Nachkommen (Bevölkerung nach Einwanderungsgeschichte) umgesetzt (Canan/Petschel, 2023). Personen mit Einwanderungsgeschichte sind entweder seit 1950 selbst in das heutige Gebiet Deutschlands eingewandert (Eingewanderte) oder ihre beiden Elternteile sind seit 1950 in das heutige Gebiet Deutschlands zugezogen (Nachkommen). Personen, bei denen ein Elternteil seit 1950 eingewandert ist, bilden die eigenständige Kategorie „Mit einseitiger Einwanderungsgeschichte“, zählen aber nicht zur Bevölkerung mit Einwanderungsgeschichte.

Nachdem das Konzept in einem ersten Schritt auf Personenebene umgesetzt worden war, veröffentlicht das Statistische Bundesamt seit dem Jahr 2024 auch Angaben zu Haushalten, Familien und anderen Lebensformen<sup>1</sup> mit Einwanderungsgeschichte. Die aus dem vorhandenen Bestand der Erhebungsmerkmale des Mikrozensus abgeleiteten Merkmale stellen jeweils die Einwanderungsgeschichte von Haushalten beziehungsweise von Lebensformen zusammenfassend dar.<sup>2</sup>

Betrachtet werden bei der Klassifizierung die erwachsenen Haushaltsmitglieder (Haushalte) beziehungsweise Elternteile (Familien beziehungsweise Lebensformen). Da für die soziale Lage von Haushalten und Lebensformen jeweils vor allem die Einwanderungsgeschichte der erwachsenen Personen ausschlaggebend ist, werden die im Haushalt beziehungsweise der Lebensform lebenden Kinder nicht berücksichtigt. Unterschieden werden Haushalte danach, ob alle Mitglieder bezie-

hungsweise ein Teil der (erwachsenen) Mitglieder eines Haushalts eingewandert sind, ob sie Nachkommen von Eingewanderten sind, oder ob sie eine einseitige Einwanderungsgeschichte haben. Bei den Lebensformen wird analog verfahren. Die zusammengefassten Kombinationen, die im Mikrozensus standardmäßig für Auswertungen zur Verfügung stehen, zeigt [Übersicht 1](#).

Ergänzend können für spezifische Fragestellungen Merkmale mit allen möglichen Kombinationen des Einwanderungsstatus bei den Haushaltsmitgliedern beziehungsweise Elternteilen bereitgestellt werden. Haushalte und Lebensformen mit einseitiger Einwanderungsgeschichte werden dabei nicht bei den Haushalten und Lebensformen mit Einwanderungsgeschichte subsumiert, sondern als eigenständige Kategorie nachgewiesen. Soll eine dichotome Aufteilung in Haushalte beziehungsweise Lebensformen mit oder ohne Einwanderungsgeschichte gewählt werden, sollten die Haushalte mit einseitiger Einwanderungsgeschichte entsprechend den Empfehlungen der Fachkommission zu den Rahmenbedingungen der Integrationsfähigkeit den Haushalten beziehungsweise Lebensformen ohne Einwanderungsgeschichte zugerechnet werden.

Das Konzept von Haushalten und Lebensformen nach Einwanderungsgeschichte eignet sich insbesondere für vergleichende Analysen zu Haushalts- und Familienstrukturen. So weisen Untersuchungen in der demografischen und sozialwissenschaftlichen Literatur häufig darauf hin, dass die Kinderzahl in Familien Eingewanderter höher ausfällt als bei Familien ohne Einwanderungsgeschichte. Ebenso stellen sie die Frage, ob sich die Fertilität der zweiten Einwanderergeneration (Nachkommen) von derjenigen der ersten Generation unterscheidet (zum Beispiel Naderi, 2015; Jähnert, 2020). Auch die Notwendigkeit einer verstärkten Berücksichtigung des Haushaltskontextes wird dabei betont (Ette und andere, 2013).

Der vorliegende Beitrag stellt erstmals Ergebnisse unter Anwendung des neuen Konzepts dar. Die Kapitel 2 und 3 betrachten zunächst Struktur und Größe von Haushalten und Familien mit und ohne Einwanderungsgeschichte sowie deren zeitliche Entwicklung seit dem Jahr 2005. Ergebnisse zur Bevölkerung nach Lebensformen im Zeitvergleich enthält Kapitel 4. Auf den Unterschieden bei Haushalts- und Familienzusammensetzung baut im

1 Das seit 2005 im Mikrozensus verwendete Lebensformenkonzept gruppiert die zum Zeitpunkt der Befragung im Haushalt lebenden Personen zu Lebensformen: Paare mit Kind(ern), Paare ohne Kind, Alleinerziehende und Alleinstehende.

2 Die Gruppen sind in diesem Beitrag aus Gründen der Übersichtlichkeit teilweise weiter zusammengefasst.

# Haushalte, Familien und Lebensformen mit Einwanderungsgeschichte – Auswirkungen auf die Erwerbsbeteiligung von Eltern

## Übersicht 1

Zusammengefasste Kombinationen von aus Erhebungsmerkmalen des Mikrozensus abgeleiteten Merkmalen zur Einwanderungsgeschichte

Klassifizierung A: Lebensformen/Familien nach Einwanderungsgeschichte		
1	Lebensform/Familie Eingewanderter	Die Bezugsperson und gegebenenfalls der Partner/die Partnerin sind seit 1950 auf das heutige Gebiet Deutschlands eingewandert.
2	Lebensform/Familie von Nachkommen	Die Bezugsperson und gegebenenfalls der Partner/die Partnerin sind Nachkommen von Eltern, die beide seit 1950 auf das heutige Gebiet Deutschlands eingewandert sind.
3	Lebensform/Familie mit Eingewanderten und Nachkommen	Paar, bei dem ein Partner/eine Partnerin Eingewanderte(r) und der oder die andere Nachkomme ist.
4	Lebensform/Familie mit Partnern mit und ohne Einwanderungsgeschichte	Paar, bei dem ein Partner/eine Partnerin Eingewanderte(r) oder Nachkomme ist und der oder die andere nicht.
5	Lebensform/Familie ohne Einwanderungsgeschichte	Weder die Bezugsperson noch der Partner/die Partnerin ist Eingewanderte(r) oder Nachkomme.
Klassifizierung B: Haushalte nach Einwanderungsgeschichte		
1	Haushalt mit Einwanderungsgeschichte	Alle (erwachsenen) Haushaltsmitglieder sind entweder Eingewanderte oder Nachkommen.
2	Haushalt mit teilweiser Einwanderungsgeschichte	Ein Teil der (erwachsenen) Haushaltsmitglieder sind entweder Eingewanderte oder Nachkommen.
3	Haushalt mit einseitiger Einwanderungsgeschichte	Alle (erwachsenen) Haushaltsmitglieder haben eine einseitige Einwanderungsgeschichte.
4	Haushalt mit teilweise einseitiger Einwanderungsgeschichte	Ein Teil der (erwachsenen) Haushaltsmitglieder hat eine einseitige Einwanderungsgeschichte.
5	Haushalt ohne Einwanderungsgeschichte	Alle (erwachsenen) Haushaltsmitglieder sind ohne Einwanderungsgeschichte.
Klassifizierung C: Haushalte nach Einwanderungsgeschichte (mit Unterscheidung Eingewanderter und Nachkommen)		
1	Haushalt Eingewanderter	Alle (erwachsenen) Haushaltsmitglieder sind eingewandert.
2	Haushalt zum Teil eingewandert	Ein Teil der Haushaltsmitglieder ist eingewandert.
3	Haushalt von Nachkommen	Alle (erwachsenen) Haushaltsmitglieder sind Nachkommen.
4	Haushalt zum Teil mit Nachkommen	Ein Teil der Haushaltsmitglieder sind Nachkommen.
5	Haushalt mit einseitiger Einwanderungsgeschichte	Alle (erwachsenen) Haushaltsmitglieder haben eine einseitige Einwanderungsgeschichte.
6	Haushalt zum Teil mit einseitiger Einwanderungsgeschichte	Ein Teil der Haushaltsmitglieder hat eine einseitige Einwanderungsgeschichte.
7	Haushalt ohne Einwanderungsgeschichte	Alle (erwachsenen) Haushaltsmitglieder sind ohne Einwanderungsgeschichte.

Anschluss die Analyse der Erwerbstätigkeit von Müttern und Vätern mit und ohne Einwanderungsgeschichte auf. Anhand weiterer Merkmale, wie dem Alter des jüngsten Kindes und der Zahl der Kinder im Haushalt, wird der Frage nachgegangen, inwieweit Unterschiede bei der Zusammensetzung von Haushalten und Lebensformen als Erklärung für das unterschiedliche Erwerbsverhalten herangezogen werden können (Kapitel 5). In Kapitel 6 werden die Arrangements untersucht, mit denen Paare

bezahlte Arbeit untereinander aufteilen. Der Artikel schließt mit einem Fazit (Kapitel 7).

## 2

## Haushalte mit Einwanderungsgeschichte

Die Einführung des Konzepts der Bevölkerung nach Einwanderungsgeschichte ermöglicht eine Differenzierung von Haushalten nach Einwanderungsgeschichte und einen Vergleich der Größe von Haushalten mit und ohne Einwanderungsgeschichte. Als Haushalte mit Einwanderungsgeschichte werden hier Haushalte bezeichnet, in denen mindestens ein erwachsenes Haushaltsmitglied eine Einwanderungsgeschichte hat (Gruppen 1 und 2 der Klassifizierung B in Übersicht 1). Haushalte, in denen Personen mit einseitiger Einwanderungsgeschichte leben, werden dabei zu Haushalten ohne Einwanderungsgeschichte gezählt, wenn sich keine Eingewanderten oder Nachkommen im Haushalt befinden. Die Haushalte ohne Einwanderungsgeschichte umfassen hier die Gruppen 3, 4 und 5 der Klassifizierung B in Übersicht 1.

Haushalte mit Einwanderungsgeschichte unterscheiden sich hinsichtlich der Haushaltsgröße deutlich von Haushalten ohne Einwanderungsgeschichte: Im Jahr 2023 lebten in knapp 80 % der Haushalte ohne Einwanderungsgeschichte eine oder zwei Personen, was nur auf gut 60 % der Haushalte mit Einwanderungsgeschichte zutraf. Demgegenüber betrug der Anteil von Haushalten mit vier oder mehr Personen bei Haushalten mit Einwanderungsgeschichte rund 23 %, gegenüber 10 % bei solchen ohne Einwanderungsgeschichte. Hintergrund hierfür ist, dass Familien mit Einwande-

rungsgeschichte im Durchschnitt eine höhere Kinderzahl aufweisen als Familien ohne Einwanderungsgeschichte.

Seit dem Jahr 2005 lassen sich zudem deutliche Veränderungen beobachten: Die Zahl der Haushalte mit Einwanderungsgeschichte wuchs um etwa zwei Drittel von 6,5 Millionen im Jahr 2005 auf 10,9 Millionen im Jahr 2023. Demgegenüber nahm die Zahl der Haushalte ohne Einwanderungsgeschichte in diesem Zeitraum um 5 % auf 30,5 Millionen ab. Die Gesamtzahl aller Hauptwohnsitzhaushalte in Deutschland stieg per saldo von 38,5 Millionen auf 41,3 Millionen. Zu berücksichtigen ist, dass die zeitliche Vergleichbarkeit des Mikrozensus hier aufgrund folgender methodischer Veränderungen eingeschränkt ist: 2013 – Umstellung der Hochrechnung auf die Bevölkerungsfortschreibung auf Basis des Zensus 2011; 2017 – vollständige Erfassung elterlicher Informationen zur Einwanderungsgeschichte; 2020 – Neu-

Tabelle 1

Haushalte mit und ohne Einwanderungsgeschichte

	Insgesamt	Haushaltsgröße				
		1 Person	2 Personen	3 Personen	4 Personen	5 und mehr Personen
1 000	%					
mit Einwanderungsgeschichte						
2005	6 457	24,7	32,2	18,6	16,1	8,5
2009	6 724	27,3	31,5	17,6	15,2	8,3
2013	7 008	28,1	31,5	17,1	15,4	7,8
2017	9 180	31,9	30,0	16,2	14,3	7,6
2018	9 568	32,1	30,3	15,9	14,1	7,5
2019	9 728	32,3	29,9	16,1	14,0	7,7
2020	9 507	31,4	29,7	16,2	15,1	7,6
2021	9 875	32,2	29,3	15,9	14,8	7,8
2022	10 360	31,5	29,6	15,8	14,8	8,3
2023	10 851	32,0	29,4	15,7	14,5	8,3
ohne Einwanderungsgeschichte						
2005	32 067	39,2	34,6	13,3	9,9	3,0
2009	32 862	41,6	35,1	12,0	8,7	2,6
2013	32 402	42,5	35,3	11,7	8,0	2,4
2017	31 541	43,9	34,9	11,0	7,9	2,2
2018	31 237	44,2	35,2	10,9	7,7	2,1
2019	31 174	44,7	34,6	10,8	7,7	2,2
2020	31 034	43,5	35,3	10,9	8,1	2,2
2021	31 098	44,7	34,8	10,6	7,8	2,2
2022	30 605	44,3	35,0	10,6	7,8	2,3
2023	30 479	44,4	35,0	10,6	7,8	2,3

Ergebnisse des Mikrozensus – Bevölkerung in Hauptwohnsitzhaushalten.

gestaltung des Mikrozensus sowie Beeinträchtigung der Erhebungsdurchführung infolge der Corona-Pandemie.  
[Tabelle 1](#)

## 3

### Familien mit Einwanderungsgeschichte

Familien sind im Mikrozensus als Eltern-Kind-Gemeinschaften definiert, das heißt gemischtgeschlechtliche und gleichgeschlechtliche Ehepaare/Lebensgemeinschaften sowie alleinerziehende Mütter und Väter mit Kind(ern) im Haushalt. Das Alter des Kindes oder der Kinder spielt dabei erst einmal keine Rolle (Nöthen, 2005; Hochgürtel/Wilke, 2022). Nach dieser Definition ermittelt der Mikrozensus für das Jahr 2023 insgesamt 12,0 Millionen Familien, im Jahr 2005 waren es noch 12,7 Millionen.

Die folgende Analyse betrachtet Familien mit mindestens einem minderjährigen Kind, das ist die für Untersuchungen zur Vereinbarkeit von Familie und Beruf relevante

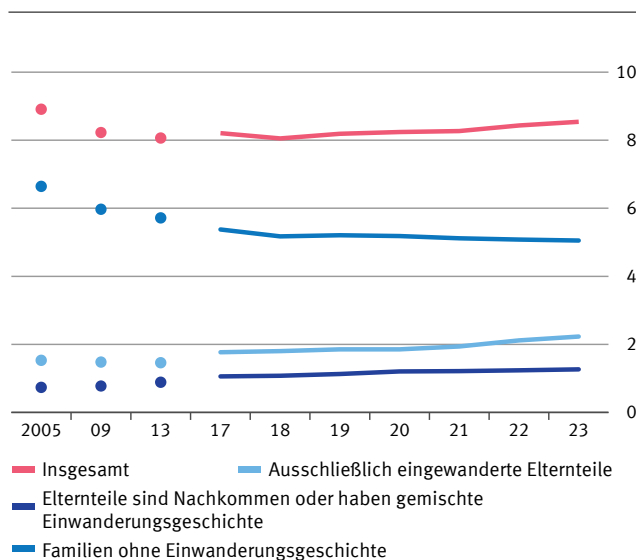
Gruppe.<sup>13</sup> Deren Zahl ist von 2005 bis 2013 zunächst zurückgegangen und etwa ab Mitte der 2010er-Jahre bis 2023 wieder leicht angestiegen. [Grafik 1](#) Differenziert nach der Einwanderungsgeschichte ergibt sich dabei folgendes Bild: Die Zahl der Familien mit ausschließlich eingewanderten Elternteilen sowie der Familien mit Nachkommen als Elternteilen oder gemischter Einwanderungsgeschichte nahm im Betrachtungszeitraum kontinuierlich zu, während die Zahl der Familien ohne Einwanderungsgeschichte nahezu stetig zurückging. Familien mit eingewanderten Elternteilen machten im Jahr 2005 etwa ein Sechstel aller Familien aus; bis 2023 ist dieser Anteil auf ein gutes Viertel angestiegen.

[Grafik 2](#) stellt die Familien, bei denen die Eltern Nachkommen sind, und solche mit gemischter Einwanderungsgeschichte tiefer differenziert dar. Hier zeigt sich wie bei den Familien mit eingewanderten Elternteilen im Zeitvergleich ebenfalls ein Anstieg. Familien

3 Familien ohne Einwanderungsgeschichte umfassen hier die Gruppe 5 der Klassifizierung A in Übersicht 1, Familien mit eingewanderten Elternteilen die Gruppe 1 und Familien mit Nachkommen oder gemischter Einwanderungsgeschichte die Gruppen 2, 3 und 4.

**Grafik 1**

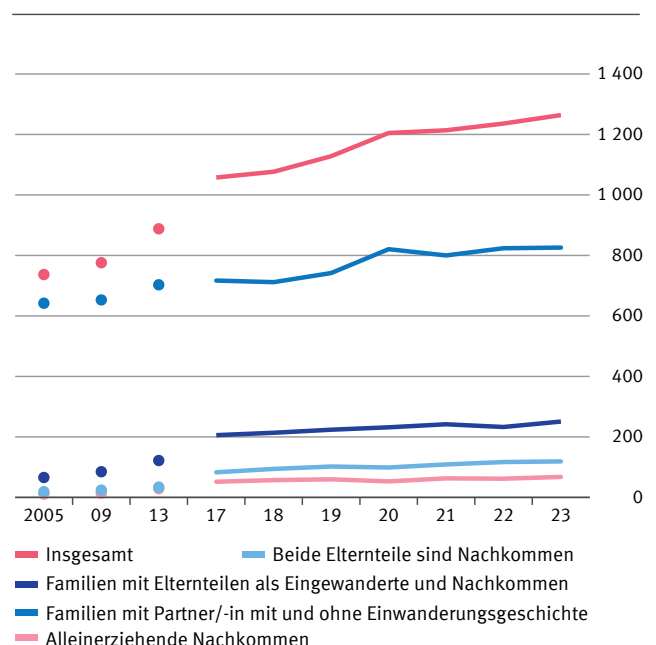
Familien mit minderjährigen Kindern nach Einwanderungsgeschichte in Mill.



Ergebnisse des Mikrozensus. – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.

**Grafik 2**

Familien mit Nachkommen oder mit gemischter Einwanderungsgeschichte in 1 000



Ergebnisse des Mikrozensus. – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.

mit gemischter Einwanderungsgeschichte umfassen dabei sowohl Familien, bei denen ein Elternteil ein Nachkomme und der andere eingewandert ist, als auch Familien mit Elternteilen mit und ohne Einwanderungsgeschichte. Tatsächlich stellen Familien, in denen ein Elternteil keine Einwanderungsgeschichte vorweist und ein Elternteil selbst eingewandert oder Nachkomme von Eingewanderten ist, den größten Anteil der betrachteten Familientypen.

Deutlich unterscheiden sich Familien mit und ohne Einwanderungsgeschichte hinsichtlich der Zahl der Kinder. Im Jahr 2023 hatten 27,1 % der Familien mit zwei eingewanderten Elternteilen oder eingewanderten Alleinerziehenden drei und mehr Kinder, dies traf nur auf 13,4 % der Familien ohne Einwanderungsgeschichte zu. Bei Familien mit gemischter Einwanderungsgeschichte lag der Anteil mit 17,2 % dazwischen. Das Umgekehrte galt für den Anteil von Ein-Kind-Familien, der bei Familien ohne Einwanderungsgeschichte mit 42,1 % im Jahr 2023 am höchsten war. Er betrug bei Familien mit zwei eingewanderten Elternteilen 32,9 % und bei Familien mit gemischter Einwanderungsgeschichte 37,5 %. Diese Strukturen sind im Zeitvergleich recht stabil. [↗ Grafik 3](#)

Dagegen haben sich die absoluten Zahlen abhängig von der Einwanderungsgeschichte stark verändert. Lag das Verhältnis von Familien ohne Einwanderungsgeschichte zu Familien mit eingewanderten Elternteilen mit jeweils einem Kind beziehungsweise zwei Kindern im Jahr 2005 noch bei mindestens fünf zu eins, so hat sich dieses Verhältnis bis 2023 auf drei zu eins reduziert. Bei größeren Familien (mit drei Kindern) verringerte sich der Unterschied zusehends und ab vier Kindern überwogen im Jahr 2023 Familien mit Einwanderungsgeschichte.

[↗ Tabelle 2](#)

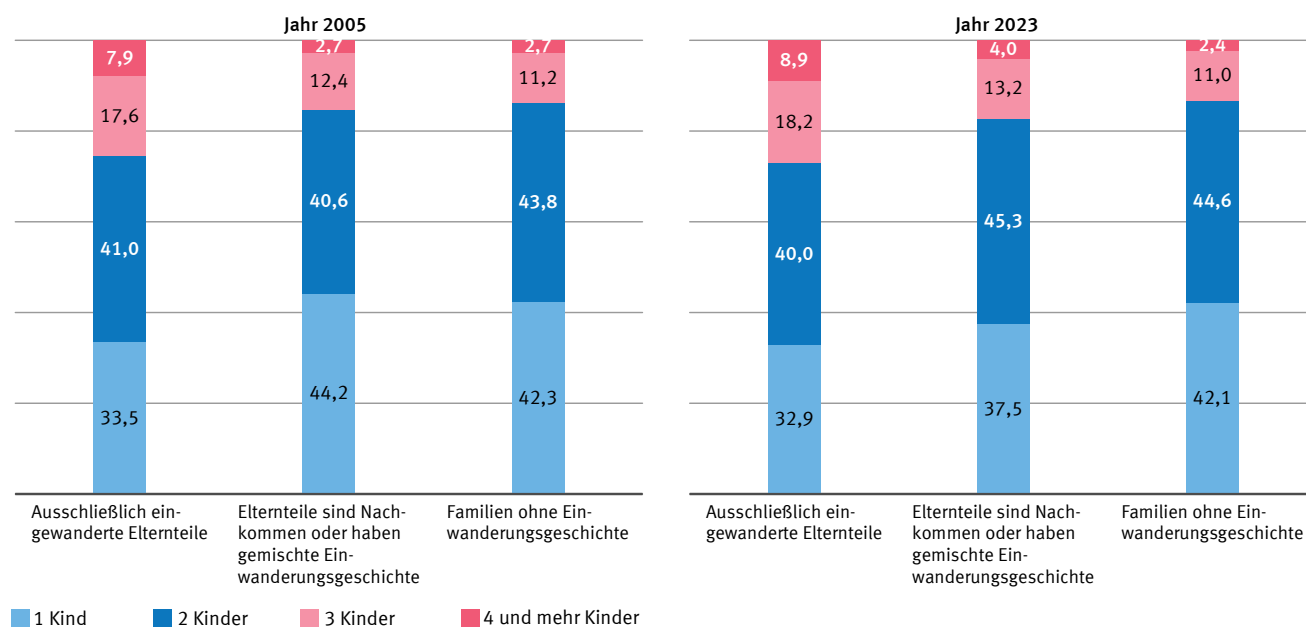
## 4

### Bevölkerung in Familien und Lebensformen

Die folgende Analyse bezieht sich auf die Personen, die in Familien und anderen Lebensformen leben. Wie viele Personen mit und ohne Einwanderungsgeschichte leben als Elternteile oder Kinder in Familien? Wie viele leben in Lebensformen ohne Kind, beispielsweise als Paare ohne Kind, wie viele sind Alleinstehende?

**Grafik 3**

Familien mit mindestens einem minderjährigen Kind nach Zahl der Kinder in %



Ergebnisse des Mikrozensus – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.

Tabelle 2  
Familien mit minderjährigen Kindern nach Einwanderungsgeschichte

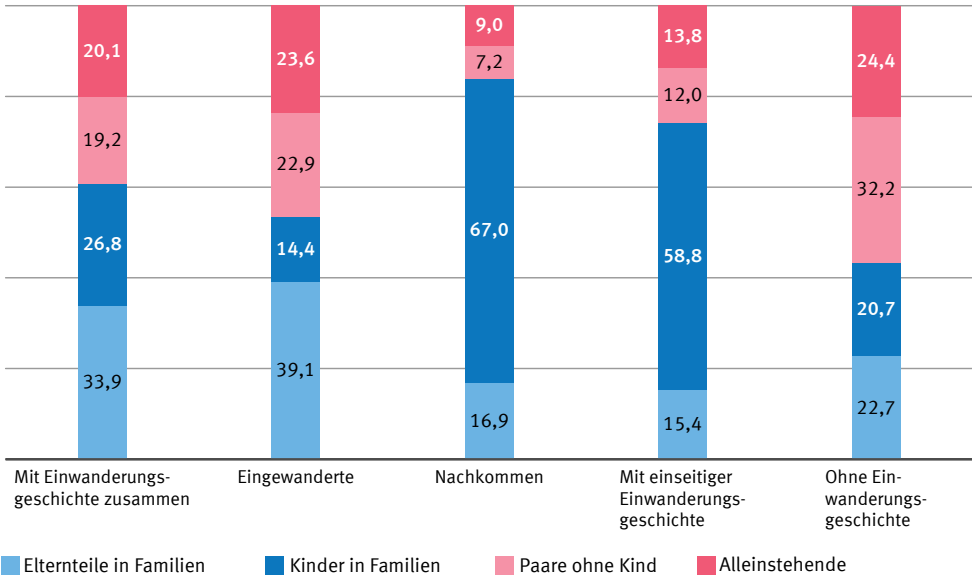
	Familien mit ...							
	1 Kind		2 Kindern		3 Kindern		4 und mehr Kindern	
	1 000	%	1 000	%	1 000	%	1 000	%
2023								
Familien mit eingewanderten Elternteilen	734	32,9	892	40,0	407	18,2	198	8,9
Elternteile sind Nachkommen oder haben gemischte Einwanderungsgeschichte	473	37,5	572	45,3	167	13,2	50	4,0
Familien ohne Einwanderungsgeschichte	2 124	42,1	2 250	44,6	553	11,0	122	2,4
2005								
Familien mit eingewanderten Elternteilen	512	33,5	628	41,0	269	17,6	121	7,9
Elternteile sind Nachkommen oder haben gemischte Einwanderungsgeschichte	325	44,2	299	40,7	91	12,4	20	2,7
Familien ohne Einwanderungsgeschichte	2 808	42,3	2 911	43,8	747	11,2	178	2,7

Ergebnisse des Mikrozensus – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.

Wie nach den Ergebnissen zu Haushalten und Familien zu erwarten, unterscheidet sich auch die Verteilung von Personen hinsichtlich ihrer Lebensform deutlich in Abhängigkeit von der Einwanderungsgeschichte. Im Jahr 2023 lebten 60,7% der Personen mit Einwanderungsgeschichte entweder als Elternteile oder als Kinder in Familien, was nur auf 43,4% der Personen ohne Einwanderungsgeschichte zutraf. ➡ Grafik 4 Dies ist auch

Ausdruck der unterschiedlichen Altersstruktur: Der Anteil Älterer ist bei Personen ohne Einwanderungsgeschichte deutlich höher als bei Personen mit Einwanderungsgeschichte (Grigoriev/Körner, 2024). Entsprechend sind Personen ohne Einwanderungsgeschichte häufiger alleinstehend (24,4% gegenüber 20,1% mit Einwanderungsgeschichte) oder leben als Paare ohne Kind im Haushalt (32,2% gegenüber 19,2% mit Einwanderungs-

Grafik 4  
Verteilung von Personen hinsichtlich ihrer Lebensformen nach Einwanderungsgeschichte 2023 in %



Ergebnisse des Mikrozensus – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.



geschichte). Deutlich wird ebenfalls die starke Besetzung jüngerer Altersgruppen bei Nachkommen und Personen mit einseitiger Einwanderungsgeschichte – als Kinder in Familien lebten 67 % (Nachkommen) beziehungsweise 58,8 % (einseitige Einwanderungsgeschichte).

Die Zusammensetzung der Bevölkerung in Familien und Lebensformen nach der Einwanderungsgeschichte hat sich seit dem Jahr 2005 deutlich verändert. Im Jahr 2023 lebten mit nahezu 84 Millionen Menschen 2,1 Millionen Menschen mehr in Deutschland als 2005.<sup>14</sup> Unter anderem aufgrund der Alterung der Bevölkerung ist der Anteil der in Familien lebenden Personen insgesamt zurückgegangen, von 53 % (2005) auf 49 % (2023). Die Zuwanderung hat dabei den Rückgang etwas gedämpft: Jedes dritte Elternteil (33,9 %) wies 2023 eine Einwanderungsgeschichte auf (2005: 20,8 %), der Anteil der Kinder mit einer Einwanderungsgeschichte ist von 17,9 % (2005) auf 28,1 % (2023) angestiegen. Ebenfalls zugenommen, und zwar von 11,7 % im Jahr 2005 auf 22,3 % im Jahr 2023, hat der Anteil alleinstehender Menschen mit Einwanderungsgeschichte. Demgegenüber fiel der Anstieg bei Paaren ohne Kind (+ 5 Prozentpunkte) moderater aus. ➔ **Tabelle 3**

4 Die Ergebnisse des Mikrozensus 2005 wurden an Eckwerten auf Basis der Volkszählung 1987, die des Mikrozensus 2023 an denen auf Basis des Zensus 2011 hochgerechnet, was die Vergleichbarkeit der absoluten Bevölkerungszahlen beeinträchtigt. Die Ergebnisse des Zensus 2022 wurden bei der Hochrechnung der Ergebnisse des Mikrozensus 2023 noch nicht berücksichtigt.

## 5

### Erwerbstätigkeit von Müttern und Vätern mit und ohne Einwanderungsgeschichte

Die gleichberechtigte Teilhabe am Arbeitsmarkt ist eine wesentliche Voraussetzung für die Gleichstellung von Frauen und Männern, gleichzeitig ist sie von großer Bedeutung für Integration und Fortentwicklung der Gesellschaft. Vor dem Hintergrund des Fach- und Arbeitskräftebedarfs ist es für den Arbeitsmarkt ausschlaggebend, das Potenzial an Personen, die als Erwerbstätige zur Verfügung stehen, auszuschöpfen.

Eine weiterführende Analyse für die Bevölkerung in Familien und Lebensformen teilt alle Personen nach den Merkmalen Erwerbstätigkeit<sup>15</sup> und Geschlecht sowie ihren Familienzusammenhang und ihre Einwanderungsgeschichte ein. Unterschieden wird dabei einerseits zwischen Müttern und Vätern sowie andererseits zwischen Frauen und Männern ohne Kind im Haushalt. Zu den

5 Im Folgenden wird das Konzept der realisierten Erwerbstätigkeit zugrunde gelegt. Erwerbstätig sind demnach alle Personen, die nach den Vorgaben der Internationalen Arbeitsorganisation (ILO, 2023) als erwerbstätig gezählt werden, sofern diese nicht ihre Tätigkeit in der Berichtswoche wegen Mutterschutz oder Elternzeit unterbrochen haben (Hochgürtel, 2018; Keller/Körner, 2023).

**Tabelle 3**

#### Bevölkerung nach Lebensformen und Einwanderungsgeschichte

	Insgesamt	Bevölkerung mit Einwanderungsgeschichte			Mit einseitiger Einwanderungsgeschichte	Ohne Einwanderungsgeschichte
		zusammen	Eingewanderte	Nachkommen		
	1 000	%				
2023						
Insgesamt	83 822	25,2	19,3	6,0	4,8	70,0
Elternteile	21 104	33,9	29,9	4,0	2,9	63,1
Kinder	20 169	28,1	11,5	16,6	11,7	60,2
Paare ohne Kind	23 424	17,3	15,8	1,5	2,1	80,6
Alleinstehende	19 125	22,3	19,9	2,4	2,9	74,8
2005						
Insgesamt	81 732	16,0	12,9	3,1	2,2	81,8
Elternteile	22 790	20,8	20,1	0,7	0,3	78,9
Kinder	20 868	17,9	7,3	10,6	7,6	74,5
Paare ohne Kind	22 629	12,3	11,9	0,3	0,2	87,5
Alleinstehende	15 444	11,7	11,1	0,6	0,6	87,7

Ergebnisse des Mikrozensus – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.



Personen ohne Kind zählen Menschen, die in Partnerschaften ohne Kind oder als Alleinstehende leben. Kinder im erwerbsfähigen Alter zählen ebenfalls zur Gruppe der Personen ohne Kind. Zunächst erfolgt eine Betrachtung der realisierten Erwerbstätigenquoten, anschließend wird näher auf Unterschiede bei der Arbeitszeit eingegangen. Dabei wird jeweils die Gruppe der Eingewanderten mit Personen ohne Einwanderungsgeschichte verglichen. Nachkommen und Personen mit einseitiger Einwanderungsgeschichte bleiben unberücksichtigt, da diese beiden Gruppen vor allem Menschen jüngerer Alters umfassen, was Vergleiche zum Erwerbsverhalten erschwert.

### 5.1 Erwerbstätigenquoten von Eltern

---

Bei den Vätern hat die Lebensphase der Familiengründung offenkundig nur einen geringen Einfluss auf die Erwerbstätigenquote, die als Anteil der Erwerbstätigen an der jeweiligen Bevölkerungsgruppe definiert ist. Nach dem Eintritt in den Arbeitsmarkt bleibt die Erwerbstätigenquote bei Vätern ebenso wie bei Männern ohne Kind konstant auf hohem Niveau. Dieses Bild zeigt sich zunächst unabhängig von der Einwanderungsgeschichte. Bei Vätern ohne Einwanderungsgeschichte lag die Erwerbstätigenquote für alle Altersjahre von 25 bis 59 Jahren über 90% und damit durchschnittlich knapp 7 Prozentpunkte über der Erwerbstätigenquote von Männern ohne Kind und ohne Einwanderungsgeschichte.

Bei der Analyse der Zahlen eingewanderter Väter sowie eingewanderter Männer ohne Kind zeigt sich zunächst, dass die Erwerbstätigenquote in allen Altersjahren etwas niedriger liegt als bei den entsprechenden Gruppen ohne Einwanderungsgeschichte. Die Unterschiede zwischen eingewanderten Vätern und eingewanderten Männern ohne Kind waren dabei in den Altersjahren bis etwa 40 Jahre gering und die Erwerbstätigenquoten lagen auf einem ähnlichen Niveau von rund 85%. Bei älteren eingewanderten Vätern lag die Erwerbstätigenquote rund 5 Prozentpunkte über der von eingewanderten Männern ohne Kind. ➔ Grafik 5 auf Seite 48

Frauen ohne Kind und Mütter unterscheiden sich in ihrer Erwerbstätigkeit deutlich stärker voneinander als Männer ohne Kind und Väter: Frauen ohne Kind erreichten im Jahr 2023 im Alter von 30 Jahren die höchste Erwerbstätigenquote (ohne Einwanderungsgeschichte: 92,0%;

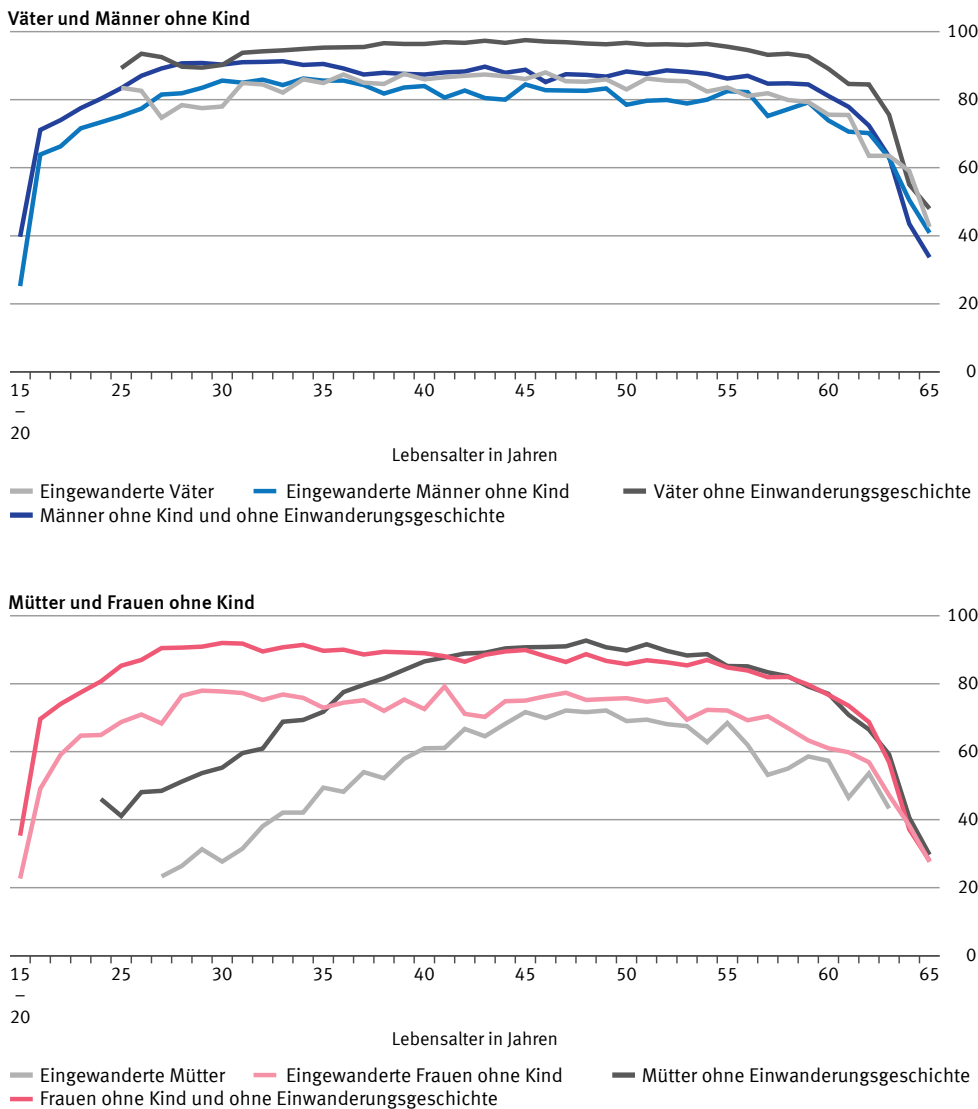
Einwanderte: 77,7%), in älteren Altersjahren lag diese um bis zu 10 Prozentpunkte niedriger. Auch hier war die Erwerbstätigkeit von eingewanderten Frauen ohne Kind stets um knapp 15 Prozentpunkte geringer als diejenige von Frauen ohne Kind und ohne Einwanderungsgeschichte.

Eine andere Entwicklung zeigt die Erwerbstätigkeit von Müttern: Diese stieg ab dem Alter von 30 Jahren mit einem Wert von gut 55% (Mütter ohne Einwanderungsgeschichte) beziehungsweise 28% (eingewanderte Mütter) bis zum Alter von rund 50 Jahren jeweils um etwa 40 Prozentpunkte an. Dabei fällt auf, dass die Differenz der Erwerbstätigenquoten bei eingewanderten Müttern und Müttern ohne Einwanderungsgeschichte mit bis zu 28 Prozentpunkten fast doppelt so groß ist wie bei den entsprechenden Gruppen der Frauen ohne Kind. Zudem übersteigt die Erwerbstätigenquote der Mütter ohne Einwanderungsgeschichte ab dem Alter von 42 Jahren die der Frauen ohne Kind und ohne Einwanderungsgeschichte, was bei den Ergebnissen für die entsprechenden Gruppen eingewanderter Frauen nie der Fall ist. Eine mögliche Erklärung hierfür könnten die Unterschiede in der Anzahl der Kinder sein (siehe unten).

Der Vergleich von Männern und Frauen zeigt, dass sich Männer und Frauen ohne Kind und ohne Einwanderungsgeschichte im Erwerbsverhalten am ähnlichsten sind. Väter ohne Einwanderungsgeschichte erreichen die höchste realisierte Erwerbstätigkeit überhaupt, während eingewanderte Mütter die geringsten Erwerbstätigenquoten aufweisen.

**Grafik 5**

Realisierte altersspezifische Erwerbstätigenquoten 2023  
in %



Ergebnisse des Mikrozensus – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten. Fehlende Werte sind darauf zurückzuführen, dass der Zahlenwert aufgrund der geringen Fallzahl (70 oder weniger) nicht sicher genug ist (relativer Standardfehler durchschnittlich über 15 %).

## 5.2 Normalerweise geleistete Arbeitszeit

Mit der Erwerbstätigenquote ist keine Aussage über den Umfang der Tätigkeit verbunden, sie gibt lediglich Auskunft darüber, wie hoch der Anteil einer Bevölkerungsgruppe ist, die einer bezahlten Tätigkeit nachgeht. Nach der Definition der Internationalen Arbeitsorganisation (ILO, 2023) wird jede Tätigkeit ab einem Umfang von

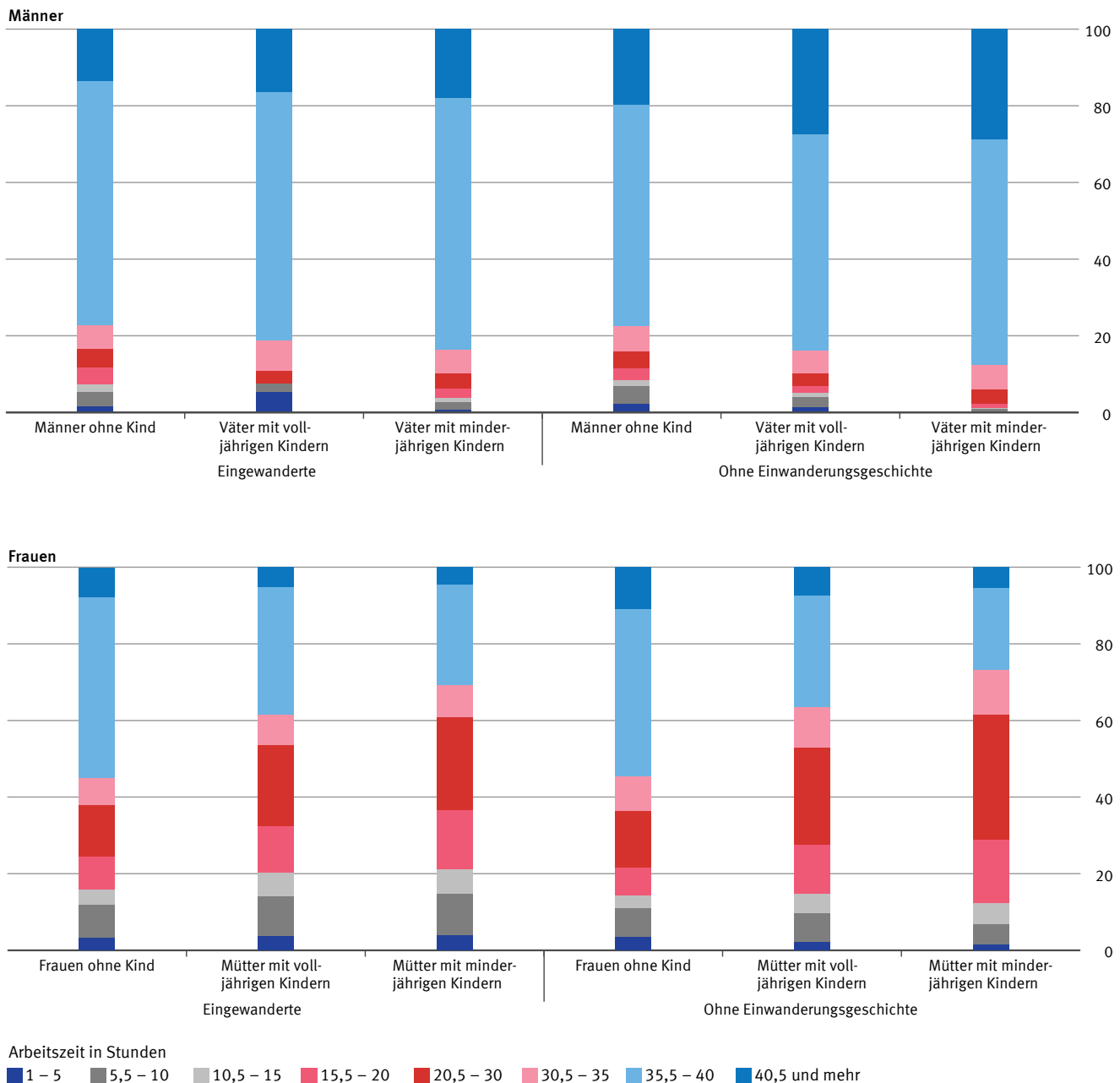
einer Stunde als Erwerbstätigkeit erfasst. Daher ist es zur weiteren Analyse der Erwerbstätigen wichtig, auch den Umfang der wöchentlichen Arbeitszeit zu betrachten.

➤ Grafik 6 zeigt, dass auch innerhalb der Gruppe der Erwerbstätigen Mütter und Frauen ohne Kind je Woche normalerweise deutlich weniger Stunden arbeiten als Väter und Männer ohne Kind. Betrachtet nach der Einwanderungsgeschichte, fallen die Unterschiede für Väter

# Haushalte, Familien und Lebensformen mit Einwanderungsgeschichte – Auswirkungen auf die Erwerbsbeteiligung von Eltern

**Grafik 6**

Normalerweise geleistete Arbeitszeit von realisiert Erwerbstätigen 2023  
in %



Ergebnisse des Mikrozensus – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.

und Männer ohne Kind vor allem bei Arbeitszeiten über 35 Stunden auf: Eingewanderte Männer ohne Kind und eingewanderte Väter arbeiteten häufiger 35 bis 40 Stunden je Woche, während die entsprechenden Gruppen

ohne Einwanderungsgeschichte etwas häufiger mehr als 40 Stunden tätig sind. Bei den Tätigkeiten mit einem Umfang von bis 35 Stunden sind die Unterschiede dagegen nur gering.

Frauen ohne Kind und Mütter verteilen ihre Arbeitszeit etwas gleichmäßiger auf die unterschiedlichen Stufen als die Vergleichsgruppen der Männer und Väter. Nach der Einwanderungsgeschichte sind auch hier die Unterschiede insgesamt gering. Am häufigsten arbeiteten Frauen ohne Kind in einer Tätigkeit mit mehr als 35 Wochenstunden; das trifft sowohl für Eingewanderte als auch für Frauen ohne Einwanderungsgeschichte zu. Fast jede neunte Frau ohne Einwanderungsgeschichte hatte dabei eine Wochenarbeitszeit von mehr als 40 Stunden, bei den eingewanderten Frauen war es jede Dreizehnte.

Die Wochenarbeitszeit erwerbstätiger Mütter lag unter der von Frauen ohne Kind, insbesondere Mütter mit Kindern unter 18 Jahren waren häufiger in den Kategorien von 15,5 bis 20 Stunden und von 20,5 bis 30 Stunden Wochenarbeitszeit vertreten. Eingewanderte Mütter arbeiteten dagegen häufiger von 35,5 bis 40 Stunden als Mütter ohne Einwanderungsgeschichte. Ein Grund dafür ist möglicherweise, dass Mütter ohne Einwanderungsgeschichte häufiger erwerbstätig sind als solche mit Einwanderungsgeschichte, dazu jedoch häufiger eine Teilzeittätigkeit mit weniger Arbeitsstunden wählen.

Frauen und Mütter arbeiteten generell häufiger bis zu 15 Stunden je Woche als Männer und Väter. Während 2023 nicht einmal jeder zehnte Mann ohne Kind einer realisierten Erwerbstätigkeit unter 15 Stunden die Woche nachging, war es bei den Frauen etwa jede sechste. Bei den eingewanderten Müttern ging gut jede fünfte einer Tätigkeit bis 15 Wochenstunden nach, bei Müttern ohne Einwanderungsgeschichte war es etwa jede achte. Für Väter spielt diese Form der Erwerbstätigkeit kaum eine Rolle.

### 5.3 Erwerbstätigkeit und Arbeitszeit – welchen Einfluss hat das Alter des jüngsten Kindes?

Um die Frage der Vereinbarkeit von Familie und Beruf zu beantworten ist von entscheidender Bedeutung, ob und wie sich das Erwerbsverhalten von Männern und Frauen mit der Geburt eines Kindes verändert (siehe beispielsweise Feldhoff, 2021). Bislang dominiert in Deutschland nach wie vor das modifizierte männliche Ernährermodell (male breadwinner model). Bei diesem baut der Vater seine Erwerbstätigkeit nach der Geburt eines Kindes eher noch aus, während die Mutter ihre Erwerbs-

tätigkeit unterbricht und danach (häufig in Teilzeit) auf den Arbeitsmarkt zurückkehrt (Trappe und andere, 2015; Keller/Körner, 2023). Um die Auswirkungen der Geburt von Kindern auf die Erwerbstätigkeit von Frauen und Männern zu untersuchen, liefert eine Analyse der Erwerbstätigenquoten nach dem Alter des jüngsten Kindes wichtige Hinweise.<sup>6</sup>

Das Alter des jüngsten Kindes hat so gut wie keine Auswirkung auf die realisierte Erwerbstätigkeit von Vätern. Väter weisen unabhängig vom Alter des jüngsten Kindes eine konstant hohe Erwerbstätigkeit auf. Bei Vätern ohne Einwanderungsgeschichte liegt die Erwerbstätigenquote 2023 bei rund 95 %, bei eingewanderten Vätern rund 10 Prozentpunkte darunter. Etwas geringer ist die Erwerbstätigenquote in beiden Fällen nur bei Vätern von Kindern im Alter von unter einem Jahr sowie bei Vätern von Kindern im Alter über 18 Jahren<sup>7</sup>. [➤ Grafik 7](#)

Bei den Müttern zeigen sich – abhängig von der Einwanderungsgeschichte – dagegen deutlich stärkere Unterschiede. Mütter von Kindern unter einem Jahr wiesen in beiden Fällen eine niedrige realisierte Erwerbstätigenquote auf (Eingewanderte: 8,6 %; ohne Einwanderungsgeschichte: 13,2 %) als Mütter älterer Kinder. Während jedoch bei Müttern ohne Einwanderungsgeschichte mit zunehmendem Alter des jüngsten Kindes eine höhere Erwerbstätigenquote zu beobachten ist, fällt der Anstieg bei eingewanderten Müttern weniger ausgeprägt aus. Bei Kindern ab dem Alter von sechs Jahren nähert sich die Erwerbstätigenquote von Müttern ohne Einwanderungsgeschichte wieder der der Väter an, während die Erwerbstätigenquote von eingewanderten Müttern zwischen 13 und 25 Prozentpunkten unter der von eingewanderten Vätern liegt.

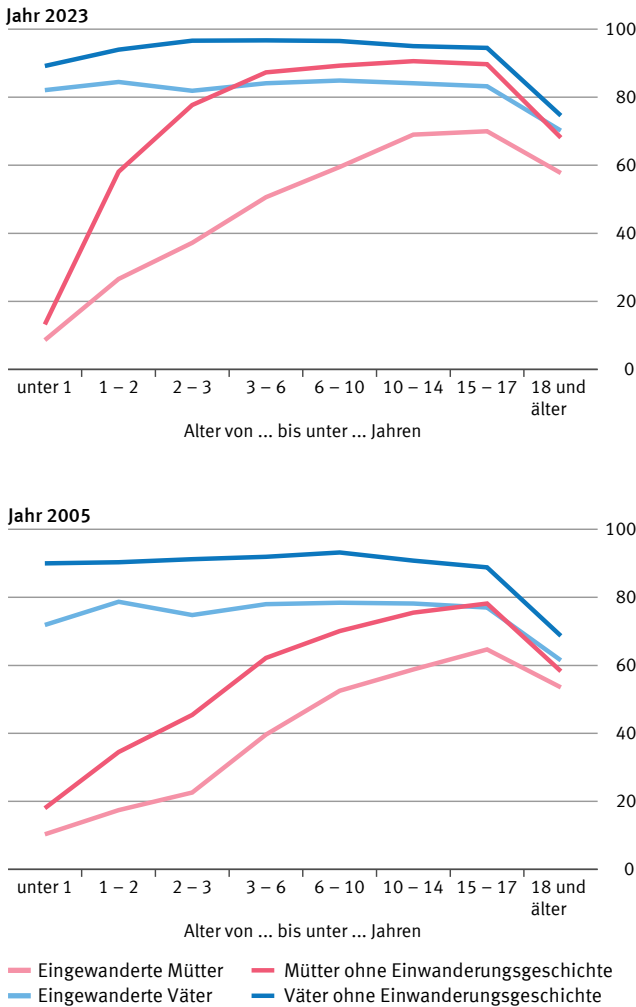
Im Vergleich der Ergebnisse von 2023 mit denen für das Jahr 2005 ergibt sich für alle Gruppen eine etwas höhere realisierte Erwerbsbeteiligung. Der deutlichste Anstieg zeigt sich bei Müttern ohne Einwanderungsgeschichte und eingewanderten Müttern, was auch auf familienpolitische Maßnahmen wie den Ausbau der Kinder-

6 Zu beachten ist dabei, dass die folgenden Darstellungen nicht einzelne Kohorten vergleichen, sondern die Situation von Müttern mit Kindern in unterschiedlichen Altersklassen im Berichtsjahr 2023 darstellen. So gehören Frauen von Kindern ab 18 Jahren vielfach einer anderen Alterskohorte an als Frauen von Kindern im Alter von einem Jahr. Das Erwerbsverhalten der Kohorten kann sich auch aus anderen Gründen als der Geburt von Kindern unterscheiden.

7 Vermutlich befinden sich diese teilweise schon im Ruhestand.

**Grafik 7**

Erwerbstätigenquoten von Müttern und Vätern nach Alter des jüngsten Kindes in %



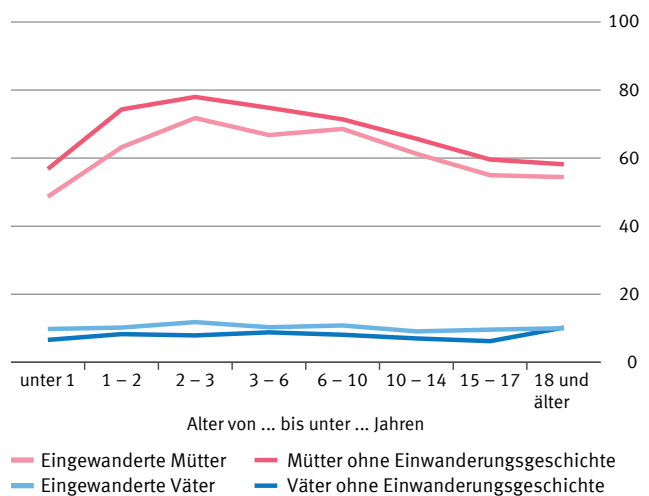
tagesbetreuung und die Einführung des Elterngelds zurückzuführen sein dürfte. Der Anstieg fällt dabei bei den Müttern ohne Einwanderungsgeschichte etwas stärker aus, das heißt die Erwerbslücke zwischen Müttern ohne Einwanderungsgeschichte und eingewanderten Müttern hat sich sogar noch etwas vergrößert.

Auch hier empfiehlt es sich, den Vergleich der Erwerbstätigenquote um eine Darstellung der Arbeitszeit zu ergänzen, da sich der Umfang der Erwerbstätigkeit stark unterscheiden kann. [Grafik 8](#) zeigt die realisierten Teilzeitquoten für Mütter und Väter mit und ohne Einwande-

rungsgeschichte. Dabei offenbaren sich deutliche Unterschiede in den Arbeitszeitmodellen von Müttern und Vätern, und zwar größtenteils unabhängig von der Einwanderungsgeschichte. Ein Vergleich der Väter ohne Einwanderungsgeschichte mit eingewanderten Vätern ergibt, dass deren Verhalten nahezu identisch ist – etwa jeder zehnte Vater hat seine Arbeitszeit auf Teilzeit reduziert.<sup>18</sup>

**Grafik 8**

Teilzeitquoten realisiert erwerbstätiger Mütter und Väter nach dem Alter des jüngsten Kindes 2023 in %



Bei Müttern gibt es geringe Unterschiede zwischen Müttern ohne Einwanderungsgeschichte und eingewanderten Müttern, die Teilzeitquote liegt jeweils deutlich höher als bei Vätern. Mütter ohne Einwanderungsgeschichte hatten im Jahr 2023 eine Teilzeitquote von 57 % bei Kindern bis zu einem Jahr, da viele Mütter nach einer Geburt ihre Erwerbstätigkeit ganz unterbrechen. Bei Kindern im Kindergartenalter betrug die Teilzeitquote rund 75 %, bei Kindern im Schulalter lag sie zwischen 60 und 70 %. Auch bei Kindern ab 18 Jahren bleibt die Teilzeitquote bei über 50 %. Die Teilzeitquote von eingewanderten Müttern weist eine ähnliche Verteilung auf, liegt aber jeweils rund 5 bis 10 Prozentpunkte unter der

8 Als teilzeitbeschäftigt gilt im Mikrozensus, wer normalerweise weniger Stunden als in einer in der Branche beziehungsweise dem Betrieb üblichen Vollzeitbeschäftigung arbeitet. Die Erfassung erfolgt als Selbsteinschätzung der Befragten.

von Müttern ohne Einwanderungsgeschichte (siehe Grafik 8). Hintergrund dürfte auch hier sein, dass insgesamt deutlich weniger eingewanderte Mütter einer bezahlten Erwerbstätigkeit nachgehen.

## 5.4 Familiengröße und Erwerbsbeteiligung

Beim Vergleich der Erwerbstätigkeit von Müttern ohne Einwanderungsgeschichte mit eingewanderten Müttern sind nicht zuletzt die Unterschiede in den Strukturen der Familien zu berücksichtigen. Dass die Zahl der Kinder in Familien mit Einwanderungsgeschichte höher ist als bei Familien ohne Einwanderungsgeschichte, könnte eine Erklärung für die geringere Erwerbstätigkeit von eingewanderten Müttern sein. Denn die Betreuung von mehreren minderjährigen Kindern lässt auch bei Nutzung von Betreuungsangeboten weniger Raum für eine Erwerbstätigkeit.

Eine Differenzierung der Erwerbstätigenquote von Müttern und Vätern nach der Zahl der minderjährigen Kinder im Haushalt liefert Hinweise darauf, inwiefern die Unterschiede bei der Erwerbstätigkeit mit unterschiedlichen Familienstrukturen zusammenhängen können.<sup>9</sup>

9 Tatsächlich ist die Betreuungsquote von Kindern mit Einwanderungsgeschichte deutlich geringer als die von Kindern ohne Einwanderungsgeschichte (Heilmann, 2021), was auch auf Benachteiligungen bei der Vergabe von Betreuungsplätzen zurückgeführt wird (Schmitz und andere, 2023; Lokhande, 2023; Jessen und andere, 2020).

So liegt die Erwerbstätigenquote von eingewanderten Müttern beziehungsweise Vätern stets unter der von Müttern beziehungsweise Vätern ohne Einwanderungsgeschichte: Bezogen auf Haushalte mit mindestens einem minderjährigen Kind gab es in Deutschland 2023 rund 2,2 Millionen eingewanderte Väter und 2,6 Millionen eingewanderte Mütter; 4,3 Millionen Väter und 4,9 Millionen Mütter hatten keine Einwanderungsgeschichte. Väter ohne Einwanderungsgeschichte erreichen eine realisierte Erwerbstätigenquote von rund 95 %, die Kinderzahl scheint nur gering mit der Erwerbsbeteiligung zu korrelieren. Die Erwerbstätigenquoten von eingewanderten Vätern mit einem Kind, zwei oder drei Kindern lagen 2023 gut 10 Prozentpunkte unter derjenigen von Vätern ohne Einwanderungsgeschichte. Mit vier und mehr Kindern waren eingewanderte Väter noch deutlich seltener erwerbstätig (70,7 %). [Tabelle 4](#)

Bei den Müttern hat die Zahl der Kinder einen stärkeren Einfluss auf die Erwerbsbeteiligung als bei den Vätern. Bei Müttern ohne Einwanderungsgeschichte lag die realisierte Erwerbstätigenquote bei einem, zwei oder drei minderjährigen Kindern im Haushalt jeweils bei rund 80 %. Deutlich geringer war die Erwerbstätigenquote bei eingewanderten Müttern: Betrug diese bei einem oder zwei Kindern noch gut 55 %, so lag sie bei drei Kindern bei 43,6 % und bei vier und mehr Kindern noch bei 23,4 %. Mütter ohne Einwanderungsgeschichte und mit vier und mehr Kindern im Haushalt waren dagegen zu 63,9 % erwerbstätig.

**Tabelle 4**

**Väter und Mütter nach Einwanderungsgeschichte, Zahl der Kinder<sup>1</sup> und Erwerbstätigenquoten 2023**

	Insgesamt	Elternteile mit ...			
		1 Kind	2 Kindern	3 Kindern	4 und mehr Kindern
	1 000				
Eingewanderte Väter	2 169	666	902	405	196
Eingewanderte Mütter	2 612	876	1 078	451	207
Väter ohne Einwanderungsgeschichte	4 336	1 707	2 031	493	105
Mütter ohne Einwanderungsgeschichte	4 930	2 034	2 219	552	125
	Erwerbstätigenquoten <sup>1,2</sup> in %				
Eingewanderte Väter	84,2	85,4	87,3	82,1	70,7
Eingewanderte Mütter	52,2	55,7	58,4	43,6	23,4
Väter ohne Einwanderungsgeschichte	95,6	94,6	96,5	95,9	93,0
Mütter ohne Einwanderungsgeschichte	79,9	78,7	82,7	76,4	63,9

<sup>1</sup> Ohne Altersbeschränkung der Kinder.

<sup>2</sup> Von Müttern und Vätern mit minderjährigen Kindern.

Ergebnisse des Mikrozensus – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.



Die Erwerbstätigkeit von eingewanderten Müttern liegt durchgehend unter der von Müttern ohne Einwanderungsgeschichte. Unterschiede in der Familienstruktur können daher offenkundig keinen größeren Beitrag zur Erklärung der geringeren Erwerbstätigenquoten von eingewanderten Müttern liefern.

## 6

### Erwerbsarrangements von Paaren mit Kindern

Von besonderem Interesse für die Frage, wie gleichmäßig Paare Erwerbsarbeit und unbezahlte Sorgearbeit untereinander aufteilen, ist die Analyse der Arbeitsteilung innerhalb von Paarhaushalten. Mithilfe des Mikrozensus ist es möglich, die Zeit zu untersuchen, die die in einem Haushalt lebenden Personen in Erwerbsarbeit verbringen.<sup>10</sup> Somit kann die Analyse zumindest indirekt Tendenzaussagen zur Teilung der unbezahlten Sorgearbeit liefern (Keller/Körner, 2023; Schäper und andere, 2023). Interessant ist in diesem Kontext insbesondere die Frage, ob bei Paaren abhängig von der Einwanderungsgeschichte Unterschiede bezüglich des dominierenden Ernährermodells deutlich werden.

Auch bei Paaren soll zunächst die Erwerbstätigkeit der beiden Partner betrachtet werden. Hier zeigen sich

deutliche Unterschiede: Im Jahr 2023 waren bei 47,4 % der Paarfamilien mit zwei eingewanderten Elternteilen beide Elternteile erwerbstätig, bei Paarfamilien ohne Einwanderungsgeschichte traf dies auf 78,1 % zu. Entsprechend war bei Paarfamilien Eingewanderter in 36 % der Fälle ausschließlich der Vater erwerbstätig gegenüber 18 % bei Paarfamilien ohne Einwanderungsgeschichte. Schließlich war der Anteil von Paarfamilien, bei denen weder der Vater noch die Mutter erwerbstätig waren, bei zwei eingewanderten Elternteilen mit 13 % deutlich größer als bei Paaren ohne Einwanderungsgeschichte (1,8 %). Die Erwerbsarrangements von Familien mit gemischter Einwanderungsgeschichte bewegen sich zwischen denen der beiden anderen Gruppen, ähneln hinsichtlich des Ernährermodells aber eher Familien ohne Einwanderungsgeschichte. [↗ Grafik 9](#)

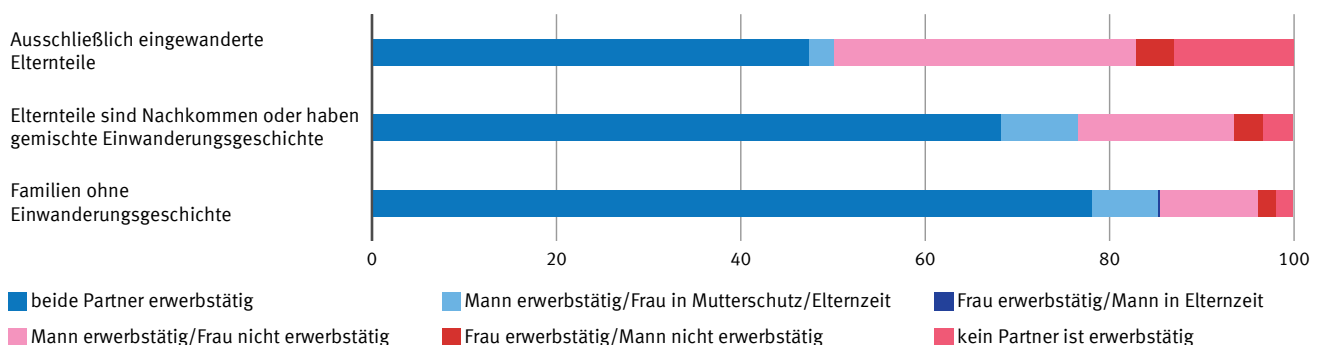
Die Unterschiede in Familien, bei denen Vater und Mutter erwerbstätig sind, fallen gering aus: Der Anteil der Paarfamilien, bei denen beide Elternteile Vollzeit arbeiten, lag 2023 bei Familien Eingewanderter mit 31,1 % etwas höher als bei Familien ohne Einwanderungsgeschichte (25,8 %). Dagegen war das Arrangement, in dem der Mann Vollzeit und die Frau Teilzeit arbeitet, bei Familien ohne Einwanderungsgeschichte mit 66,7 % etwas stärker vertreten als bei Paaren Eingewanderter (60,8 %). Eine Teilzeitbeschäftigung beider Elternteile war bei allen drei Gruppen mit rund 5 % gleichermaßen die Ausnahme. [↗ Grafik 10 auf Seite 54](#)

Insgesamt ist festzuhalten, dass bei Paarfamilien mit Einwanderungsgeschichte das klassische männliche Ernährermodell, in dem nur der Vater erwerbstätig ist,

10 Für die Aufteilung der unbezahlten Arbeit stehen die [Ergebnisse der Zeitverwendungserhebung 2022](#) zur Verfügung.

#### Grafik 9

Erwerbsarrangements von Paaren mit minderjährigen Kindern 2023 in %

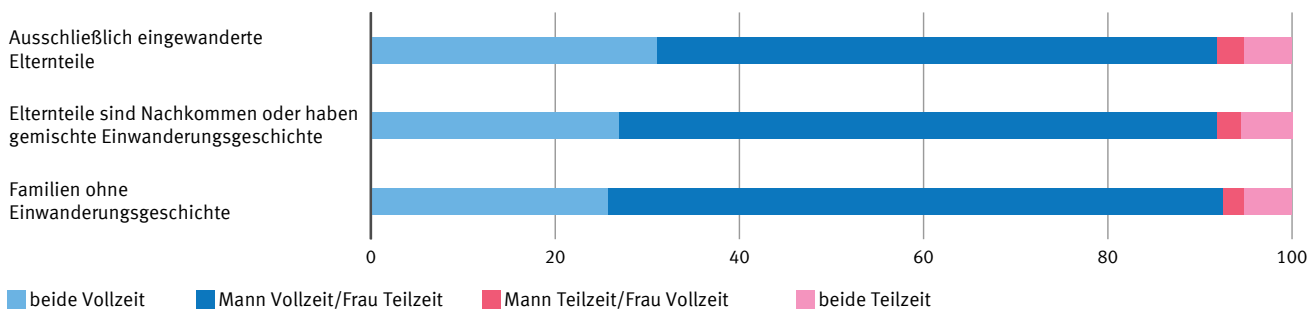


Ergebnisse des Mikrozensus – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.



**Grafik 10**

Erwerbsarrangements von realisiert erwerbstätigen Paaren mit minderjährigen Kindern 2023  
in %



Ergebnisse des Mikrozensus – Bevölkerung in Familien und Lebensformen in Hauptwohnsitzhaushalten.

stärker verbreitet ist als in Familien ohne Einwanderungsgeschichte. In diesen wiederum dominiert das modifizierte männliche Ernährermodell, bei dem der Mann Vollzeit und die Frau Teilzeit arbeitet. Sind bei Familien mit Einwanderungsgeschichte allerdings beide Elternteile erwerbstätig, so ist der Anteil der Paare, in denen Mutter und Vater Vollzeit arbeiten, etwas höher als in Familien ohne Einwanderungsgeschichte.

## 7

### Fazit

Die Struktur von Haushalten, Familien und anderen Lebensformen unterscheidet sich abhängig von deren Einwanderungsgeschichte teilweise deutlich. Die in Ergänzung zum Konzept der Einwanderungsgeschichte eingeführte Kategorisierung des Statistischen Bundesamtes liefert die Grundlage für einen differenzierten Vergleich von Haushalten, Familien und Lebensformen.

Hervorzuheben sind zunächst die Unterschiede bei der Haushaltsgröße: Haushalte mit Einwanderungsgeschichte sind im Durchschnitt größer als Haushalte ohne Einwanderungsgeschichte. Insbesondere der Anteil der Haushalte mit vier und mehr Personen ist bei Vorliegen einer Einwanderungsgeschichte mehr als doppelt so groß wie bei Haushalten ohne Einwanderungsgeschichte, die häufiger aus ein oder zwei Personen bestehen.


Hinzu kommt, dass die Zahl der Haushalte und Lebensformen mit Einwanderungsgeschichte seit dem Jahr

2005 – insbesondere im Zusammenhang mit der Einwanderung Schutzsuchender ab dem Jahr 2015 – deutlich angestiegen ist. Im gleichen Zeitraum stagnierten die Zahlen der Haushalte und Lebensformen ohne Einwanderungsgeschichte oder sind zurückgegangen, während die Zahl der Haushalte und Familien mit Einwanderungsgeschichte kontinuierlich angestiegen ist. Diese Entwicklung führt beispielsweise dazu, dass ein zunehmender Anteil der Familien mit drei und mehr Kindern eine Einwanderungsgeschichte hat.

Haushalte mit Einwanderungsgeschichte haben häufiger und mehr Kinder als Haushalte ohne Einwanderungsgeschichte, was sich auch auf das Erwerbsverhalten auswirken dürfte. Tatsächlich ist die Erwerbstätigkeit von eingewanderten Müttern und Vätern geringer als die von Vätern und Müttern ohne Einwanderungsgeschichte. Insbesondere bei eingewanderten Müttern ist der Abstand deutlich, was im Hinblick auf gleichstellungspolitische Themen sowie den Arbeits- und Fachkräftemangel von Bedeutung ist. Ein Vergleich der Erwerbskonstellationen von Paaren veranschaulicht, dass bei Paarfamilien mit Einwanderungsgeschichte das traditionelle männliche Ernährermodell noch stärker verbreitet ist. Die (relativ wenigen) Mütter mit Einwanderungsgeschichte, die eine bezahlte Tätigkeit ausüben, tun dies dafür mit einem etwas höheren Stundenumfang als Mütter ohne Einwanderungsgeschichte.

Die Unterschiede bei der Kinderzahl als einen möglichen Erklärungsansatz für die geringere Erwerbstätigkeit von eingewanderten Müttern konnte die vorgelegte Analyse allenfalls teilweise bestätigen. Zwar steigen die Unterschiede mit wachsender Kinderzahl an, was

aber nicht bedeutet, dass diese mit unterschiedlichen Familienstrukturen erklärt werden können. Als weitere Erklärung werden in der Literatur etwa die geringere Inanspruchnahme von Angeboten der Kinderbetreuung, unterschiedliche Voraussetzungen bei Qualifikation und Arbeitsmarktzugang, aber auch Diskriminierung bei der Arbeitssuche genannt. Gerade bei eingewanderten Frauen stellt sich dabei die Frage, ob Mehrfachdiskriminierungen vorliegen könnten, die den Unterschied zu eingewanderten Vätern erklären (für einen Überblick siehe Heath/Cheung, 2007; Kalter/Granato, 2018; Kalter, 2022). Zur Relevanz dieser möglichen Ursachen lassen sich auf Basis der hier untersuchten Ergebnisse allerdings kaum Aussagen treffen.

Ein sinnvoller Ansatzpunkt für weitere Untersuchungen mit dem Mikrozensus wäre, im Rahmen der durch den Stichprobenumfang gesetzten Möglichkeiten stärker nach Herkunftsländern beziehungsweise Herkunftsregionen sowie der Aufenthaltsdauer zu differenzieren. Damit könnten die Gruppen besser identifiziert werden, bei denen die Erwerbstätigkeit besonders niedrig ist. 

## LITERATURVERZEICHNIS

---

Canan, Coşkun/Petschel, Anja. *Die Umsetzung des Konzepts „Einwanderungsschichte“ im Mikrozensus 2022*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2023, Seite 61 ff.

Ette, Andreas/Mundil-Schwarz, Rabea/Sauer, Lenore/Sulak, Harun. *Ein neues Bild der Migration: Sozioökonomische Struktur und Arbeitsmarktintegration von Neuzuwanderern aus Drittstaaten in Deutschland*. In: Bevölkerungsforschung Aktuell 02/2013, Seite 2 ff. [Zugriff am 21. Juni 2024]. Verfügbar unter: [www.verband-binationaler.de](http://www.verband-binationaler.de)

Fachkommission Integrationsfähigkeit. *Gemeinsam die Einwanderungsgesellschaft gestalten*. Bericht der Fachkommission der Bundesregierung zu den Rahmenbedingungen der Integrationsfähigkeit. 2021. [Zugriff am 28. Juni 2024]. Verfügbar unter: [www.fachkommission-integrationsfaehigkeit.de](http://www.fachkommission-integrationsfaehigkeit.de)

Feldhoff, Charlotte H. *The Child Penalty: Implications of Parenthood on Labour Market Outcomes for Men and Women in Germany*. SOEPpapers on Multidisciplinary Panel Data Research Nr. 1120/2021. [Zugriff am 28. Juni 2024]. Verfügbar unter: [www.diw.de](http://www.diw.de)

Grigoriev, Olga/Körner, Thomas. *Eingewanderte und ihre Nachkommen*. In: Sozialbericht. Bundeszentrale für politische Bildung. 2024 (in Vorbereitung).

Heath, Anthony F./Cheung, Sin Y. *The Comparative Study of Ethnic Minority Disadvantage*. In: Heath, Anthony F./Cheung, Sin Y. (Herausgeber). *Unequal Chances: Ethnic Minorities in Western Labour Markets*. London 2007. DOI: [10.5871/bacad/9780197263860.003.0001](https://doi.org/10.5871/bacad/9780197263860.003.0001)

Heilmann, Heike. *Kindertagesbetreuung*. In: [Datenreport 2021. Ein Sozialbericht für die Bundesrepublik Deutschland](#). Herausgeber: Statistisches Bundesamt, Wissenschaftszentrum Berlin für Sozialforschung, Bundesinstitut für Bevölkerungsforschung. Bonn 2021, Seite 65 ff.

Hermes, Henning/Lergetporer, Philipp/Mierisch, Fabian/Peter, Frauke/Wiederhold, Simon. *Discrimination in Universal Social Programs? A Nationwide Field Experiment on Access to Child Care*. In: CESifo Working Paper. Ausgabe 10368/2024. [Zugriff am 28. Juni 2024]. Verfügbar unter: [www.cesifo.org](http://www.cesifo.org)

Hochgürtel, Tim. *Realisierte Erwerbstätigkeit zur Messung des Vereinbarkeitsarrangements von Familie und Beruf*. In: WISTA Wirtschaft und Statistik. Ausgabe 1/2018, Seite 23 ff.

Hochgürtel, Tim/Wilke, Clarissa. *Methodische Weiterentwicklungen in der Lebensformenberichterstattung auf Grundlage des Mikrozensus ab dem Berichtsjahr 2020*. In: WISTA Wirtschaft und Statistik. Ausgabe 1/2022, Seite 97 ff.

ILO (International Labour Organization). *Resolution concerning statistics of work, employment and labour underutilization (including amendments)*. 2023. [Zugriff am 26. Juni 2024]. Verfügbar unter: [www.ilo.org](http://www.ilo.org)

## LITERATURVERZEICHNIS

---

- Jähnert, Alexandra. *Lebenswelten von jungen Menschen mit Migrationshintergrund: Familialer Alltag und Freizeitgestaltung*. In: Lochner, Susanne/Jähnert, Alexandra (Herausgeber). DJI-Kinder- und Jugendmigrationsreport 2020. Datenanalyse zur Situation junger Menschen in Deutschland. Bielefeld 2020. Seite 48 ff. [Zugriff am 21. Juni 2024]. Verfügbar unter: [www.dji.de](http://www.dji.de)
- Jessen, Jonas/Schmitz, Sophia/Waights, Sevrin. *Understanding Day Care Enrolment Gaps*. Deutsches Institut für Wirtschaftsforschung – DIW Berlin. Discussion Paper 1808. 2020. [Zugriff am 27. Juni 2024]. Verfügbar unter: [www.diw.de](http://www.diw.de)
- Kalter, Frank/Granato, Nadia. *Migration und ethnische Ungleichheit auf dem Arbeitsmarkt*. In: Abraham, Martin/Hinz, Thomas (Herausgeber). Arbeitsmarktsoziologie. 3. Auflage. Wiesbaden 2018. Seite 355 ff. DOI: [10.1007/978-3-658-02256-3\\_10](https://doi.org/10.1007/978-3-658-02256-3_10)
- Kalter, Frank. *Integration in migration societies*. In: Gërkhani, Klarita/de Graaf, Nan Dirk/Raub, Werner (Herausgeber). Handbook of Sociological Science: Contributions to Rigorous Sociology. Cheltenham 2022. Seite 135 ff. DOI: [10.4337/9781789909432.00016](https://doi.org/10.4337/9781789909432.00016).
- Keller, Matthias/Körner, Thomas. *Closing the gap? Erwerbstätigkeit und Arbeitszeit von Müttern und Vätern nach 15 Jahren Elterngeld*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2023, Seite 88 ff.
- Lokhande, Mohini. *Integrationsmotor Kita. Wie gut ist die frühkindliche Betreuung auf den Normalfall Vielfalt eingestellt?* SVR-Kurzinformation 4-2023. Berlin 2023. [Zugriff am 27. Juni 2024]. Verfügbar unter: [www.svr-migration.de](http://www.svr-migration.de)
- Naderi, Robert. *Kinderzahl und Migrationshintergrund*. In: Zeitschrift für Familienforschung. Jahrgang 27. Heft 3/2015, Seite 322 ff. [Zugriff am 21. Juni 2024]. Verfügbar unter: [www.bib.bund.de](http://www.bib.bund.de)
- Nöthen, Manuela. *Von der „traditionellen Familie“ zu „neuen Lebensformen“*. In: Wirtschaft und Statistik 1/2005. Seite 25 ff.
- Schäper, Clara/Schrenker, Annekatrin/Wrohlich, Katharina. *Gender Pay Gap und Gender Care Gap steigen bis zur Mitte des Lebens stark an*. In: DIW Wochenbericht. Ausgabe 9/2023, Seite 100 ff. [Zugriff am 28. Juni 2024]. Verfügbar unter: [www.diw.de](http://www.diw.de)
- Schmitz, Sophia/Spieß, C. Katharina/Huebener, Mathias. *Weiterhin Ungleichheiten bei der Kita-Nutzung*. In: Bevölkerungsforschung aktuell. Jahrgang 44. Ausgabe 2/2023, Seite 3 ff. [Zugriff am 27. Juni 2024]. Verfügbar unter: [www.bib.bund.de](http://www.bib.bund.de)
- Trappe, Heike/Pollmann-Schult, Matthias/Schmitt, Christian. *The Rise and Decline of the Male Breadwinner Model: Institutional Underpinnings and Future Expectations*. In: European Sociological Review. Jahrgang 31. Ausgabe 2/2015, Seite 230 ff. DOI: [10.1093/esr/jcv015](https://doi.org/10.1093/esr/jcv015)

#### Dr. Alejandro Donado

ist promovierter Volkswirt und wissenschaftlicher Mitarbeiter im Referat „Investitionen“ des Statistischen Bundesamtes. Zu seinen Aufgaben gehört die methodische Weiterentwicklung der Berechnung von Investitionen in Daten als eigenständiges Wirtschaftsgut.

#### Benedikt Kuckelkorn

ist Volkswirt und war bis Ende Mai 2023 wissenschaftlicher Mitarbeiter im Referat „Vermögensrechnung, Internationale VGR Methodik“ des Statistischen Bundesamtes. Seit Juni 2023 ist er Referent im Referat „Makroökonomische Globalisierungsfragen, Bruttonationaleinkommen“.

#### Yannik Weigelt

ist M. Sc. der Wirtschaftswissenschaften und Referent im Referat „Investitionen“ des Statistischen Bundesamtes. Seine Tätigkeitsschwerpunkte liegen in der konzeptionellen und methodischen Weiterentwicklung der Berechnung von Investitionen in sonstige Anlagen sowie in der Modernisierung der Volkswirtschaftlichen Gesamtrechnungen.

#### Daniel Seeger

ist Diplom-Volkswirt und leitet das Referat „Investitionen“ des Statistischen Bundesamtes. Sein Aufgabenschwerpunkt liegt auf der konzeptionellen und methodischen Weiterentwicklung der Investitionsrechnungen als Bestandteil des Bruttoninlandsproduktes.

# DATEN ALS EIGENSTÄNDIGES VERMÖGENSGUT IN DEN VOLKSWIRTSCHAFTLICHEN GESAMTRECHNUNGEN – EIN WERKSTATTBERICHT

Alejandro Donado, Benedikt Kuckelkorn, Yannik Weigelt, Daniel Seeger

🔗 **Schlüsselwörter:** Digitalisierung – SNA 2025 – Bruttoinlandsprodukt – Investitionen – geistiges Eigentum

## ZUSAMMENFASSUNG

In einer zunehmend digitalisierten Welt gewinnen Daten als eigenständiges Vermögensgut erheblich an Bedeutung. Die aktuelle Revision des System of National Accounts der Vereinten Nationen empfiehlt, Daten als Vermögensgut in das Bruttoinlandsprodukt zu integrieren. Das Statistische Bundesamt untersucht, wie Investitionen in Daten und darauf anfallende Abschreibungen erfasst werden könnten. Neben einem möglichen Berechnungsansatz werden auch die Grenzen und Schwächen des aktuell empfohlenen Rechenmodells diskutiert, die eine Integration in das Bruttoinlandsprodukt und damit die Volkswirtschaftlichen Gesamtrechnungen herausfordernd machen.

🔗 **Keywords:** digitalisation – SNA 2025 – gross domestic product – investments – intellectual property

## ABSTRACT

*In our increasingly digitalised world, data are growing in importance as an asset in their own right. The recommendation to integrate data as an asset in the gross domestic product (GDP) has arisen from the current revision of the United Nations System of National Accounts. The Federal Statistical Office is currently investigating how investments in data and associated depreciation could be captured. In addition to a possible calculation approach, the article also discusses the limitations of the currently recommended calculation model. These weaknesses represent the challenges involved in integrating the assets in GDP and national accounts.*

## 1

### Einleitung

Mit der stetig zunehmenden Digitalisierung wächst die Bedeutung von elektronisch gehaltenen Daten als eigenständiges Vermögensgut. Gelegentlich werden Daten sogar als das neue Öl des 21. Jahrhunderts bezeichnet (The Economist, 2017) und damit als der treibende Wirtschaftsfaktor überhaupt. Daten spielen eine wesentliche Rolle bei der Ausrichtung der Produktionsprozesse, der Lieferketten und der Erschließung von neuen Geschäftsfeldern. Als Anlagegüter werden sie langfristig und wiederholt in Produktionsprozessen eingesetzt und generieren über einen längeren Zeitraum wirtschaftlichen Nutzen. Daher investieren Unternehmen erhebliche Ressourcen, um Zugriff auf für sie relevante Daten zu bekommen oder eigene Daten zu produzieren.

Auch die Umsätze durch die Nutzung von Daten wachsen stetig. Im Jahr 2015 erwirtschaftete beispielsweise Google einen Jahresumsatz in Höhe von rund 75 Milliarden US-Dollar. Im Jahr 2023 stieg dieser auf rund 307 Milliarden US-Dollar an. Mehr als 77% dieser Einnahmen stammen aus dem Werbegeschäft (Alphabet, 2024). Durch die Analyse von Nutzendendaten kann Google personalisierte Werbeanzeigen schalten, die erfolgreicher sind als solche ohne entsprechende Personalisierung. Insgesamt führt der rasante Ausbau von Netzwerktechnologien, mobilen Datenangeboten oder auch Social-Media-Aktivitäten zu einer unaufhaltsamen Zunahme an Datenquellen und -mengen. Einer Schätzung der International Data Corporation zufolge betrug das Gesamtvolumen der im Jahr 2012 erzeugten Daten 6,5 Zettabyte.<sup>1</sup> Im Jahr 2020 erreichte die Datenproduktion bereits ein Volumen von 64,2 Zettabyte, im Jahr 2025 soll es 180 Zettabyte überschreiten (Reichen- tal, 2022).

Aufgrund dieser wachsenden Bedeutung von elektronischen Daten hat die Statistische Kommission der Vereinten Nationen das Thema auf die Forschungsagenda der laufenden Revision des System of National Accounts (SNA – System der Volkswirtschaftlichen Gesamtrechnungen) gesetzt. Dahinter steht die Frage, wie der Produktionsfaktor Daten in den Volkswirtschaftlichen

Gesamtrechnungen (VGR) als neuer Baustein des sogenannten geistigen Eigentums (neben beispielsweise Forschung und Entwicklung [FuE] oder Urheberrechten) explizit abgebildet werden kann. Im März 2024 hat die Statistische Kommission der Vereinten Nationen entschieden, das Thema zur Aufnahme in das neue SNA 2025 zu empfehlen.

Das Statistische Bundesamt führt seit 2022 mit finanzieller Unterstützung durch das Statistische Amt der Europäischen Union (Eurostat) Testrechnungen zu dieser Thematik durch. Die Bedeutung dieser Arbeiten ist durch die voraussichtliche Aufnahme in das SNA 2025 gestiegen, denn damit stellt sich auch die Frage nach der Aufnahme in das revidierte Europäische System Volkswirtschaftlicher Gesamtrechnungen (ESVG). Das ESGV wiederum würde die deutschen VGR verpflichten, entsprechende Rechenbausteine zu integrieren.

Der Beitrag skizziert zunächst die derzeitige Vorgehensweise im Bereich des geistigen Eigentums in den deutschen VGR (Kapitel 2). Die Berechnungen von Investitionen in sowie von Abschreibungen und dem Anlagevermögen von Daten sind Thema in Kapitel 3. Danach stellt Kapitel 4 die Projektergebnisse vor. Abschließend erfolgen eine kritische Bewertung der bisherigen Untersuchungsergebnisse sowie ein Ausblick auf das weitere Vorgehen.

## 2

### Aktuelle Berechnung des geistigen Eigentums

Als Kernkomponente der sonstigen Anlagen im Bereich der Bruttoanlageinvestitionen umfasst das geistige Eigentum derzeit Investitionen in FuE, Software und Datenbanken, Urheberrechte sowie Suchbohrungen.

➤ **Grafik 1.** Die Investitionen in FuE haben dabei die größte Bedeutung, deren Anteil am geistigen Eigentum beträgt rund drei Viertel.

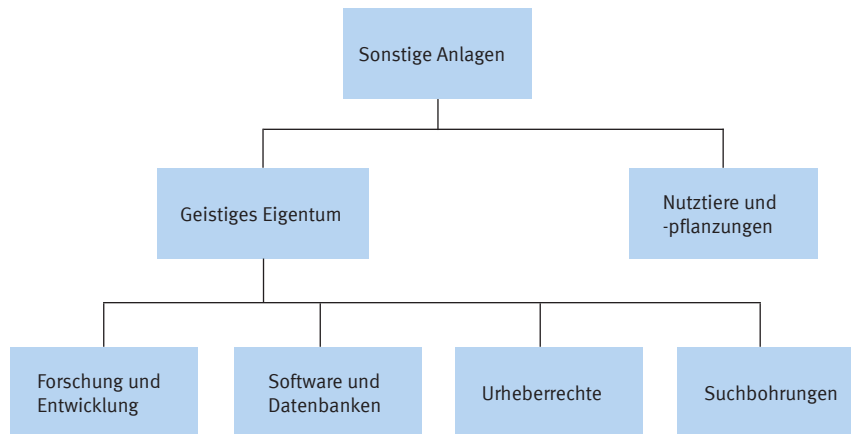
Die Berechnung der **Investitionen in FuE** folgt dem Kostensummenansatz (Adler und andere, 2014). Als zentrale Datenquelle dient die FuE-Erhebung des Stifterverbands für die Deutsche Wissenschaft. Da die FuE-Statistik unter anderem detaillierte Informationen zu internen und externen FuE-Aufwendungen sowie FuE-

1 Ein Zettabyte entspricht einer Billion Gigabytes.



**Grafik 1**

Sonstige Anlagen und geistiges Eigentum mit seinen Bestandteilen



Personal zur Verfügung stellt, lassen sich die relevanten Kostenkomponenten identifizieren. Nach der darauf aufbauenden Ermittlung der Kostensummen werden zudem die Käufe und Verkäufe an FuE-Leistungen zwischen den Wirtschaftszweigen sowie dem Ausland betrachtet, um den Wert der Investitionen in FuE zu bestimmen.<sup>12</sup>

**Investitionen in Software und Datenbanken** werden für selbsterstellte und erworbene Software separat betrachtet. Als Grundlage für die Messung und Bewertung der selbsterstellten Software werden im Rechenmodell zunächst die Beschäftigten in relevanten Berufsgattungen identifiziert. Für die so identifizierten Berufsgattungen werden Annahmen getroffen, zu welchem prozentualen Anteil ihrer Arbeitszeit sie investive Software erstellen. Diese Anteile werden an die Bruttogehälter der Beschäftigten angelegt, um die anfallenden Personalkosten zu ermitteln. Ergänzt um einen Gewinnzuschlag werden auch hier die Investitionen in selbsterstellte Software dem Kostensummenansatz folgend berechnet. Für den Erwerb von Software stehen dagegen für viele Wirtschaftszweige Marktpreise als Erhebungsdaten zur Verfügung.

Bei der Berechnung von FuE und Software ist das Problem einer möglichen Doppelzählung zu berücksichtigen: Einerseits kann FuE durchgeführt werden, um Software zu produzieren. Andererseits kann Software verwendet werden, um FuE zu realisieren. Um eine

Doppelzählung von FuE und selbsterstellter Software zu vermeiden, werden etwa 10 % der selbsterstellten Software von den internen FuE-Aufwendungen abgezogen.

**Investitionen in Urheberrechte** umfassen unter anderem Originalfilme und Tonträgeroriginale sowie die Urheberrechte von Schriftstellerinnen und Schriftstellern. Die Ermittlung des Werts der Urheberrechte basiert auf einer Vielzahl unterschiedlicher amtlicher und nicht amtlicher Quellen wie den Geschäftsberichten verschiedener Verwertungsgesellschaften.

Weiterhin zählen **Suchbohrungen zur Erforschung von Bodenschätzen** zum geistigen Eigentum. Als Wert der Investitionen werden die Kosten der eigentlichen Suchbohrungen sowie die Nebenkosten, wie Luftbilder oder Vermessungsarbeiten, einbezogen.

Die Hinzunahme von Daten als Vermögensgut kann zu weiteren Überschneidungen im Bereich des geistigen Eigentums führen, wodurch die Gefahr von Doppelerfassungen steigt. Diese Sachverhalte sind bei der Ermittlung von Investitionen in Daten zu beachten und bestmöglich zu vermeiden. Insbesondere bei der Abgrenzung von Software sowie Datenbanken und Daten ist darüber hinaus eine Änderung der bestehenden Definitionen und Abgrenzungen notwendig.

<sup>12</sup> Die Berechnungsschritte folgen den Vorgaben des Frascati-Handbuchs (OECD, 2018).



### 3

## Methodik zur Berechnung von Investitionen in Daten

### 3.1 Bewertung von Daten als Vermögensgut

Ausgangspunkt der durchgeführten Untersuchungen stellte die sogenannte Guidance Note DZ.06 der Arbeitsgruppe zur Digitalisierung der Statistischen Kommission der Vereinten Nationen dar (Vereinte Nationen, 2023). Darin werden Daten definiert als „Informationsinhalte, die durch den Zugang zu und die Beobachtung von Phänomenen sowie durch die Aufzeichnung, Organisation und Speicherung von Informationselementen aus diesen Phänomenen in einem digitalen Format erzeugt werden, die bei der Verwendung in produktiven Tätigkeiten einen wirtschaftlichen Nutzen erbringen“.

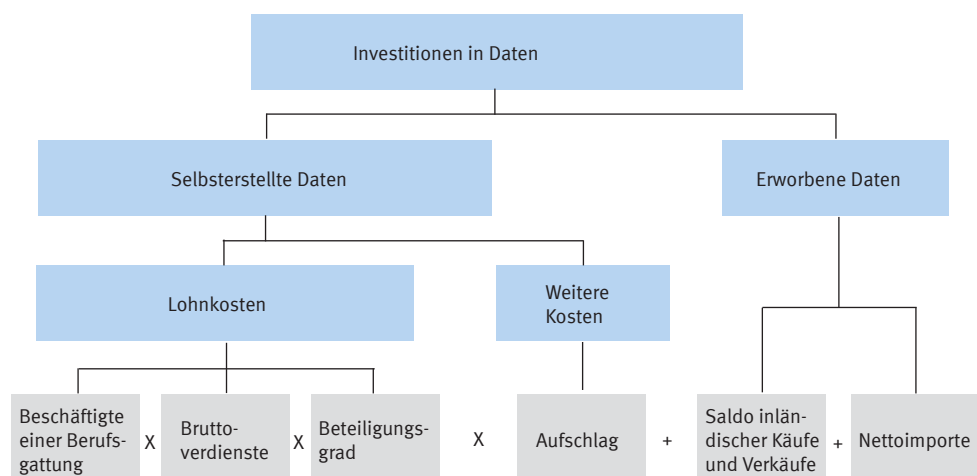
Für die Bewertung von Vermögensgütern sieht das SNA drei zulässige Methoden vor: den Kostensummenansatz, die Barwertmethode oder tatsächliche Marktpreise. Da keine ausreichenden statistischen Informationen zu Transaktionen in Daten vorliegen und Daten häufig selbst von Unternehmen erstellt werden, empfiehlt die Guidance Note, Daten anhand des Kostensummenansatzes zu bewerten (Vereinte Nationen, 2023). Der Kostensummenansatz wird in der deutschen und

der internationalen VGR bereits unter anderem für die Berechnung der Investitionen in FuE und Software verwendet (siehe Kapitel 2). Die Ableitung der einzelnen Kostenfaktoren ist für die Produktion von Daten jedoch um einiges schwieriger: Einerseits mangelt es an einer vergleichbaren Quelle wie der FuE-Statistik, andererseits werden Daten nicht wie Software vor allem gezielt von Beschäftigten in klassischen IT-Berufen aktiv produziert. Datenproduktion findet heutzutage in nahezu allen Berufsfeldern statt, oftmals als Nebenprodukt der Kern-tätigkeiten. Auch werden Daten zu einem großen Anteil automatisiert durch verschiedenste Anwendungen oder KI-Systeme produziert. Beispielsweise setzt die Deutsche Bahn KI-basierte Systeme ein, die kontinuierlich Daten über den Zustand und die Leistung der Züge und Infrastruktur generieren, um vorausschauende Wartung zu ermöglichen (Deutsche Bahn, 2024). Selbst unter der Annahme, dass ein Großteil der generierten Daten nicht zu Produktionszwecken eingesetzt wird, stellt der Kostensummenansatz von den nach SNA zulässigen Methoden zwar den aktuell empfohlenen Ansatz dar. Die dadurch vorgegebene Beschränkung auf Arbeitskosten, die durch gezielte menschliche Datenerzeugung entstehen, kann das tatsächliche Produktionspotenzial von Daten jedoch kaum erfassen.

Das auf dem Kostensummenansatz basierende Berechnungsmodell geht davon aus, dass sich Investitionen in Daten aus der Summe der im Inland selbstgestellten Daten zuzüglich der erworbenen Daten zusammensetzen. [➤ Grafik 2](#)

**Grafik 2**

Auf dem Kostensummenansatz basierendes Berechnungsmodell zur Berechnung von Daten als Vermögensgut



### 3.2 Selbsterstellte Daten

---

Die Ermittlung des Produktionswertes selbsterstellter Daten nach dem empfohlenen Kostensummenansatz erfolgt in mehreren Schritten.

Zur Berechnung der Lohnkosten ist es erforderlich, alle an der Produktion von Daten beteiligten Beschäftigten zu identifizieren. Dazu wurde die in Deutschland gültige Klassifikation der Berufe (KldB) verwendet (Bundesagentur für Arbeit, 2021). Die KldB klassifiziert Berufe in etwa 1 300 Berufsgattungen und beschreibt die jeweiligen Tätigkeiten und Fähigkeiten. Mithilfe qualitativer Analysen der einzelnen Berufsbeschreibungen wurden 249 Berufsgattungen ausgewählt, die mit Datenproduktion ganz oder zum Teil verbunden scheinen. Die Anzahl der in den ausgewählten Berufen Beschäftigten stammt aus dem jährlichen Mikrozensus (Statistische Ämter des Bundes und der Länder, 2024). Für den durchschnittlichen Bruttoverdienst der Beschäftigten wurde die Verdienststrukturerhebung des Statistischen Bundesamtes genutzt (Statistisches Bundesamt, 2024). Die zuvor nur alle vier Jahre durchgeführte Verdienststrukturerhebung wurde im Jahr 2022 durch die monatliche Verdiensterhebung ersetzt (Finke und andere, 2023). Für die Modellierung wurden die Daten der Jahre 2014 und 2018 verarbeitet, für die übrigen Jahre wurden Interpolations- und Extrapolationsverfahren angewendet.

Da Beschäftigte nicht die komplette Arbeitszeit für die Erstellung von Daten verwenden können, sieht das Modell einen sogenannten Beteiligungsgrad vor, der den Anteil der Datenproduktion an der Arbeitszeit widerspiegeln soll. Der Beteiligungsgrad ist ein zentraler Baustein des Rechenmodells, er hat maßgeblichen Einfluss auf das Gesamtvolumen selbsterstellter Daten. Aufgrund des Mangels an quantitativen Daten wurden die Beteiligungsgrade für jede Berufsgattung ausschließlich auf Basis einer qualitativen Analyse der Berufsbeschreibungen und Tätigkeitsprofile festgelegt. Diese schwache Fundierung führt daher zu großen Unsicherheiten in weiteren Analysen.

Die Lohnkosten der Datenproduktion werden ermittelt, indem die Anzahl der Beschäftigten, ihre durchschnittlichen Bruttoverdienste und die angenommenen Beteiligungsgrade für jede Berufsgattung miteinander multipliziert werden. Anschließend werden diese Werte über alle Berufsgattungen hinweg aggregiert.

Neben den Lohnkosten für die Beschäftigten sind Ausgaben der Unternehmen für Sozialabgaben, Vorleistungen, Kapitalkosten gemessen durch Abschreibungen und Nettoproduktionsabgaben im Kostensummenansatz zu berücksichtigen. Da auch für diese Größen keine statistischen Informationen vorliegen, werden sie gemeinsam als Aufschlag auf die Lohnkosten erfasst. Unter der Annahme, dass Produktionsprozesse im IT-Sektor eine vergleichbare Kostenstruktur zur Datenproduktion aufweisen, wurde der Aufschlag durch das Verhältnis der genannten weiteren Kosten zur gezahlten Lohnsumme in den Wirtschaftszweigen 62 „Erbringung von Dienstleistungen der Informationstechnologie“ und 63 „Informationsdienstleistungen“ ermittelt.<sup>13</sup> Bei Marktproduzenten wird zusätzlich ein Gewinnzuschlag auf der Grundlage der von der Deutschen Bundesbank veröffentlichten Jahresabschlussstatistik der Unternehmen berücksichtigt (Deutsche Bundesbank, 2023).

### 3.3 Vermeidung von Doppelerfassungen

---

Investitionen in Daten stellen eine Erweiterung der Investitionen in geistiges Eigentum dar. Die Datenproduktion ist sehr eng mit der Erstellung von Software und auch mit FuE verbunden. Wie zuvor beschrieben, musste bereits die Einführung von FuE zur Generalrevision 2014 der Volkswirtschaftlichen Gesamtrechnungen von Maßnahmen begleitet werden, die Doppelzählungen im Bereich des geistigen Eigentums vermeiden. Das Risiko von Doppelerfassungen erhöht sich bei der Betrachtung von Daten.

Das Rechenmodell der selbsterstellten Software diene als Ausgangspunkt der Untersuchungen zur Abbildung von selbsterstellten Daten. Um eine Doppelerfassung zu vermeiden, wurde bei Berufen, die sowohl der Software- als auch der Datenproduktion zugerechnet werden können, darauf geachtet, dass die Summe beider Beteiligungsgrade unter 100 % liegt.

Weiterhin wurden Doppelerfassungen zwischen der Datenproduktion und FuE untersucht. Dabei wurden 30 Berufe identifiziert, die beiden Bereichen zugerechnet werden können. Unter Berücksichtigung der jeweiligen Beteiligungsgrade errechnet sich für diese Berufe ein Investitionsvolumen in Daten, das etwa 8 % der bishe-

---

3 Klassifikation der Wirtschaftszweige, Ausgabe 2008 (Statistisches Bundesamt, 2008).

gen Investitionen in FuE entspricht und künftig eher der Datenproduktion zuzurechnen wäre.

Die für das SNA 2025 vorgesehene Aufteilung des geistigen Eigentums in FuE, Software, Daten und Datenbanken, Urheberrechte und Suchbohrungen erhöht aufgrund der zum Teil sehr schwierigen Abgrenzung das Risiko von Doppelerfassungen weiter. Die gemeinsame Betrachtung der Teilkomponenten FuE, Software sowie Daten und Datenbanken als ein Investitionsbereich könnte hier Abhilfe schaffen, da weitere, wenig fundierte Annahmen zur Aufteilung vermieden würden.

### 3.4 Erworbene Daten

---

Erworbene Daten ergeben sich aus dem Saldo von inländischen Käufen und Verkäufen sowie den Nettoimporten von Daten.

Im Gegensatz zu Investitionen in Software, für die beispielsweise statistische Angaben aus der Investitions-erhebung im Verarbeitenden Gewerbe vorliegen, fehlen Informationen für inländische Käufe und Verkäufe von Daten. Sinnvolle Schätzungen waren nicht möglich, wodurch eine Lücke in der Modellrechnung besteht. Hier bedarf es künftig zwingend belastbarer Erhebungsdaten. Es ist jedoch wahrscheinlich, dass der Zugang zu externen Daten hauptsächlich über Lizenzgebühren erfolgt, die nicht als Investitionen gelten. Tatsächlich zeigt eine Erhebung von Eurostat, dass 15,3 % der Großunternehmen in Europa Datenanalysen nutzen, aber nur 1,3 % diese Daten kaufen (OECD, 2022).

Die Nettoimporte ergeben sich aus der Differenz zwischen den aus dem Ausland erworbenen und den ins Ausland verkauften Daten. Als Obergrenze für die Datenimporte und -exporte dienen in der Modellrechnung die Importe und Exporte von IT-Dienstleistungen (Deutsche Bundesbank, 2024). Dabei wurde angenommen, dass der Saldo der Nettoimporte von IT-Dienstleistungen dem der Daten entspricht.

### 3.5 Abschreibungen und Anlagevermögen

---

Daten als Vermögensgut zu behandeln impliziert, dass Daten Bestandteil des Anlagevermögens sind und abzuschreiben sind. Abschreibungen in den VGR messen

die Wertminderung von Anlagegütern durch normalen Verschleiß, wirtschaftliches Veralten und Unfallschäden (Eurostat, 2014). In den VGR werden Vermögensbestände und Abschreibungen in der Regel mithilfe der Kumulationsmethode (Perpetual Inventory Method – PIM) berechnet. Dabei werden Investitionsströme unter Annahme einer durchschnittlichen ökonomischen Nutzungsdauer sowie geeigneter Abschreibungs- und Abgangsfunktionen über die Zeit akkumuliert, um den Wert der Bestände an und Abschreibungen auf Vermögensgüter zu bestimmen. Dazu werden lange Investitionsreihen benötigt.

Im Gegensatz zu materiellen Vermögenswerten unterliegen Daten keinem physischen Verschleiß. Grundsätzlich ist es möglich, Daten unbegrenzt zu speichern und zu nutzen. Der Wert der Daten kann dennoch im Lauf der Zeit durch abnehmende Relevanz, Datenverfall oder Datenverlust sinken. Gleichzeitig kann der Wert bestimmter Datensätze durch Kombination mit anderen Datensätzen oder durch innovative Geschäftsmodelle ansteigen. Weiterhin gibt es verschiedene Arten von Daten, deren ökonomische Nutzungsdauern stark variieren können. Diese Merkmale stellen besondere Herausforderungen dafür dar, ein Abschreibungsprofil, also Nutzungsdauern und Abschreibungsfunktion, für Daten festzulegen.

Daher wurde nicht eine Kombination von Parametern und Funktion festgelegt, um die Kumulationsmethode anzuwenden, sondern es wurden verschiedene Modellspezifikationen getestet, um die Sensitivität der Ergebnisse in Bezug auf die Parameter zu bestimmen.

Das erste Modell kombiniert lineare Abschreibungen mit einer glockenförmigen Abgangsfunktion, genauer gesagt der Dichtefunktion der Gammaverteilung. Dieses Modell folgt den Vorgaben des ESVG 2010 und ist die Standardspezifikation in den deutschen VGR (Schmalwasser/Schidlowski, 2006).

Bei den weiteren Modellen handelt es sich um verschiedene Spezifikationen geometrisch-degressiver Abschreibungsmodelle. Bei diesen Modellen ergibt sich der jährliche Abschreibungsbetrag als fester Prozentsatz des Restwerts des Vermögenswerts am Anfang des Jahres. Der Abschreibungsbetrag nimmt also jedes Jahr ab, da der Restwert über die Zeit sinkt. Im Gegensatz hierzu wird bei linearen Abschreibungen jährlich ein fester Prozentsatz, gegeben durch  $1/\text{Nutzungsdauer}$  des Vermö-

gensgegenstands, des Anfangswertes abgeschrieben. Bei Modell 2 entspricht der Abschreibungsprozentsatz dem doppelten des linearen Abschreibungsprozentsatzes. Modell 3 schreibt mit dem 1,65-Fachen des linearen Prozentsatzes ab. Diese Rate nutzen das U.S. Bureau of Economic Analysis und Statistics Canada, um die Abschreibungen für Software und Datenbanken sowie FuE zu berechnen (Statistics Canada, 2015). Das geometrisch-degressive Abschreibungsmodell kombiniert Abschreibungen und Abgänge, sodass keine Abgangsfunktion definiert werden muss. Das ESVG 2010 lässt geometrische Abschreibungen ebenfalls zu.

Da zur ökonomischen Nutzungsdauer von Daten ebenfalls keine Informationen vorliegen, wurden durchschnittliche Nutzungsdauern von zwei, fünf und zehn Jahren angenommen. Diese stehen im Einklang mit den Nutzungsdauern anderer immaterieller Investitionsgüter in den deutschen VGR. Die sehr kurze Nutzungsdauer von zwei Jahren drückt die Möglichkeit aus, dass ein großer Teil der Daten weniger als ein Jahr in der Produktion genutzt werden könnte und daher nicht als Kapitalbildung erfasst werden sollte. Die Annahme einer derart kurzen Nutzungsdauer kann daher dazu beitragen, den Vermögenswert an Daten nicht zu überschätzen. Für Software und Datenbanken wird aktuell eine Nutzungsdauer von fünf Jahren angenommen; diese Option spiegelt die Tatsache wider, dass Daten und Datenbanken intrinsisch verbunden sind. Die Nutzungsdauer von zehn Jahren stellt die Obergrenze für die meisten immateriellen Vermögensgüter in den deutschen VGR dar.

## 4

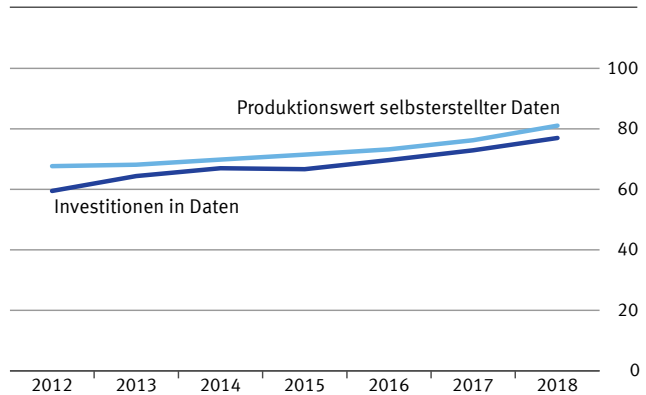
### Projektergebnisse

#### 4.1 Investitionen in Daten

Die Ergebnisse der bisherigen Modellrechnungen sind in Tabelle 1 und den Grafiken 3 bis 5 dargestellt. Die hellblaue Kurve in [Grafik 3](#) zeigt den Produktionswert selbsterstellter Daten zwischen 2012 und 2018. Unter den gewählten Annahmen für die relevanten Berufsgattungen und Beteiligungsgrade wurden im Jahr 2018 selbsterstellte Daten im Wert von 81,07 Milliarden Euro produziert. Die Beteiligungsgrade haben einen großen Einfluss auf die Ergebnisse des Rechenmodells. Eine

**Grafik 3**

**Produktionswert selbsterstellter Daten und Investitionen in Daten**  
Mrd. EUR



Erhöhung der Beteiligungsgrade über alle Berufsgattungen um 1 % würde den Produktionswert der Daten auch um 1 % steigern.

Die dunkelblaue Kurve in Grafik 3 zeigt die Investitionen in Daten. Sie resultiert aus der Addition der Nettoimporte zum Produktionswert der selbsterstellten Daten. Da die Nettoimporte in allen Jahren zwischen 2012 und 2018 negativ sind (es also weniger Importe als Exporte von Daten stattfanden), ergibt sich, dass die Investitionssumme unter dem Produktionswert liegt. Beispielsweise ergaben sich für die Investitionen im Jahr 2018 unter den vorgenommenen Annahmen 77 Milliarden Euro, während die Produktion 81 Milliarden Euro betrug. Somit wurden im Jahr 2018 rund 4 Milliarden Euro mehr Daten exportiert als im Gegenzug importiert, das heißt aus dem Ausland bezogen.

Die mithilfe des Modells berechneten Investitionen in Daten ermöglichen einen Vergleich mit den aktuellen Bruttoanlageinvestitionen in Deutschland. Die ermittelten Investitionen in Daten machen etwa 11 % der gesamten Bruttoanlageinvestitionen aus. [Tabelle 1](#)

**Tabelle 1**

Bruttoanlageinvestitionen und Investitionen in Daten  
in jeweiligen Preisen

	Bruttoanlage- investitionen	Investitionen in Daten	
	Mrd. EUR		% <sup>1</sup>
2012	557,88	59,46	11
2013	559,50	64,38	12
2014	586,67	66,97	11
2015	605,84	66,68	11
2016	636,30	69,65	11
2017	666,88	72,91	11
2018	708,94	77,00	11

1 Anteil an den Bruttoanlageinvestitionen.

## 4.2 Abschreibungen und Anlagevermögen

Wie in Abschnitt 3.5 bereits erläutert, wurden verschiedene Modelle zur Berechnung der Abschreibungen auf Daten und des Anlagevermögens ausgewählt, um die Sensitivität der Ergebnisse in Bezug auf verschiedene Parameter zu untersuchen. Diese Modelle umfassen lineare und geometrisch-degressive Abschreibungsfunktionen sowie verschiedene Annahmen zur Nutzungsdauer.

Die unterstellten Nutzungsdauern (zwei, fünf und zehn Jahre) werden mit den drei Abschreibungsmodellen kombiniert. Die resultierenden Abschreibungen und das Nettoanlagevermögen zeigt [Grafik 4](#) auf Seite 66. Die Ergebnisse legen dar, dass die Wahl des Abschreibungsmodells nur marginalen Einfluss auf die Höhe der Abschreibungen (in der Grafik links dargestellt) haben. Entscheidend für das Niveau ist vielmehr die angenommene durchschnittliche Nutzungsdauer. Auch auf die Veränderungsrate hat die Wahl des Modells im Vergleich zur Nutzungsdauer nur einen geringen Effekt.

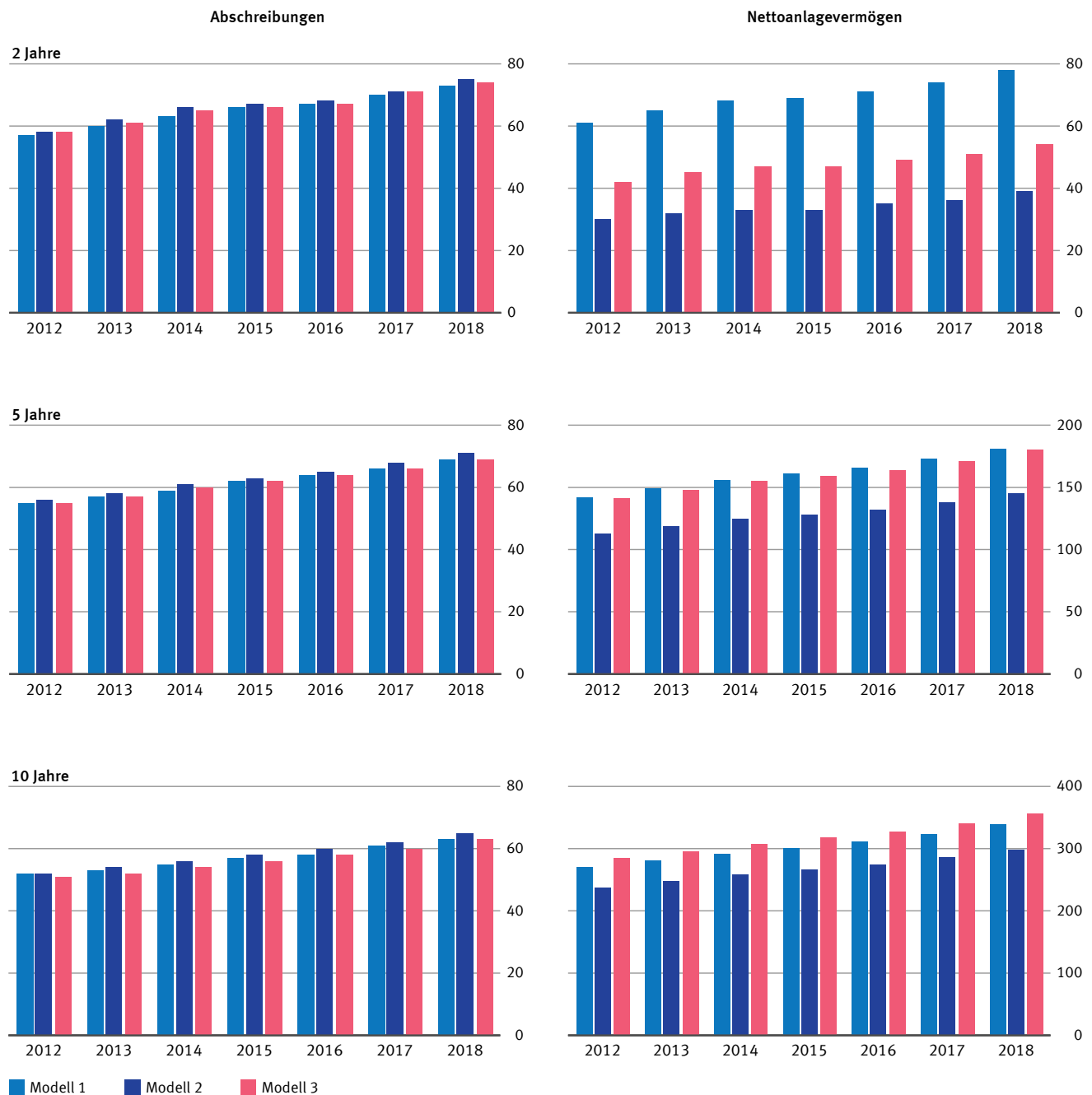
Anders sieht es für das Nettoanlagevermögen aus (in Grafik 4 rechts dargestellt). Das geometrisch-degressive Modell 2 generiert für alle angenommenen Nutzungsdauern signifikant niedrigere Vermögensbestände im Vergleich zu den anderen betrachteten Modellen. Die Differenzen werden mit steigender Nutzungsdauer allerdings geringer. Modell 3 generiert für eine Nutzungsdauer von zwei Jahren ebenfalls niedrigere Bestände im Vergleich zum linearen Abschreibungsmodell 1.

Bei einer Nutzungsdauer von fünf Jahren ergeben sich keine größeren Unterschiede zwischen den Modellen 1 und 3. Bei einer Nutzungsdauer von zehn Jahren erzielt Modell 3 sogar ein höheres Vermögen als Modell 1. Die Differenzen ergeben sich vor allem durch die Modellierung der Abgänge im linearen Modell und die sich daraus ergebenden Unterschiede im Kapitalaufbau in der Phase, bevor die Investitionszeitreihe eine ausreichende Länge erreicht hat.<sup>4</sup> Grafik 4 verdeutlicht allerdings ebenfalls, dass die angenommene Nutzungsdauer den deutlich größeren Effekt im Vergleich zur Abschreibungsmodellierung hat. Im Gegensatz zum Niveau des Vermögens ist der Effekt der Modellwahl auf die Entwicklung des Vermögens zu vernachlässigen. Hier hat wieder die Nutzungsdauer den stärkeren Effekt. Dies ist besonders relevant, da das Vermögen häufig zu Produktivitätsanalysen genutzt wird. Entscheidend ist dabei aber in der Regel die Veränderungsrate, nicht das Niveau des Vermögens.

<sup>4</sup> Unter anderem wird im linearen Abschreibungsmodell davon ausgegangen, dass es im Jahr der Investition keine Abgänge gibt. Da die geometrisch-degressiven Modelle Abschreibungen und Abgänge kombinieren, kann diese Annahme hier nicht getroffen werden. Eine ausführliche Beschreibung des linearen Modells findet sich in Schmalwasser/Schidlowski (2006).

Grafik 4

Abschreibungen auf Daten (in jeweiligen Preisen) und Nettoanlagevermögensbestand an Daten (in Wiederbeschaffungspreisen) für verschiedene durchschnittliche Nutzungsdauern und Abschreibungsfunktionen  
Mrd. EUR



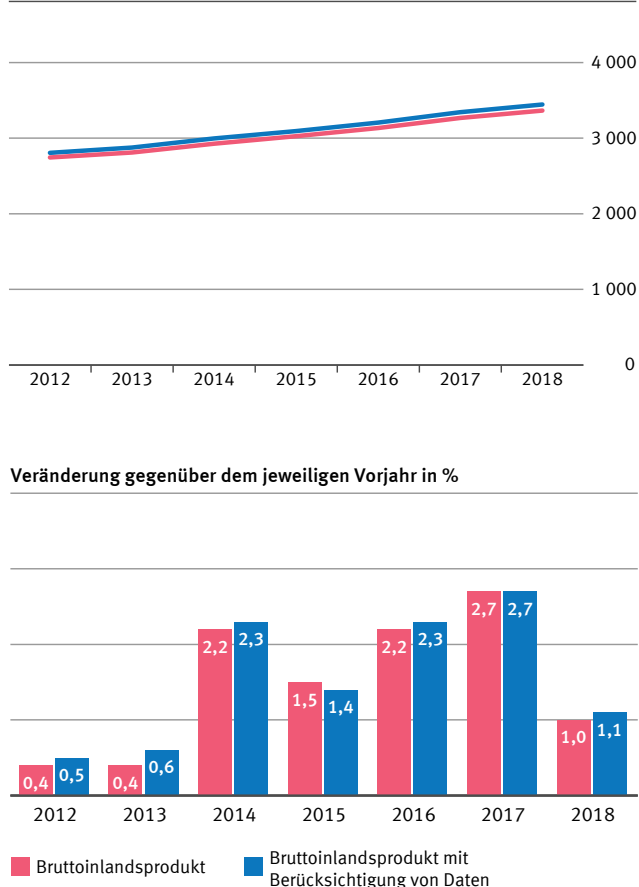


## 4.3 Auswirkungen auf das Bruttoinlandsprodukt

Es ist von entscheidender Bedeutung für die Bewertung der ökonomischen Rolle von Dateninvestitionen, wie sich diese unter den beschriebenen Modellannahmen auf das Bruttoinlandsprodukt (BIP) auswirken würden. Im oberen Teil von [Grafik 5](#) zeigen die beiden Kurven das durch die Modellrechnung ermittelte Niveau des Bruttoinlandsprodukts sowohl mit als auch ohne Berücksichtigung von Daten (gemessen in Milliarden Euro) in den Jahren von 2012 bis 2018. Die Säulen im unteren Teil der Grafik stellen die jeweiligen Veränderungsrate des preisbereinigten Bruttoinlandsprodukts dar.

Grafik 5 veranschaulicht somit zwei zentrale Ergebnisse: Erstens würde die Einbeziehung von Dateninvestitionen

**Grafik 5**  
Bruttoinlandsprodukt in jeweiligen Preisen mit und ohne Investitionen in Daten  
Mrd. EUR



unter den vorgenommenen Annahmen zu einem durchschnittlichen Anstieg des BIP-Niveaus um 2,3 % führen. Zweitens würde die Veränderungsrate des preisbereinigten Bruttoinlandsprodukts im Durchschnitt um 0,07 Prozentpunkte steigen. Beide Effekte bewegen sich in der Größenordnung der Ergebnisse anderer Staaten, die vergleichbare Studien zur Bewertung von Daten durchgeführt haben (Vereinte Nationen, 2023). Dabei ist die uneinheitliche Wirkung von Investitionen in Daten auf die Veränderungsrate des Bruttoinlandsprodukts vor allem auf die Volatilität der Nettoimporte von Daten zurückzuführen. So ist der Einfluss auf die Veränderungsrate stets positiv, außer 2015, als die Nettoimporte stark zurückgingen.

Es ist zu beachten, dass die Ergebnisse auf der umfassenden Erfassung relevanter Berufe und der unterstellten Beteiligungsgrade beruhen und stark von der genauen Modellspezifikation abhängen. Trotz vieler Unsicherheiten verdeutlichen die Berechnungen den potenziell erheblichen Einfluss von Dateninvestitionen auf die Messung der Wirtschaftsleistung.

## 5

### Fazit und Ausblick

Im Rahmen von Testrechnungen hat das Statistische Bundesamt untersucht, wie Investitionen in Daten für Deutschland erfasst werden könnten. Das gewählte Berechnungsmodell basiert auf dem von der Statistischen Kommission der Vereinten Nationen empfohlenen Kostensummenansatz. Die Untersuchungen haben gezeigt, dass Investitionen in Daten erhebliche Auswirkungen auf die Bruttoanlageinvestitionen und somit auch auf das Bruttoinlandsprodukt haben können. Die Investitionen in Daten belaufen sich schätzungsweise auf etwa 11 % der aktuellen Bruttoanlageinvestitionen. Darüber hinaus würde die Integration dieser Investitionen in Daten zu einer durchschnittlichen Erhöhung des BIP-Niveaus um 2,3 % beitragen.

Da die Arbeitskosten den Ausgangspunkt der Berechnungen über den Kostensummenansatz bilden, ist die Methode jedoch kaum in der Lage, die Dynamik und das Ausmaß der Datenproduktion adäquat abzubilden. Die zunehmende wirtschaftliche Bedeutung von Daten lässt sich vielmehr auf technologische Fortschritte zurückführen.




ren, die es ermöglichen, Daten durch installierte Anwendungen oder KI-Systeme automatisiert zu sammeln und auszuwerten.

Über die Frage der grundsätzlichen Eignung des aktuellen Rechenmodells hinaus verfügt dieses über entscheidende Stellschrauben mit beträchtlichem Spielraum. Diese machen das Ergebnis fast vollständig von den getroffenen Annahmen abhängig und gefährden dadurch die internationale Vergleichbarkeit der VGR-Ergebnisse. Eine Harmonisierung der Annahmen wäre daher zwingend notwendig. Insbesondere die Ermittlung der Beteiligungsgrade, die Auswahl der relevanten Berufsgattungen und die Berechnung des Aufschlags auf die Lohnkosten sind international noch nicht harmonisiert.

Um aussagekräftigere Ergebnisse zu erhalten ist es unabdingbar, statistische Erhebungen auf beziehungsweise auszubauen, um etwa systematisch Käufe und Verkäufe von Daten sowie interne Aufwendungen zur Datenproduktion zu erfassen. Solche Erhebungen könnten dazu beitragen, die wirtschaftliche Bedeutung von Daten präziser abzubilden und die Qualität der statistischen Analysen zu verbessern.

Dazu wäre allerdings zunächst die statistische Definition von Daten zu überarbeiten. Rückmeldungen von Verbänden deuten darauf hin, dass mit der aktuellen Definition keine belastbaren Erhebungsdaten zu erwarten wären. Ebenso unklar gestaltet sich die Trennung von produktiven und nicht produktiven Daten. Nur für den Produktionsprozess relevante Daten sind als Investition zu buchen. Und das auch nur dann, wenn diese mindestens ein Jahr im Produktionsprozess verwendet werden. Zu beiden Bedingungen, die das Investitionsvolumen in Daten potenziell stark verringern können, liegen keine statistischen Informationen vor; sie können in der vorgestellten Studie nicht berücksichtigt werden.

International harmonisierte Leitlinien und Empfehlungen zu den Annahmen, die der Kumulationsmethode für die Berechnung des Anlagevermögens und der Abschreibungen zugrunde liegen, sind ebenfalls nötig. Die Studienergebnisse zeigen, dass insbesondere die durchschnittliche Nutzungsdauer Einfluss auf die Höhe der Abschreibungen und des Nettoanlagevermögens hat, während die Wahl der Abschreibungsfunktion eine eher untergeordnete Rolle spielt.

Weitere Projekte werden die Untersuchungen fortsetzen, um gezielt Schwachstellen im gewählten Rechenmodell anzugehen. Insbesondere für die Fundierung des Beteiligungsgrades sind inzwischen weitere interessante Datenquellen bekannt. Zudem werden die bisherigen und in diesem Beitrag beschriebenen Ergebnisse in einer von Eurostat und dem Internationalen Währungsfonds koordinierten internationalen Expertengruppe vorgestellt und diskutiert. Diese Diskussionen sollen helfen, einen harmonisierten und konsistenten Rechenansatz zu entwickeln. Ziel der Expertengruppe ist, ein Handbuch zur Bewertung von Daten zu erarbeiten und damit sicherzustellen, dass künftige Berechnungen durch die statistischen Ämter verlässliche und international vergleichbare Ergebnisse liefern. 

## LITERATURVERZEICHNIS

---

Adler, Walther/Gühler, Nadine/Oltmanns, Erich/Schmidt, Daniel/Schmidt, Pascal/Schulz, Ingeborg. *Forschung und Entwicklung in den Volkswirtschaftlichen Gesamtrechnungen*. In: Wirtschaft und Statistik. Ausgabe 12/2014, Seite 703 ff.

Alphabet Inc. *Annual Report 2023*. [Zugriff am 3. Juli 2024]. Verfügbar unter: [abc.xyz/assets/52/88/5de1d06943cebc569ee3aa3a6ded/goog023-alphabet-2023-annual-report-web-1.pdf](https://abc.xyz/assets/52/88/5de1d06943cebc569ee3aa3a6ded/goog023-alphabet-2023-annual-report-web-1.pdf)

Bundesagentur für Arbeit. *Klassifikation der Berufe 2010 – überarbeitete Fassung 2020 – Band 2: Definitorischer und beschreibender Teil*. Nürnberg 2021. [Zugriff am 3. Juli 2024]. Verfügbar unter: [statistik.arbeitsagentur.de](https://statistik.arbeitsagentur.de)

Deutsche Bahn. *Künstliche Intelligenz bei der DB*. [Zugriff am 3. Juli 2024]. Verfügbar unter: [www.deutschebahn.com](https://www.deutschebahn.com)

Deutsche Bundesbank. *Jahresabschlussstatistik (Hochgerechnete Angaben) Dezember 2023*. Frankfurt am Main 2023. [Zugriff am 18. Juli 2024]. Verfügbar unter: [www.bundesbank.de](https://www.bundesbank.de)

Deutsche Bundesbank. *Zahlungsbilanzstatistik*. 2024. [Zugriff am 4. Juli 2024]. Verfügbar unter: [www.bundesbank.de](https://www.bundesbank.de)

Eurostat. *Europäisches System Volkswirtschaftlicher Gesamtrechnungen: ESVG 2010*. Luxemburg 2014. [Zugriff am 4. Juli 2024]. Verfügbar unter: [ec.europa.eu](https://ec.europa.eu)

Finke, Claudia/Geisler, Susanna/Überschaer, Anja. *Aus drei mach eins: die neue Verdiensterhebung*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2023, Seite 58 ff.

OECD (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung). *Frascati-Handbuch 2015: Leitlinien für die Erhebung und Meldung von Daten über Forschung und experimentelle Entwicklung. Messung von wissenschaftlichen, technologischen und Innovationstätigkeiten*. Paris 2018. DOI: [10.1787/9789264291638-de](https://doi.org/10.1787/9789264291638-de)

OECD (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung). *Measuring the value of data and data flows*. In: OECD Digital Economy Papers. Nr. 345. 2022. [Zugriff am 4. Juli 2024]. Verfügbar unter: [www.oecd.org](https://www.oecd.org)

Reichental, Jonathan. *Data Governance for Dummies*. Hoboken, New Jersey 2022.

Schmalwasser, Oda/Schidlowski, Michael. *Kapitalstockrechnung in Deutschland*. In: Wirtschaft und Statistik. Ausgabe 11/2006, Seite 1107 ff.

Statistics Canada. *An Update on Depreciation Rates for the Canadian Productivity Accounts*. 2015. [Zugriff am 4. Juli 2024]. Verfügbar unter: [www150.statcan.gc.ca](https://www150.statcan.gc.ca)

Statistische Ämter des Bundes und der Länder. *Informationen zum Mikrozensus*. [Zugriff am 3. Juli 2024]. Verfügbar unter: [www.mikrozensus.de](https://www.mikrozensus.de)

Statistisches Bundesamt. *Verdienststrukturerhebung*. 2024. [Zugriff am 3. Juli 2024]. Verfügbar unter: [www.destatis.de](https://www.destatis.de)

## LITERATURVERZEICHNIS

---

Statistisches Bundesamt. *Klassifikation der Wirtschaftszweige, Ausgabe 2008* (WZ 2008). Wiesbaden 2008.

The Economist. *The world's most valuable resource is no longer oil, but data*. 2017. [Zugriff am 3. Juli 2024]. Verfügbar unter: [www.economist.com](http://www.economist.com)

Vereinte Nationen. *System of National Accounts 2008*. New York 2009. [Zugriff am 5. Juli 2024]. Verfügbar unter: [unstats.un.org](http://unstats.un.org)

Vereinte Nationen. *DZ.6 Recording of Data in the National Accounts*. 2023. [Zugriff am 3. Juli 2024]. Verfügbar unter: [unstats.un.org](http://unstats.un.org)

# DIE NUTZENDENUMFRAGE ZUR KULTURSTATISTIK – KONZEPT UND ERGEBNISSE

Marlon Suárez, Lieselotte Hofmann

📌 **Schlüsselwörter:** Kulturlandschaft – Kreativbranche – Kulturverbände – Publikumsverhalten – kulturstatistisches Berichtssystem

## ZUSAMMENFASSUNG

Das Statistische Bundesamt ist seit dem Jahr 2014 an einem Projekt zur Etablierung einer fundierten, bundesweiten Kulturstatistik beteiligt, aktuell in der dritten Projektphase (2023 bis 2027) mit dem Namen „Kontinuierliche bundesweite Kulturstatistik“.

Der Artikel stellt Methodik und Ergebnisse einer Umfrage bei potenziellen Nutzenden kulturstatistischer Daten vor. Ermittelt wurden unter anderem bislang unerfüllte Bedarfe nach Kulturdaten sowie die Zufriedenheit der Nutzenden und die Art und Häufigkeit der Nutzung des kulturstatistischen Datenangebots des Statistischen Bundesamtes. Die Umfrageergebnisse unterstützen dabei, die künftige inhaltliche Ausrichtung der Kulturstatistik zu verbessern und Handlungsempfehlungen für die weitere Entwicklung der bundesweiten Kulturstatistik zu geben.

📌 **Keywords:** cultural landscape – creative sector – cultural associations – audience behaviour – cultural statistics reporting system

## ABSTRACT

*Since 2014, the Federal Statistical Office has been taking part in a project aimed at establishing a well-founded, nationwide cultural statistics system. The third phase of the project – known as “continuous nationwide cultural statistics” – is currently underway and is scheduled to run from 2023 to 2027.*

*This article presents the methodology and results of a survey conducted among potential users of cultural statistics data. The survey collected information, for example, on cultural data needs that are currently unmet, user satisfaction and the type and frequency of use of the cultural statistics data provided by the Federal Statistical Office. The survey results will help improve the future content of cultural statistics and provide recommendations for the further development of nationwide cultural statistics.*



**Marlon Suárez**

hat Internationale Beziehungen, Politische Ökonomie und Politikwissenschaften an der Johannes Gutenberg-Universität Mainz studiert. Seit 2018 ist er im Statistischen Bundesamt beschäftigt. Nach mehreren Stationen in den Bereichen der Konjunktur- und Strukturstatistiken sowie der europäischen und internationalen Koordination ist er seit 2023 wissenschaftlicher Mitarbeiter im Projekt „Bundesweite Kulturstatistik“.

**Lieselotte Hofmann**

hat Theaterwissenschaften und Wirtschaftswissenschaften an der Johannes Gutenberg-Universität Mainz studiert. Seit Sommer 2023 arbeitet sie im Referat „Forschung, Kultur“ des Statistischen Bundesamtes und unterstützt dort das Team der Kulturstatistik.

## 1

### Einleitung

Die vom Deutschen Bundestag eingesetzte Enquete-Kommission „Kultur in Deutschland“ führte von 2003 bis 2007 eine umfassende Untersuchung der deutschen Kulturlandschaft durch. Ihr zentrales Anliegen war es, Vorschläge zu erarbeiten, um die Kulturlandschaft zu stärken und weiterzuentwickeln sowie die Situation der Kulturschaffenden zu verbessern. Der Abschlussbericht der Enquete-Kommission unterstrich, dass eine verlässlichere Datenbasis über das kulturelle Leben in Deutschland dringend notwendig sei. Eine Überprüfung der damals vorhandenen kulturstatistischen Daten offenbarte eine unübersichtliche, kaum vergleichbare und sehr heterogene Datenlage, die sich als unzureichend für fundierte kulturpolitische Entscheidungen erwies (Deutscher Bundestag, 2007). Auch auf europäischer Ebene ist eine verbesserte Kulturstatistik notwendig, um unter anderem die Leistung der Kultur- und Kreativbranche besser überwachen und sie faktenbasiert unterstützen zu können oder die während der Corona-Pandemie entdeckten Schwachstellen aufzuzeigen und gegensteuern zu können. Hierfür bedarf es europaweit vergleichbarer Kulturstatistiken (Europäisches Amtsblatt, 2022, hier: Seite 4 ff.).

Nachdem die Enquete-Kommission im Jahr 2007 ihren Abschlussbericht veröffentlicht hatte, vergingen zunächst sieben Jahre, bevor im Jahr 2014 ein Bund-Länder-Projekt zum Aufbau einer bundesweiten Kulturstatistik gestartet wurde. Das Projekt hatte das Ziel, eine fundierte Datenbasis für kulturpolitische Entscheidungen zu schaffen. Die Beauftragte der Bundesregierung für Kultur und Medien sowie die Kultusministerkonferenz finanzieren seitdem diese Bund-Länder-Initiative. Das Statistische Bundesamt führt das Projekt als Hauptauftragnehmer gemeinsam mit dem Hessischen Statistischen Landesamt, als Patenland für den Statistikbereich Kultur, durch. Derzeit befindet sich das Projekt in der dritten Phase, die den Namen „Kontinuierliche bundesweite Kulturstatistik“ trägt und den Zeitraum von 2023 bis 2027 umfasst (Kultusministerkonferenz, 2022).<sup>1</sup>

---

<sup>1</sup> Die erste Phase lief von 2014 bis 2016 unter dem Projektnamen „Aufbau einer bundeseinheitlichen Kulturstatistik“, die zweite Phase von 2017 bis 2022 unter dem Namen „Bundesweite Kulturstatistik“.

Seit Beginn der ersten Projektphase im Jahr 2014 begleitet der Arbeitskreis Kulturstatistik die Aktivitäten des Projekts (Liersch/Asef, 2018, hier: Seite 38). Der Arbeitskreis fungiert als steuerndes Gremium, das die Fortschritte und Entwicklungen im laufenden Projekt koordiniert, Zielsetzungen definiert und Arbeitsergebnisse abnimmt. Zum Stand Juli 2024 gehörten dem Arbeitskreis folgende Organisationen an:

- › die Beauftragte der Bundesregierung für Kultur und Medien,
- › der Deutsche Städtetag,
- › das Hessische Ministerium für Wissenschaft und Kunst,
- › das Hessische Statistische Landesamt,
- › das Institut für Kulturpolitik der Kulturpolitischen Gesellschaft e. V.,
- › das Sekretariat der Kultusministerkonferenz,
- › das Niedersächsische Ministerium für Wissenschaft und Kultur,
- › das Statistische Bundesamt sowie
- › das Ministerium für Wissenschaft, Forschung und Kultur des Landes Brandenburg, welches derzeit den Interims-Vorsitz des Arbeitskreises Kulturstatistik innehat.

Der Arbeitskreis entschied zu Beginn der dritten Projektphase Anfang 2023, prüfen zu lassen, welche fachlichen Anforderungen für eine bundesweite Kulturstatistik erforderlich sind und ob Datenbedarfe bestehen, die das bisherige Angebot der amtlichen Statistik nicht abdeckt. Dazu beauftragte der Arbeitskreis das Statistische Bundesamt, eine umfassende Nutzendenumfrage durchzuführen. Sie hatte das Ziel, das bestehende bundesweite Angebot der Kulturstatistik zu evaluieren, mögliche Datenlücken aufzudecken und die Kooperation mit den datenproduzierenden Stellen zu überprüfen.

Nach intensiven Beratungen im Arbeitskreis Kulturstatistik von Juli bis Oktober 2023 wurde die Umfrage um weitere Zielsetzungen ergänzt. Der nun erweiterte Fragenkatalog sollte zusätzlich den Bekanntheitsgrad und die Nutzungshäufigkeit der vom Statistischen Bundesamt für den Kulturbereich bereitgestellten Veröffentlichungsformate ermitteln (siehe auch Übersicht 1 in

Abschnitt 3.1) sowie die Nutzendenzufriedenheit bewerten. Außerdem zielte die Umfrage darauf ab zu untersuchen, wie kulturstatistische Daten außerhalb des Statistischen Bundesamtes genutzt werden und welche Rolle sie für kulturpolitische Entscheidungen spielen.

Um zudem einen umfassenden Überblick über das verfügbare kulturstatistische Datenangebot zu erhalten, wurden Forschungsinstitute, die Statistischen Ämter der Länder ebenso wie Kulturverbände nach eigenen Studien, Forschungsarbeiten oder Erhebungen im Bereich Kultur befragt. Wichtig war außerdem, den Datenbedarf zu erfassen und abzufragen, wie zufrieden die Kulturverbände in Bezug auf die Zusammenarbeit mit dem Statistischen Bundesamt sind. Weitere Fragen untersuchten, inwieweit die Kulturverbände sowie Verwaltungen auf Gemeinde- und Kreisebene zur zusätzlichen Datenerhebung und -lieferung bereit wären, da dies für den Ausbau des amtlichen kulturstatistischen Datenangebots entscheidend ist.

Kapitel 2 stellt zunächst das Konzept der Umfrage vor, insbesondere die Zielgruppen der Umfrage und das Umfrage-Design. Außerdem beschäftigt sich dieses Kapitel mit der Methodik der Datenerhebung und -analyse. Die zentralen Ergebnisse der umfassenden Datenauswertung präsentiert Kapitel 3. Das Fazit fasst die Kernergebnisse zusammen und erörtert abschließend, wie die Umfrageergebnisse die künftige Gestaltung und die strategische Ausrichtung der bundesweiten Kulturstatistik unterstützen könnten.

## 2

### Konzeption und Methodik der Umfrage

---

#### 2.1 Befragtenkreis

---

In seiner konstituierenden Sitzung zur aktuellen Projektphase im Januar 2023 beschloss der Arbeitskreis Kulturstatistik zunächst, die Umfrage unter den mit dem Statistischen Bundesamt kooperierenden Kulturverbänden und weiteren Datenproduzenten durchzuführen. Die Hauptziele waren dabei, mögliche Datenbedarfe, die die Datenproduzierenden mit ihren eigenen Erhebungen bisher nicht abdecken, zu ermitteln und zu klären, wie die bisherige Zusammenarbeit intensiviert und verbessert werden könnte.

Das Projektteam erweiterte während der Beratungen auf Anregung des Arbeitskreises den Kreis der Befragten. Zusätzlich zu den bereits kooperierenden Datenproduzierenden kamen weitere relevante Gruppen hinzu, die als potenzielle Nutzende der Kulturstatistik eingeschätzt wurden. Dazu zählen neben den Kulturverwaltungen auf Bundes- und Länderebene auch wissenschaftliche Institute an Hoch- und Fachhochschulen, die Statistischen Ämter der Länder sowie Politikberatungsunternehmen im Bereich Kultur. Der Befragtenkreis wurde außerdem um sämtliche kreisfreie Städte und Landkreise sowie um alle Städte ab 50 000 Einwohnerinnen und Einwohnern erweitert, um so auch die kulturstatistischen Bedarfe und Wünsche der kommunalen Kulturverwaltungen abbilden zu können. Darüber hinaus beschloss der Arbeitskreis Kulturstatistik, auch Verbände im Kulturbereich einzubeziehen, die bislang nicht mit dem Statistischen Bundesamt zusammenarbeiten. Aus diesem Grund wurden sämtliche Kulturverbände, die Mitglied im Deutschen Kulturrat sind, zum Befragtenkreis hinzugefügt. Insgesamt wurden 1 102 Einheiten zur Teilnahme an der freiwilligen Umfrage angeschrieben (siehe Tabelle 1 in Abschnitt 3.1).

Der Arbeitskreis Kulturstatistik unterstützte das Projektteam des Statistischen Bundesamtes dabei, alle Nutzendengruppen präzise zu erreichen. Er stellte unter anderem E-Mail-Verteilerlisten bereit und formulierte und verbreitete Infoschreiben des Deutschen Städtetags sowie des Deutschen Landkreistags vor der Umfrage, um die Aufmerksamkeit für die Umfrage auf kommunaler Ebene zu erhöhen. Die E-Mail-Adressen der einzelnen Hochschul- und Fachhochschulinstitute sowie der kommunalen Kulturverantwortlichen, die direkt für die Umfrage kontaktiert wurden, hat das Projektteam des Statistischen Bundesamtes durch Eigenrecherchen herausgefunden. Die Ermittlung sämtlicher kreisfreier Städte und Landkreise sowie der Städte mit mindestens 50 000 Einwohnerinnen und Einwohnern erfolgte anhand der Daten des Gemeindeverzeichnis-Informationssystems (GV-ISys) zum Gebietsstand 31. Dezember 2022.

#### 2.2 Zeitplan

---

Die technische Vorbereitung der Umfrage begann mit einer Pretest-Phase im Oktober 2023. Die Befragten erhielten Anfang November 2023 ein Vorabanschreiben per E-Mail, das über den Zweck und Ablauf der Umfrage



informierte. Dies ermöglichte dem Projektteam, die zuvor angelegten E-Mail-Verteilerlisten nachzupflegen, da durch den Versand des Vorabanschreibens ungültige E-Mail-Adressen identifiziert und angepasst werden konnten. Dieses Vorgehen verbesserte den erreichbaren Adressatenkreis der E-Mail-basierten Umfrage.

Die eigentliche Umfrage startete am 8. November 2023. Um eine möglichst hohe Teilnahme zu erreichen, wurden drei Erinnerungsrunden durchgeführt: Die erste Erinnerungsrunde erfolgte am 23. November 2023, die zweite am 6. Dezember 2023 und die letzte Erinnerungsrunde am 18. Dezember 2023. Diese Erinnerungen sollten Befragte, die noch nicht geantwortet hatten, zur Teilnahme motivieren. Die Umfrage endete schließlich am 20. Dezember 2023. Alle bis zu diesem Datum eingegangenen Antworten flossen in die Auswertung ein.

## 2.3 Umfrageinstrument

---

Für die Durchführung wurde die Umfragefunktion des vom ITZBund<sup>2</sup> bereitgestellten, webbasierten Dienstes „Basic Support for Cooperative Work – BSCW“ verwendet. Diese Plattform wurde aufgrund ihrer datenschutzrechtlichen Sicherheitsmerkmale gewählt, da sie ermöglicht, personalisierte Umfragelinks über zuvor angelegte E-Mail-Adresslisten zu generieren und zu verschicken, um so die Befragten gezielt zu erreichen. Eine vorherige Anmeldung oder Registrierung ist für das Ausfüllen des Fragebogens nicht erforderlich – ein zusätzlicher Vorteil, um die Rückmeldungsschwelle niedrig zu halten. Ferner unterstützt die Plattform die Identifizierung von Befragten, die noch keine Rückmeldung gesendet oder den Umfragebogen nicht abgeschlossen haben. Auf diese Weise war es möglich, mehrere Erinnerungsrunden gezielt durchzuführen, ohne den gesamten Befragtenkreis erneut zu adressieren.

Die Konzeption des Fragebogens beziehungsweise des Fragenkatalogs wurde eng mit dem Arbeitskreis Kulturstatistik und dem Fachbereich für Design und Weiterentwicklung von Erhebungsinstrumenten des Statistischen Bundesamtes abgestimmt. Um alle Zielsetzungen der Umfrage zu erreichen und differenzierte Erkenntnisse zu erlangen, war es notwendig, eigene Fragebogen für jede Nutzendengruppe zu konzipieren. Diese bestan-

den aus allgemeinen Fragen, welche allen Nutzendengruppen in gleicher Weise gestellt wurden, sowie gruppenspezifischen Fragen. Um statistisch quantifizierbare Antworten zu erhalten, wurden größtenteils einfache geschlossene Fragen (Ja-Nein-Fragen), erweiterte geschlossene Fragen mit vorgegebenen Antwortkategorien sowie Skalenfragen verwendet. Diese sind in der amtlichen Statistik die dominierenden Fragetypen (Statistische Ämter des Bundes und der Länder, 2023, hier: Seite 32 ff.). Zur Ermittlung von Datenbedarfen, bei der Abfrage von kulturstatistischen Erhebungen und weiteren Anforderungen hinsichtlich der Kooperation mit dem Statistischen Bundesamt ergänzten zudem offene Fragen den Fragenkatalog.

Nachdem der Fragenkatalog erstellt und intern abgestimmt war, wurde ein Pretest unter sämtlichen Nutzendengruppen durchgeführt. Diese Pretest-Phase erstreckte sich über einen Zeitraum von eineinhalb Wochen. Außer der technischen Zugänglichkeit, Handhabbarkeit und dem benötigten Zeitaufwand bei der Anwendung des Online-Tools wurde auch anhand der Reaktionen der Testpersonen geprüft, wie verständlich und geeignet die Fragen waren. Ziel war es, die Validität und Reliabilität des Umfrageverfahrens sicherzustellen, indem das Verfahren sowie potenzielle Antwortfehler (response set), basierend auf den Rückmeldungen der Testteilnehmenden, intensiv geprüft wurden (Porst, 2014, hier: Seite 17).

## 2.4 Erhebungs- und Analysemethoden

---

Die Ergebnisse der quantifizierbaren, geschlossenen Fragen wurden überwiegend auf der Grundlage von Häufigkeitsverteilungen deskriptiv ausgewertet.

Die offenen Fragen beziehungsweise Freitextfelder mit umfangreichen Einzelinformationen wurden mithilfe des KI-Modells ChatGPT-4<sup>3</sup> kategorisiert. Vor der Anwendung von ChatGPT-4 wurden die Einzelangaben der Umfrageteilnehmenden anonymisiert. Die Daten wurden anschließend bei deaktiviertem Trainingsmodus des KI-Modells im Zuge eines internen Tests zur Anwendung generativer Künstlicher Intelligenz (KI) im Statistischen Bundesamt verarbeitet.

---

<sup>2</sup> Das Informationstechnikzentrum Bund (ITZBund) erbringt IT-Dienstleistungen für 200 Behörden der Bundesverwaltung.

<sup>3</sup> ChatGPT-4 ist ein KI-Modell beziehungsweise Chatbot, das von Nutzenden eingegebene Texte versteht und auch eigenständig Texte generieren kann. Es kann kommunizieren, Inhalte erstellen und Aufgaben in natürlicher Sprache ausführen.



Zielgerichtete Prompts<sup>4</sup> veranlassten das KI-Modell, die anonymisierten Einzeldaten in thematische Cluster einzuordnen. Diese vorgeschaltete Kategorisierung erwies sich als hilfreich für das Projektteam, das anschließend manuell ein Kategoriensystem erstellte.

Die qualitative Auswertung erfolgte nach den Prinzipien der „inhaltlich strukturierenden Inhaltsanalyse“ nach Mayring (Mayring, 2015, hier: Seite 97 ff.). Dies ermöglichte eine systematische und transparente Analyse der offenen Fragen. Um die qualitative Inhaltsanalyse durchzuführen, werden dabei zunächst die gesammelten Daten systematisch codiert und kategorisiert. Hierfür wird ein Kategoriensystem gebildet, das auf den Themen und Inhalten basiert, welche sich aus den Umfragedaten ergeben. Das Kategoriensystem dient als Leitfaden für die Analyse und ermöglicht, die relevanten Informationen strukturiert zu erfassen und zusammenzufassen. Mehrere unabhängige Codierende bearbeiten die Einzelangaben, indem sie die Inhalte der Freitextfelder gemäß des festgelegten Kategoriensystems zuordnen. Dies ist ein wichtiger Schritt in der qualitativen Auswertung von Umfragen. Die Codierenden bearbeiten dabei die Umfragedaten unabhängig voneinander. Im Anschluss vergleichen und diskutieren sie die Ergebnisse, um mögliche Unterschiede in der Interpretation zu klären und so einen Konsens zu erreichen. Der Einsatz mehrerer unabhängiger Codierenderinnen und Codierer erhöhte damit die Objektivität und Reliabilität der Analyse.

Zusammenfassend lässt sich festhalten, dass die Bildung eines Kategoriensystems und der Einsatz mehrerer Codierenderinnen und Codierer die Qualität und Validität der Analyse sicherten, was zu fundierten und aussage-

kräftigen Ergebnissen führte. Die KI-gestützte Methode unterstützte dabei die systematische und effiziente Analyse der offenen Fragen und trug dazu bei, die Vielzahl an Einzelinformationen zu strukturieren und zu interpretieren.

### 3

## Ergebnisse

### 3.1 Nutzung und Nutzungshäufigkeit des kulturstatistischen Angebots

Insgesamt haben sich 299 der 1 102 angeschriebenen Einheiten an der Umfrage zur Weiterentwicklung der Kulturstatistik beteiligt. Dies entspricht einer Teilnahmequote von 27 %, was einen mittleren Wert bei E-Mail-beziehungsweise webbasierten Umfragen darstellt (Manfreda und andere, 2008, hier: Seite 90 ff.). Die Mitwirkung der vergleichsweise kleinen Nutzendengruppe der Kulturverwaltungen sticht dabei mit einer Quote von 66 % besonders hervor. ➡ **Tabelle 1**

Bei allen Nutzendengruppen wurden zunächst die Bekanntheit des kulturstatistischen Angebotes des Statistischen Bundesamtes sowie Details zu dessen Nutzung erfragt. Die Auswertungen dazu ergaben, dass 43 % aller Umfrageteilnehmenden die kulturstatistischen Daten des Statistischen Bundesamtes bereits genutzt haben. Gemeinden und Kommunen greifen seltener auf das Datenangebot zurück (30 %), die Vertreterinnen und Vertreter der Kulturverwaltungen auf Bundes- und Länderebene nutzen es am häufigsten (66 %). Dabei geben 30 % der Vertretungen von Kulturverwaltungen an, die Daten konkret bei der Vorberei-

<sup>4</sup> Ein Prompt bei generativen KI-Modellen ist eine Eingabeaufforderung, der die Künstliche Intelligenz oder den Chatbot anleitet, eine entsprechende Antwort oder Fortsetzung zu generieren.

**Tabelle 1**

**Nutzendengruppen und Teilnahmequoten der Nutzendenumfrage zur Kulturstatistik**

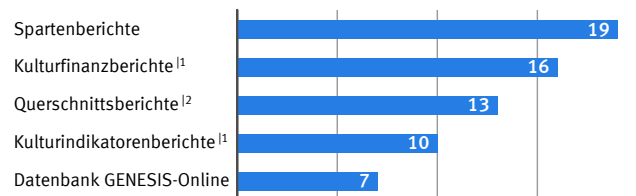
	Angeschriebene Einheiten	Rückmeldungen	Teilnahmequote gerundet
	Anzahl		%
Kulturverwaltung Länder und Bund	50	33	66
Institute und Statistische Ämter der Länder	251	47	19
Verbände	298	86	29
Gemeinden und Landkreise	503	133	26
Insgesamt	1 102	299	27

tung kulturpolitischer Entscheidungen heranzuziehen. Sie nutzen die Daten beispielsweise, um die kulturelle Lage in Deutschland zu analysieren, einen Überblick über die einzelnen kulturellen Sektoren zu erhalten, Argumentationen für kulturpolitische Entscheidungen auf eine Faktenbasis zu stellen oder die Ziele politischer Entscheidungen besser beurteilen zu können. Dagegen greifen 46 % bei der politischen Entscheidungsfindung nicht auf die Daten der bundesweiten Kulturstatistik zurück. Die Gründe dafür sind mitunter, dass das Angebot nicht bekannt beziehungsweise für Entscheidungen nicht relevant ist. Zudem gab diese Nutzendengruppe an, direkt auf Verbandsdaten zurückzugreifen, da diese meist spezifischer und detaillierter seien.

➤ Übersicht 1 bildet das vielfältige Angebot auf Ebene der einzelnen Veröffentlichungsformate ab. ➤ Grafik 1 zeigt, dass die unterschiedlichen Berichte, welche das Statistische Bundesamt auf seiner Webseite anbietet, von mehr Umfrageteilnehmenden genutzt werden als die Tabellen der Datenbank [GENESIS-Online](#). Dies betrifft vor allem die Spartenberichte, die die einzelnen Kultursparten jeweils einmal ausführlich vorstellen und sich mit deren Besonderheiten auseinandersetzen. Es folgen der alle zwei Jahre erscheinende Kulturfinanz-

Grafik 1

Nutzung der einzelnen Veröffentlichungsformate des Statistischen Bundesamtes zur bundesweiten Kulturstatistik in %



n = 299. Angaben zur mindestens einmaligen Nutzung der verschiedenen Veröffentlichungsformate. Mehrfachnennungen waren möglich. Ohne "Keine Angabe".

1 Gemeinschaftsveröffentlichungen der Statistischen Ämter des Bundes und der Länder.

2 Querschnittsbericht Musikfestivals und Musikfestspiele: Gemeinschaftsveröffentlichung der Statistischen Ämter des Bundes und der Länder.

bericht, verschiedene Querschnittsberichte sowie die jährlich veröffentlichten Kulturindikatorenberichte. Auf die einzelnen Veröffentlichungen und kulturstatistischen Berichte wurde überwiegend jährlich (74 %) und monatlich (23 %) zugegriffen, eine häufigere Nutzung kommt nur selten vor.

Insgesamt zeigt sich, dass ein Großteil der Umfrageteilnehmenden, die das kulturstatistische Angebot des

## Übersicht 1

### Veröffentlichungsformate des Statistischen Bundesamtes zur bundesweiten Kulturstatistik

	Erscheinungsmodus/-jahr
Datenbank GENESIS-Online, Tabellen zum Kulturbereich	jährlich
Kulturfinanzbericht <sup>1</sup>	zweijährlich, zuletzt 2022
Kulturindikatoren auf einen Blick <sup>1</sup>	jährlich, zuletzt 2024
Spartenberichte:	
Spartenbericht Literatur und Presse	einmalig, 2022
Spartenbericht Darstellende Kunst	einmalig, 2021
Spartenbericht Bildende Kunst	einmalig, 2021
Spartenbericht Soziokultur und Kulturelle Bildung	einmalig, 2020
Spartenbericht Film, Fernsehen und Hörfunk	einmalig, 2019
Spartenbericht Baukultur, Denkmalschutz und Denkmalpflege	einmalig, 2018
Spartenbericht Museen, Bibliotheken und Archive	einmalig, 2017
Spartenbericht Musik	einmalig, 2016
Querschnittsberichte:	
Musikfestivals und Musikfestspiele in Deutschland <sup>1</sup>	einmalig, 2017
Zeitverwendung für Kultur und kulturelle Aktivitäten in Deutschland – Sonderauswertung der Zeitverwendungserhebung 2012/2013	2016, neue Ausgabe für 2024 (mit Ergebnissen der Zeitverwendungserhebung 2022) geplant
Erwerbstätige in Kultur und Kulturwirtschaft – Sonderauswertung aus dem Mikrozensus	2015, 2019

1 Gemeinschaftsveröffentlichung der Statistischen Ämter des Bundes und der Länder.

Statistischen Bundesamt nutzen, sehr zufrieden (13 %) oder eher zufrieden (69 %) mit dem Angebot sind. Lediglich 16 % sind eher unzufrieden und 2 % sehr unzufrieden mit den bereitgestellten kulturstatistischen Daten. Dabei ist jedoch auch zu berücksichtigen, dass 57 % das Angebot des Statistischen Bundesamtes nicht nutzen beziehungsweise das Angebot nicht kennen.

Um Verbesserungsmaßnahmen ermitteln zu können ist es wichtig, die Gründe zu identifizieren, warum einzelne Umfrageteilnehmende das Angebot des Statistischen Bundesamtes nicht nutzen oder nicht damit zufrieden sind. Daher werden im Folgenden die Anforderungen und Bedarfe der Nutzenden tiefergehend analysiert.

### 3.2 Datenbedarfe, Datenlücken und Anforderungen

Ein zentrales Ziel der Umfrage ist es, die Anforderungen und Bedürfnisse der Nutzenden an das kulturstatistische Angebot zu erfassen, insbesondere jene, die bislang nicht erfüllt werden. Daher wurden alle Nutzendengruppen gefragt, ob aus ihrer Sicht zusätzliche Anforderungen oder Bedarfe an kulturstatistischen Daten bestehen. Diese spezifische Frage haben 32 % aller an der Umfrage Teilnehmenden beantwortet.

Auf wahrgenommene Datenlücken beziehungsweise den Wunsch, diese Datenlücken zu schließen, beziehen sich 71 % der geäußerten Anforderungen. Dabei nennen die Befragten zum einen ganze Themenfelder und Bereiche,

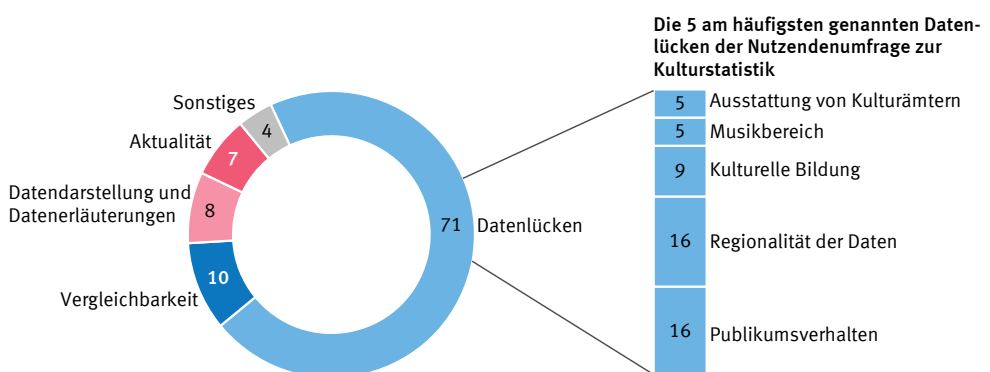
zu denen keine beziehungsweise keine zufriedenstellende Datenlage vorhanden ist. Zum anderen machen sie auf einzelne Merkmale aufmerksam, die bei bereits bestehenden Statistiken ergänzt werden sollten. Es werden aber auch vermeintliche Datenlücken genannt, wie beispielsweise Finanzdaten im Kulturbereich, obwohl ein entsprechendes Angebot mit dem Kulturfinanzbericht bereits besteht. Dort sind entsprechende Angaben sowohl für den Bund als auch für die Länder detailliert dargestellt. Die Antworten deuten darauf hin, dass das bereitgestellte Angebot nicht ausreichend bekannt ist oder nicht den Wünschen der Nutzenden entspricht.

Die benannten Datenlücken weisen zwei Schwerpunkte aus: Zum einen zeigt sich ein großer Bedarf an Daten zum Nutzungsverhalten des Publikums. Dabei interessieren sich die Auskunftgebenden nicht nur dafür, wie viele Menschen welche kulturellen Angebote nachfragen, sondern insbesondere, wer bestimmte Angebote nachfragt und welche soziodemografische Struktur, wie Alter, Geschlecht, Bildungsstand und Herkunft, das Publikum aufweist. [➤ Grafik 2](#)

Zum anderen wird häufig auf ein Fehlen regionaler Daten verwiesen, da sich das bestehende kulturstatistische Angebot und alle Veröffentlichungsformate des Statistischen Bundesamtes auf die Bundes- und Länderebene beschränken. Insbesondere die Nutzendengruppe der Gemeinde- und Landkreisverwaltungen wünscht sich regional tiefer gegliederte Daten, um mehr Informationen über das kulturelle Angebot, die kulturellen Einrichtungen, die Organisationsstrukturen oder die Ausgaben

**Grafik 2**

**Datenbedarfe und weitere Anforderungen der Nutzenden an das kulturstatistische Angebot in %**



n = 96. Mehrfachnennungen waren möglich.

für Kultur in anderen Landkreisen oder Kommunen zu erhalten.

Weitere Datenlücken bestehen aus Sicht der Auskunftgebenden im Bereich der kulturellen Bildung und Musik, bezüglich der finanziellen und personellen Ausstattung von Kulturrämtern oder der Erwerbstätigkeit und sozialen Lage von Kulturschaffenden.

Neben der Schließung von Datenlücken wünschen sich 10 % der Teilnehmenden, dass kulturstatistische Daten besser vergleichbar sind. Dies betrifft die Vergleichbarkeit von Daten über einen bestimmten Zeitraum hinweg, beispielsweise in Form von Zeitreihen, die Entwicklungen und Trends aufzeigen. Zudem äußern 8 % der Teilnehmenden Bedarf nach einer Harmonisierung von Darstellungsformen, Methodik und Inhalten bei den Statistiken; dies soll Vergleichbarkeit auf kommunaler Ebene, auf Länder- oder Bundesebene, aber auch international ermöglichen. Besonders herausfordernd sind hierbei unterschiedliche Abgrenzungen beziehungsweise Zuordnungen von Kultur insgesamt sowie einzelner Kulturbereiche.

Auch die Darstellungsformen der Kulturdaten sollte das Statistische Bundesamt nach Meinung einiger Teilnehmenden überarbeiten. Informationen sollten beispielsweise stärker gebündelt werden, Daten für eine leichtere Interpretation und Einordnung ausführlich erläutert sowie in verarbeitbaren Formaten digital bereitgestellt werden. Dabei zeigt sich, dass nur wenige das Angebot der Datenbank GENESIS-Online nutzen, das auf eine erleichterte Weiterverarbeitung der Daten ausgelegt ist. Außerdem werden eine höhere Aktualität und schnellere Verfügbarkeit von Daten und Publikationen gefordert.

### 3.3 Zusammenarbeit mit Datenproduzierenden

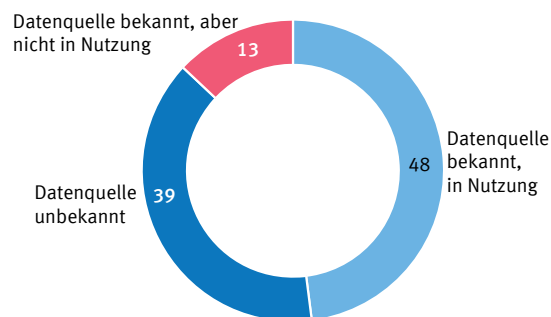
Grundlage des bestehenden kulturstatistischen Datenangebots des Statistischen Bundesamtes sind bereits verfügbare Daten von Kulturverbänden und weiteren Datenproduzierenden, die im Rahmen der Projekte gesichtet, zusammengetragen und zu Kennzahlen verdichtet werden. Überwiegend stellen kooperierende Verbände die Daten auf freiwilliger Basis zur Verfügung, weswegen eine gute Zusammenarbeit mit diesen essenziell ist, um das kulturstatistische Angebot bereitzustellen

und weiterzuentwickeln. Die Umfrage sollte daher klären, wie die bisherige Zusammenarbeit mit dem Statistischen Bundesamt bewertet wird, ob es mögliche Verbesserungen gibt und ob (weiterhin) eine Bereitschaft zur künftigen Zusammenarbeit besteht.

Mehr als die Hälfte aller Verbände (53 %) gibt an, eigene Statistiken zu erstellen oder Daten anderer Verbände und Institutionen zu sammeln, zu bündeln und zu veröffentlichen. Von diesen Primär- und Sekundärstatistiken sind dem Statistischen Bundesamt bereits 48 % bekannt und werden auch für Veröffentlichungen genutzt. Weitere 13 % sind zwar bekannt, aber zur Weiterverarbeitung nicht geeignet und werden daher nicht genutzt. Die übrigen 39 % der genannten Statistiken und Erhebungen waren im Statistischen Bundesamt dagegen nicht bekannt; sie werden im Nachgang der Nutzenumfrage gesichtet und geprüft, um bei beiderseitigem Interesse gegebenenfalls in das kulturstatistische Angebot des Statistischen Bundesamtes aufgenommen zu werden. ➡ Grafik 3

Grafik 3

Bekanntheit externer kulturstatistischer Datenquellen im Statistischen Bundesamt in %



Die Kontaktaufnahme mit neuen Datenproduzierenden, aber auch die Pflege und Intensivierung der bereits bestehenden Kooperationen ist Grundlage dafür, das Angebot auszuweiten und zu verbessern. Von den Verbänden geben 83 % an, dass sie es sehr beziehungsweise eher für wahrscheinlich halten, eigene Daten für die Nutzung durch das Statistische Bundesamt zu erheben beziehungsweise bereitzustellen. Dies verdeutlicht, dass die Verbände ebenfalls an einer (weiterführenden) Zusammenarbeit interessiert sind.

Insgesamt geben 85 % der Verbände, mit denen bereits zusammengearbeitet wurde, an, sehr zufrieden beziehungsweise eher zufrieden mit der Zusammenarbeit zu sein, 14 % teilen mit, eher unzufrieden zu sein. Um auch künftig eine gute beziehungsweise eine bessere Zusammenarbeit mit diesen Verbänden zu gewährleisten, wurden sie um Vorschläge gebeten, wie die Kooperation verbessert werden kann. Am häufigsten schlugen sie dabei vor, in Beratungen und Gremien einbezogen zu werden. Zudem wünschen sich die Verbände einen verbesserten Datenaustausch, beispielsweise bezüglich der Organisation der Datenübermittlung oder der Datenlieferformate. Weiterhin nennen sie Kooperationen bei der Datenerhebung, zum Beispiel durch gemeinsame Datensammlung und eine methodische Unterstützung.

Interessiert an einer Kooperation ist auch die Nutzengruppe der Landkreis- und Gemeindeverwaltungen, mit welcher bisher keine Zusammenarbeit bestand. So geben 80 % der 119 Landkreise und Gemeinden, die auf diese Frage geantwortet haben, an, sich an Datenerhebungen beteiligen zu wollen, falls dies erforderlich wäre. Ein Grund für dieses Engagement liegt mitunter darin, dass die Kommunen und Landkreise es mit einem Anteil von 77 % befürworten, wenn das Datenangebot von der Bundes- beziehungsweise Länderebene auf die Ebene der Gemeinden ausgeweitet wird.

Die positiven Rückmeldungen zeigen, dass die Verbände und Gemeinden bereit sind, kulturstatistische Daten für ein bundesweites Angebot zu erheben beziehungsweise bereitzustellen. Sie wünschen sich aber auch eine verbesserte Kommunikation und stärkere Einbeziehung bei der Etablierung eines kontinuierlichen, bundesweiten kulturstatistischen Angebots.

## 4

---

### Fazit und Ausblick

---


Die Umfrage bei bekannten und potenziellen Nutzengruppen der bundesweiten Kulturstatistik lieferte wichtige Erkenntnisse zur Nutzung und zu den Anforderungen an das kulturstatistische Angebot des Statistischen Bundesamtes. Positiv zu bewerten ist, dass mehr als zwei Fünftel der Teilnehmenden das kulturstatistische Angebot des Statistischen Bundesamtes bereits genutzt haben, wobei es Kulturverwaltungen auf Bundes- und Länderebene verhältnismäßig am meisten nutzten. Die allgemeine Zufriedenheit mit dem Angebot ist zwar hoch, jedoch ist mehr als der Hälfte der Teilnehmenden das kulturstatistische Datenangebot des Statistischen Bundesamtes nicht bekannt.

Die Analyse der Datenbedarfe und Datenlücken offenbart, dass insbesondere detaillierte Informationen zum Publikumsverhalten sowie regional tiefer gegliederte Daten fehlen. Auch der Bedarf an Daten zur kulturellen Bildung, zur finanziellen und personellen Ausstattung von Kulturämtern sowie zur sozialen Lage von Kulturschaffenden wurde hervorgehoben. Zudem wünschen sich die Befragten besser vergleichbare Daten über Zeiträume hinweg sowie harmonisierte Daten und Konzepte, um internationale und regionale Vergleiche zu erleichtern. Weiterhin sollen die Darstellungsformen überarbeitet werden (beispielsweise vermehrt digital verarbeitbare Produkte).

Die Zusammenarbeit zwischen dem Statistischen Bundesamt und den datenproduzierenden Verbänden und Institutionen wird überwiegend positiv bewertet, wobei in den Bereichen Kommunikation und Datenaustausch Verbesserungspotenziale gesehen werden. Viele Verbände sind zur weiteren Kooperation und Datenerhebung bereit, sodass eine solide Basis besteht, um das kulturstatistische Angebot auszuweiten und zu verbessern. Auch die Gemeinden und Landkreise signalisierten ihr Interesse an einer intensiveren Zusammenarbeit, insbesondere im Hinblick auf eine detailliertere regionale Datenerhebung.

Für die Zukunft der Kulturstatistik lassen sich aus den Umfrageergebnissen mehrere Handlungsempfehlungen ableiten:

- › Das bestehende kulturstatistische Angebot sollte bekannter gemacht und leichter zugänglich werden, um eine breitere Nutzung zu gewährleisten.
- › Es ist dringend angeraten, die identifizierten Datenlücken möglichst zu schließen und das Angebot um relevante, neue Datenbereiche zu erweitern, insbesondere bezüglich des Publikumsverhaltens und regional gegliederter Kulturdaten.
- › Es sollten Maßnahmen ergriffen werden, um besser mit den datenproduzierenden Stellen zusammenzuarbeiten und zu kommunizieren und damit eine effizientere und umfassendere Datenerhebung zu ermöglichen.

Um eine langfristige und kontinuierliche Datenversorgung im Bereich der Kulturstatistik wie von der Enquete-Kommission „Kultur in Deutschland“ empfohlen sicherzustellen, bedarf es allerdings einer rechtlichen Grundlage. Dies könnte die erforderliche Verlässlichkeit und Stabilität der Datenerhebung und -bereitstellung gewährleisten. Eine Rechtsgrundlage würde zudem die Zusammenarbeit mit den Datenproduzenten stärken und sicherstellen. Entsprechende Untersuchungen sind Teil des laufenden Projektes. Mit weiterführenden Erkenntnissen dazu wird Ende 2024 gerechnet. 

### LITERATURVERZEICHNIS

---

Deutscher Bundestag. *Schlussbericht der Enquete-Kommission „Kultur in Deutschland“*. BT-Drucksache 16/7000. 2007. [Zugriff am 2. Juli 2024]. Verfügbar unter: [dserver.bundestag.de](https://dserver.bundestag.de)

Europäisches Amtsblatt (Amtsblatt der Europäischen Union). *Entschließung des Rates zum Arbeitsplan für Kultur 2023-2026*. 7. Dezember 2022. C 466/1. [Zugriff am 2. Juli 2024]. Verfügbar unter: [eur-lex.europa.eu](https://eur-lex.europa.eu)

Kultusministerkonferenz. *Bundesweite Kulturstatistik wird fortgeführt*. Pressemitteilung vom 29. Dezember 2022. [Zugriff am 2. Juli 2024]. Verfügbar unter: [www.kmk.org](https://www.kmk.org)

Liersch, Anja/Asef, Dominik. *Aufbau einer bundesweiten Kulturstatistik*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2018, Seite 37 ff.

Manfreda, Katja Lozar/Bosnjak, Michael/Berzelak, Jernej/Haas, Iris/Vehovar, Vasja. *Web surveys versus other survey modes: a meta-analysis comparing response rates*. In: International Journal of Market Research. Jahrgang 50. Ausgabe 1/2008, Seite 79 ff. DOI: [10.1177/147078530805000107](https://doi.org/10.1177/147078530805000107)

Mayring, Philipp. *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 12., überarbeitete Auflage. Weinheim und Basel 2015.

Porst, Rolf. *Fragebogen. Ein Arbeitsbuch*. 4., erweiterte Auflage. Wiesbaden 2014.

Statistische Ämter des Bundes und der Länder. *Handbuch zur Erstellung von Erhebungsinstrumenten der amtlichen Statistik. Vereinbarungen und Leitlinien*. Version 2.1.7. Stand Juni 2023. [Zugriff am 2. Juli 2024]. Verfügbar unter: [www.statistischebibliothek.de](https://www.statistischebibliothek.de)



**Herausgeber**  
Statistisches Bundesamt (Destatis), Wiesbaden

---

**Schriftleitung**  
Dr. Daniel Vorgrimler  
Redaktion: Ellen Römer

---

**Ihr Kontakt zu uns**  
[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

---

**Erscheinungsfolge**  
zweimonatlich, erschienen im August 2024, Seite 76 korrigiert am 16.09.2024  
Ältere Ausgaben finden Sie unter [www.destatis.de](http://www.destatis.de) sowie in der [Statistischen Bibliothek](#).

---

Artikelnummer: 1010200-24004-4, ISSN 1619-2907

---

© Statistisches Bundesamt (Destatis), 2024  
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.