
WISTA

Wirtschaft und Statistik

Georg Thiel | Walter Krämer

Dorothea Klumpen | Christina
Schliffka | Dominique Polus

Marius Bellingen | Simon
Felgendreher | Johannes Oehrlein |
Simon Schürz | Stephan Arnold

Jasmin Arnold | Patric Brandt |
Heike Gerighausen

Marcus Schmidt | Jana Lucia Stein
Anja Überschaer

Verleihung des Gerhard-Fürst-Preises 2021

**Die Europäische Datenstrategie – ein Binnenmarkt
für Daten**

**Ökosystemgesamtrechnungen – Flächenbilanzierung
der Ökosysteme (Extent Account)**

**Erprobung der satellitengestützten Ertragsschätzung
für die Agrarstatistik – Projekt SatAgrarStat**

Die Hochrechnung im Mikrozensus ab 2020

**Der interaktive Gehaltsvergleich – eine profilbasierte
Schätzung von Verdiensten**

6 | 2021

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 in der Online-Ausgabe hinterlegt.

INHALT

3	Editorial
4	Kennzahlen – wichtige Indikatoren zur Coronakrise
8	Informationsangebote zur Coronakrise
10	Kurznachrichten
17	Georg Thiel, Walter Krämer Verleihung des Gerhard-Fürst-Preises 2021 <i>The 2021 Gerhard Fürst Award</i>
22	Dorothea Klumpen, Christina Schliffka, Dominique Polus Die Europäische Datenstrategie – ein Binnenmarkt für Daten <i>The European data strategy – A single market for data</i>
31	Marius Bellingen, Simon Felgendreher, Johannes Oehrlein, Simon Schürz, Stephan Arnold Ökosystemgesamtrechnungen – Flächenbilanzierung der Ökosysteme (Extent Account) <i>Ecosystem accounting – Ecosystem extent account</i>
43	Jasmin Arnold, Patric Brandt, Heike Gerighausen Erprobung der satellitengestützten Ertragsschätzung für die Agrarstatistik – Projekt SatAgrarStat <i>Testing of satellite-based yield estimation for agricultural statistics – The SatAgrarStat project</i>

INHALT

- | | |
|----|---|
| 54 | Marcus Schmidt, Jana Lucia Stein
Die Hochrechnung im Mikrozensus ab 2020
<i>Extrapolation in the microcensus from 2020 onwards</i> |
| 67 | Anja Überschaer
Der interaktive Gehaltsvergleich – eine profilbasierte Schätzung von Verdiensten
<i>Interactive salary comparison – Profile-based estimation of earnings</i> |

EDITORIAL

Dr. Georg Thiel



LIEBE LESERIN, LIEBER LESER,

am 18. und 19. November 2021 konnten wir das [30. Wissenschaftliche Kolloquium](#) „Von der Umweltstatistik zur nachhaltigen Entwicklung“ als hybride Veranstaltung durchführen – mit einer begrenzten Anzahl von Vortragenden, Diskutanten und Gästen in Präsenz in Wiesbaden und einer großen Zahl an Teilnehmenden im Livestream. So war es mir auch möglich, in diesem Rahmen die diesjährige [Verleihung des Gerhard-Fürst-Preises](#) persönlich vorzunehmen. Die Laudationes auf die insgesamt vier Preisträgerinnen und Preisträger können Sie in dieser Ausgabe nachlesen; im nächsten Jahr werden diese ihre prämierten Arbeiten in WISTA vorstellen.

Kurz nach dem Ende der 26. Klimakonferenz der Vereinten Nationen in Glasgow hat das Kolloquium ebenfalls verschiedene Umweltthemen aufgegriffen und dabei nicht nur deutsche, sondern auch europäische Entwicklungen in den Fokus gerückt. Auf einem der Vorträge basiert ein Beitrag in diesem Heft: Er stellt die Flächenbilanzierung der Ökosysteme Deutschlands dar, die den Ausgangspunkt der Arbeiten zur Berechnung der Ökosystemgesamtrechnungen bildet. Der neue [Ökosystematlas](#) des Statistischen Bundesamtes visualisiert die Resultate der Flächenbilanzierung in interaktiven Online-Karten. Ein weiterer Artikel berichtet über die Erprobung der satellitengestützten Ertragsschätzung für die Agrarstatistik. In einem Pilotprojekt haben das Statistische Bundesamt, die Statistischen Ämter von vier Bundesländern sowie das Julius Kühn-Institut Möglichkeiten untersucht, ob durch die Verknüpfung von fernerkundlich gewonnenen Vegetationsparametern und Pflanzenwachstumsmodellen Aussagen zum Ertragspotenzial möglich sind.

Erneut hat die Corona-Pandemie ein Jahr lang das gesellschaftliche und wirtschaftliche Leben geprägt. Trotz der nach wie vor notwendigen Einschränkungen wünsche ich Ihnen erholsame Festtage und einen guten Start in das Jahr 2022. Bleiben Sie gesund!

A handwritten signature in black ink, reading "Georg Thiel". The signature is written in a cursive, flowing style.

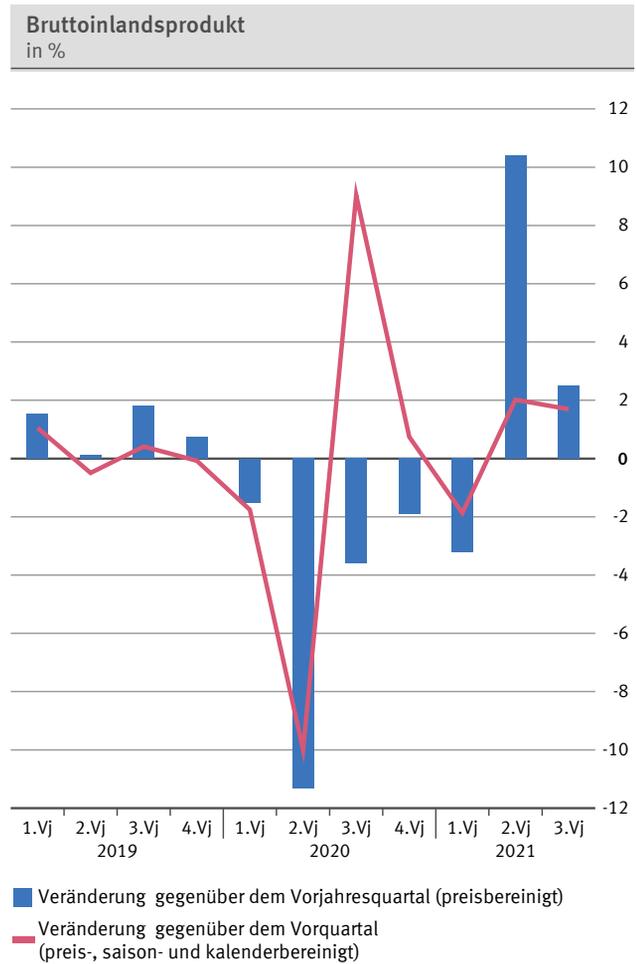
Präsident des Statistischen Bundesamtes

Auswirkungen der Corona-Pandemie auf Wirtschaft und Gesellschaft

Die globale Ausbreitung des SARS-CoV-2-Virus und der dadurch verursachten Erkrankung COVID-19 seit dem Frühjahr 2020 trifft Gesellschaft und Wirtschaft stark. Die langfristigen Folgen der Corona-Pandemie sind noch nicht überschaubar.

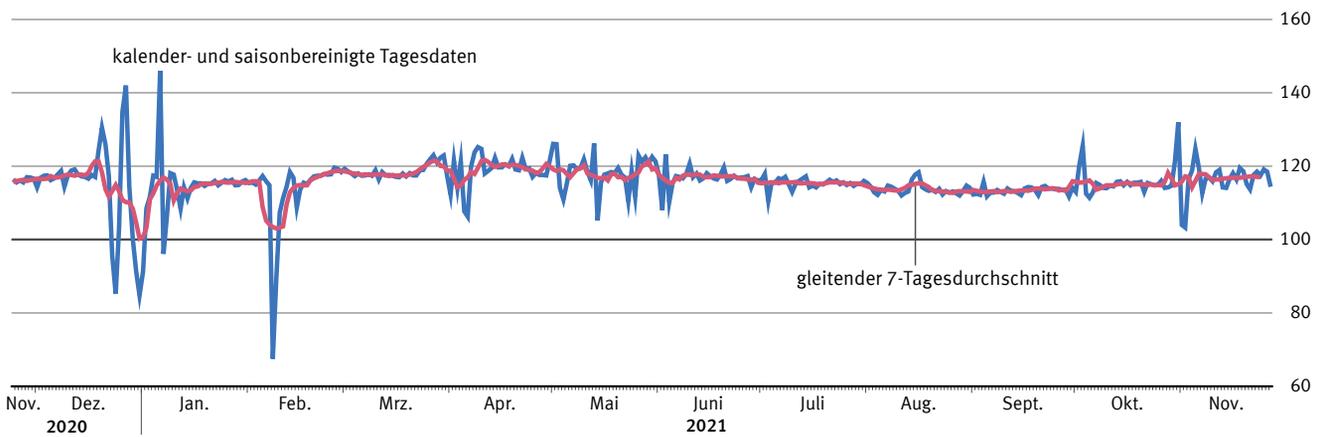
Die Kennzahlen auf dieser und den folgenden Seiten zeigen die Entwicklung einer Auswahl wichtiger Indikatoren, bei denen sich Auswirkungen zeigen.

Am 29. Oktober 2021 hat das Statistische Bundesamt das Bruttoinlandsprodukt (BIP) in seiner Schnellschätzung veröffentlicht. Mit der BIP t+30-Schnellschätzung stehen hochaktuelle faktenbasierte Wirtschaftsdaten, die in Krisenzeiten besondere Relevanz haben, üblicherweise spätestens 30 Tage nach Quartalsende zur Verfügung.



Lkw-Maut-Fahrleistungsindex ab November 2020

2015 = 100

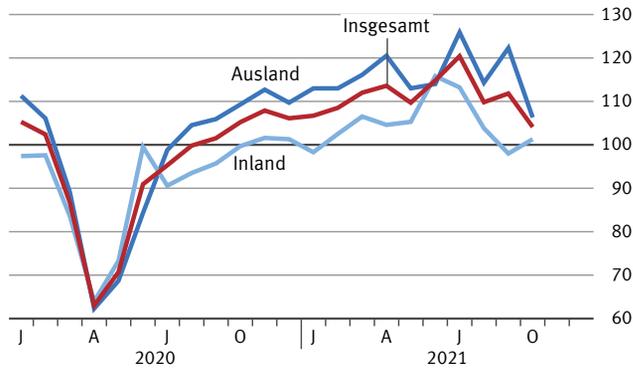


Quellen: Bundesamt für Güterverkehr, Deutsche Bundesbank, Statistisches Bundesamt

Stand: 10.12.2021

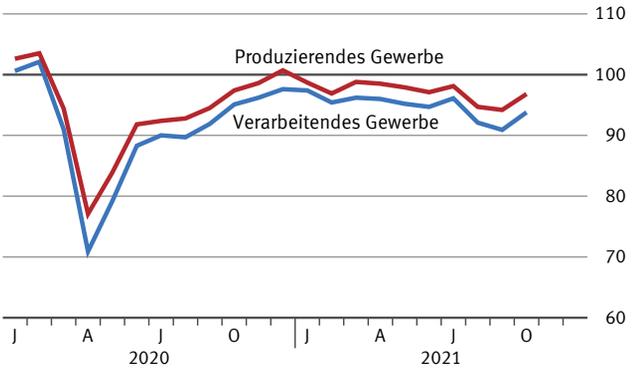
Kennzahlen – wichtige Indikatoren zur Coronakrise

Auftragseingang im Verarbeitenden Gewerbe
Volumenindex 2015 = 100



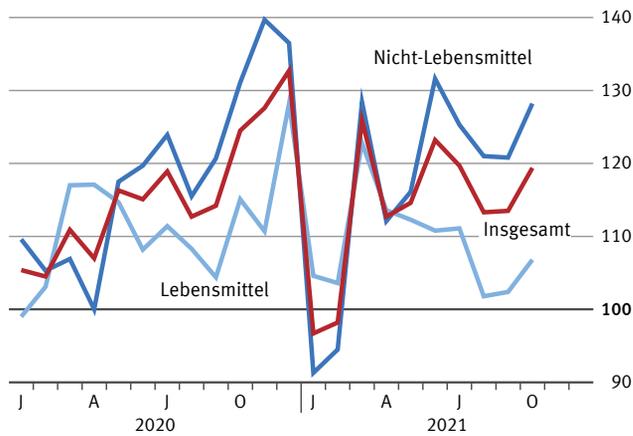
Kalender- und saisonbereinigter Wert nach dem Verfahren X13 JDemetra+ – Vorläufiges Ergebnis.

Produktion im Produzierenden und Verarbeitenden Gewerbe
Index 2015 = 100



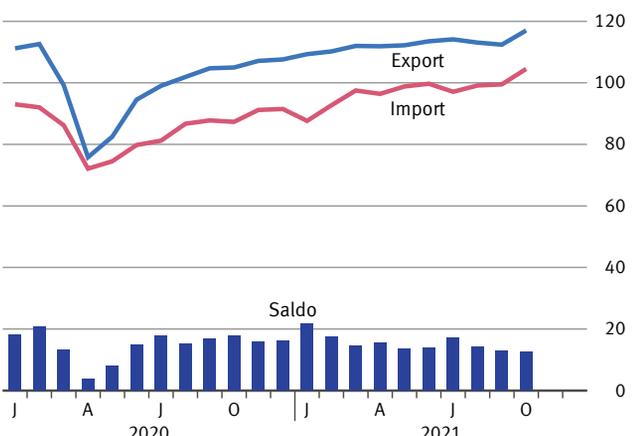
Kalender- und saisonbereinigte Werte nach dem Verfahren X13 JDemetra+ – Vorläufiges Ergebnis.

Umsatz im Einzelhandel
2015 = 100



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

Außenhandel
in Mrd. EUR



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

Stand: 10.12.2021

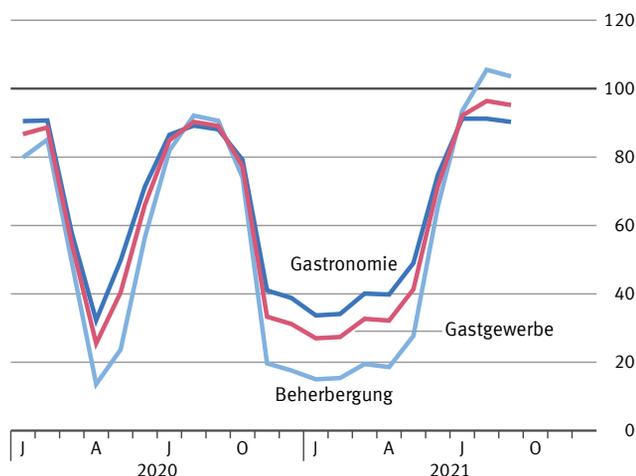
Kennzahlen – wichtige Indikatoren zur Coronakrise

Verbraucherpreisindex 2015 = 100

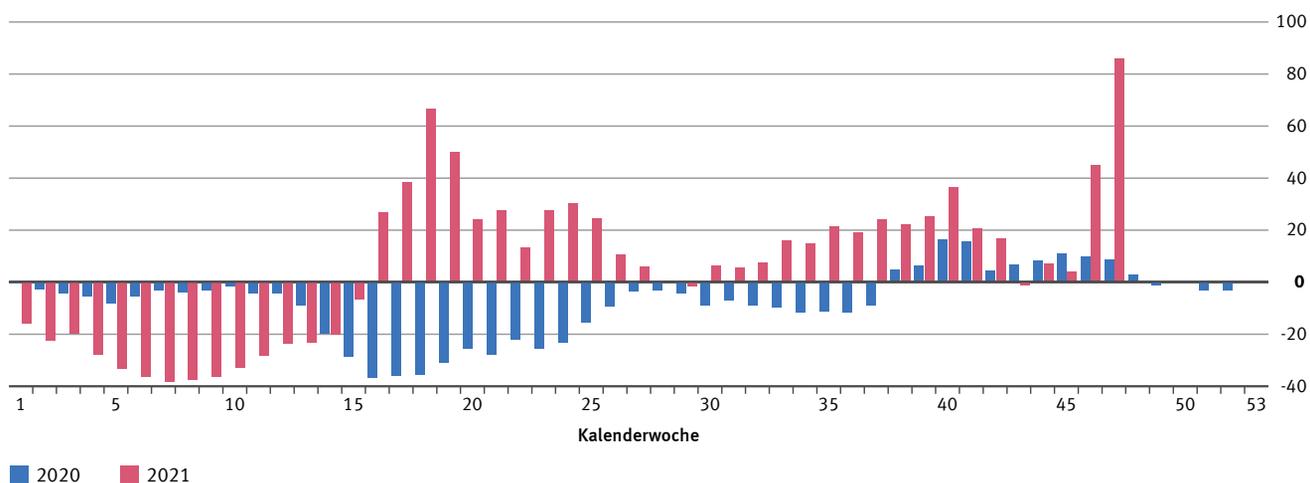
2020		2021	
Januar	105,2	Januar	106,3
Februar	105,6	Februar	107,0
März	105,7	März	107,5
April	106,1	April	108,2
Mai	106,0	Mai	108,7
Juni	106,6	Juni	109,1
Juli	106,1	Juli	110,1
August	106,0	August	110,1
September	105,8	September	110,1
Oktober	105,9	Oktober	110,7
November	105,0	November	110,5
Dezember	105,5		

↑ 5,2 %
Veränderung

Umsatz des Gastgewerbes in konstanten Preisen (real), Originalwert, 2015 = 100



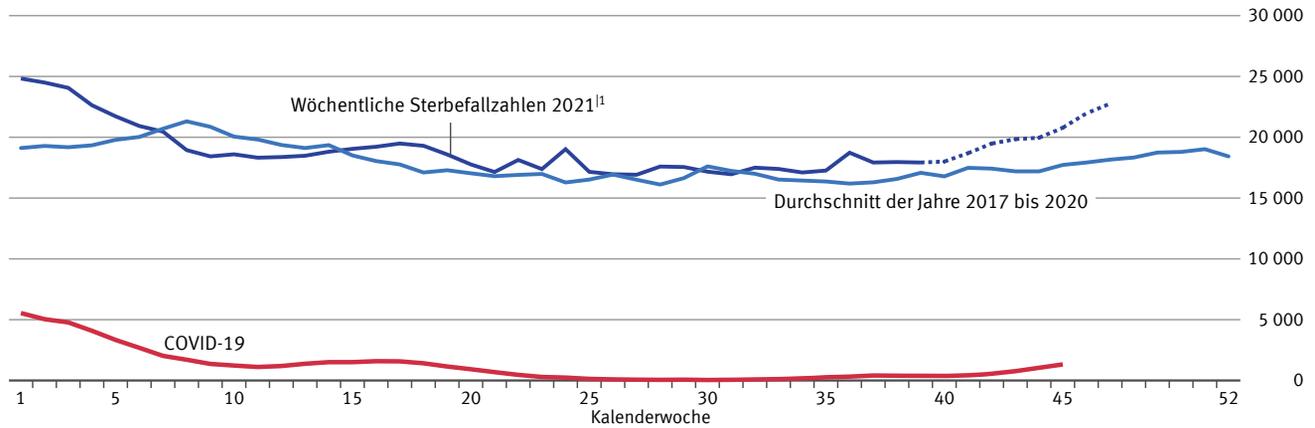
Neue Kreditverträge nach Kalenderwochen Veränderung gegenüber der entsprechenden Vorjahreswoche in %



Stand: 10.12.2021

Kennzahlen – wichtige Indikatoren zur Coronakrise

Wöchentliche Sterbefallzahlen in Deutschland 2021

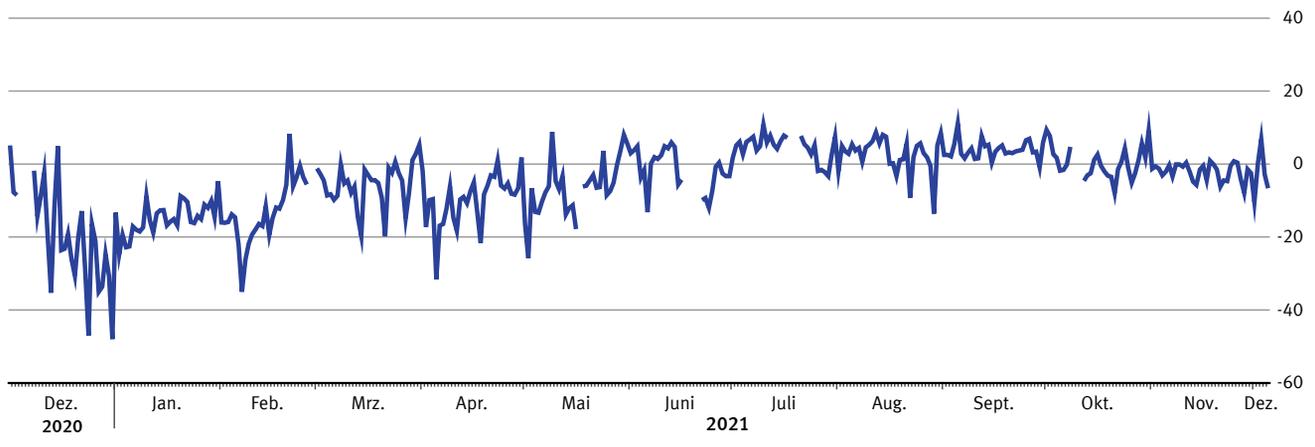


Gestrichelte Werte enthalten Schätzanteil.

1 Sonderauswertung der vorläufigen Sterbefallzahlen.

Quellen: Statistisches Bundesamt (Sterbefallzahlen insgesamt), Robert Koch-Institut (COVID-19-Todesfälle)

Veränderung der Mobilität gegenüber 2019 in %



Datenlücken entstehen in der Regel aufgrund von technischen Problemen beim Mobilfunkanbieter.

Experimentelle Daten; Quelle: Teralytics

Stand: 10.12.2021



Corona-Statistiken – Auswirkungen der COVID-19-Pandemie

Die globale Ausbreitung des SARS-CoV-2-Virus und die damit verbundenen Maßnahmen zur Eindämmung der Pandemie treffen Gesellschaft und Wirtschaft stark. Die langfristigen Folgen der Corona-Pandemie sind noch nicht überschaubar. Auf der Sonderseite „Corona-Statistiken“ werden statistische Ergebnisse präsentiert, in denen sich bereits die Auswirkungen zeigen oder in denen Auswirkungen zu erwarten sind. Dazu gehören auch neue Statistiken und Auswertungen, die das Statistische Bundesamt kurzfristig seit Beginn der Coronakrise bereitgestellt hat, um den Bedarf an Zahlen am aktuellen Rand zu decken:

↳ www.destatis.de/corona



Corona-Daten Deutschland

Mit Beginn der SarS-CoV2-Pandemie ist die Relevanz aktueller Daten und Statistiken noch einmal deutlich gestiegen. Vor diesem Hintergrund hat das Statistische Bundesamt gezielt sein Informationsangebot um aktuelle und regional tief gegliederte Daten zum Infektionsgeschehen, zur Krankenhaussituation, zur Mobilität und zu den ergriffenen Eindämmungsmaßnahmen der Regierungen erweitert. Die im Rahmen eines im Jahr 2020 vom Bundesministerium für Wirtschaft und Energie initiierten Projekts aufgebaute Datensammlung ist nun für die gesamte Öffentlichkeit frei zugänglich.

↳ www.corona-daten-deutschland.de



EXDAT – Experimentelle Daten

In der Rubrik „EXDAT – Experimentelle Daten“ veröffentlicht das Statistische Bundesamt regelmäßig neue, innovative Projektergebnisse. Sie entstehen auf der Grundlage neuer Datenquellen und Methoden. Im Reifegrad unterscheiden sie sich von amtlichen Statistiken, insbesondere in Bezug auf Harmonisierung, Erfassungsbereich und Methodik. Doch die Corona-Pandemie hat gezeigt, dass experimentelle Daten gerade in Krisenzeiten eine wertvolle und unverzichtbare Ergänzung zum amtlichen Datenangebot sind, weil sie eine schnellere Berichterstattung ermöglichen.

↳ www.destatis.de/exdat



Datenportal: Dashboard Deutschland

Das vom Statistischen Bundesamt neu entwickelte Datenportal präsentiert hochaktuelle und hochfrequente Daten aus verschiedenen Themenbereichen, wie Gesundheit, Mobilität und Wirtschaft, auf anschauliche Art und Weise. Es trägt damit zu einem faktenbasierten demokratischen Diskurs der Öffentlichkeit und zur evidenzbasierten Entscheidungsfindung durch Politik und Verwaltung bei. In der aktuellen Lage dient das Dashboard Deutschland auch als Gradmesser für die Coronakrise mit aktuellen Zahlen zu gewährten Corona-Hilfen für Unternehmen, täglich aktualisierten Mobilitätsindikatoren und den wichtigsten Informationen des Robert Koch-Instituts zur aktuellen Pandemiesituation.

↳ www.dashboard-deutschland.de



EU-Monitor COVID-19

Die COVID-19-Pandemie hat in allen EU-Staaten schwerwiegende Folgen für Wirtschaft und Gesellschaft verursacht. Die Viruserkrankung und die damit einhergehenden Sicherheitsmaßnahmen werden auch noch längere Zeit spürbare Auswirkungen haben. Dieser Monitor zeigt anhand von interaktiven Grafiken, welche EU-Staaten besonders unter den Folgen leiden und wie sich die Lage entwickelt. Dargestellt werden verschiedene Indikatoren aus den Bereichen Wirtschaft und Preise, Industrie und Baugewerbe, Arbeitsmarkt, Dienstleistungen, Exportwirtschaft sowie Bevölkerung und Gesundheit.

↳ www.destatis.de/corona



European Statistical Recovery Dashboard

Das während der deutschen EU-Ratspräsidentschaft im zweiten Halbjahr 2020 gemeinsam entwickelte europäische Dashboard bildet die wirtschaftlichen und sozialen Entwicklungen infolge der Corona-Pandemie in der Europäischen Union insgesamt und ihren einzelnen Mitgliedstaaten ab. Mehr als 20 Indikatoren ermöglichen Vergleiche zwischen den EU-Mitgliedstaaten. Die interaktive Anwendung erlaubt eine personalisierte Darstellung sowie das Herunterladen und Weiterverarbeiten der statistischen Daten.

↳ ec.europa.eu/eurostat

KURZNACHRICHTEN

IN EIGENER SACHE

Neues Webangebot zum Thema Klima

Vom 31. Oktober bis 12. November 2021 fand in Glasgow die 26. Weltklimakonferenz statt. Der Klimawandel ist eine der größten Herausforderungen unserer Zeit. Die Folgen der globalen Erwärmung wie Dürre oder Extremwetter-Ereignisse sind auch in Deutschland bereits sichtbar.

Die Fragestellungen rund um den Klimaschutz sind längst in alle Bereiche der Gesellschaft vorgedrungen. Wie beeinflusst unsere derzeitige Lebens- und Wirtschaftsweise das Klima? Wie wirkt sich die Umstellung hin zu mehr Klimaschutz gesamtgesellschaftlich aus? Und was bedeutet sie für unseren Alltag – vom Weg zur Arbeit bis zum aktuellen Strompreis? Wo zeigen sich die Folgen des Klimawandels schon jetzt?

Das Statistische Bundesamt bündelt in einem neuen Webangebot sein umfassendes und sich ständig erweiterndes Angebot an Daten und Fakten zum Thema Klima, Klimawandel und Klimaschutz:

↳ www.destatis.de

Statistische Angebote rund ums Thema Inflation

Die derzeit hohen Inflationsraten stehen im Fokus der Verbraucherinnen und Verbraucher. Aktuelle Zahlen und Fakten sowie vielfältige weiterführende Informationen bietet das Statistische Bundesamt in seinem Internet-

angebot auf der [Themenseite zur Inflationsrate auf Endverbraucherebene](#). Das Video „[Verbraucherpreisindex und Inflation kurz erklärt](#)“ ermöglicht einen kurzen, kompakten Einstieg ins Thema. Hintergrundinformationen und Analysen zu den Folgen der Corona-Pandemie für die Preisentwicklung liefert der Podcast „[StatGespräch](#)“. Und wer wissen möchte, wie sehr ihre oder seine persönliche von der amtlichen Teuerungsrate abweicht, nutzt dazu den persönlichen [Inflationsrechner](#).

AUS EUROPA

106. DGINS-Konferenz

Die Konferenz der Generaldirektorinnen und Generaldirektoren europäischer Statistikämter (DGINS) hat am 27. und 28. Oktober 2021 in Warschau das Thema Erdbeobachtung für Zwecke der amtlichen Statistik behandelt.

Die Bedeutung der Nutzung von Erdbeobachtungsdaten für statistische Zwecke wächst rasant und das wissenschaftliche Potenzial dieser Daten ist eine wichtige Quelle für künftige innovative Statistikprodukte. Viele staatliche Dienste und die Industrie verwenden bereits Erdbeobachtungsdaten. Auch nationale statistische Ämter untersuchen die Nutzung von Satellitenbilddaten und entwickeln amtliche Statistiken zu einer Vielzahl von Themen, wie Landwirtschaft, Umwelt (einschließlich Notfallmanagement) oder Verkehr. Diese Daten eröffnen Möglichkeiten, Statistiken schneller zur Verfügung zu stellen, einen stärkeren Detaillierungsgrad von Ergebnissen zu erreichen und den Beantwortungsaufwand für Auskunftspflichtige zu reduzieren.

Das Hauptziel der Konferenz bestand darin, Chancen und Herausforderungen zu identifizieren und die zentrale Frage zu klären, wie diese neuen Datenquellen und Technologien in die reguläre amtliche Statistikproduktion integriert werden könnten.

Neben Vertreterinnen und Vertretern des Europäischen Statistischen Systems (ESS) berichteten Fachleute aus der Raumfahrt und anderen Fachgebieten über Programme wie Copernicus, die Nutzung von Cloud-Technologien für die Auswertung von Satellitendaten und nutzerfreundlichen Anwendungen sowie die regionale beziehungsweise lokale Analyse von Satellitendaten.

Deutschland beteiligte sich mit einem Vortrag zur Schätzung der landwirtschaftlichen Produktion mithilfe von Satellitenfernerkundung (siehe den Beitrag Arnold und andere in dieser Ausgabe auf Seite 43 ff.).

Die weiteren Arbeiten werden die zuständigen strategischen und fachlichen Gremien des ESS in den kommenden Jahren behandeln.

47. Sitzung des AESS

Der Ausschuss für das Europäische Statistische System (AESS) hat in seiner Sitzung am 29. Oktober 2021 den Europäischen Grünen Deal behandelt. Das Statistische Amt der Europäischen Union (Eurostat) legte einen aktualisierten Aktionsplan über Statistiken für den Europäischen Grünen Deal vor, der die laufenden statistischen Aktivitäten in diesem Zusammenhang koordiniert und priorisiert. Der AESS billigte im Grundsatz diesen Aktionsplan. Gleichwohl betonten zahlreiche Mitgliedstaaten die Notwendigkeit, ausreichende Ressourcen für eine erfolgreiche Umsetzung des Aktionsplans bereitzustellen. Überdies soll Eurostat im Zuge der Evaluation bestehender EU-Rechtsverordnungen auch negative Prioritäten benennen, um eine zusätzliche Mehrbelastung sowohl für die Mitgliedstaaten als auch für die Auskunftgebenden zu vermeiden. Inhaltlich werden sich die Direktorengruppen mit praktischen Aspekten und technischen Fragen des Aktionsplans befassen, Eurostat wird regelmäßig über die Fortschritte bei der Umsetzung des Aktionsplans informieren.

Der AESS hat den revidierten „Leitlinien für die Bewertung von Forschungseinrichtungen, Forschungsanträgen und Zugangseinrichtungen“ zugestimmt. Diese regeln Grundsätze und Details, um einen Zugriff von Forschungseinrichtungen aus ihren Räumlichkeiten auf Mikrodaten des Europäischen Statistischen Systems (ESS) zu ermöglichen. Ferner wurde betont, auch die europäischen Statistikämter als nationale Kontaktpunkte zu definieren, damit auch sie die neue Datenzugriffsmöglichkeit nutzen können. Deutschland wies darauf hin, dass auf nationaler Ebene die Klärung rechtlicher Fragen noch nicht abgeschlossen ist.

Die Europäische Zentralbank berichtete über die Überarbeitung ihrer geldpolitischen Strategie und den damit verbundenen Ansatz, die Kosten des selbst genutzten Wohneigentums künftig im europäischen Verbraucherpreisindex zu berücksichtigen. Es ist vorgesehen, hierbei zunächst einen vierteljährlichen Indikator auf experimenteller Basis zu erstellen. Grundsätzlich begrüßt Deutschland dieses Vorgehen, sieht aber hinsichtlich der methodischen Umsetzung noch weiteren Klärungsbedarf.

Eurostat legte einen Vorschlag zur Änderung der bisherigen ESVG-Verordnung 549/2013 vor, der aufgrund fehlender Ressourcen auf Bedenken stieß. So sprach sich Deutschland gegen eine Ausweitung der Verfügbarkeit von Statistiken über die Staatsfinanzen im Rahmen des neuen Lieferprogramms aus.

Das ESS hat ein System von Reviews wichtiger Statistikbereiche gestartet, um künftig besser Prioritäten zu definieren und Ressourcen zu planen. Das erste Projekt befasste sich mit der Überprüfung der Statistiken zum Klimawandel. Den Bericht Eurostats über die vorläufigen Ergebnisse nahm der AESS zur Kenntnis. Ferner billigte der AESS gesundheitsbezogene Statistiken als Thema für die nächste Überprüfung, an der Deutschland, Spanien, Frankreich, Ungarn, die Niederlande und Portugal teilnehmen wollen.

Dem AESS lag zudem der erste Entwurf für eine Innovationsagenda für das ESS vor. Ihr sollen wichtige Digitalisierungsprojekte, beispielsweise Gaia-X, ein europäisches Projekt zum Aufbau einer Dateninfrastruktur in Europa, hinzugefügt werden.

AUS DEM INLAND

Neue Statistik zu untergebrachten wohnungslosen Personen

Auf Ebene des Bundes und für die meisten Bundesländer liegen bislang keine aussagekräftigen Daten zu Umfang und Ausmaß von Wohnungslosigkeit und den hiervon betroffenen Personen vor. Für sozialpolitisch fundierte Entscheidungen sind jedoch belastbare Informationen auf Bundes-, Länder- und kommunaler Ebene erforderlich. Diese Datenlücke wird in Zukunft geschlossen: Das Wohnungslosenberichterstattungsgesetz (WoBerichtsG) vom 4. März 2020 sieht die Einführung einer Wohnungslosenberichterstattung sowie einer Statistik über untergebrachte wohnungslose Personen vor.

Durchgeführt wird die neue Statistik vom Statistischen Bundesamt jährlich zum Stichtag 31. Januar. Dies geschieht erstmals zum 31. Januar 2022. Für die in ihrem örtlichen Zuständigkeitsbereich untergebrachten wohnungslosen Personen sind diejenigen Stellen auskunftspflichtig, die nach Landesrecht für die polizei- und ordnungsrechtliche Unterbringung zuständig sind. Außerdem können Stellen als auskunftspflichtig benannt werden, die Räume zu Wohnzwecken überlassen oder Übernachtungsmöglichkeiten zur Verfügung stellen.

Für die Statistik werden bei Ordnungs- oder Sozialämtern, aber auch bei auskunftspflichtigen Einrichtungen der Wohnungslosenhilfe Daten über untergebrachte wohnungslose Personen erhoben. Zu den Erhebungsmerkmalen gehören Angaben zur untergebrachten Person wie Alter, Geschlecht und Staatsangehörigkeit, aber auch Angaben zu Art, Ort und Angebot der Unterbringung. Die Veröffentlichung der ersten Ergebnisse erfolgt Mitte 2022 auf der Themenseite:

↳ www.destatis.de/wohnungslosigkeit

VERANSTALTUNGEN

Call for Abstracts: Forschungsprojekte auf Basis amtlicher Mikrodaten der Forschungsdatenzentren des Bundes und der Länder

Forschen mit amtlichen Mikrodaten – so lautet der Titel einer zweitägigen Konferenz der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder am 17. und 18. Mai 2022 in Düsseldorf.

Über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder stehen unterschiedlichste Mikrodaten der amtlichen Statistik für die wissenschaftliche Nutzung zur Verfügung. Nachdem die Konferenzen in den vergangenen Jahren jeweils den Fokus auf einzelne Themen beziehungsweise Statistiken legten, steht bei der kommenden Veranstaltung die Vielfalt der wissenschaftlichen Forschung im Mittelpunkt.

Die Konferenz soll ein breites Spektrum wissenschaftlichen Arbeitens mit amtlichen Mikrodaten widerspiegeln: Ungleichheit, Migration, Gesundheit, Wirtschaft, Recht, Finanzen, Demografie ... Der Call for Abstracts wendet sich an Wissenschaftlerinnen und Wissenschaftler aus dem In- und Ausland, die bereits mit Mikrodaten der amtlichen Statistik arbeiten oder gearbeitet haben. Auf Basis dieser Daten sollen aktuelle Forschungsergebnisse diskutiert werden. Zugleich bietet die Konferenz Raum für Erfahrungsaustausch, Diskussionen und Netzwerken zwischen Forschenden und Datenproduzierenden.

Der Call for Abstracts schließt am 28. Februar 2022.

↳ www.forschungsdatenzentrum.de

DAGStat-Konferenz 2022

Die sechste Konferenz der Deutschen Arbeitsgemeinschaft Statistik (DAGStat) wird vom 28. März bis 1. April 2022 in Hamburg stattfinden. Alle Informationen zu Ablauf und Programm enthält die Webseite der Konferenz:

↳ www.dagstat2022.uni-hamburg.de

Q2022 in Vilnius

Die European Conference on Quality in Official Statistics findet vom 7. bis 10. Juni 2022 in der litauischen Hauptstadt Vilnius statt.

Ziel der Konferenz ist es, Statistik, Wissenschaft und Interessenvertretungen zusammenzubringen, um die Verbreitung von Wissen und neuer Forschung zu neuen Fragen der Qualität in der amtlichen Statistik zu stärken. Darüber hinaus soll sie die Zusammenarbeit und den Austausch zwischen der amtlichen Statistik und der wissenschaftlichen Gemeinschaft verbessern und ein Forum für den Austausch bewährter Verfahren bei der Bewältigung neuer Herausforderungen in der amtlichen Statistik bieten.

Die Q2022 wird sich thematisch auf die institutionellen Herausforderungen des Qualitätsmanagements, der Qualitätssicherung im entstehenden Datenökosystem und neue Herausforderungen bei der Schaffung und Aufrechterhaltung von Vertrauen unter den Nutzern sowie auf neue Möglichkeiten angesichts globaler Krisen konzentrieren.

Ebenso soll sie die Aufmerksamkeit staatlicher Stellen auf die Bedeutung qualitativ hochwertiger, zeitnahe und detaillierter Statistiken lenken und andere Interessengruppen für die Herausforderungen sensibilisieren, denen sich Produzenten amtlicher Statistiken insbesondere in Krisenzeiten gegenübersehen.

Die offizielle Website enthält alle wichtigen und nützlichen Informationen rund um die Q2022:

↘ <https://q2022.stat.gov.lt>

20. D-A-CH-Demografietreffen

Vom 27. bis 29. Oktober 2021 fand das 20. deutsch-österreichisch-schweizerische Demografie-Treffen (D-A-CH) als Online-Konferenz statt. Die D-A-CH Demografie-Treffen werden alle zwei Jahre abwechselnd in einem der drei Länder ausgerichtet. Sie dienen dem Austausch zwischen der amtlichen Bevölkerungsstatistik und der soziodemografischen Forschung. Die diesjährige Veranstaltung mit rund 60 Teilnehmenden aus 14 Institutionen wurde vom Statistischen Bundesamt und dem Bundesinstitut für Bevölkerungsforschung organisiert.

Der thematische Schwerpunkt des diesjährigen Treffens lag auf der Corona-Pandemie und ihren Auswirkungen auf die demografische Entwicklung in den D-A-CH-Ländern, weitere Themen waren gesunde Lebenserwartung, Mobilität und Migration, Diversität sowie methodische Fragen.

Das Programm der Tagung und die einzelnen Präsentationen sind online verfügbar:

↘ www.bib.bund.de/DACH21

„Berliner Demografiegespräch“ zur Geburtenentwicklung während der COVID-19-Pandemie

Die Vortragsreihe „Berliner Demografiegespräche“ für die Zielgruppe Politik wurde am 2. November 2021 mit einer Veranstaltung zum Thema „Geburtenknick oder Baby-Boom? Die Covid-19-Pandemie und die Geburtenentwicklung“ fortgeführt.

Das dritte „Berliner Demografiegespräch“ stieß auf großes Interesse, etwa 100 Personen aus dem Deutschen Bundestag und verschiedenen Bundesministerien nahmen an der Webkonferenz teil. Die Vorträge und die Diskussion stehen als Aufzeichnung auf dem [YouTube-Kanal des Bundesinstituts für Bevölkerungsforschung](#) zur Verfügung.

30. Wissenschaftliches Kolloquium: Von der Umweltstatistik zur nachhaltigen Entwicklung

Renommierte europäische und deutsche Expertinnen und Experten berichteten am 18. und 19. November 2021 über aktuelle Handlungsfelder, Instrumente und Möglichkeiten, wie Daten und Statistiken bei der Bewältigung der Klimakrise unterstützen können. Erstmals vorgestellt wurde in diesem Rahmen auch der Ökosystematlas des Statistischen Bundesamtes.

Die Vorträge und Diskussionen der Veranstaltung stehen unter [↘ www.destatis.de](http://www.destatis.de) zur Verfügung.

NEUE INTERAKTIVE ANWENDUNGEN

Neuer digitaler Atlas zeigt die Vielfalt der Ökosysteme in Deutschland

Deutschland verfügt über eine Vielzahl verschiedener Ökosysteme (wie Siedlungsflächen, Agrarland, Wälder und Gewässer), deren Verbreitung regional stark variiert. Ihre Bedeutung für den Klimaschutz und die Biodiversität nimmt mehr und mehr zu. Der neue Ökosystematlas des Statistischen Bundesamtes zeigt die Vielfalt und räumliche Verteilung der Ökosysteme in interaktiver Weise. In verschiedenen thematischen Karten werden für alle Gemeinden beziehungsweise Gemeindeverbände die Flächen der Ökosysteme und deren Veränderung detailliert visualisiert.

↳ www.destatis.de

Neuer Revisionskalender online

Die amtliche Statistik befindet sich in einem Spannungsfeld zwischen Aktualität und Genauigkeit: Nutzerinnen und Nutzer fordern einerseits hochaktuelle Daten, andererseits werden aber auch genaue und belastbare Statistiken benötigt. Um diesen beiden Qualitätsanforderungen gerecht zu werden, werden in manchen Statistiken zunächst vorläufige Ergebnisse veröffentlicht. Revisionen werden hierdurch notwendig.

Der neue Revisionskalender des Statistischen Bundesamtes steht nun online zur Verfügung. Er dient dazu, den Prozess der Revisionen allen Nutzerinnen und Nutzern transparent zu machen und gibt einen Überblick darüber, bei welchen Statistiken (geplante) Revisionen durchgeführt werden und nach welchem Zyklus diese erfolgen.

↳ www.destatis.de

Neue Version des interaktiven Tourismusatlas

Bis einschließlich Februar 2020 war der Tourismus in Deutschland über lange Jahre hinweg stabil im Aufwärtstrend. Dann setzte die Corona-Pandemie ein und

die deutsche Tourismusbranche geriet in schweres Fahrwasser. Der erweiterte interaktive Tourismusatlas der Statistischen Ämter des Bundes und der Länder zeigt anhand von Geodaten erstmals auch kleinräumige Auswirkungen der Corona-Pandemie im Tourismusbereich mit aktuellen Daten für 2020.

↳ gis-hsl.hessen.de

NEUERSCHEINUNGEN

Transformation gestalten: Bildung, Digitalisierung und Nachhaltigkeit

Das Jahresgutachten 2021/22 des Sachverständigenrats zur Begutachtung der gesamtwirtschaftlichen Entwicklung trägt den Titel „Transformation gestalten: Bildung, Digitalisierung und Nachhaltigkeit“. Es wurde am 10. November 2021 an die Bundesregierung überreicht und steht auf der Webseite des Sachverständigenrats zur Verfügung:

↳ www.sachverstaendigenrat-wirtschaft.de

Health at a Glance 2021: OECD Indicators

Wie ist es um die Gesundheit der Menschen in Zeiten von COVID-19 bestellt? Die Pandemie hat beispielsweise direkt oder indirekt zu einem Anstieg der erwarteten Zahl der Todesfälle um rund 16 % im OECD-Raum beigetragen. Die Lebenserwartung sank in 24 von 30 Ländern, für die vergleichbare Daten vorliegen. Das Minus liegt dabei in Deutschland bei 0,3 Jahren, in Österreich und der Schweiz bei 0,7 beziehungsweise 0,8 Jahren; in den Vereinigten Staaten verringerte sich die Lebenserwartung sogar um 1,6 Jahre. Stark zugenommen haben dagegen Angststörungen und Depressionen.

Der OECD-Gesundheitsbericht „Health at a Glance 2021“ bietet einen umfassenden Satz von Indikatoren zur Gesundheit der Bevölkerung und zur Leistungsfähigkeit der Gesundheitssysteme in den OECD-Mitgliedsländern und den wichtigsten Schwellenländern. Der Schwer-

punkt dieser Ausgabe liegt auf den Auswirkungen von COVID-19.

↳ www.oecd-ilibrary.org

Learning from the Past, Looking to the Future

Die COVID-19-Krise hat die Unzulänglichkeiten und Ungerechtigkeiten in den Bildungssystemen noch verstärkt. Da uns die Zukunft weiterhin überraschen wird, wird die Bedeutung von Widerstandsfähigkeit, Anpassungsfähigkeit und Fairness nur noch zunehmen. Gerechte Schulbildung bedeutet dabei mehr als Gleichbehandlung: Bildung sollte sich an die Unterschiede von Schülerinnen und Schülern anpassen.

Die OECD-Veröffentlichung „Learning from the Past, Looking to the Future: Excellence and Equity for all“ geht der Frage nach, wie sich Bildungssysteme auf künftige Veränderungen vorbereiten können. Dabei werden unter anderem Lernziele der verschiedenen Länder vorgestellt, das Bildungsziel einer ganzheitlichen Entwicklung des Kindes – also ein Gleichgewicht von kognitiver, psychologischer, sozialer und physischer Entwicklung – untersucht und dargelegt, wie sich die Chancengleichheit verbessern lässt.

↳ www.oecd-ilibrary.org

COVID-19 and Well-being: Life in the Pandemic

Über die körperliche Gesundheit hinaus hat die COVID-19-Pandemie gravierende Auswirkungen auf fast alle Lebensbereiche. Sehr viele Menschen haben in der Pandemie Angehörige verloren. Gleichzeitig nahmen Depressionen und Einsamkeit zu. Auch Angst vor Arbeitsplatzverlust und finanzieller Not war für viele ein bestimmendes Thema.

Die OECD-Veröffentlichung „COVID-19 and Well-being: Life in the Pandemic“ untersucht insgesamt elf Dimensionen, die maßgeblich mitbestimmen, ob es Menschen gut geht: Einkommen und Wohlstand, Arbeit und Arbeitsplatzqualität, Wohnsituation, bürgerschaftliches Engagement, Gesundheit, Umwelt, Sicherheit, subjektives

Wohlbefinden, soziale Beziehungen, Vereinbarkeit von Beruf und Privatleben sowie Wissen und Fähigkeiten. Die Studie betrachtet außerdem soziale, gesamtwirtschaftliche und umweltbezogene Nachhaltigkeitsfaktoren. Daten aus den ersten 12 bis 15 Pandemienmonaten zeigen in all diesen Bereichen deutliche Veränderungen.

↳ www.oecd-ilibrary.org

VERLEIHUNG DES GERHARD-FÜRST-PREISES 2021

Georg Thiel, Walter Krämer

↳ **Schlüsselwörter:** Gerhard-Fürst-Preis – amtliche Statistik – Nachwuchspreis – Bevölkerungsindikatoren – Imputationsalgorithmen – Tax Payer Panel – „Unternehmenssparen“ im Nichtfinanzsektor

ZUSAMMENFASSUNG

Mit dem Gerhard-Fürst-Preis des Statistischen Bundesamtes werden Arbeiten in den Kategorien Dissertationen und Master- beziehungsweise Bachelorarbeiten ausgezeichnet, die theoretische Themen mit einem engen Bezug zum Aufgabenspektrum der amtlichen Statistik behandeln oder empirische Fragestellungen unter intensiver Nutzung von Daten der amtlichen Statistik untersuchen.

Das Statistische Bundesamt möchte mit diesen jährlichen wissenschaftlichen Auszeichnungen die Zusammenarbeit zwischen Wissenschaft und amtlicher Statistik weiter intensivieren. Zugleich sollen die Preise junge Wissenschaftlerinnen und Wissenschaftler ermutigen, das vielfältige Datenangebot der amtlichen Statistik für ihre empirischen Forschungen ausgiebig zu nutzen.

↳ **Keywords:** Gerhard Fürst Award – official statistics – prize for young researchers – population indicators – imputation algorithms – tax payer panel – “business savings” in the non-financial sector

ABSTRACT

The Federal Statistical Office grants the Gerhard Fürst Awards to doctoral, Master’s and Bachelor’s theses which deal with theoretical topics with a close link to the scope of duties of official statistics, or examine empirical questions with the intense use of data from official statistics.

The annual scientific awards of the Federal Statistical Office are aimed at intensifying the cooperation between the scientific community and the official statistical agencies. At the same time, the awards are intended to encourage junior scientists to make extensive use in their empirical research of the wide variety of data offered by official statistics.



Dr. Georg Thiel

studierte Rechtswissenschaften und promovierte an der Universität zu Köln. Er ist seit 2017 Präsident des Statistischen Bundesamtes und Bundeswahlleiter. Zuvor war er Vizepräsident des Statistischen Bundesamtes und Stellvertreter des Bundeswahlleiters, ständiger Vertreter der Abteilungsleiterin O „Verwaltungsmodernisierung, Verwaltungsorganisation“ im Bundesministerium des Innern und Präsident der Bundesanstalt Technisches Hilfswerk.



Prof. Dr. Walter Krämer

ist Direktor des Sonderforschungsbereichs „Statistical modelling of nonlinear dynamic processes“ an der Fakultät Statistik der Technischen Universität Dortmund. Seit dem Jahr 2014 ist er Vorsitzender des unabhängigen Gutachtergremiums, das die wissenschaftlichen Arbeiten zur Auszeichnung mit dem Gerhard-Fürst-Preis des Statistischen Bundesamtes empfiehlt.

Einleitung

Auf Empfehlung eines unabhängigen Gutachtergremiums zeichnet das Statistische Bundesamt seit 1999 jährlich herausragende wissenschaftliche Arbeiten mit einem engen Bezug zur amtlichen Statistik mit dem Gerhard-Fürst-Preis aus. Für den Gerhard-Fürst-Preis 2021 hat das Gutachtergremium eine Arbeit in der Kategorie „Dissertationen“ sowie drei Arbeiten in der Kategorie „Master-/Bachelorarbeiten“ als gleichermaßen herausragend und preiswürdig erachtet. Das Statistische Bundesamt hat am 18. November 2021 somit vier wissenschaftliche Arbeiten mit dem Gerhard-Fürst-Preis 2021 prämiert:

- › Für seine an der Universität Trier bei Professor Dr. Ralf Münnich entstandene Dissertation zum Thema „Regularization Methods for Statistical Modelling in Small Area Estimation“ wurde Dr. Joscha Krause ausgezeichnet. Die präzise Schätzung von Bevölkerungsindikatoren auf regionaler Ebene bei geringen Stichprobenumfängen sahen die Gutachterinnen und Gutachter als besonders förderwürdig an.
- › Maria Thurow erhielt den Preis für ihre Bachelorarbeit „Optionen zur Bemessung des Abstandes zweier Verteilungen in der Praxis“. Ihre Abschlussarbeit, die sich mit der Anwendung und vergleichenden Bewertung von Imputationsalgorithmen befasst, hat eine große praktische Relevanz für statistische Ämter. Ihre Arbeit entstand an der Technischen Universität Dortmund bei Prof. Dr. Markus Pauly.
- › David Kläffling nahm den Förderpreis für seine Bachelorarbeit „Corporate Governance and Corporate Saving in Advanced Economies“ in Empfang. In seiner Abschlussarbeit zeigt David Kläffling das Phänomen des „Unternehmenssparens“ im Nichtfinanzsektor auf. Die Arbeit entstand bei Dr. Michael Schwan an der Universität zu Köln.
- › Christina Meyer bekam den Förderpreis für ihre Masterarbeit zum Thema „Gender-specific retirement saving behavior. Evidence from the German Tax Payer Panel“. Ihre Auswertung von Daten mit ökonomischen Methoden demonstriert nachdrücklich den Wert des Tax Payer Panel (TPP) für die angewandte Wirtschaftsforschung. Christina Meyer verfasste ihre Arbeit an der Ludwig-Maximilians-Universität München bei Prof. Dr. Joachim Winter.

Die Preisverleihung fand im Rahmen des 30. Wissenschaftlichen Kolloquiums in Wiesbaden statt. Der Präsident des Statistischen Bundesamtes, Dr. Georg Thiel, überreichte den Preisträgerinnen und den Preisträgern ihre Urkunden. Die Laudationes auf die prämierten Arbeiten hielt der Vorsitzende des Gutachtergremiums, Professor Dr. Walter Krämer (Technische Universität Dortmund). Die Preisträgerinnen und Preisträger berichten im Jahr 2022 in dieser Zeitschrift ausführlich über ihre Arbeiten.

Die [Kurzfassungen](#) der prämierten Arbeiten sowie weitere Details zur Verleihung des Gerhard-Fürst-Preises sind im [Internetangebot](#) des Statistischen Bundesamtes zu finden.

Laudationes Gerhard-Fürst-Preis 2021

Dr. Georg Thiel

Sehr geehrte Damen und Herren,

ich begrüße Sie zur Verleihung des Gerhard-Fürst-Preises. Die Laudationes wird der Vorsitzende des Gutachtergremiums, Herr Prof. Dr. Walter Krämer, übernehmen, den ich hier herzlich willkommen heiße. Zuvor möchte ich Ihnen als Präsident des Statistischen Bundesamtes die diesjährigen Preisträgerinnen und Preisträger kurz vorstellen.

In der Kategorie „Dissertationen“ wird der mit 5 000 Euro dotierte Gerhard-Fürst-Preis an Herrn **Dr. Joscha Krause** von der Universität Trier verliehen.

Er hat seinen Dokortitel der Rechts- und Staatswissenschaften im Schwerpunkt „Computergestützte Statistik, robuste Statistik, statistische Modellierung“ summa cum laude (1,0) abgeschlossen. Zuvor hat Herr Dr. Krause – ebenfalls an der Universität Trier – zwei „Master of Science“-Abschlüsse jeweils mit der Note 1,4 bestanden – einen Master in „Ökonometrie, Wohlfahrtsökonomik“ und den anderen in „Computergestützte Statistik, statistische Modellierung“. Auch seinen „Bachelor of Science in Economics“ hat er zuvor mit 1,9 absolviert.

Aktuell arbeitet Herr Dr. Krause als Post-Doctoral-Researcher am Lehrstuhl Wirtschafts- und Sozialstatistik an der Universität Trier. Seine Schwerpunkte sind Data Science, maschinelles Lernen und statistische Modellierung.

Herr Prof. Krämer, Sie haben nun das Wort für die erste Laudatio.

Prof. Dr. Walter Krämer

Sehr geehrte Damen und Herren,

traditionsgemäß gehört der Abschluss des ersten Tages unseres jährlichen wissenschaftlichen Kolloquiums dem Gerhard-Fürst-Preis. Der ist jetzt 22 Jahre alt und damit fast so alt wie das Kolloquium. Über den größten Teil dieser Zeit war ich selbst als Jurymitglied und zuletzt als Juryvorsitzender aktiv in die Preisverleihung eingebunden, und es war mir immer wieder eine Freude zu sehen, auf welch furchtbaren Boden diese Initiative des Statistischen Bundesamtes gefallen ist. Ich glaube, der erste Präsident des Statistischen Bundesamtes Gerhard Fürst, nach dem der Preis benannt ist, hätte seine Freude daran gehabt. Vergeben wird der Preis für akademische Abschlussarbeiten, die entweder an zentraler Stelle mit Daten der Amtsstatistik arbeiten oder die für die Amtsstatistik wichtige Methoden weiterentwickeln.

Dieses Jahr hatten wir wie fast immer ein Luxusproblem: Es gab mehr gute Arbeiten als Preise. Wir vergeben heute vier, zwei an Männer, zwei an Frauen. Das war nicht so geplant, kann aber gerne als unser Beitrag zum Internationalen Jahrzehnt der Gleichberechtigung gewertet werden.

Der Gerhard-Fürst-Preis für die beste Doktorarbeit geht dieses Jahr an Joscha Krause vom Fachbereich Wirtschafts- und Sozialwissenschaften, Mathematik und Informatikwissenschaften der Universität Trier. Hier geht es um ein Methodenproblem. Auch solche Arbeiten kommen für den Gerhard-Fürst-Preis infrage, sofern diese Methoden für die Amtsstatistik von Bedeutung sind. Das ist hier eindeutig der Fall. Wann immer die Amtsstatistik mit Stichproben arbeitet und zugleich auch regional hoch aufgelöste Daten ausweisen will, kommt man um mathematisch-statistische Hilfsmittel, gemeinhin unter Small-Area-Methoden zusammengefasst, nicht herum. Denn der Stichprobenfehler wäre auf Regionalebene dann einfach zu groß. Um den zu reduzieren, baut man Informationen aus vergleichbaren Regionen in ein Modell zur Schätzung der interessierenden Zielvariablen ein. Der Beitrag von Herrn Krause zu diesem hochinteressanten Thema besteht nun darin, die Parameterschätzung mit der Modellwahl zu vereinen (das firmiert in der Mathematischen Statistik unter

Regularisierung) und dabei durchaus neue Wege einzuschlagen. Insbesondere sind die hier vorgeschlagenen Schätzer robust gegen Messfehler, wie sie insbesondere dann zu erwarten sind, wenn Social-Media-Daten aus dem Internet in die Parameterschätzung und Modellanpassung eingehen. Damit schlägt Herr Krause zugleich eine Brücke zum Thema Big Data und internetbasierte Hilfestellung für die Amtsstatistik, das ja im Moment immer wichtiger wird.

Herr Krause, herzlichen Glückwunsch, Sie haben diesen Preis verdient!

Dr. Georg Thiel

In der Kategorie „Master-/Bachelorarbeiten“ geht der diesjährige mit 2 500 Euro dotierte Gerhard-Fürst-Preis an Frau **Maria Thurow** für ihre Bachelorarbeit an der Technischen Universität Dortmund.

Sie hat ihr Bachelorstudium an der Technischen Universität Dortmund abgeschlossen.

Im Herbst 2020 hat Frau Thurow ihr Masterstudium im Fach Statistik mit dem Nebenfach Informatik an der Technischen Universität Dortmund begonnen.

Seit Anfang des Jahres 2021 arbeitet sie an der Fakultät Statistik der TU Dortmund als wissenschaftliche Hilfskraft.

Für die zweite Laudatio übergebe ich an Herrn Prof. Krämer.

Prof. Dr. Walter Krämer

Nun zu den Bachelor- und Masterarbeiten. Hier haben mehrere Beiträge die Juroren begeistert, und einer ganz besonders. Es geht um fehlende Werte bei Erhebungen aller Art. Davon weiß ja auch die Amtsstatistik ein eher trauriges Lied zu singen. Wie füllt man diese Leerstellen auf optimale Weise auf, sodass sich die resultierende Verteilung möglichst wenig von der unbekanntenen wahren Verteilung der Daten unterscheidet? Wer dazu etwas wissen will, findet keine bessere Quelle als die Bachelorarbeit von Maria Thurow von der Fakultät Statistik der TU Dortmund. Auf einem Niveau, das für eine Bachelorarbeit absolut ungewöhnlich ist, fasst sie hier den Stand der Methodenforschung zusammen und wendet diese dann auf zwei Datensätze aus der Amtsstatistik an, die

CAMPUS-Files der DRG-Statistik¹ und den Arbeitnehmerdatensatz der VSE²-Statistik. Für die Expertinnen und Experten hier im Saal: Frau Thurow zeigt, dass die sogenannten Random-Forest-Verfahren hier besonders gut geeignet sind.

Frau Thurow, das haben Sie gut gemacht, herzlichen Glückwunsch zu diesem Preis!

Dr. Georg Thiel

Zudem hat die Jury 2021 die Vergabe von zwei Förderpreisen in der „Kategorie Abschlussarbeiten Master/Bachelor“ mit einem Preisgeld von jeweils 1000 Euro beschlossen.

Förderpreis Nummer 1 vergibt die Jury an Herrn David Kläffling, der im Jahr 2020 an der Universität zu Köln in Volkswirtschaftslehre und Politikwissenschaften sein Bachelorstudium summa cum laude (1,0) abgeschlossen hat.

Im vergangenen Herbst hat Herr Kläffling sein Masterstudium „Public Economics“ an der Freien Universität Berlin aufgenommen.

Herr Prof. Krämer, sie haben nun das Wort für die dritte Laudatio.

Prof. Dr. Walter Krämer

Leider gibt es nur einen Gerhard-Fürst-Preis, aber es gibt die Möglichkeit für einen Förderpreis. Und dieses Jahr war die Qualitätsdichte so hoch, wir vergeben sogar zwei. Der erste geht an David Kläffling von der FU Berlin. Seine preisgekrönte Bachelorarbeit hat er allerdings an der Uni Köln geschrieben. Darin geht es um ein heißes Thema der aktuellen Wirtschaftstheorie wie auch der empirischen Wirtschaftsforschung: Was ist bloß mit unseren Unternehmen los? Während traditionellerweise Haushalte sparen und Unternehmen diese Ersparnisse investieren, ist der Unternehmenssektor in vielen entwickelten Volkswirtschaften selbst fast schon zu einem Nettosparer geworden. Herr Kläffling diskutiert zunächst auf einem für eine Bachelorarbeit wiederum ungewöhnlich anspruchsvollen Niveau verschiedene theoretische Ansätze zur Erklärung dieses Sachverhalts und schätzt

1 DRG(Diagnosis Related Groups)-Statistik=fallpauschalenbezogene Krankenhausstatistik.

2 VSE=Verdienststrukturerhebung.

dann mit Daten von Eurostat eine Verhaltensgleichung mit länderspezifischen Anteilen der Unternehmensersparnis am Sozialprodukt, mit dem Ergebnis, dass die Unternehmensersparnisse ganz wesentlich von der Art der Unternehmensüberwachung (auf Deutsch Corporate Governance) abhängen: Je mehr Rechte und Eingriffsmöglichkeiten Aktionäre haben, desto weniger wird gespart (sondern stattdessen als Dividende an die Eigentümer ausgezahlt).

Das ist vielleicht ex post nicht allzu überraschend, aber es ist doch immer wieder schön, wenn die ökonomische Wirklichkeit sich so verhält, wie die Theorie es uns erwarten lässt. Und das haben Sie, Herr Kläffling, sehr kompetent gezeigt. Herzlichen Glückwunsch zu dem Förderpreis!

Dr. Georg Thiel

Den zweiten Förderpreis vergibt die Jury an Frau Christina Meyer, Ludwig-Maximilians-Universität München.

Sie hat ihren „Master of Science in Economics“ an der Ludwig-Maximilians-Universität München mit der Note 1,2 abgeschlossen. Zuvor erwarb Frau Meyer Ihren „BA Philosophy & Economics“ an der Universität Bayreuth mit der Abschlussnote 1,2.

Seit Frühjahr 2020 ist Frau Meyer als wissenschaftliche Mitarbeiterin am Leibniz-Zentrum für Europäische Wirtschaftsforschung im Bereich Internationale Finanzmärkte tätig.

Seit September 2020 arbeitet sie an der GESS³ an der Universität Mannheim an ihrem PhD am Center for Doctoral Studies in Economics.

Wir freuen uns auf die vierte Laudatio. Herr Prof. Krämer, ich darf Sie noch einmal bitten.

Prof. Dr. Walter Krämer

Der zweite Förderpreis geht an Frau Christina Meyer von der Ludwig-Maximilians-Universität München für ihre Masterarbeit zur Riester-Rente und wie sich unser Verhalten bei geänderten Gesetzen ebenfalls ändert. Vor 2006 war zum Beispiel das Geschlecht ein wichtiges Kriterium der Risikobewertung von Anbietern von Riester-Produkten. Danach durfte zwischen Männern

3 GESS=Graduate School of Economic and Social Sciences.

und Frauen nicht mehr unterschieden werden. Das hatte zum Beispiel zur Folge, dass Männer nach 2006 für die gleiche Beitragszahlung geringere Erträge erhalten als Männer, die vor der Reform einen Vertrag abgeschlossen haben. Dies sollte für Männer dazu führen, dass sie seltener als zuvor einen Vertrag abschließen. Diese und andere Konsequenzen werden sorgfältig theoretisch abgeleitet und anhand von Daten der Amtsstatistik überprüft. Ein Musterbeispiel für eine theoretisch und praktisch höchst anspruchsvolle Masterarbeit. Frau Meyer, herzlichen Glückwunsch.

Dr. Georg Thiel

Die Ehrung dieser herausragenden jungen Wissenschaftlerinnen und Wissenschaftler war uns eine Freude. Nun freue ich mich, unserem Laudator Herrn Prof. Dr. Walter Krämer unsere Wertschätzung auszusprechen. Herr Prof. Dr. Krämer, seit 2002 setzen Sie sich engagiert für die Förderung des Statistiknachwuchses im Rahmen des Gerhard-Fürst-Preises ein, seit 2014 bringen Sie Ihre Expertise und Erfahrungen als Vorsitzender des Gutachtergremiums mit ein. Wir möchten Ihnen herzlich für Ihren verlässlichen und leidenschaftlichen Einsatz über so viele Jahre hinweg danken. Und wir hoffen, Sie bleiben der Community gewogen. Wir freuen uns darauf, Sie im nächsten Jahr als interessierten Teilnehmer des Kolloquiums willkommen heißen zu dürfen. Dann können Sie sich zurücklehnen und die Verleihung der Gerhard-Fürst-Preise aus einer bequemerer Perspektive verfolgen.

Nach diesem feierlichen Akt und der Würdigung der diesjährigen Preisträgerinnen und Preisträger und der Verabschiedung unseres geschätzten Laudators Prof. Dr. Krämer erlaube ich mir den Hinweis auf eine Besonderheit im kommenden Jahr 2022:

Vor dem Hintergrund der Corona-Pandemie und der daraus entstandenen Relevanz aktueller Daten und Statistiken zu diesem Thema ist für den Gerhard-Fürst-Preis 2022 die Vergabe eines Corona-Sonderpreises in Planung. 

Dorothea Klumpen

ist Juristin (Ass. Jur.) und als Referentin im Referat „Europäische und internationale Koordination“ des Statistischen Bundesamtes tätig.

Christina Schliffka

ist Politikwissenschaftlerin und wissenschaftliche Mitarbeiterin im Referat „Europäische und internationale Koordination“ des Statistischen Bundesamtes.

Dominique Polus

ist Ökonomin und arbeitet als Referentin im Leitungsstab des Statistischen Bundesamtes.

DIE EUROPÄISCHE DATENSTRATEGIE – EIN BINNENMARKT FÜR DATEN

Dorothea Klumpen, Christina Schliffka, Dominique Polus

📌 **Schlüsselwörter:** Europäische Datenstrategie – europäischer Binnenmarkt für Daten – gemeinsame Datenräume – Europäisches Statistisches System

ZUSAMMENFASSUNG

Mit der Europäischen Datenstrategie hat sich die Europäische Kommission zum Ziel gesetzt, einen europäischen Binnenmarkt für Daten zu etablieren. Sie ist ein wichtiger Baustein der Antwort der Europäischen Union auf die Herausforderungen der Digitalisierung. Die Europäische Datenstrategie soll fundierte Entscheidungen auf der Grundlage von Daten ermöglichen und den Austausch und die Nutzung von Daten optimieren. An ihrer Umsetzung ist das Statistische Bundesamt als Teil des Europäischen Statistischen Systems sowohl als Produzent als auch als Nutzer von Daten beteiligt. Dieser Beitrag richtet den Blick auf geplante Inhalte und Zielsetzungen der Europäischen Datenstrategie sowie deren Umsetzung in konkreten Rechtsakten. Zudem stellt er Bedeutung und Chancen für die amtliche Statistik sowie deren Beitrag zur Umsetzung der Europäischen Datenstrategie vor.

📌 **Keywords:** European data strategy – European single market for data – common data spaces – European Statistical System

ABSTRACT

With the European data strategy, the European Commission has set itself the goal of establishing a European single market for data. The strategy constitutes an important element of the European Union's response to the challenges of digitalisation. The European data strategy is to enable sound decisions on the basis of data and optimise the exchange and use of data. Being part of the European Statistical System, the Federal Statistical Office is involved in implementing the strategy as both a data producer and data user. This paper takes a look at the planned content and objectives of the European data strategy and its implementation by specific legal acts. It also describes the strategy's importance, the opportunities it holds for official statistics and how official statistics authorities can contribute to implementing the European data strategy.

1

Einleitung

„Wir wollen den europäischen Weg ins Digitalzeitalter gehen – basierend auf unseren Werten, unserer Stärke und unseren globalen Ambitionen“ (Europäische Kommission, 2020a, hier: Seite 15). Diese Vision der digitalen Zukunft Europas zeichnet Ursula von der Leyen, die Präsidentin der Europäischen Kommission, in ihrer Rede zur Lage der Union im September 2020. Daten stehen im Mittelpunkt des digitalen Wandels. Deshalb hat die Europäische Kommission die Europäische Datenstrategie verabschiedet. Diese soll einen europäischen Binnenmarkt für Daten etablieren. Die Europäische Union (EU) soll sich so zu einer der attraktivsten, sichersten und agilsten Datenwirtschaften im internationalen Vergleich entwickeln. Die Europäische Datenstrategie soll

- › fundierte Entscheidungen auf der Grundlage von Daten ermöglichen und das Leben aller Europäerinnen und Europäer verbessern,
- › den Austausch und die Nutzung von Daten optimieren und gleichzeitig
- › die hohen Datenschutz-, Sicherheits- und Ethikstandards der EU wahren.

Derzeit arbeiten verschiedene Gremien des Europäischen Statistischen Systems (ESS) an der konkreten Ausgestaltung der Europäischen Datenstrategie mit. Insbesondere die COVID-19-Pandemie hat gezeigt, wie wichtig relevante und aktuelle statistische Daten von höchster Qualität sind. Die amtliche Statistik spielt hierbei eine zentrale Rolle. Um neuen Informationsbedarfen auch in Zukunft gerecht werden zu können, muss die amtliche Statistik die Nutzung neuer digitaler Datenquellen weiter ausbauen. Die Europäische Datenstrategie kann hierzu entscheidend beitragen. Auch das Statistische Bundesamt bringt sich als Teil des ESS in die Umsetzung der Europäischen Datenstrategie ein. Vor diesem Hintergrund beschreibt der Beitrag in Kapitel 2 Inhalte und Zielsetzungen der Europäischen Datenstrategie. Die Umsetzung dieser Ziele in konkreten Rechtsakten stellt Kapitel 3 dar. Bedeutung und Chancen der Europäischen Datenstrategie für die amtliche Statistik sowie deren Beitrag zur Umsetzung der Europäischen Datenstrategie sind Thema in Kapitel 4.

2

Ziele und Inhalte der Europäischen Datenstrategie

2.1 Ein Binnenmarkt für Daten – Zielsetzung

Die Digitalisierung hat in den letzten Jahren die Gesellschaft, die Wirtschaft und das tägliche Leben aller Europäerinnen und Europäer verändert. Daten stehen im Mittelpunkt dieses Wandels. Bei der Sammlung und Verwendung der immer größer werdenden Datenmengen sollten die Interessen der Einzelnen im Vordergrund stehen, im Einklang mit den europäischen Werten und Grundrechten.

Die Europäische Datenstrategie formuliert politische Maßnahmen und notwendige Investitionen für den Aufbau einer europäischen Dateninfrastruktur, um eine international führende Rolle der EU im Bereich der Datenwirtschaft sicherzustellen. Dabei verfolgt die Europäische Datenstrategie das Ziel, einen Binnenmarkt für Daten zu schaffen, im Sinne eines einheitlichen europäischen Datenraums.

Diesem Ziel liegen unter anderem folgende Prämissen zugrunde:

- › Auf der Grundlage von Daten wird der Wirtschaft, dem öffentlichen Sektor sowie den Bürgerinnen und Bürgern der EU ermöglicht, fundierte und dadurch bessere Entscheidungen zu treffen.
- › Bessere Entscheidungen führen zu einer Steigerung der Produktivität und zu einer Verbesserung des sozialen und wirtschaftlichen Wohlergehens innerhalb der EU.
- › Daten verändern die Art und Weise zu produzieren, zu verbrauchen und zu leben. Diese Potenziale gilt es zum Vorteil aller zu erschließen, angefangen bei einem effizienteren Energieverbrauch über die Rückverfolgbarkeit von Produkten bis zu einem gesünderen Leben und einer optimierten Gesundheitsversorgung.

2.2 Ein Binnenmarkt für Daten – Herausforderungen und Umsetzung

Um diese ehrgeizigen Ziele zu erreichen gilt es, einige Hürden zu überwinden. Besonders die Fragmentierung in den Mitgliedstaaten gefährdet die Verwirklichung eines gemeinsamen europäischen Datenraums. Des Weiteren sollen die Verfügbarkeit und der Austausch von Daten zwischen allen Beteiligten verbessert werden. Ein weiteres Problem stellt die ungleiche Verteilung der Marktmacht beim Zugang und der Nutzung von Daten dar. Diese Marktmacht könnte es großen Marktteilnehmern, beispielsweise großen Online-Plattformen, ermöglichen, die Regeln für die Datennutzung festzulegen und einen allgemeinen Zugang zu erschweren. Um Daten aus verschiedenen Quellen möglichst sektorenübergreifend kombinieren zu können, muss die Interoperabilität und Qualität der Daten gewährleistet sein. Schlussendlich sollen Defizite im Bereich der Datenkompetenzen, der relevanten Technologien und der Cybersicherheit behoben und die Position der Einzelperson bei der Ausübung ihrer Rechte gestärkt werden.

Um die beschriebenen Herausforderungen zu überwinden, basiert die Umsetzung der Europäischen Datenstrategie auf den folgenden vier Säulen:

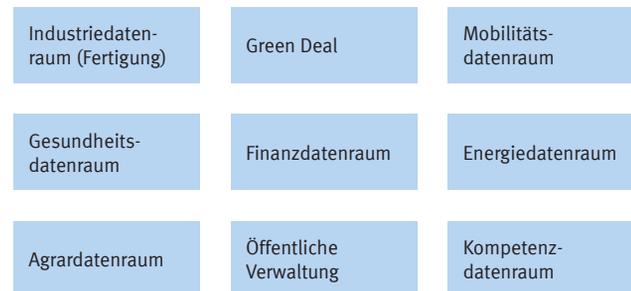
1. Schaffung gemeinsamer europäischer Datenräume
2. Schaffung eines Rechtsrahmens für Datenzugang und Datennutzung
3. Investitionen in die Stärkung der europäischen Kapazitäten und Infrastrukturen für das Hosting, die Verarbeitung und die Nutzung von Daten sowie die Sicherstellung der Interoperabilität der Daten
4. Investitionen in den Kompetenzaufbau und dadurch Stärkung der Handlungskompetenz der Einzelnen

Um einen Binnenmarkt für Daten zu etablieren, fördert die Europäische Kommission die Entwicklung gemeinsamer europäischer Datenräume in strategisch wichtigen Wirtschaftszweigen sowie in Bereichen von besonderem öffentlichen Interesse. [↪ Grafik 1](#)

Folgende Voraussetzungen gelten für die Schaffung der sektorspezifischen Datenpools: ein entsprechender Rechtsrahmen, eine geeignete Infrastruktur, die Interoperabilität der Daten sicherstellen sowie die sektorübergreifende Nutzung der Daten ermöglichen. Grund-

Grafik 1

Bereiche, in denen gemeinsame europäische Datenräume eingerichtet werden sollen



2021 - 0489

sätzlich behält sich die Kommission vor, künftig weitere gemeinsame europäische Datenräume für ergänzende Bereiche zu etablieren.

3

Rechtliche Verankerung

Drei neue Rechtsakte sind zur Umsetzung der Europäischen Datenstrategie vorgesehen, die eine Harmonisierung der Rechtslage in den Mitgliedstaaten schaffen sollen. Mit diesem Rechtsetzungspaket nimmt die Europäische Kommission zentrale rechtliche Fragen der sich entwickelnden Dateninfrastrukturen in Angriff, unter anderem die Etablierung von Zugangsrechten zu privat gehaltenen Daten (Europäische Kommission, 2017)¹.

3.1 Data Governance Act

Der Vorschlag für einen Data Governance Act hat zum Ziel, das Teilen von Daten zu erleichtern. Dazu soll ein Rechtsrahmen für die vorgesehenen gemeinsamen Datenräume geschaffen werden. Ferner sollen EU-weit Strukturen und Prozesse etabliert werden, damit nicht-persönliche Daten zwischen Behörden, Unternehmen sowie Bürgerinnen und Bürgern sicher und im Interesse des Gemeinwohls genutzt werden können. Die inhaltlichen Ziele umfassen die Erschließung des Werts von Daten öffentlicher Stellen, die Verringerung der Kosten der Datennutzung durch technische Interoperabilität und Standards sowie die Reduzierung der Transaktions-

¹ Siehe hierzu im Detail Wiebe (2017).

kosten beim Teilen von Daten (Europäische Kommission, 2020b). Der Entwurf des Data Governance Act setzt sich aus vier thematischen Blöcken zusammen:

System zur Weiterverwendung geschützter Daten öffentlicher Stellen²

Die Europäische Kommission strebt an, ein System zur Weiterverwendung von geschützten, das heißt nicht öffentlich zugänglichen Daten öffentlicher Stellen zu schaffen. Der Vorschlag soll die sogenannte PSI-Richtlinie³ (Open Data Richtlinie) für nicht öffentlich zugängliche Daten ergänzen. Der Data Governance Act sieht Mechanismen für eine Weiterverwendung von Daten für kommerzielle und nicht kommerzielle beziehungsweise unternehmerische oder wissenschaftliche Zwecke vor.

Dienste für die gemeinsame Datennutzung („data sharing services“/Datenintermediäre)

Der Legislativvorschlag enthält einen Regelungsrahmen für Dienste für die gemeinsame Datennutzung, das heißt für Datentreuhänder (Trusted Intermediaries) zwischen Datenbereitstellern und Datennutzern. Ziel ist, durch neutrale Akteure das Vertrauen zwischen Datenanbietern und Datennachfragern zu stärken. Datenintermediäre sollen die Datenräume für Unternehmen und Einzelpersonen organisieren, ohne eine geballte Marktmacht zu schaffen.

Datenaltruismus (Datenspende)

Daten als öffentliches Gut freiwillig zur Verfügung zu stellen soll einfacher gestaltet werden. Datenspendender sollen ihre Zustimmung über ein standardisiertes Einwilligungsdokument geben.

Schaffung eines European Data Innovation Board (Expertengruppe)

Das Gremium soll die Europäische Kommission hinsichtlich einer Standardisierung der Datenräume und hinsichtlich der Interoperabilität von Daten beraten. Dem

Board gehören Vertreterinnen und Vertreter der Europäischen Kommission an, ebenso nationale Fachleute aus den einzelnen Mitgliedstaaten und auf Empfehlung der Kommission Vertreterinnen und Vertreter verschiedener Bereiche oder Datenräume. Diese Bereiche sollen Kommission, Rat und Europäisches Parlament in weiteren Verhandlungen festlegen.

3.2 Durchführungsverordnung zu hochwertigen Datensätzen (High Value Data Sets)

Diese Durchführungsverordnung soll die Bereitstellung von hochwertigen Datensätzen in der amtlichen Statistik ermöglichen. Die PSI-Richtlinie sieht Durchführungsrechtsakte zur Bestimmung sogenannter hochwertiger Datensätze vor, die sechs Themenbereiche betreffen – darunter die amtliche Statistik. Ob Datensätze hochwertig sind, entscheidet sich anhand der PSI-Richtlinie primär danach, wie hoch deren potenzieller Mehrwert im Falle einer offen lizenzierten und technisch guten Bereitstellung ist (BMW, 2021). In einem Durchführungsrechtsakt der Kommission sind nun die hochwertigen Datensätze für den Bereich der amtlichen Statistik zu spezifizieren. Eine erste Festlegung der zu liefernden Datensets erfolgte durch die ESS-Direktorengruppen⁴, und zwar in den Bereichen Unternehmensstatistiken, Makroökonomische Statistiken und Sozialstatistiken.

3.3 Data Act

Ziel des Data Act ist es, Regelungen zum Teilen von Daten zwischen Unternehmen sowie zwischen Unternehmen und öffentlichen Stellen im öffentlichen Interesse festzulegen. Die Regelungen sollen einen fairen und verlässlichen Zugang zu neuen digitalen Datenquellen ermöglichen. Diese Rechtsetzungsinitiative vervollständigt den Vorschlag zum Data Governance Act durch eine ausdrückliche Regelung des Datenzugangs. Sie schließt

2 Siehe hierzu Rat der EU (2021).

3 Die Richtlinie (EU) 2019/1024 über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors bildet die horizontale Rechtsgrundlage für die Weiterverwendung von Daten des öffentlichen Sektors unter Einbeziehung der öffentlichen Unternehmen. Ziel der Richtlinie ist, im öffentlichen Sektor vorhandene Informationen der Öffentlichkeit möglichst unbürokratisch zugänglich zu machen.

4 Die ESS-Direktorengruppen bilden die Verbindungen zwischen der strategischen Ebene des AESS (Ausschuss für das Europäische Statistische System) und der operationalen Ebene (Arbeitsgruppen und Task Forces). Sie bereiten die Diskussionen und Entscheidungen im AESS vor und unterstützen die Kommission in der Vorbereitung von EU-Rechtsvorschriften. Insgesamt gibt es sechs Direktorengruppen in den Bereichen Volkswirtschaftliche Gesamtrechnung, Unternehmen, Soziales, Methodik/IT, Umwelt und Landwirtschaft.

personenbezogene und nicht personenbezogene Daten ein. Öffentliche und private Akteure sollen dabei von Entwicklungen wie Big Data und Machine Learning profitieren können. Es sollen Impulse zur Generierung von Daten gesetzt werden. Angestrebt wird ein ausgewogenes Verhältnis zwischen dem Recht auf Zugang zu Daten und Anreizen für Investitionen in Daten, ohne dabei die geltenden Regelungen des Datenschutzes zu ändern. Hierbei sollen die Empfehlungen einer von der Kommission eingesetzten Expertengruppe zum Teilen von Daten berücksichtigt werden (Europäische Kommission, 2020c). Diese Expertengruppe empfiehlt unter anderem die Schaffung eines EU-weiten Rechtsrahmens für den Zugang zu privat gehaltenen Daten. Der Rechtsrahmen sollte ein Teilen von Daten zwischen Unternehmen und öffentlichen Stellen im öffentlichen Interesse ermöglichen und erleichtern sowie zu einem konsistenten Ansatz der Mitgliedstaaten beim Teilen von Daten führen. Ein erster Textentwurf des Data Act wird voraussichtlich Anfang 2022 durch die Europäische Kommission verabschiedet.

4

Bedeutung der Europäischen Datenstrategie für die amtliche Statistik

4.1 Gemeinsame Ziele für eine digitale Zukunft

Die Europäische Datenstrategie definiert gleich zu Beginn ein zentrales Ziel: „Die Bürger sollten in die Lage versetzt werden, bessere Entscheidungen auf der Grundlage von Erkenntnissen zu treffen, die aus nicht-personenbezogenen Daten gewonnen werden. Diese Daten sollten allen zugänglich sein – ob öffentlich oder privat, ob groß oder klein, ob Start-up oder Gigant.“ (Europäische Kommission, 2020d, hier: Seite 1). Diese Zielsetzung entspricht der Kernaufgabe der amtlichen Statistik in Deutschland, Politik und Gesellschaft neutrale, objektive und fachlich unabhängige Statistiken zur Verfügung zu stellen. Dies ermöglicht, dass Bürgerinnen und Bürger fundierte Informationen erhalten und demokratische Entscheidungen auf Basis von hochwertigen amtlichen Daten getroffen werden können (Thiel/Meinke, 2017).

Angesichts des großen Bedarfs an aussagekräftigen, aktuellen und qualitativ hochwertigen Daten aus Wirtschaft und Gesellschaft ist diese Zielsetzung von zentraler Bedeutung, insbesondere während der derzeitigen COVID-19-Pandemie. Das Statistische Bundesamt hat bereits früh neue digitale Wege eingeschlagen, um die Aktualität der amtlichen Daten zu erhöhen und den wachsenden Datenbedarf von Wirtschaft und Gesellschaft zu decken. Die Europäische Datenstrategie unterstützt die amtliche Statistik bei ihrer Fortentwicklung zu einem digitalen Datenmanager und kundenorientierten Informationsdienstleister. Gleichzeitig steuert die amtliche Statistik als Datenproduzent wesentliche Kompetenzen und Leistungen bei, die benötigt werden, um die Zielsetzung der Europäischen Datenstrategie zu erreichen. Die folgenden Abschnitte beleuchten zentrale Aspekte des Zusammenwirkens von Europäischer Datenstrategie und amtlicher Statistik und zeigen, wie beide voneinander profitieren können.

4.2 Chancen für die amtliche Statistik – Beitrag der amtlichen Statistik

Nutzung privat gehaltener Daten durch Behörden

Die Nutzung neuer Datenquellen, auch privat gehaltener Daten, eröffnet eine Reihe von Weiterentwicklungsmöglichkeiten für die amtliche Statistik. Sie kann zur Entwicklung neuer statistischer Produkte und Dienstleistungen beitragen und bietet die Möglichkeit, zusätzliche Informationsbedarfe zu decken und Informationen am aktuellen Rand zu veröffentlichen. Gleichzeitig eröffnet sich hierdurch die Chance, eine Reduzierung der Belastungen von Auskunftgebenden zu erreichen. Die amtliche Statistik ist daher auf eine klare Regelung über den Zugang zu privat gehaltenen Daten angewiesen. Vor diesem Hintergrund wäre es im Sinne des ESS, mit dem Data Act einen allgemeinen Rahmen für den Datenaustausch zwischen Unternehmen und Behörden zu schaffen. Die amtliche Statistik ist ein öffentliches Gut und ein Grundpfeiler der demokratischen Gesellschaft. Der Data Act sollte aus Sicht des ESS daher klare Verpflichtungen zur gemeinsamen Nutzung privater Daten festlegen, die für Zwecke von hohem öffentlichen Interesse erforderlich sind. Zu diesem Zweck sollte der Rechtsakt einen Verweis auf die Entwicklung, Erstellung und Verbreitung amtlicher Statistiken enthalten.

Gemeinsame Datennutzung zwischen Behörden

Die Europäische Datenstrategie soll zudem die gemeinsame Datennutzung zwischen Behörden verbessern. Dabei wird insbesondere auf den Mehrwert für Politikgestaltung und öffentliche Dienstleistungen hingewiesen, aber auch auf Entlastungen für Unternehmen im Sinne des Once-only-Prinzips. Auch der Austausch von Verwaltungsdaten kann einen großen Beitrag zur Reduzierung der Berichtspflichten beitragen.

Hochwertige Datensätze

Die amtliche Statistik ist ein zentraler Anbieter hochwertiger Datensätze.¹⁵ Alle Daten der amtlichen Statistik erfüllen die Anforderungen der PSI-Richtlinie, indem sie zum Beispiel als Open Data über die Datenbank des Statistischen Bundesamtes zugänglich sind. Somit kann die amtliche Statistik als Anbieter hochwertiger Datensätze in erheblichem Maße zu den Zielen und der Umsetzung der Europäischen Datenstrategie beitragen. Auch kann die amtliche Statistik von hochwertigen Datensätzen anderer öffentlicher Stellen und öffentlicher Unternehmen profitieren, wenn sie diese für ihre Zwecke nutzen kann.

Gemeinsame Datenräume und Governancestrukturen

Die statistischen Ämter des ESS sind sowohl Nutzer als auch Anbieter von Daten in den geplanten gemeinsamen Datenräumen. Sie verfügen über einen gesetzlichen Auftrag zur Sammlung sowie Verarbeitung von Daten, über langjährige Erfahrung im Umgang mit großen Datenmengen zum Wohle der Allgemeinheit und genießen in dieser Rolle ein hohes Maß an Vertrauen. Diese Erfahrungen können die statistischen Ämter in die geplanten gemeinsamen Datenräume einbringen. Aufgrund ihrer fachlichen Kompetenz können sie auch zur Entwicklung gemeinsamer technischer Standards im Hinblick auf die Beschreibung, Qualität und Interoperabilität der Daten beitragen. Dies ist erforderlich, um die gemeinsame Nutzung von Daten über Sektoren und Bereiche hinweg zu ermöglichen. Dafür ist es notwendig, dass das

¹⁵ Maßgeblich ist der Verhaltenskodex für europäische Statistiken, eine Selbstverpflichtung der statistischen Ämter auf nationaler und europäischer Ebene, die deren Unabhängigkeit, Integrität und Rechenschaftspflicht festlegt sowie die Einhaltung der Qualitätskriterien Relevanz, Genauigkeit und Zuverlässigkeit, Aktualität und Pünktlichkeit, Kohärenz und Vergleichbarkeit, Zugänglichkeit und Klarheit (Eurostat, 2017).

ESS aufgrund seiner Expertise an den im Zuge des Data Governance Acts einzurichtenden Gremien beteiligt ist, insbesondere im vorgestellten European Data Innovation Board.

Dateninfrastruktur

Immer größere Datenmengen stellen die amtliche Statistik vor Herausforderungen. Die Nutzung von Clouds, die den strengen Datenschutzbestimmungen der amtlichen Statistik genügen und statistische Geheimhaltung gewährleisten, ist daher ein wichtiger Ansatzpunkt auf dem Weg zum digitalen Datenmanager. Die europäische Datenstrategie betont ebenfalls die Vorteile der Nutzung von Cloud-Diensten, insbesondere mit Blick auf die Schaffung von technischen Voraussetzungen, um Technologien wie Künstliche Intelligenz (KI) nutzen zu können. Für die Umsetzung sieht die Europäische Datenstrategie mehrere Maßnahmen vor: Vereinbarungen mit den Mitgliedstaaten über einen Cloud-Zusammenschluss, Einrichtung eines Marktplatzes für Cloud-Dienste für Nutzer aus dem privaten und dem öffentlichen Sektor sowie die Erarbeitung eines EU-Cloud-Regelwerks (Europäische Kommission, 2020d, hier: Seite 23).

5

Fazit

Die Schaffung eines europäischen Binnenmarkts für Daten soll Bürgerinnen und Bürgern, Unternehmen, Forschenden und öffentlichen Verwaltungen die Möglichkeit eröffnen, Entscheidungen auf einer fundierten Faktengrundlage treffen zu können. Die amtliche Statistik kann mit ihrer langjährigen Erfahrung im Umgang mit großen Datenmengen zum Wohle der Allgemeinheit entscheidend zur Umsetzung der Europäischen Datenstrategie beitragen, indem sie sich beispielsweise proaktiv in die Gestaltung der gemeinsamen Datenräume einbringt. Die Umsetzung der Europäischen Datenstrategie kann umgekehrt die amtliche Statistik dabei unterstützen, ihren Weg hin zu einem digitalen Datenmanager und Informationsdienstleister erfolgreich zu beschreiben. Insbesondere die geplante Schaffung gesetzlicher Regeln für den Datenzugang und die Weiterverwendung von Daten sind für die amtliche Statistik von großer Bedeutung. Von den vorgesehenen Investitionen in die Infrastrukturen für die Speicherung und Verarbeitung

von Daten sowie der Nutzung gemeinsamer Datenräume wird die amtliche Statistik in Deutschland und Europa ebenfalls profitieren können. Um den Bedarfen an aktuellen und hochwertigen Informationen gerecht zu werden, wird die Umsetzung der Europäischen Datenstrategie zentrale Bedingung für das Voranbringen der Modernisierung der amtlichen Statistik und wichtiger Innovationen durch das Statistische Bundesamt und das ESS sein. Das Statistische Bundesamt unterstützt daher die Bestrebung des ESS, die Ausgestaltung und Umsetzung der Europäischen Datenstrategie im Sinne der amtlichen Statistik mitzugestalten. 

LITERATURVERZEICHNIS

Bundesministerium für Wirtschaft und Energie (BMWi). *Hochwertige Datensätze in Deutschland*. Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie. 2021. [Zugriff am 1. November 2021]. Verfügbar unter: www.bmwi.de

Europäische Kommission. *Aufbau einer Europäischen Datenwirtschaft*. Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen. 2017. KOM(2017)9 final.

Europäische Kommission. *Lage der Union 2020. Rede zur Lage der Union 2020*. 2020a. [Zugriff am 1. November 2021]. Verfügbar unter: ec.europa.eu

Europäische Kommission. *Gemeinsame Datennutzung in der EU – gemeinsame europäische Datenräume (neue Regeln)*. 2020b. [Zugriff am 1. November 2021]. Verfügbar unter: ec.europa.eu

Europäische Kommission. *Towards a European strategy on business-to-government data sharing for the public interest*. Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing. 2020c. [Zugriff am 1. November 2021]. Verfügbar unter: ec.europa.eu

Europäische Kommission. *Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen. Eine europäische Datenstrategie*. 2020d. KOM(2020)66 final

Eurostat. *Verhaltenskodex für Europäische Statistiken. Für die nationalen statistischen Ämter und Eurostat (statistisches Amt der EU)*. 2017. [Zugriff am 1. November 2021]. Verfügbar unter: ec.europa.eu

Rat der EU. *EU will gemeinsame Datennutzung vereinfachen: Rat legt Standpunkt zum Daten-Governance-Gesetz fest*. Pressemitteilung vom 1. Oktober 2021. [Zugriff am 1. November 2021]. Verfügbar unter: www.consilium.europa.eu

Thiel, Georg/Meinke, Irina. *Gut aufgestellt für die Zukunft – ein Dank an Dieter Sarreither*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2017, Seite 9 ff.

Wiebe, Andreas. *Von Datenrechten zu Datenzugang – Ein rechtlicher Rahmen für die europäischen Datenwirtschaft. Überblick und erste Bewertung zur Mitteilung der EU-Kommission vom 10.1.2017*. In: Computer und Recht. Band 33. Ausgabe 2/2017, Seite 87 ff.

RECHTSGRUNDLAGEN

Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen. Aufbau einer Europäischen Datenwirtschaft. KOM(2017)9 final

Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen. Eine europäische Datenstrategie. KOM(2020)66 final

Richtlinie (EU) 2019/1024 des Europäischen Parlaments und des Rates vom 20. Juni 2019 über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors (Amtsblatt der EU Nr. L 172, Seite 56).

ÖKOSYSTEMGESAMTRECHNUNGEN – FLÄCHENBILANZIERUNG DER ÖKOSYSTEME (EXTENT ACCOUNT)

Marius Bellinghen, Simon Felgendreher, Johannes Oehrlein,
Simon Schürz, Stephan Arnold

↳ **Schlüsselwörter:** Ökosystemgesamtrechnungen – Flächenbilanzierung –
Umweltökonomische Gesamtrechnungen – Ökosystemleistungen – SEEA EA

ZUSAMMENFASSUNG

Im Statistischen Bundesamt begannen im August des Jahres 2020 die Arbeiten zur Berechnung der Ökosystemgesamtrechnungen. Deren Ziel ist eine bundesweit räumlich und zeitlich konsistente Datengrundlage zu Ausmaß, Zustand und Leistungen der Ökosysteme für den Menschen, um politische Entscheidungsfindungen effektiv unterstützen zu können. Die Flächenbilanzierung, welche die Ökosysteme Deutschlands erfasst, klassifiziert und in Kartenform darstellt, bildet den Ausgangspunkt des Berichtssystems. Der Vergleich zwischen den Flächenbilanzen verschiedener Erhebungsjahre bietet zudem die Möglichkeit, Veränderungen zwischen Ökosystemen auszuweisen.

↳ **Keywords:** *ecosystem accounting – extent account – environmental economic accounts – ecosystem services – SEEA-EA*

ABSTRACT

At the Federal Statistical Office, work on the compilation of ecosystem accounts began in August 2020. The aim of this work is to create a nationwide, spatially and temporally consistent data basis on the extent and condition of the ecosystems and the services they provide to humans, in order to effectively support political decision-making. The extent account, in which Germany's ecosystems are recorded, classified and mapped, constitutes the basis of the reporting system. The comparison between the extent accounts of different survey years also offers the possibility to identify changes between ecosystems.

Marius Bellinghen

studierte Geographie und Management natürlicher Ressourcen. Er ist als wissenschaftlicher Mitarbeiter im Referat „Monetäre Umweltökonomische Gesamtrechnungen, Materialflussrechnungen, Nachhaltigkeitsindikatoren“ des Statistischen Bundesamtes für die ökologische Perspektive der Ökosystemgesamtrechnungen in Deutschland tätig.

Dr. Simon Felgendreher

ist Volkswirt und im selben Referat als Referent für die Koordinierung der Ökosystemgesamtrechnungen in Deutschland zuständig.

Dr. Johannes Oehrlein

hat Mathematik studiert und im Fachbereich Geoinformatik promoviert. Als wissenschaftlicher Mitarbeiter im oben genannten Referat betreut er die technische Konzeption und Umsetzung der Ökosystemgesamtrechnungen in Deutschland.

Dr. Simon Schürz

ist Volkswirt mit universitären Stationen in Wien, Bologna und Göteborg. Im selben Referat wie die Vorgenannten ist er wissenschaftlicher Mitarbeiter und mit der monetären Bewertung von Ökosystemleistungen innerhalb der Ökosystemgesamtrechnungen für Deutschland befasst.

Stephan Arnold

ist Diplom-Geograph und im Referat „Pflanzliche Erzeugung und Flächennutzung“ des Statistischen Bundesamtes für die Flächenstatistik tätig. Derzeit befasst er sich mit deren Neukonzeptionierung hinsichtlich der Auswertung objektstrukturierter Geobasisdaten zu den Themen Landbedeckung und Landnutzung.

1

Einleitung

Im März 2021 hat die Statistische Kommission der Vereinten Nationen beschlossen, die Kapitel eins bis sieben des System of Environmental Economic Accounting – Ecosystem Accounting (SEEA EA) als statistischen Standard zu bestätigen. Damit steht nun ein anerkanntes Rahmenwerk zum Aufbau der Ökosystemgesamtrechnungen zur Verfügung und die Arbeiten zur Umsetzung werden im Statistischen Bundesamt wie auch international intensiviert. Die Ökosystemgesamtrechnungen bilden dabei einen weiteren Teil der bereits etablierten Umweltökonomischen Gesamtrechnungen (UGR). Die UGR erfassen die Wechselwirkungen zwischen Umwelt und Ökonomie, beispielsweise die Umweltbelastungen durch die Wirtschaft, den Ertrag von Naturressourcen und den Wert von Umweltschutzmaßnahmen. Ziel der Ökosystemgesamtrechnungen ist, das Ausmaß, den Zustand und die Leistungen von Ökosystemen für Gesellschaft und Wirtschaft im Ganzen zu betrachten. Dabei sollen jene Funktionen der Ökosysteme, die über eine Entnahme und Abgabe von Materialflüssen an die Umwelt hinausgehen, explizit räumlich erfasst werden. Beispiele hierfür sind die Luftreinhaltung, der Hochwasserschutz oder auch der Erholungswert der Natur. Die Ökosystemgesamtrechnungen ermöglichen im Zeitverlauf eine Analyse der Interaktionen zwischen Mensch und Natur, sodass Rückkopplungseffekte in der Natur, ausgelöst durch anthropogene Verhaltensweisen, sichtbar werden.

Für die Messung der Ökosystemleistungen ist zunächst eine flächendeckende Erfassung der Ökosysteme erforderlich. Hierbei wird die Fläche Deutschlands einschließlich der deutschen Ausschließlichen Wirtschaftszone (AWZ)¹ in Nord- und Ostsee lückenlos und überschneidungsfrei verschiedenen Ökosystemtypen zugeordnet. Die so entstandene Flächenbilanz (Extent Account) ermöglicht es, das Ausmaß der verschiedenen Ökosysteme sowie deren Veränderungen im Zeitverlauf zu erfassen.

Neben der Flächenbilanz ist auch die Bestimmung des Zustands der Ökosysteme eine Voraussetzung für die

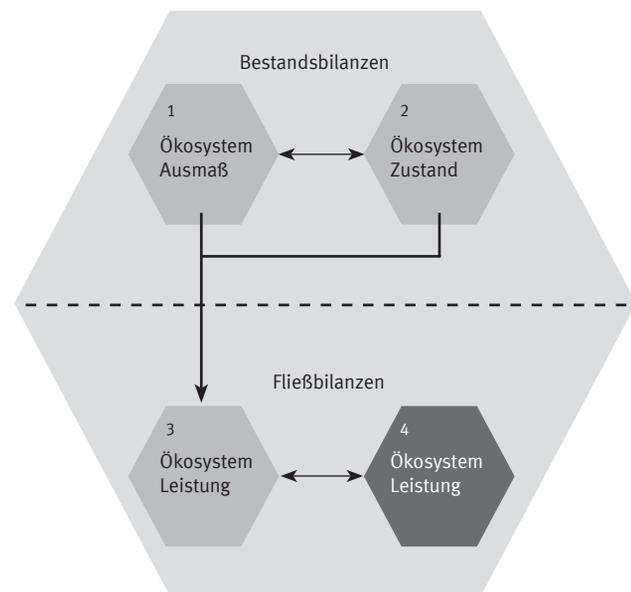
1 Die Ausschließliche Wirtschaftszone (AWZ) ist das Meeresgebiet seewärts des Küstenmeeres (12-Seemeilen-Zone) bis maximal zur 200-Seemeilen-Grenze. Die AWZ gehört nicht zum Hoheitsgebiet des angrenzenden Küstenstaates, jedoch hat dieser exklusive Nutzungsrechte (BMI, 2021).

Ableitung von Ökosystemleistungen. Innerhalb der Zustandsbilanzierung (Condition Accounting) der Ökosysteme werden deshalb Variablen, die die biotischen, abiotischen und landschaftlichen Charakteristiken der jeweiligen Ökosysteme bestmöglich widerspiegeln, erfasst und deren zeitliche Veränderungen festgehalten. Somit sind Rückschlüsse auf die Leistungsfähigkeit und Stabilität von Ökosystemen möglich.

In der Ökosystemleistungsbilanz (Service Account) werden dann Leistungen wie die Kohlenstoffspeicherung oder der Überflutungsschutz auf Basis der Flächen- und Zustandsbilanz in physischen Einheiten gemessen oder modelliert. Durch die räumlich hochauflösende Verortung der Ökosystemflächen, werden die Ökosystemleistungen ebenfalls flächengenau berechnet. Somit kann beispielsweise die lokale Kühlungsleistung von Grünflächen bei Hitzewellen in dichtbesiedelten Gebieten von jener in ländlichen Gebieten unterschieden werden. Die monetäre Inwertsetzung bestehender Ökosystemleis-

Grafik 1

Schematischer Aufbau der Ökosystemgesamtrechnungen



Accounts:
 1 Ecosystem Extent
 2 Ecosystem Condition
 3 } Ecosystem Services
 4 }

● Physische Bilanzen
 ● Monetäre Bilanzen

Eigene Abbildung nach dem Rahmenwerk System of Environmental Economic Accounting – Ecosystem Accounting (SEEA EA) der Vereinten Nationen.

2021 - 0502

tungen vervollständigt die Ökosystemgesamtrechnungen. Deren Aufbau mit ihren verschiedenen Bilanzen verdeutlicht [↘ Grafik 1](#).

Ergebnis der Ökosystemgesamtrechnungen wird eine Datengrundlage sein, die es politischen Entscheidungsträgern ermöglicht, Maßnahmen zum Erhalt oder zur Steigerung von Ökosystemleistungen effektiv und zielgenau zu gestalten sowie die Bedeutung der Ökosysteme bei politischen und wirtschaftlichen Entscheidungen adäquat zu berücksichtigen. Durch die international standardisierten Methoden sind die Bilanzen länderübergreifend vergleichbar.

Erste Ergebnisse zur Flächenbilanz der Ökosysteme wurden Mitte Oktober 2021 für die Referenzjahre 2015 und 2018 veröffentlicht. Das Vorgehen und die Methodik dieser Bilanz werden in den weiteren Kapiteln dieses Aufsatzes ausführlich erläutert: Kapitel 2 beschreibt die nationale Klassifikation der Ökosysteme. Kapitel 3 befasst sich mit den damit verbundenen Anforderungen an die Datenquellen. Die technische Umsetzung der Flächenbilanzierung im Einzelnen ist in Kapitel 4 dargestellt, deren Ergebnisse werden in Kapitel 5 beschrieben. Nach einem Fazit zur Flächenbilanzierung gibt Kapitel 6 einen Ausblick auf die derzeit laufenden Arbeiten zur Erstellung der Zustandsbilanz beim Statistischen Bundesamt und die im darauffolgenden Arbeitsschritt beginnende Berechnung der Ökosystemleistungen.

2

Nationale Klassifikation der Ökosysteme

Die neu erstellte nationale Ökosystemklassifikation ist eine hierarchisch aufgebaute Typologie, die es erlaubt, alle Flächen Deutschlands eindeutig einer bestimmten Ökosystemklasse zuzuweisen. Sie ist sowohl strukturell als auch räumlich vollständig, deckt also alle relevanten ökologischen Bereiche und das gesamte Staatsgebiet einschließlich der Ausschließlichen Wirtschaftszone (AWZ) in Nord- und Ostsee ab. Durch die Orientierung an Landbedeckung und -nutzung sowie ökologischer und struktureller Variablen ist die Klassifikation leicht verständlich und ermöglicht eine effektive Kommunikation und Visualisierung der Ökosystemdaten.

Alle Ökosysteme, die

- › im nationalen Kontext Deutschlands derzeit existieren oder in naher Zukunft existieren könnten,
- › durch ausreichend detaillierte und bundesweit flächendeckende Datengrundlagen erfassbar sind und
- › im Hinblick auf erbrachte Ökosystemleistungen relevant sind,

werden durch eigene Ökosystemklassen repräsentiert. Diese stellen die kleinste Untergliederung der Ökosystemklassifikation dar und können für die nationale und internationale Berichterstattung zu Ökosystemgruppen, -abteilungen und -abschnitten aggregiert werden. Die beiden Nationalen Ökosystemabschnitte (National Ecosystem Sections) „Terrestrischer Bereich“ und „Gewässer“ umfassen sechs Nationale Ökosystemabteilungen (National Ecosystem Divisions), welche sich in insgesamt 21 Nationale Ökosystemgruppen (National Ecosystem Groups) und schließlich in 74 Nationale Ökosystemklassen (National Ecosystem Classes) untergliedern.

Die Klassifikation definiert für jede Ökosystemklasse eindeutige Kriterien, nach denen eine Fläche einer bestimmten Klasse zugeteilt wird. Diese Kriterien stützen sich auf viele verschiedene Variablen, wie Landbedeckung, Höhe über Normalhöhennull oder Baudichte. Eine Fläche wird zum Beispiel der Klasse „Almen und Bergwiesen“ zugeteilt, wenn sie die Landbedeckung „Inhomogenes Grünland“ und die Landnutzung „Extensive Bewirtschaftung“ aufweist und über 900 Meter Höhe über Normalhöhennull im Alpenraum beziehungsweise über 600 Meter Höhe über Normalhöhennull in den Mittelgebirgen liegt.

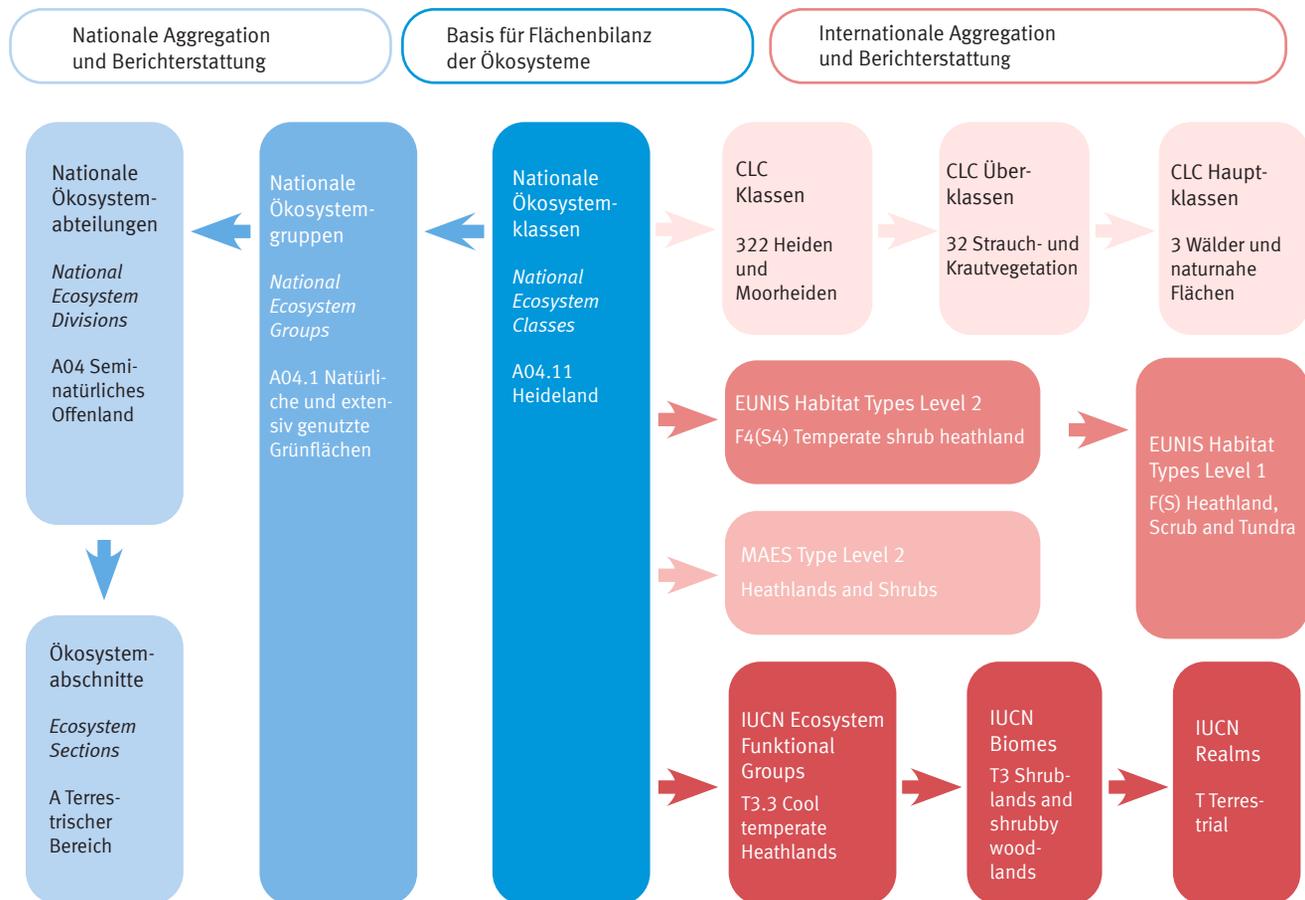
Um aufkommende internationale Berichtspflichten, beispielsweise seitens der Europäischen Union (EU), erfüllen zu können, gewährleisten Überführungstabellen, sogenannte Crosswalks, die Anbindung an international anerkannte Klassifikationen. So kann zum Beispiel die Nationale Ökosystemklasse A04.11 „Heideland“ dem IUCN Typ T3.3 „Cool temperate heathlands“, dem MAES Typ „Heathlands and shrubs“ sowie der CLC Klasse 322 „Heiden und Moorheiden“ zugeordnet werden.¹²

[↘ Grafik 2](#)

2 IUCN = International Union for Conservation of Nature; MAES = Mapping and Assessment of Ecosystems and their Services; CLC = Corine Land Cover.

Grafik 2

Überführung der Ökosystemklassen in international anerkannte Klassifikationen am Beispiel der Ökosystemklasse A04.11 Heideland



CLC = Corine Land Cover; EUNIS = European Nature Information System; MAES = Mapping and Assessment of Ecosystems and their Services; IUCN = International Union for Conservation of Nature

2021 - 0503

3

Räumliche und zeitliche Konsistenz – eine Herausforderung an die Daten- quellen

Die primäre Herausforderung zur vollständigen, lücken- losen und hochaufgelösten Flächenbilanzierung der Ökosysteme liegt darin, valide Daten in hinreichendem Umfang zusammenzutragen, um die in der Klassifikation festgelegte Kriterienmatrix anwenden zu können. Um alle Flächen mit den nötigen Informationen auszustat- ten, werden für die Flächenbilanzierung der Ökosysteme eine Vielzahl von räumlichen Datenquellen kombiniert.

Die Basis hierzu bildet das vom Bundesamt für Kartogra- phie und Geodäsie (BKG) erstellte Landbedeckungs- modell für Deutschland (BKG, 2020).

Das Landbedeckungsmodell für Deutschland (LBM-DE) hat eine Mindestkartierfläche³ von einem Hektar und wird derzeit in einem dreijährlichen Turnus erstellt. Neben den Hauptvariablen Landbedeckung und Land- nutzung weist es weitere Merkmale wie Vegetations- anteil und Versiegelungsgrad aus. Datengrundlagen des LBM-DE sind neben Daten aus dem Amtlichen Topographisch-Kartographischen Informationssystem

3 Dies reflektiert den Mindestanspruch der Flächenbilanzierung, alle Ökosystemflächen über einer Größe von einem Hektar zu erfassen und zu klassifizieren. Im Gegenzug bedeutet dies, dass auch kleinere Flächen abgebildet werden können, sofern es die Datenlage zulässt.

(ATKIS) der Bundesländer auch Fernerkundungsdaten. Die Begriffe Landbedeckung und Landnutzung sind eng miteinander verwoben und bedingen sich teilweise gegenseitig. Jedoch muss bei der Konzipierung eines Klassifikationssystems darauf geachtet werden, diese Aspekte getrennt voneinander und gleichermaßen umfassend zu betrachten. Neben der Art der eigentlichen Landbedeckung und der Landnutzung lassen sich noch weitere Eigenschaften beobachten und erfassen, die eine noch detailliertere Beschreibung der Flächen ermöglichen. Eine dreigliedrige semantische Betrachtung der Landbedeckung und Landnutzung sowie darüber hinausgehender Eigenschaften ist für die Klassifikation der Ökosysteme notwendig.

Zu diesem Zweck werden zusätzlich zum LBM-DE eine Vielzahl von ökologischen und strukturellen Informationen benötigt, um alle relevanten Ökosysteme, zum Beispiel „Streuobstwiesen“ oder „Auen-“ und „Moorwäldern“, ausweisen zu können. Hierzu werden neben dem Digitalen Geländemodell und dem Digitalen Basis-Landschaftsmodell (Basis-DLM) auch Datensätze wie die Bodenübersichtskarte (BÜK200), Daten der Europä-

ischen Wasserrahmenrichtlinie, Daten des Copernicus-Satellitendatenprogramms sowie Biototypkartierungen der Bundesländer genutzt und räumlich verknüpft. Wichtige Kriterien zur Auswahl der verwendeten Quellen sind eine angemessene räumliche und zeitliche Auflösung sowie eine hohe Qualität der Daten. Daher wurde, sofern möglich, auf amtliche Daten zurückgegriffen. Des Weiteren fanden qualitativ hochwertige Daten wie die des [europäischen Copernicus-Programms](#) Verwendung. Dieses Programm verknüpft Daten verschiedener europäischer Institutionen, unter anderem die der europäischen Weltraumorganisation (ESA). Ein Beispiel für Daten aus dem Copernicus-Programm stellt der [Datensatz der Uferzonen](#) (Riparian Zones) dar, der das Ergebnis einer Berechnung aus verschiedenen hydrologischen und geomorphologischen Parametern ist. Zur Sicherstellung einer hohen Qualität wurde hier im Rahmen des Copernicus-Programms ein qualitativer Expertenbewertungsansatz gewählt.

➤ **Übersicht 1** zeigt alle zusätzlich zum LBM-DE genutzten Datensätze (Bellingen und andere, 2021). Die Kombination aus einer Vielzahl von heterogenen Datensät-

Übersicht 1

Zusätzliche Datenquellen zur Flächenbilanzierung der Ökosysteme

Abgeleitete Informationen	Datenquellen
Alpine Großlandschaft und Mittelgebirge	Landschaften Deutschlands (BfN)
Höhe über Normalhöhennull in Metern	Digitales Geländemodell (BKG)
Auenwälder/Uferzonen	Copernicus Riparian Zones High Resolution Layer
Moorwald	Basis-DLM (BKG)
Moortypen	Bodenübersichtskarte (BGR)
Riffe	Biotopkartierung (BfN)
Sandbänke	Biotopkartierung (BfN), Daten der Landesämter
Seegras- und Algenbestände (Grünalgen, Seegras, Muscheln)	Nationalparkverwaltung Niedersächsisches Wattenmeer, Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (Niedersachsen), Landesbetrieb für Küstenschutz, Nationalpark und Meeresschutz (Schleswig-Holstein), Schubert und andere (2015)
Erheblich veränderte und künstliche Binnengewässer	Wasserrahmenrichtlinie (BfG)
Küstennahe marine Gewässer	Wasserrahmenrichtlinie (BfG)
Meeresfläche	Ausschließliche Wirtschaftszone (BSH) + Wasserrahmenrichtlinie (BfG)
Fließgewässer	Basis-DLM (BKG) + Wasserrahmenrichtlinie (BfG)
Feldhecken	Basis-DLM (BKG) + Copernicus Small Woody Features High Resolution Layer
Straßen	Basis-DLM (BKG)
Schienen	Basis-DLM (BKG)
Küstendünen	Biototypkartierungen der Länder
Baudichte	LoD1-DE (BKG)
Streuobstwiesen	Biototypkartierungen der Länder

BfN = Bundesamt für Naturschutz, BKG = Bundesamt für Kartographie und Geodäsie, BGR = Bundesanstalt für Geowissenschaften und Rohstoffe, BfG = Bundesamt für Gewässerkunde, BSH = Bundesamt für Seeschifffahrt und Hydrographie

zen ermöglicht einerseits die angestrebte Detailtiefe der Flächenbilanzierung. Sie hat aber andererseits zur Folge, dass potenzielle Fehler in den Eingangsdaten nicht immer erkannt und korrigiert werden können.

4

Technische Umsetzung der Flächenbilanzierung

Die technische Umsetzung der Flächenbilanzierung der Ökosysteme erfolgt in den folgenden drei Schritten: Vorverarbeitung, Klassifizierung der Flächen und Nachverarbeitung. Nach einer einmaligen manuellen Harmonisierung der Eingangsdatensätze werden alle Schritte in Geographischen Informationssystemen (GIS) mit Werkzeugen zur Geodatenverarbeitung automatisch über ein Python-Skript abgewickelt. In der Vorverarbeitung wird das LBM-DE mit weiteren räumlichen Datenquellen kombiniert und erweitert. Dazu werden drei verschiedene Arten von Geodatenverarbeitung durchgeführt.

Erstens werden flächenhafte Geometrien von Ökosystemen, die bei der Erstellung des LBM-DE keine hinreichende Rolle spielen, aus anderen Datenquellen abgeleitet. Dies ist unter anderem für Hecken, Straßen und Flüsse notwendig, da diese Daten nur als linienhafte Objekte vorliegen. Sie werden anhand von zusätzlichen Breiteninformationen gepuffert und finden so als Flächen Eingang in die Flächenbilanzierung. Für die Meeresflächen des Staatsgebiets sowie die der Ausschließlichen Wirtschaftszone wird eine gänzlich neue Flächenstruktur aufgebaut, die sich an der Landbedeckung und den Messflächen der Wasserrahmenrichtlinie orientiert.

Zweitens umschließt die Vorverarbeitung das Zerschneiden von Flächen des LBM-DE zur weiteren Ausdifferenzierung der Ökosysteme. So wird beispielsweise die Ökosystemklasse Auenwald durch die Verschneidung von Flächen mit Landbedeckung Laubwald und dem Copernicus-Datensatz der Uferzonen (Riparian Zones) ermittelt.

Drittens werden bestehende Geometrien aus dem LBM-DE mit Daten angereichert, ohne ihre Geometrie selbst zu verändern. Hierbei sind nur Prozesse zugelassen, die sich nicht auf die finale Fläche der Geometrie

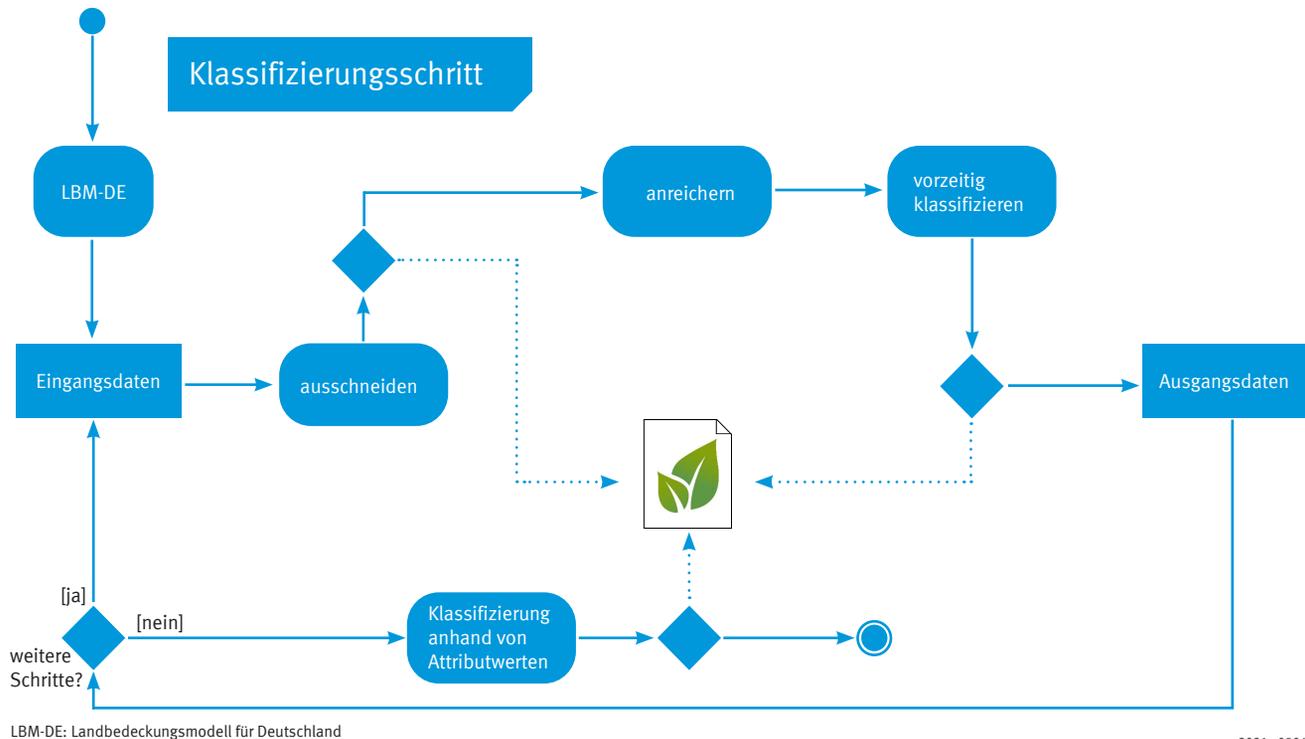
stützen. So ist die räumliche Verbindung von Standgewässern aus dem LBM-DE mit der Gewässermorphologie der Seen aus der Wasserrahmenrichtlinie ein zulässiger Schritt, die Berechnung der mittleren Höhe über Normalhöhennull einer Geometrie jedoch nicht.

Nachdem alle Vorverarbeitungsschritte erfolgt sind, beginnt die eigentliche Klassifizierung und damit die Flächenbilanzierung der Ökosysteme. Zur Verarbeitung des umfangreichen Datenvolumens wird jedes Bundesland sowie die Nord- und Ostsee einschließlich der AWZ separat klassifiziert. Die Klassifizierung erfolgt durch eine Abfolge von Klassifizierungsschritten für jeweils eine oder mehrere Ökosystemklassen. Ausgehend von der initialen Arbeitsdatei, dem LBM-DE, werden in jedem Schritt bestimmte Flächen klassifiziert und aus der Arbeitsdatei ausgeschnitten. Die verbleibenden Flächen sind Datengrundlage für den nächsten Klassifizierungsschritt. [↘ Grafik 3](#) stellt die Struktur des Algorithmus zur Klassifizierung schematisch dar. Jeder Klassifizierungsschritt hat demnach folgenden Ablauf:

- › Zunächst können vollständig vorverarbeitete Ökosysteme aus der aktuellen Arbeitsdatei ausgeschnitten und der Ergebnisdatei zugespielt werden. Dabei ist auf die Einschränkung zu achten, dass nur Flächen, die in der Arbeitsdatei enthalten sind, Eingang in die Ergebnisdatei finden. So wird eine erneute Klassifizierung bereits endgültig klassifizierter Flächen verhindert.
- › Gibt es Verarbeitungsschritte, die für die Vorverarbeitung der Daten unzulässig sind, werden diese nun zur Verknüpfung von Informationen mit den aktualisierten Geometrien durchgeführt. Beispielsweise wird durch das Zuspieldes Digitalen Geländemodells durch zonale Statistik die mittlere Höhe über Normalhöhennull für jede Ökosystemfläche berechnet. Diese Information ermöglicht unter anderem, montane und subalpine Nadelwälder von sonstigen Nadelwäldern zu unterscheiden.
- › Ausgewählte Flächen werden anhand der Klassifizierungsmatrix einer Klasse zugeteilt. Anschließend wird die Arbeitsdatei aktualisiert und alle in diesem Schritt klassifizierten Flächen werden aus der Arbeitsdatei entfernt, um eine spätere erneute Klassifikation zu verhindern.

Grafik 3

Ablauf des iterativen Klassifizierungsprozesses



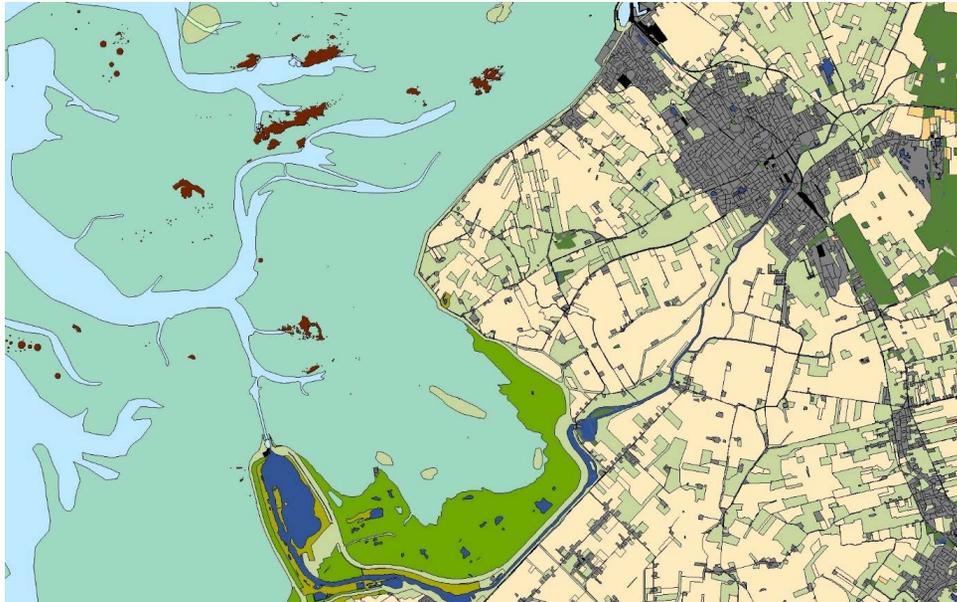
Abschließend bereinigt die Nachverarbeitung den errechneten Datensatz. Hierbei werden die Vollständigkeit und Eindeutigkeit der Klassifizierung kontrolliert und Geometrien repariert. Jede Ökosystemfläche und jedes zugeschpielte Attribut kann einem Vorprozessierungs- und einem Klassifikationsschritt zugeordnet werden. Diese Protokollierung ermöglicht darüber hinaus eine Rekonstruktion, welche Datengrundlage für welche Klassifizierung verantwortlich ist. Die Qualitätskontrolle der klassifizierten Ökosystemflächen erfolgt unter anderem durch stichprobenartige Kontrollen mit digitalen Orthofotos. Zuletzt werden die hochaufgelösten räumlichen Daten zur Erstellung der endgültigen Bilanzen und Karten auf verschiedene Verwaltungs- und Klassifizierungsebenen aggregiert.

↳ Grafik 4 zeigt am Beispiel der niedersächsischen Küste die Detailtiefe des Datensatzes, der der Flächenbilanzierung der Ökosysteme Deutschlands zugrunde liegt. Jede einzelne hier sichtbare Flächengeometrie ist einer Ökosystemklasse zugeteilt, enthält aber auch alle weiteren Informationen, die in diesen Prozess eingeflossen sind. Die Genauigkeit der Flächenbilanzierung

ermöglicht es, Daten und Werte aus den nachfolgenden Bilanzierungen flächenscharf zuzuweisen. Dies eröffnet Entscheidungsträgern in Politik und Gesellschaft in Zukunft die Möglichkeit, den Zustand und die Leistungsfähigkeit der einzelnen Ökosystemflächen effektiv zu berücksichtigen.

Grafik 4

Detailtiefe des errechneten Datensatzes zur Flächenbilanzierung der Ökosysteme



Die einzelnen Datenquellen zur Flächenbilanzierung der Ökosysteme sind im Aufsatz in Übersicht 1 aufgeführt.

2021 - 0505

5

Ergebnisse der Flächenbilanzierung

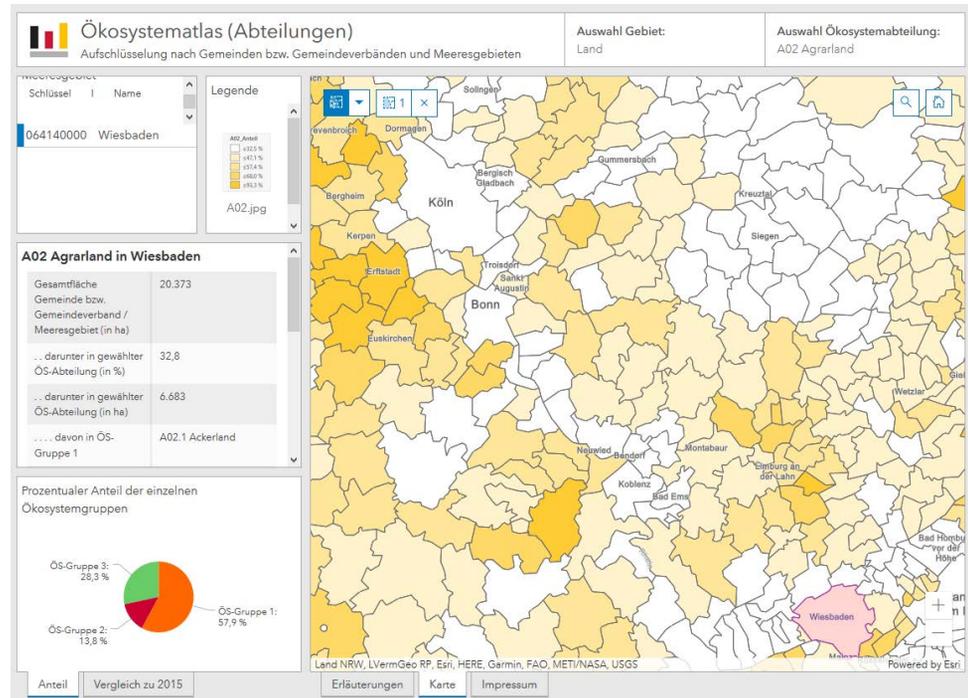
Der neue [Ökosystematlas](#) visualisiert die Resultate der Flächenbilanzierung in interaktiven Online-Karten. Nutzerinnen und Nutzer können dort auf Ebene der Gemeinden beziehungsweise Gemeindeverbände⁴ (im weiteren Text: Gemeindeverband) die Verteilung der Ökosysteme deutschlandweit auf Ökosystemabteilungs- und -gruppenebene erkunden. Über das interaktive Dashboard können für jeden Gemeindeverband auch Flächendaten und -grafiken auf der detailreicheren nächsten Hierarchieebene, das heißt für Ökosystemgruppen und -klassen, aufgerufen werden. [Grafik 5](#) zeigt einen Screenshot des digitalen Ökosystematlas für das Beispiel „Agrarland“. Gemeindeverbände sind gemäß dem Anteil des Agrarlands an der Gesamtfläche eingefärbt. Bei Auswahl eines bestimmten Gemeindeverbands zeigt das Dashboard, wie sich dieses Agrarland in der entsprechenden Verwaltungseinheit auf „Ackerland“, „Dauerkulturen“, „Grünland“ und „Feldhecken“ verteilt.

⁴ Für Hinweise zur Gliederung der Gemeinden beziehungsweise der Gemeindeverbände siehe <https://sg.geodatenzentrum.de>; der Ausweis erfolgt auf Basis der Verwaltungsgliederung Stand 2018.

Die gesammelten Ergebnisse der Flächenbilanzierung werden zusätzlich auch in Bilanzen, dem eigentlichen Kern des Berichtssystems, veröffentlicht. Dies geschieht auf verschiedenen administrativen Aggregationsebenen, also separat für Deutschland insgesamt, die Bundesländer- und Gemeindeverbandsebene, für Nord- und Ostsee sowie für die AWZ. [Tabelle 1](#) zeigt die bundesweite Flächenbilanz der Ökosysteme 2015 bis 2018 auf Ökosystemabteilungs- und -gruppenebene einschließlich der Ausweisung von Veränderungen und stabilem Bestand. Die Ökosystemabteilungen „Agrarland“ (50%) und „Wälder und Gehölz“ (30%) dominieren das terrestrische Landschaftsbild Deutschlands. Beim Agrarland liegen die flächenhaften Abnahmen prozentual für den Zeitraum von 2015 bis einschließlich 2018 bei 0,2% des Anfangsbestands und stellen damit abteilungsübergreifend die größte Abnahme dar. Der größte Flächenzuwachs ist mit 0,9% für den Zeitraum von 2015 bis einschließlich 2018 in der Abteilung „Siedlungsflächen und Verkehrsinfrastruktur“ wiederzufinden. Einen Flächenzuwachs von jeweils 0,2% haben die Abteilungen „Semi-natürliches Offenland“ und „Binnengewässer“ erfahren. In den Abteilungen „Wälder und Gehölz“ und „Meeresgewässer“ finden sich keine nennenswerten Flächenveränderungen hin zu anderen Abteilungen.

Grafik 5

Screenshot des digitalen Ökosystematlas am Beispiel der Abteilung A02 Agrarland



Quellen: Statistisches Bundesamt, GDI-TH, Esri, HERE, Garmin, FAO, NOAA, USGS

2021 - 0506

Tabelle 1

Flächenbilanz der Ökosysteme und deren Veränderung zwischen 2015 und 2018 auf Abteilungsebene

	Ökosystem- abschnitt A Terrestrischer Bereich	Ökosystemabteilungen				Ökosystem- abschnitt B Gewässer	Ökosystemabteilungen	
		A01 Siedlungs- flächen und Verkehrs- infrastruktur	A02 Agrarland	A03 Wälder und Gehölz	A04 Semi- natürliches Offenland		B01 Binnen- gewässer	B02 Meeres- gewässer
		ha						
Anfangsbestand (2015)	35 144 810	4 005 247	19 037 976	11 578 218	523 369	2 994 762	609 277	2 385 485
Zunahme	1 783	44 999	9 412	4 987	4 329	2 460	2 434	45
Abnahme	2 460	8 606	41 866	10 560	3 372	1 783	1 475	328
Zunahme + Abnahme (= Turnover)	4 243	53 605	51 279	15 547	7 700	4 243	3 910	373
Stabiler Bestand								
Anfangsbestand – Abnahme	35 142 350	3 996 641	18 996 109	11 567 658	519 998	2 992 979	607 802	2 385 157
in % des Anfangsbestands	100,0	99,8	99,8	99,9	99,4	99,9	99,8	100,0
Nettoveränderung ¹								
Zunahme – Abnahme	- 677	36 393	- 32 454	- 5 573	957	677	959	- 282
in % des Anfangsbestands	0,0	0,9	- 0,2	0,0	0,2	0,0	0,2	0,0
< Mindestgröße	- 3 667	4 756	- 4 006	- 3 778	- 639	3 671	3 667	4
Endbestand (2018)	35 140 466	4 046 396	19 001 515	11 568 867	523 688	2 999 109	613 903	2 385 207

¹ Ziel der Ausweisung der Nettoveränderung ist die Erfassung von realen Veränderungen. Deshalb werden Veränderungen unter einer Mindestgröße (bezogen auf Fläche und Kompaktheit) in der Zeile „< Mindestgröße“ gesammelt ausgewiesen und als technische Ungenauigkeit der Geodatenanalyse interpretiert. Flächen beziehen sich auf das deutsche Staatsgebiet (ohne deutsche Ausschließliche Wirtschaftszone in Nord- und Ostsee).

Ein Blick darauf, welche Flächen von einer Ökosystemklasse in eine andere übergehen und in welchem Ausmaß dies geschieht, lohnt sich. Je länger die Zeitreihe der Flächenbilanzierung, desto eher lassen sich für diese Transformationen Trends abschätzen. [↪ Grafik 6](#) verdeutlicht die Veränderungsströme zwischen Ökosystemgruppen in den Jahren 2015 und 2018. So gab es beispielsweise großflächige Veränderungen von den Gruppen „Ackerland“ und „Wiesen und Weiden“ hin zu „Siedlungsflächen mit niedriger Baudichte“ sowie „Abbauflächen, Deponien und Baustellen“. Demnach fand der Ausbau von „Siedlungsflächen mit niedriger Baudichte“ vornehmlich auf Flächen der Ökosystemgruppen „Ackerland“ und „Wiesen und Weiden“ statt. Große Veränderungen in beide Richtungen gibt es zwischen Nadel-, Laub- und Mischwäldern einerseits und Wald-Strauch-Übergangsflächen andererseits. Inwiefern auch kleine flächenhafte Veränderungen bestimmter Ökosysteme Auswirkungen auf die Leistungsbilanzen dieser haben und inwiefern sich dies in den charakteristischen Ökosystemvariablen der Zustandsbilanzierung widerspiegelt, werden die nachfolgenden Bilanzen zeigen. Die detaillierten Flächenbilanzen der Ökosysteme,

disaggregiert bis auf Klassenebene, stehen online in einem Tabellenband auf der [Themenseite „Ökosystem-gesamtrechnungen“](#) zur Verfügung.

6

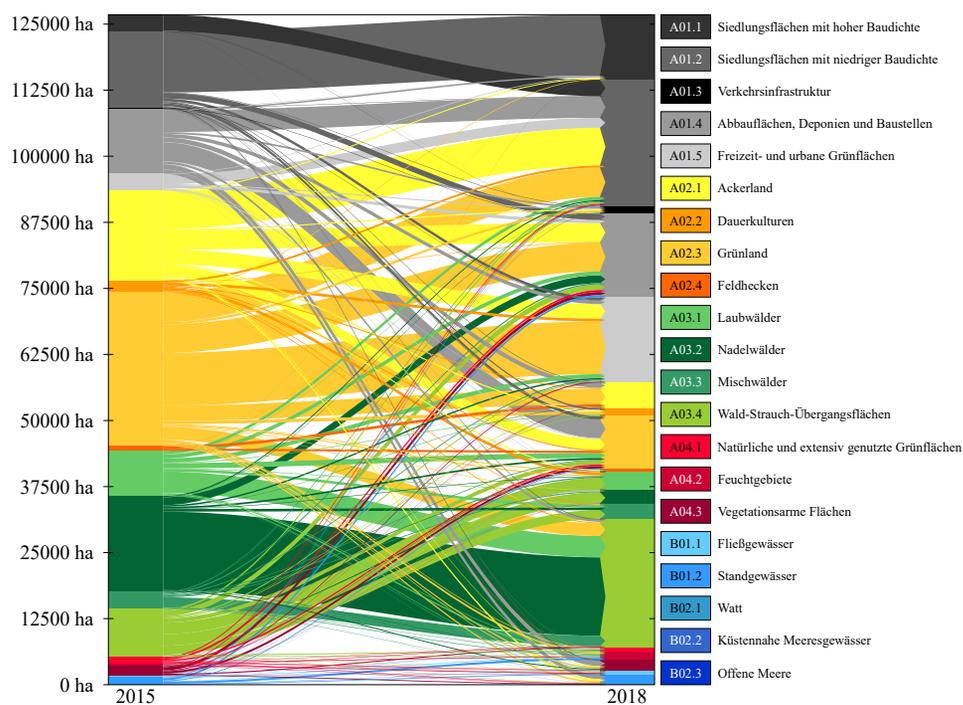
Fazit und Ausblick

Die Flächenbilanzierung der Ökosysteme baut auf vorherigen Forschungsarbeiten auf (Grunewald und andere, 2020). Sie ist der erste Schritt hin zu einer statistischen Erfassung von Ökosystemleistungen und schafft dadurch erheblichen Mehrwert für die Umweltökonomischen Gesamtrechnungen. Erstmals werden die Funktionen der Ökosysteme und damit deren Leistungen für Gesellschaft und Wirtschaft, die über eine Entnahme und Abgabe von Materialflüssen an die Umwelt hinausgehen, explizit räumlich erfasst.

Der hohe Detailgrad macht die Vielfalt der Ökosysteme in Deutschland deutlich und räumlich greifbar. Die Flächenbilanzierung kann somit dazu beitragen, politische

Grafik 6

Sankey-Diagramm der Flächenveränderungen auf Ebene der Ökosystemgruppen



2021 - 0507

Maßnahmen zielgenauer zu gestalten. Einige der erfassten Ökosystemklassen besitzen Alleinstellungsmerkmale, die nicht nur hinsichtlich der aktuellen Datenlage, sondern auch in Bezug auf künftige Analysen einen echten Mehrwert bieten. So sind erstmals bundesweit Aussagen zur Flächenbilanzierung von Streuobstwiesen sowie Auen- und Moorwäldern möglich.

Die Flächenbilanz der Ökosysteme ist konsistent mit dem internationalen Rahmenwerk SEEA EA und die Klassifikation kann in internationale Typisierungen überführt werden; dies ermöglicht länderübergreifende Vergleiche. Somit lassen sich auch künftige Berichtspflichten, beispielsweise gegenüber der EU, erfüllen.

Des Weiteren liefert die Flächenbilanzierung die Berechnungsgrundlage der Ökosystemzustands- und -leistungsbilanzierung, die die Ökosystemgesamtrechnungen vervollständigen. Beispielsweise bietet die Verbindung zwischen Bodenversiegelung und Gebäudevolumen innerhalb der Flächenbilanzierung Grundlagen zur Errechnung von Ökosystemleistungen wie lokaler Klimaregulierung. Um besser verstehen zu können, wie und warum Flächen über die Zeit von einer Ökosystemklasse in eine andere wechseln und was dies für ihre Ökosystemleistungen bedeutet, bedarf es im nächsten Arbeitsschritt der Erstellung der Zustandsbilanzierung. Die Arbeiten dazu haben in der zweiten Jahreshälfte 2021 begonnen mit einem Konzept zur Erfassung von Variablen, die die entscheidenden und spezifischen Charakteristiken der Ökosysteme bestmöglich abbilden. Durch ein Monitoring dieser Variablen kann ermittelt werden, warum bestimmte Ökosystemleistungen in ihrer Bilanz sinken, steigen oder auf gleichem Niveau verbleiben. Die Zustandsbilanzierung für die Bezugsjahre 2015 und 2018 wird Anfang 2023 fertiggestellt. Den wichtigsten Teil der Ökosystemgesamtrechnungen bilden zum Abschluss die Ermittlung der physischen Ökosystemleistungen und ihre monetäre Bewertung. Der Aufbau der Ökosystemgesamtrechnungen innerhalb der Umweltökonomischen Gesamtrechnungen gilt als fortlaufendes Produkt. 

LITERATURVERZEICHNIS

Bellinghen, Marius/Felgendreher, Simon/Oehrlein, Johannes/Schürz, Simon. *Methode der Flächenbilanzierung der Ökosysteme 2021*. Statistisches Bundesamt (Herausgeber). Umweltökonomische Gesamtrechnungen. [Zugriff am 11. November 2021]. Verfügbar unter: www.destatis.de

Bundesamt für Kartographie und Geodäsie (BKG). *Landbedeckungsmodell für Deutschland (LBM-DE)*. 2021. [Zugriff am 11. November 2021]. Verfügbar unter: www.bkg.bund.de

Bundesministerium des Innern, für Bau und Heimat (BMI). *Maritime Raumordnung – Neuer Raumordnungsplan für die AWZ*. 2021. [Zugriff am 22. November 2021]. Verfügbar unter: www.bmi.bund.de

Grunewald, Karsten/Schweppe-Kraft, Burkhard/Syrbe, Ralf-Uwe/Meier, Sophie/Krüger, Tobias/Schorcht, Martin/Walz, Ulrich. *Hierarchical classification system of Germany's ecosystems as basis for an ecosystem accounting – methods and first results*. In: One Ecosystem 5: e50648. 2020. DOI: doi.org/10.3897/oneeco.5.e50648

Schubert, Philipp R./Hukriede, Wolfgang/Karez, Rolf/Reusch, Thorsten B. H. *Mapping and modeling eelgrass *Zostera marina* distribution in the western Baltic Sea*. In: Marine Ecology Progress Series. Volume 522. 2015. Seite 79 ff. DOI: doi.org/10.3354/meps11133

SEEA Ecosystem Accounting Technical Committee. *System of Environmental-Economic Accounting – Ecosystem Accounting (SEEA-EA)*. 2021. [Zugriff am 15. November 2021]. Verfügbar unter: seea.un.org

ERPROBUNG DER SATELLITEN- GESTÜTZTEN ERTRAGSSCHÄTZUNG FÜR DIE AGRARSTATISTIK – PROJEKT SatAgrarStat

Jasmin Arnold, Patric Brandt, Heike Gerighausen

↳ **Schlüsselwörter:** Agrarstatistik – Satellitendaten – Ertragsschätzung – Copernicus – Pflanzenwachstumsmodellierung

ZUSAMMENFASSUNG

Mit dem Start des europäischen Erdbeobachtungsprogramms Copernicus ist durch die Sentinel-Satelliten eine kontinuierliche Erfassung der Erdoberfläche in bisher nicht verfügbarer zeitlicher, räumlicher und technischer Qualität gewährleistet. Aus diesen Fernerkundungsdaten lassen sich räumlich stärker differenzierte Informationen gewinnen, beispielsweise über den aktuellen Zustand landwirtschaftlicher Pflanzenbestände. Im Pilotprojekt „SatAgrarStat“ (satellitengestützte Ertragsschätzung für die Agrarstatistik) haben das Statistische Bundesamt, die Statistischen Ämter von vier Bundesländern sowie das Julius Kühn-Institut Möglichkeiten untersucht, ob durch die Verknüpfung von fernerkundlich gewonnenen Vegetationsparametern und Pflanzenwachstumsmodellen Aussagen zum Ertragspotenzial möglich sind.

↳ **Keywords:** agricultural statistics – satellite data – yield estimation – Copernicus – plant growth modelling

ABSTRACT

Since its start, the European Copernicus earth observation programme with its Sentinel satellites has ensured the continuous monitoring of the Earth's surface with an unprecedented level of temporal, spatial and technical quality. The remote sensing data offer information of greater spatial detail, such as information on the current state of agricultural crops. As part of the SatAgrarStat pilot project (satellite-based yield estimation for agricultural statistics), the Federal Statistical Office, the statistical offices of four Länder and the Julius Kühn Institute examined whether the linking of plant growth models with vegetation parameters deduced from remote sensing could open up new avenues to estimate yield potentials.

Jasmin Arnold

ist Diplom-Geografin und Referentin im Referat „Pflanzliche Erzeugung und Flächennutzung“ des Statistischen Bundesamtes. Ihr Arbeitsschwerpunkt ist die Untersuchung fernerkundungsgestützter Potenziale zur Ernte-Ertragsbestimmung für eine agrarstatistische Nutzung.

Dr. Patric Brandt

ist Nachhaltigkeitswissenschaftler und arbeitet am Institut für Pflanzenbau und Bodenkunde des Julius Kühn-Instituts. Im dortigen Forschungszentrum für landwirtschaftliche Fernerkundung beschäftigt er sich mit Ansätzen zur dynamischen Ertragsmodellierung und der Assimilation von Fernerkundungsdaten.

Dr. Heike Gerighausen

ist Diplom-Geoökologin und promovierte Geografin und leitet die Arbeitsgruppe Fernerkundung und Sensorik am Institut für Pflanzenbau und Bodenkunde des Julius Kühn-Instituts. Sie befasst sich mit der Nutzung von Fernerkundungsdaten für das Monitoring landwirtschaftlicher Pflanzenbestände und Böden, für die Abschätzung von Ernteerträgen sowie für den Einsatz dieser Technologien in der Präzisionslandwirtschaft und für die Phänotypisierung in der Züchtungsforschung.



1

Einleitung

Die jährliche Vorausschätzung und Erfassung von Erntemengen bildet eine wesentliche Grundlage zur Beschreibung des Zustands und der Entwicklung der Feldfrüchte sowie für die Erstellung von Versorgungsbilanzen auf nationaler und europäischer Ebene. Die Ermittlung der Erntemengen erfolgt derzeit im Rahmen von zwei dezentralen Bundesstatistiken, der Ernte- und Betriebsberichterstattung sowie der Besonderen Ernte- und Qualitätsermittlung.

Um die pflanzliche Produktion bestimmen zu können, schätzen freiwillige Berichterstellerinnen und Berichtersteller im Zuge der Ernte- und Betriebsberichterstattung die Ernteerträge von landwirtschaftlichen Feldfrüchten und Grünland. Mit der Besonderen Ernte- und Qualitätsermittlung werden außerdem für einige Fruchtarten objektive Ertragsmessungen auf repräsentativer Basis durchgeführt, und zwar für die Hauptgetreidesorten, Kartoffeln und Raps. Diese Informationen bilden zusammen mit den Ergebnissen der Bodennutzungshaupterhebung die Grundlagen für die Schätzung der voraussichtlichen und Bestimmung der endgültigen Erntemengen auf Bundes- und Länderebene.

Eine flächendeckende und zeitnahe räumliche Bestimmung von Ernteerträgen in tieferer regionaler Gliederung ist bisher mithilfe dieser statistischen Erhebungsverfahren nicht möglich, weil dazu die Anzahl der freiwilligen Berichterstellerinnen und Berichtersteller nicht ausreicht. Mithilfe der Fernerkundung ließen sich hingegen flächendeckende schlag- und teilschlagspezifische Informationen zum Zustand landwirtschaftlicher Pflanzenbestände ableiten. Als Schlag wird eine einheitlich bewirtschaftete, räumlich zusammenhängende und mit der gleichen Pflanzenart bestellte Fläche bezeichnet. Durch eine Verknüpfung mit Ansätzen aus der Pflanzenwachstumsmodellierung sollen regionalisierte, retrospektive Ertragserschätzungen erstellt werden, um damit die bisherigen Statistiken zu ergänzen.

Im Pilotprojekt „Prüfung einer methodisch angepassten Ernteertragsschätzung im Statistischen Verbund unter Einbezug moderner Verfahren der Fernerkundung (SatAgrarStat)“ wurden Möglichkeiten fernerkundungsbasierter Ertragserschätzungen auf Basis von Daten aus

dem europäischen Copernicus-Programm untersucht. Der methodische Ansatz sowie ausgewählte Ergebnisse werden in diesem Artikel vorgestellt: Kapitel 2 thematisiert die neuen innovativen Technologien der satellitengestützten Ertragserschätzung und ihre Potenziale für die Erntestatistiken. Es folgt in Kapitel 3 eine ausführliche Beschreibung des Pilotprojekts SatAgrarStat mit Darstellung der Ausgangslage, der Methodik, der Datengewinnung und -auswertung sowie von Ergebnissen für die Jahre 2018 und 2019. Nach einem Fazit zum Pilotprojekt zeigt das letzte Kapitel mehrere Ansatzpunkte für die weitere Entwicklung der satellitengestützten Ertragserschätzung auf.

2

Satellitengestützte Ertragserschätzung und ihr Nutzen für die Erntestatistiken

2.1 Das europäische Copernicus-Programm

Das Copernicus-Programm wurde 1998 gemeinsam von der Europäischen Kommission und der Europäischen Weltraumorganisation als unabhängiges Erdbeobachtungsprogramm initiiert. Es soll eine leistungsfähige Infrastruktur zur Erdbeobachtung und darauf basierenden Geoinformationsdiensten bereitstellen und besteht aus den Copernicus-Diensten, die Informationen in sechs thematischen Bereichen bereitstellen, einem In-situ-Beobachtungsmessnetz und einer Weltraumkomponente. Das Herz der Weltraumkomponente sind die sechs Copernicus-Sentinel-Missionen. Sie tragen unterschiedliche Technologien und Sensoren und ermöglichen damit verschiedenste Anwendungen der Erdbeobachtung, unter anderem in den Bereichen der Landüberwachung, Landwirtschaft sowie Wald- und Forstwirtschaft. Die sechs Sentinel-Missionen besitzen mittlerweile acht Satelliten im Weltall, die erfolgreich ihren Betrieb aufgenommen haben. Diese Satelliten liefern regelmäßig Daten und Informationen über den Zustand der Erdoberfläche.

Dazu zählen auch die Satelliten der Sentinel-1- und Sentinel-2-Missionen, die aufgrund ihrer Sensorsysteme und der hohen zeitlichen und räumlichen Auflösung von

besonderem Interesse für landwirtschaftliche Anwendungen sind. Sentinel-1 ist ein C-Band-Radarsystem, Sentinel-2 ein optisches, multispektrales System mit 13 Spektralbändern in einer räumlichen Auflösung von bis zu 10 Metern. Über Deutschland zeichnen diese Satelliten alle zwei bis fünf Tage Daten auf. Die offene Datenpolitik des Copernicus-Programms sichert die europäische Verordnung Nr. 1159/2013. Zudem verfolgt die Europäische Union die Strategie, das Copernicus-Programm auch nach 2020 mit Folgemissionen sowie neuen Satelliten auszurüsten. Beide Kriterien machen den Einsatz dieser Daten zur Modernisierung von Arbeitsprozessen durch Integration von erdbeobachtungsgestützten Anwendungen im behördlichen Kontext äußerst interessant.

2.2 Potenziale für die Erntestatistiken

Die Erschließung neuer digitaler Datenquellen, zum Beispiel von Fernerkundungsdaten, ist ein definiertes Ziel der Datenstrategie des Statistischen Bundesamtes. Die Nutzung der Informationen des Copernicus-Programms eröffnet Möglichkeiten, moderne Wege der Datengewinnung und Datenauswertung zu beschreiten. Die fernerkundungsgestützte Ertragsschätzung erlaubt theoretisch – anders als Stichprobennahmen und die Meldungen freiwilliger Berichterstatte(r)innen und Berichterstatte(r) – eine deutschlandweite Datenerhebung von Flächen und Erträgen sowie die Mehrfacherfassung eines Gebiets. Das Messverfahren ist dabei anhand der originalen Bilddaten stets transparent und nachvollziehbar, auch noch lange über den jeweiligen Erfassungszeitpunkt hinaus. Die Nutzung von Fernerkundung könnte in Zukunft überdies die Ausgabe tiefer regionalisierter Ergebnisse und eine verbesserte zeitliche und räum-

liche Informationsdichte über das Anbauspektrum und die erzielten Erträge einer Region ermöglichen. Damit wäre es zudem möglich, negative Entwicklungstendenzen und Gefahren durch potenzielle Ertragsrückgänge frühzeitig aufzudecken. Langfristig könnte durch den Einsatz von Fernerkundung potenziell eine Entlastung der Berichtspflichtigen und eine Verkürzung von Bereitstellungszeiten erzielt werden. Dies wäre insbesondere für Regionen mit einer sehr geringen Dichte an freiwilligen Berichterstatte(r)innen und Berichterstatte(r) vorteilhaft. Im Zusammenspiel mit anderen Quellen könnte die Fernerkundung außerdem dazu beitragen, angepasste Informationen und Handlungsempfehlungen für den Landwirtschaftssektor zu erstellen. Dies könnte vor dem Hintergrund des Klimawandels in Zukunft einen wichtigen Beitrag im Hinblick auf die Sicherung der Nahrungsmittelproduktion leisten. [↘ Übersicht 1](#)

3

Pilotprojekt SatAgrarStat

3.1 Ausgangslage

Bei dem Vorhaben handelte es sich um ein Projekt des Statistischen Verbunds¹. Beteiligt waren das Statistische Bundesamt (federführende Koordination), das Bayerische Landesamt für Statistik, das Hessische Statistische Landesamt, das Landesamt für Statistik Niedersachsen sowie für Schleswig-Holstein das Statistikamt Nord. Das Julius Kühn-Institut war ausführender

¹ Die Statistischen Ämter der Länder und das Statistische Bundesamt bilden den Statistischen Verbund.

Übersicht 1

Potenziale der Fernerkundung im Bereich der Agrarstatistiken

Merkmal	Beschreibung
Flächenabdeckung	Deutschlandweite Datenerhebung von Flächen und Erträgen
Mehrfacherfassung	Wiederholte Erhebung des gleichen Gebiets
Transparenz	Transparenz und Nachvollziehbarkeit des Messverfahrens
Regionalisierung	Ausgabe tiefer regionalisierter Ergebnisse
Aktualität	Verbesserung der zeitlichen und räumlichen Informationsdichte
Entlastung	Langfristige Entlastung von Auskunftspflichtigen
Bewirtschaftungsmaßnahmen	Bereitstellung angepasster Informationen und Handlungsempfehlungen für den Landwirtschaftssektor

Forschungs- und Entwicklungspartner. Das Pilotprojekt hatte eine Laufzeit von zwei Jahren und wurde 2018 bis 2019 umgesetzt.

Im Vorhaben SatAgrarStat wurden Ertragsschätzungen auf Basis von Sentinel-2-Satellitenbildern mit Realerträgen auf Schlägebene (In-situ-Daten) verglichen. Phänologische Erhebungen zu unterschiedlichen Zeitpunkten sowie weitere Zusatzinformationen, zum Beispiel zu Schäden im Bestand, wurden dabei zur Verifikation der Bestandsentwicklung herangezogen. In einem zweiten Schritt wurde die Qualität der fernerkundungsbasierten Ertragsschätzung durch räumliche Aggregation auf regionaler Ebene im Vergleich zur offiziellen Erntestatistik im Rahmen der Besonderen Ernte- und Qualitätsermittlung untersucht.

Die Analysen erfolgten in den vier beteiligten Bundesländern für die Fruchtarten Winterweizen, Winterraps und Sommergerste – aufgrund ihrer ackerbaulichen Relevanz und um mindestens eine Sommerfruchtart einzubeziehen. Für Winterweizen wurde auf Ergebnisse des Forschungsvorhabens Rife (Regionale fernerkundliche Erfassung des aktuellen landwirtschaftlichen Ertragspotenzials) aufgebaut (Gerighausen/Lilienthal, 2015). Für die Fruchtarten Winterraps und Sommergerste wurden sowohl Modelle zur fernerkundungsbasierten Ableitung von Vegetationsparametern als auch zur Schätzung der Erträge vollständig neu etabliert.

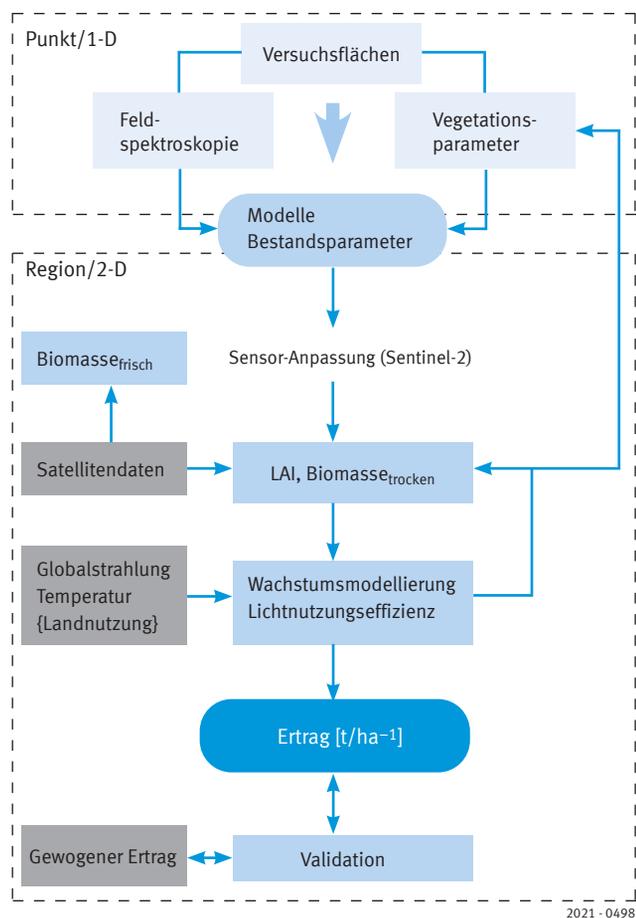
3.2 Methodik der satellitenbasierten Ertragsschätzung

Das hier angewandte Verfahren zur satellitenbasierten Ertragsschätzung von Ackerkulturen basiert auf der Verknüpfung von Daten und Methoden der Fernerkundung mit einem vereinfachten Ansatz der Wachstumsmodellierung und gliedert sich in zwei Teilschritte. ➔ Grafik 1

Vorhersagemodelle zur Schätzung von Vegetationsparametern

Auf der Basis von Feldspektren und Referenzmessungen der Bestandseigenschaften werden in einem ersten Schritt Vorhersagemodelle zur fernerkundlichen Schätzung von Vegetationsparametern etabliert. Hierzu erfolgen während der gesamten Vegetationsperiode Messungen auf den Versuchsflächen des Julius Kühn-

Grafik 1
Ablaufschema der satellitengestützten Ertragsschätzung



Instituts sowie auf landwirtschaftlich genutzten Flächen kooperierender Landwirtinnen und Landwirten. Die spektrale Reflexion des Pflanzenbestands wird mit einem Feldspektrometer ermittelt. Wichtige Bestandsgrößen, wie Blattflächenindex (LAI = leaf area index) und oberirdische Biomasse (TM), werden mithilfe destruktiver sowie nicht-destruktiver Erhebungsverfahren erhoben. Zusätzlich wird der kulturartenspezifische Ernte-Index aus dem Verhältnis geernteter Korn- und Strohmenge ermittelt. Die Kalibrierung der Vorhersagemodelle erfolgt mithilfe eines multivariaten Verfahrens, der Partial Least Squares Regression (PLSR). Hierfür werden die erhobenen Feldspektren an die spektrale Konfiguration der Sentinel-2-Satellitendaten angepasst und vorverarbeitet. Die Modellkalibrierung und -validierung erfolgt im Zuge einer Kreuzvalidierung mithilfe gemeldeter In-situ-Daten.

Wachstumsmodellierung und Ertragsschätzung

In einem zweiten Schritt schließt sich die eigentliche Ertragsschätzung an. Die auf Tagesbasis iterierende Simulation des Pflanzenwachstums basiert auf einem Ansatz von Montheith (1977), nachdem der tägliche Zuwachs an trockener Biomasse TM_i [$t\ ha^{-1}$] aus dem absorbierten Anteil der photosynthetisch aktiven Strahlung (fPAR) [$MJ\ m^{-2}$] am Tag i und der kulturartenspezifischen Lichtnutzungseffizienz ϵ [$g\ MJ^{-1}$] berechnet werden kann. fPAR wird nach Monsi und Saeki (2005) wiederum aus der täglichen Globalstrahlung, der Blattfläche [$m^2\ m^{-2}$] sowie einem dimensionslosen Extinktionskoeffizienten ermittelt. Die Verringerung des täglichen TM-Zuwachses durch mittlere Tagestemperaturen kleiner 10 Grad Celsius beziehungsweise größer 25 Grad Celsius wird durch einen Temperaturstressfaktor berücksichtigt (Zheng und andere, 2015). Die Tageswerte der Globalstrahlung sowie der Temperatur basieren auf räumlichen Daten des Deutschen Wetterdienstes. Der Ertrag wird mit Erreichen des Erntetermins schließlich durch Multiplikation der trockenen Biomasse mit dem Ernte-Index (Korn-/Stroh-Verhältnis) bestimmt. Zur Beurteilung der Güte der Ertragsschätzungen werden die ermittelten Erträge mit Realerträgen für ausgewählte Flächen (mittlerer, gewogener Ertrag je Schlag) validiert. Die geschätzten Erträge werden in einer räumlichen Auflösung von 10 Meter mal 10 Meter vom Modell ausgegeben.

3.3 Datengewinnung und Datenauswertung

Gewinnung der In-situ-Daten

Die Gewinnung von In-situ-Daten für die Validierung wurden im Rahmen einer Erhebung nach § 7 Bundesstatistikgesetz² von freiwillig teilnehmenden landwirtschaftlichen Betrieben gewonnen, von den beteiligten statistischen Ämtern aufbereitet und an das Julius Kühn-Institut übermittelt. Hier wurden sie mit den aus den Satellitenaufnahmen gewonnenen Daten zusammengeführt und ausgewertet.

In Absprache mit dem Julius Kühn-Institut wurden landwirtschaftliche Betriebe gezielt angesprochen, um die für das Projekt maßgeblichen Regionen abdecken zu können. Damit handelte es sich um eine bewusste Auswahl und keine Zufallsstichprobe. Die Regionen haben die beteiligten statistischen Ämter und das Julius Kühn-Institut gemeinsam sukzessive festgelegt.

Die Befragung der landwirtschaftlichen Betriebe erfolgte – abhängig von der Fruchtart – zu jeweils drei bis vier Berichtsterminen in den Jahren 2018 und 2019, um den aktuellen Ist-Zustand (Wachstumszustand) zu erfassen.

In den Jahren 2018 und 2019 beteiligten sich 70 beziehungsweise 64 Landwirte und Landwirtinnen mit insgesamt 144 beziehungsweise 104 Flächen am Projektvorhaben. [↘ Tabelle 1](#) Der größte Teil der gemeldeten Flächen ist mit Winterweizen und Winterraps bestellt.

² Erhebungen nach § 7 Absatz 2 Bundesstatistikgesetz dürfen zur Klärung wissenschaftlich-methodischer Fragestellungen auf dem Gebiet der Statistik Bundesstatistiken ohne Auskunftspflicht durchgeführt werden.

Tabelle 1

Am SatAgrarStat-Projekt beteiligte Landwirtinnen und Landwirte nach Fruchtart und gemeldeten Flächen sowie Landkreise, aus denen Flächen gemeldet wurden

	Bayern		Hessen		Niedersachsen		Schleswig-Holstein		Insgesamt	
	2018	2019	2018	2019	2018	2019	2018	2019	2018	2019
Projektbeteiligte Landwirtinnen und Landwirte										
Sommergerste	7	6	2	5	3	3	7	5	19	19
Winterraps	7	5	17	9	12	9	9	10	45	33
Winterweizen	9	8	19	18	12	13	10	13	50	52
Gemeldete Flächen	23	19	38	32	27	25	26	28	114	104
Landkreise, aus denen Flächen gemeldet wurden	21	18	13	11	5	5	3	4	42	38

Auf weniger als 18% der Flächen stand Sommergerste. In Schleswig-Holstein und Niedersachsen konzentrierten sich die Meldungen in beiden Jahren auf wenige Landkreise, die gemeldeten Flächen sind in diesen daher eingeschränkt repräsentativ. In Hessen und Bayern waren die gemeldeten Flächen dagegen auf zahlreiche Landkreise verstreut und daher nicht repräsentativ für den jeweiligen Landkreis.

Fernerkundungsdaten

Zur Erfassung aller im Projektvorhaben gemeldeten Schläge waren Satellitenaufnahmen von 20 immer 100 km x 100 km großen Sentinel-2-Datenkacheln erforderlich. Insgesamt wurden 292 (2018) sowie 272 (2019) Sentinel-2-Aufnahmen jeweils für den Zeitraum vom 1. April bis zum 30. Juni herangezogen sowie mithilfe einer Atmosphärenkorrektur, Wolkenmaskierung³ und einer räumlichen Stichprobenwiederholung (Resampling) auf 10 m x 10 m vorverarbeitet. Es wurden ausschließlich Satellitenbilder mit einer Wolkenbedeckung von weniger als 20% erfasst, da bei zunehmender Wolkenbedeckung die Qualität und Verwertbarkeit der Daten stark sinkt.

³ Das ist gezieltes Entfernen von Bereichen mit Wolkenbedeckung.

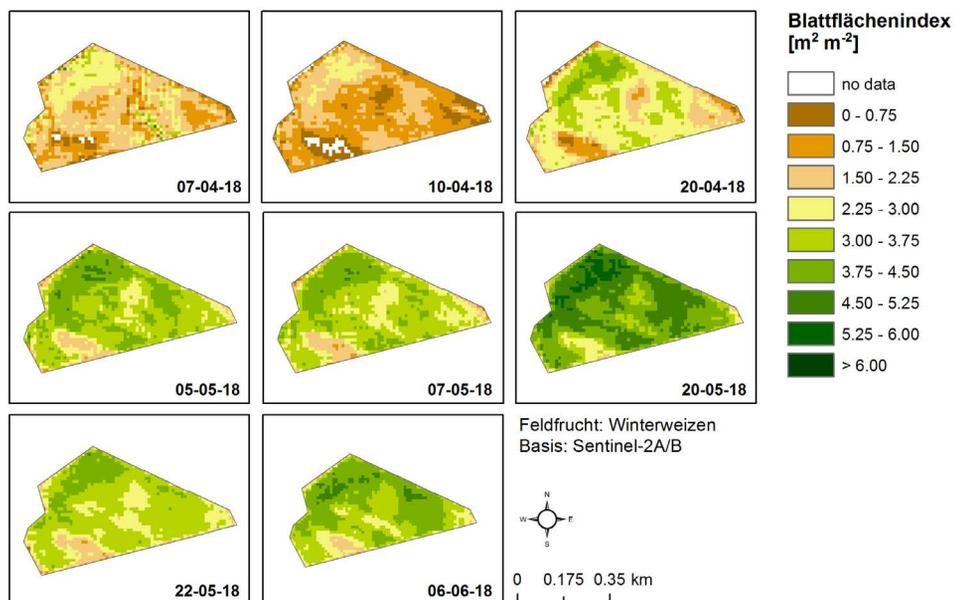
3.4 Ergebnisse

Bestandsentwicklung und Ertragsschätzung auf Schlägebene

Blattflächenindex und oberirdische Biomasse sind wichtige Parameter, mit deren Hilfe sich die Vitalität von Beständen erfassen lässt. Durch Sentinel-2 Satellitenbilder kann deren Entwicklung räumlich und zeitlich hochauflösend abgebildet werden. [↘ Grafik 2](#) zeigt exemplarisch die Entwicklung eines Weizenbestands für einen Schlag im östlichen Niedersachsen von April bis Juni 2018, abgeleitet auf Basis von Sentinel-2-Daten. Die sichtbaren Entwicklungsunterschiede im Bestand reflektieren die unterschiedlichen Relief- und Bodeneigenschaften innerhalb des Schlags (Bundesamt für Kartographie und Geodäsie, 2015; Landesamt für Bergbau, Energie und Geologie, 2018). Aufgrund der Trockenheit im Jahr 2018 steuert insbesondere die reliefbedingte Wasserverfügbarkeit die Bestandsentwicklung. Vitale Bestände befinden sich in den niederen Bereichen im nördlichen Teil des Schlags. Schlecht entwickelte Bestände stehen vor allem im westlichen Bereich auf Böden mit mittlerer Leistungsfähigkeit.

Grafik 2

Entwicklung des Blattflächenindex für einen Winterweizenschlag im östlichen Niedersachsen



2021 - 0499

Für alle in den Jahren 2018 und 2019 in den vier Bundesländern gemeldeten Schläge erfolgten Ertragsschätzungen auf Basis der Modelle. Die geschätzten Erträge zeigen für 2018 und 2019 im Vergleich zu den gemeldeten Realerträgen relative Abweichungen von 18 beziehungsweise 14 % (Winterweizen), 24 beziehungsweise 25 % (Sommergerste) und 43 beziehungsweise 42 % (Winterraps) auf. Der Fehler kann dabei abhängig von Bundesland und Fruchtart teilweise beträchtlich variieren (siehe Legende zu [Grafik 3](#)). Dennoch bewegen sich die erzielten Ergebnisse für Winterweizen im Erwartungsbereich. Für Sommergerste und Winterraps hingegen lagen keine Erfahrungswerte vor. Abgesehen von den fruchtartenspezifischen Besonderheiten bei Raps beeinträchtigten in beiden Projektjahren auftretende Wetterextreme eine Parametrisierung der Modelle für die neu hinzugekommenen Fruchtarten.

Wie die Teilgrafik B) in Grafik 3 für den bereits in Grafik 2 betrachteten Schlag in Niedersachsen zeigt, konnten auch die satellitenbasierten Schätzungen die räumliche Variabilität der Erträge innerhalb betrachteter Schläge gut abbilden. Die lokale Variabilität in den geschätzten Erträgen weist ebenso wie bei den Bestandsparametern einen starken Zusammenhang zu den relief- und bodenbedingten Standorteigenschaften auf.

Ertragsschätzung und Aggregation auf Landkreisebene

Die Ertragsschätzung auf Schlagebene stellt eine wichtige Information dar, ein potenzieller Mehrwert für die Agrarstatistik wird allerdings erst durch die Aggregation der detaillierten Ertragsschätzungen auf eine höhere, räumliche Einheit (zum Beispiel Landkreisebene) generiert. Im Pilotprojekt wurden die geschätzten Erträge für relevante Landkreise am Beispiel von Niedersachsen auf eine regionale Ebene aggregiert. [Grafik 4](#) visualisiert beispielhaft die Ergebnisse bezüglich der betrachteten Fruchtart Winterweizen für den Landkreis Helmstedt im Erntejahr 2018.

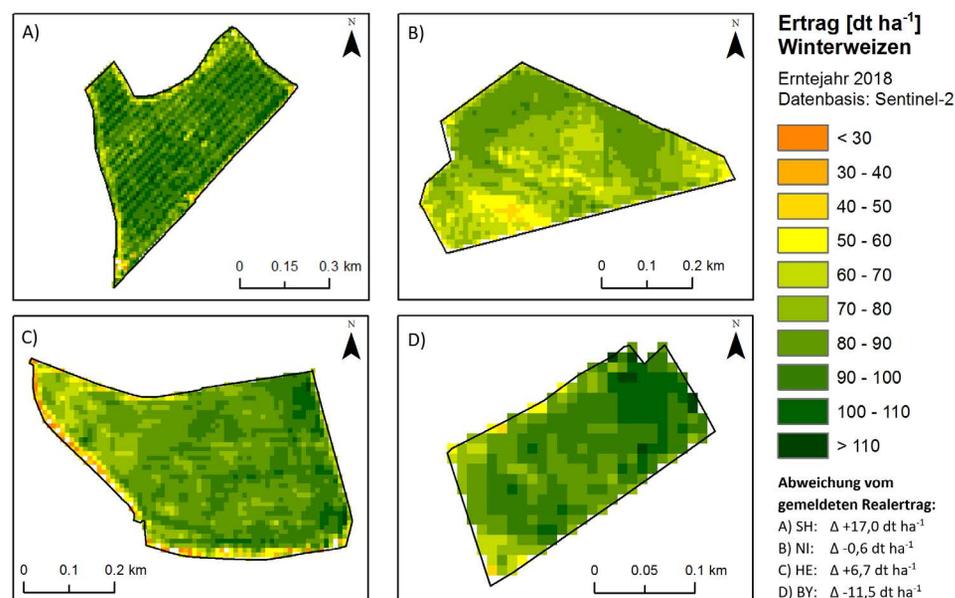
Für 2018 wurden in den über Satellit betrachteten niedersächsischen Landkreisen Winterweizenerträge zwischen 67,8 und 82,8 dt ha⁻¹ ermittelt. Für 2019 waren dies zwischen 67,9 und 80,5 dt ha⁻¹. Damit wurden im Vergleich zur Besonderen Ernte- und Qualitätsermittlung die Erträge für 2018 überschätzt und für 2019 unterschätzt. Bezogen auf den mittleren Ertrag ist dies ein relativer Fehler von 12 % jeweils für die beiden Jahre.

Für Sommergerste liegen nur für drei der betrachteten Landkreise Angaben aus der amtlichen Statistik vor.

Grafik 3

Ertragsschätzung auf Schlagebene für 2018

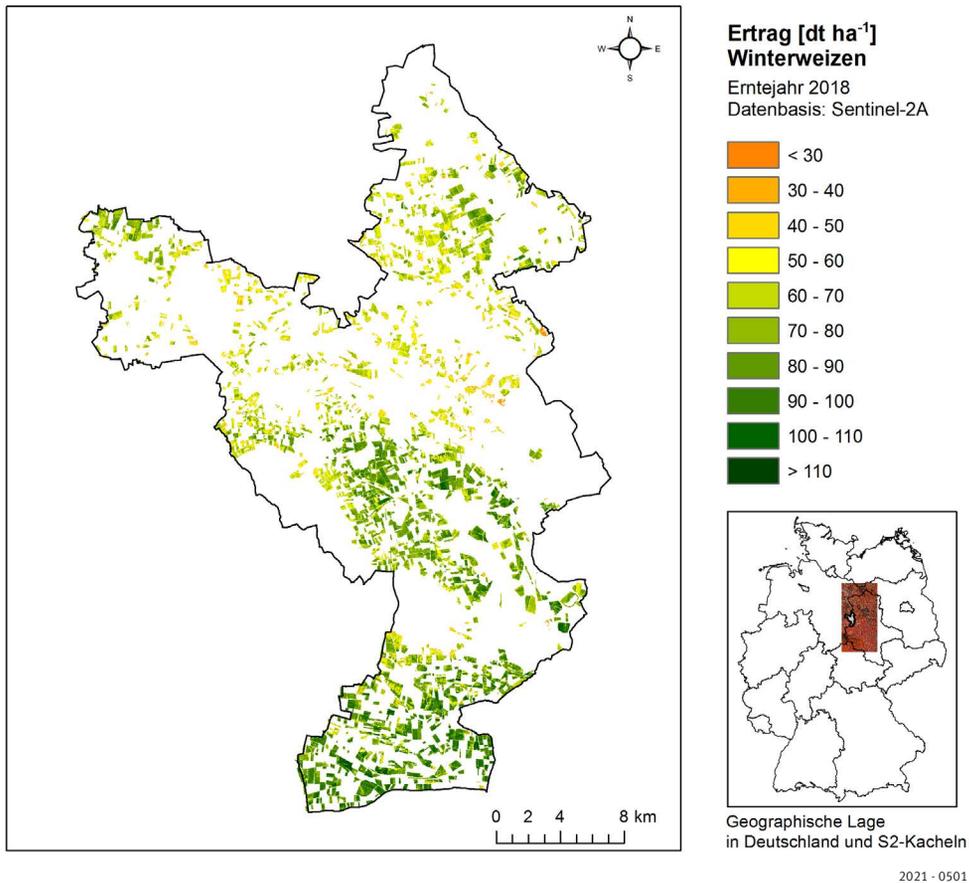
Beispiele für vier Winterweizenschläge in A) Schleswig-Holstein, B) Niedersachsen, C) Hessen und D) Bayern



2021 - 0500

Grafik 4

Ertragsschätzung für Winterweizen auf Landkreisebene für 2018: Beispiel Helmstedt im östlichen Niedersachsen



Die prozentuale Abweichung der einzelnen Landkreise beträgt für 2018 zwischen -18,9 und +35,4% und für 2019 zwischen -18,7 und +86,8%. Der relative Fehler bezogen auf die Jahresmittelwerte der amtlichen Statistik beträgt 21% für 2018 beziehungsweise 33% für 2019. Für Winterraps wurden die Erträge generell in beiden Jahren überschätzt. Die prozentuale Abweichung im Jahr 2018 liegt hier zwischen 29,6 und 44,7%, für 2019 zwischen 18,5 und 50,7%. Der mittlere relative Fehler in Bezug auf die Erntestatistiken von 2018 und 2019 beträgt jeweils 36%.

4

Fazit und Ausblick

Das Pilotprojekt SatAgrarStat konnte Möglichkeiten und Grenzen fernerkundungsbasierter Ertragsschätzungen mit Copernicus-Daten für den konkreten Anwendungsfall der amtlichen Erntestatistik aufzeigen. So belegen die Untersuchungen in den Jahren 2018 und 2019, dass eine fernerkundungsgestützte Ertragsschätzung mit dem gewählten Ansatz grundsätzlich möglich ist. Für die Feldfrüchte Winterweizen, Sommergerste und Winterraps wurden Vegetationsparametermodelle etabliert und mit den aus den Fernerkundungsdaten abgeleiteten Bestandsinformationen Erträge geschätzt sowie räumlich aggregiert. Die Qualität der Schätzungen wird dabei maßgeblich von der Verfügbarkeit von Satelliten-

bildern, der Anzahl und Qualität der In-situ-Daten und dem Vorliegen von Daten über die Verteilung des Ackerbaus beeinflusst. Modellergebnisse mit den geringsten Abweichungen zu vorliegenden Realerträgen ließen sich für die Fruchtart Winterweizen erzielen.

Die Resultate für die betrachteten Jahre deuten allerdings auch darauf hin, dass Ertragsverluste durch Trockenstress mit dem verwendeten methodischen Ansatz noch nicht beziehungsweise nicht in ausreichendem Maße erfasst werden können. Hier besteht weiterer Forschungsbedarf, beispielsweise um die Pflanzenwachstumsmodelle besser an Extremsituationen anpassen zu können. Die bisher erzielten Ergebnisse auf regionaler Ebene sind derzeit noch nicht ausreichend belastbar, um sie in die laufenden Statistiken zu integrieren. Für eine mögliche operative Nutzung von Fernerkundungsdaten im Rahmen der gesetzlich geregelten Ernteerhebungen und die dafür erforderliche Integration in den Prozess der amtlichen Statistikerstellung sind daher weitere umfangreiche Untersuchungen und Vorarbeiten erforderlich. Die Erkenntnisse aus dem Pilotprojekt bilden dabei die Grundlage für das Folgevorhaben SatAgrarStat_PLUS.

Das Projekt SatAgrarStat_PLUS wird ebenfalls im Verbund mit dem Statistischen Bundesamt (Koordination), nunmehr sechs Statistischen Landesämtern sowie dem Julius Kühn-Institut (Projektausführung) umgesetzt. Neben den vier bisher bereits beteiligten Landesämtern nehmen nun auch das Amt für Statistik Berlin-Brandenburg und der Landesbetrieb Information und Technik Nordrhein-Westfalen teil. Basierend auf Erkenntnissen aus dem Pilotprojekt wird während einer dreieinhalbjährigen Projektlaufzeit der Ansatz der satellitengestützten Ertragsschätzung in erweiterter Form erneut umgesetzt. Folgende Aspekte stehen im Fokus des Folgevorhabens:

Aufnahme der Wintergerste als zusätzliche Fruchtart

Innerhalb von SatAgrarStat_PLUS wird der bisher verwendete Modellansatz erweitert, um Erträge für die Wintergerste, eine weitere wichtige Fruchtart, zu schätzen. Wintergerste wird auf knapp 15 % des gesamten Ackerlands angebaut und ist damit die dritthäufigste Fruchtart in Deutschland (Statistisches Bundesamt, 2020). Ihr Anteil an der Anbaufläche ist in den vergangenen Jahren stetig gestiegen und wird voraussichtlich aufgrund

sich ändernder Klimabedingungen weiter ansteigen (Bundesministerium für Ernährung und Landwirtschaft, 2019). Für die Erweiterung des Modellansatzes werden zusätzliche Felderhebungen und Referenzmessungen durchgeführt und die Bestandsparametermodelle für Blattflächenindex und oberirdische Biomasse der Wintergerste entsprechend kalibriert beziehungsweise validiert.

Skalierung auf unterschiedliche räumliche Aggregationsebenen

Im Pilotprojekt wurden die Ertragsschätzungen für ausgewählte Schläge und Landkreise validiert und verfügbar gemacht. Das Folgevorhaben geht einen Schritt weiter und testet, inwiefern Ertragsschätzungen auf weiteren Aggregationseinheiten, beispielsweise auf Gemeindeebene (administratives Gliederungskriterium) oder auf Basis eines einheitlichen Gitternetzes (systematisches Gliederungskriterium), hinreichend in derzeitige agrarstatistische Erhebungsverfahren eingebettet werden können. Um eine Aggregation auf höherer, räumlicher Ebene zu ermöglichen, ist eine möglichst repräsentative Anzahl entsprechender Landnutzungs- und Anbauinformationen nötig, also schlagspezifische Angaben zur Lage und Geometrie der relevanten, landwirtschaftlich genutzten Flächen. Einige der am Folgeprojekt teilnehmenden Bundesländer werden diese Informationen im Rahmen der Agrarförderung allgemein zugänglich machen; sie sind somit im Vorhaben nutzbar.

Nutzung von „CODE-DE“

Im Folgevorhaben wird die Nutzung von „CODE-DE“ als Cloud-Computing-Plattform forciert. CODE-DE ist Bestandteil der Geoinformationsstrategie Deutschlands und dient der Prozessierung sowie Bereitstellung von Erdbeobachtungsinformationen. Die Plattform wird im Auftrag des Bundesministeriums für Verkehr und digitale Infrastruktur sowie des Deutschen Luft- und Raumfahrtzentrums bereitgestellt. Sie bietet Zugang zu umfangreichen Repositorien von Copernicus-Datenprodukten, welche flächendeckend für ganz Deutschland verfügbar sind. Dies erlaubt einerseits eine gute Skalierbarkeit der Ertragsmodellierung, andererseits verfügt CODE-DE auch über eine performante Serverarchitektur, um entsprechende Produkte hochauflösend zu berechnen und großräumig zur Verfügung zu stellen. Weitere Vorteile von CODE-DE sind die langfristige Verfügbarkeit

der Copernicus-Daten sowie der simulierten Ertrags-schätzungen. Über Web-Services können so über das Julius Kühn-Institut Projektergebnisse für die Projektbe-teiligten komfortabel und langfristig verfügbar gemacht werden.

Mehr Daten durch verbesserte Wolkenmasken

Optische Sensoren wie die von Sentinel-2-Satelliten sind grundsätzlich „blind“, wenn sich zwischen Sensor und Erdoberfläche Wolken befinden. Von Wolken bedeckte Areale können aus den Bilddaten gezielt maskiert, das heißt herausgeschnitten werden. CODE-DE verfügt über elaborierte Algorithmen, um eine saubere Wolkenmas-kierung zu ermöglichen. Dadurch lässt sich die ferner-kundliche Beobachtungsdichte letztlich sogar erhöhen, da sich Satellitenbilder selbst mit hoher Wolkenbe-deckung zumindest noch partiell für die Wachstums-modellierung nutzen lassen. Daher werden, anders als im Pilotprojekt, grundsätzlich alle zur Verfügung stehen-den Satellitenbilder genutzt. Je mehr Satellitenbeobach-tungen zur Verfügung stehen, desto häufiger lassen sich die im Laufe der phänologischen Entwicklung simulier-ten Bestandsparameter mit den vom Satelliten beobach-teten Parametern abgleichen. Dies führt zu potenziell genaueren Ertragsschätzungen.

Modellerweiterung und -optimierung

Im bisherigen Ansatz der Ertragsschätzung war es nicht möglich, Faktoren wie die niederschlagsgetriebene Was-serverfügbarkeit sowie bestimmte Bodeneigenschaften wie die Wasserhaltekapazität, die Einfluss auf das Wachstum der Kulturarten und damit auf deren Erträge haben, zu berücksichtigen. Daher wird im Folgevor-haben untersucht, ob die Integration weiterer, auf die Ertragsbildung einflussnehmender Aspekte möglich ist. Ein Beispiel dafür ist der durch den Klimawandel zunehmend an Bedeutung gewinnende Trockenstress. Darüber hinaus werden modellbasierte Sensitivitäts-analysen durchgeführt, um relevante Modellparameter fruchtartenspezifisch zu optimieren.

Intra-saisonale Ertragsprognosen

Neben der eigentlichen Ex-post-Ertragsschätzung stellt die intra-saisonale Prognose von Erträgen eine wich-tige Möglichkeit dar, Bewirtschaftungsentscheidungen ex-ante zu unterstützen. Im Folgevorhaben wird neben

der Ex-post-Schätzung daher das Potenzial von Ertrags-prognoseansätzen mithilfe von 30-jährigen meteoro-logischen Zeitreihen des Deutschen Wetterdienstes untersucht. Hierbei soll insbesondere evaluiert werden, inwiefern derartige Ansätze basierend auf einer ganzen Serie von Modellsimulationen (Ensemblesimulationen), anstatt einer einzigen Simulation, zu plausiblen Ertrags-prognosen führen. [!!!](#)

LITERATURVERZEICHNIS

Bellon-Maurel, Véronique/Fernandez-Ahumada, Elvira/Palagos, Bernard/Roger, Jean-Michel/McBratney, Alex. *Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy*. In: Trends in Analytical Chemistry. Jahrgang 29. Ausgabe 9/2010, Seite 1073 ff.

Bundesministerium für Ernährung und Landwirtschaft. *Ernte 2019. Mengen und Preise*. 2019. [Zugriff am 15. November 2021]. Verfügbar unter: www.bmel.de

Bundesamt für Kartographie und Geodäsie. *Digitales Geländemodell Gitterweite 10 m, DGM10*. 2015. [Zugriff am 15. November 2021]. Verfügbar unter: gdz.bkg.bund.de

Chang, Cheng-Wen/Laird, David A./Mausbach, Maurice J./Hurburgh, Charles R. Jr. *Near-Infrared Reflectance Spectroscopy – Principal Components Regression Analyses of Soil Properties*. In: Soil Science Society of America Journal. Jahrgang 65. Ausgabe 2/2001, Seite 480 ff.

Statistisches Bundesamt. *Feldfrüchte und Grünland: Ackerland nach Hauptfruchtgruppen und Fruchtarten*. 2020. [Zugriff am 15. November 2021]. Verfügbar unter: www.destatis.de

Gerighausen, Heike/Lilienthal, Holger. *Regionale fernerkundliche Erfassung des aktuellen landwirtschaftlichen Ertragspotenzials am Beispiel eines Testgebietes in Mecklenburg-Vorpommern*. In: Bornimer Agrartechnische Berichte. Ausgabe 88/2015, Seite 15 ff.

Landesamt für Bergbau, Energie und Geologie. *Bodenschätzungskarte von Niedersachsen im Maßstab 1:5.000 (BS5)*. 2018. [Zugriff am 15. November 2021]. Verfügbar unter: www.lbeg.niedersachsen.de

Monsi, Masami/Saeki, Toshiro. *On the factor light in plant communities and its importance for matter production*. In: Annals of Botany. Jahrgang 95. Ausgabe 3/2005, Seite 549 ff. DOI: [10.1093/aob/mci052](https://doi.org/10.1093/aob/mci052)

Monteith, John Lennox. *Climate and efficiency of crop production in Britain*. In: Philosophical Transactions of the Royal Society of London. Ausgabe 980/1977, Seite 77 ff. DOI: [10.1098/rstb.1977.0140](https://doi.org/10.1098/rstb.1977.0140)

Zheng, Bangyou/Chenu, Karine/Doherty, Alastair/Chapman, Scott. *The APSIM-Wheat Module (7.5R3008)*. 2015. [Zugriff am 15. November 2021]. Verfügbar unter: www.apsim.info

RECHTSGRUNDLAGEN

Delegierte Verordnung (EU) Nr. 1159/2013 der Kommission vom 12. Juli 2013 zur Ergänzung der Verordnung (EU) Nr. 911/2010 des Europäischen Parlaments und des Rates über das Europäische Erdbeobachtungsprogramm (GMES) durch die Festlegung von Registrierungs- und Lizenzierungsbedingungen für GMES-Nutzer und von Kriterien für die Einschränkung des Zugangs zu GMES-spezifischen Daten und Informationen der GMES-Dienste (Amtsblatt der EU Nr. L 309, Seite 1).



Marcus Schmidt

ist Diplom-Mathematiker und Referent im Referat „Mathematisch-statistische Verfahren für Bevölkerungs- und Sozialstatistiken“ des Statistischen Bundesamtes. Er ist zuständig für methodische Untersuchungen und die mathematische Entwicklung in Haushaltsstatistiken und arbeitet an der Neukonzeption und Umsetzung der Hochrechnung im Mikrozensus ab 2020.



Jana Lucia Stein

ist Master of Science Mathematik und Referentin im Referat „Mikrozensus – Methodik und Aufbereitung SILC und IKT“ des Statistischen Bundesamtes. Sie ist zuständig für Hochrechnung und Imputation im Mikrozensus und arbeitet an der Neukonzeption und Umsetzung der Hochrechnung im Mikrozensus ab 2020.

DIE HOCHRECHNUNG IM MIKROZENSUS AB 2020

Marcus Schmidt, Jana Lucia Stein

↘ **Schlüsselwörter:** Mikrozensus – Haushaltserhebung – Gewichtung – Kalibrierung – Antwortausfälle

ZUSAMMENFASSUNG

Mit dem Ziel, die amtlichen Haushaltsstatistiken zu harmonisieren, wird der Mikrozensus seit 2020 als integriertes System durchgeführt. Neben der Arbeitskräfteerhebung der Europäischen Union sind auch die europäische Haushaltsbefragung über Einkommen und Lebensbedingungen sowie ab 2021 die Statistik zur Nutzung von Informations- und Kommunikationstechnologien modular in den Mikrozensus integriert. Durch ein sogenanntes Kernprogramm, welches weiterhin über eine 1%-Stichprobe erhoben wird, sind die Erhebungsteile thematisch miteinander verknüpft. Dieser Artikel erläutert die neu konzipierte Methodik der Hochrechnung im neuen System ab 2020. Die Neukonzeption soll die Vielfalt an nationalen und europäischen Anforderungen erfüllen.

↘ **Keywords:** *microcensus – household survey – weighting – calibration – unit nonresponse*

ABSTRACT

The microcensus has been implemented as an integrated system since 2020, with the aim of harmonising the official household statistics. In addition to the labour force survey (LFS) of the European Union, the European survey of income and living conditions (SILC) is integrated as a module into the microcensus and the same applies to the statistics on the use of information and communication technologies (ICT) as of 2021. The survey components are thematically linked through the core programme of variables which continue to be collected based on a 1% sample. This article explains the redesigned methodology of extrapolation that has been used in the new system since 2020. The redesigned system is aimed to meet the variety of national and European requirements.

1

Einleitung

Auf Basis neu gestalteter Auswahl-¹ und Rechtsgrundlagen² von 2016 wurde die Durchführung des Erhebungsgeschäfts und die Aufbereitung der Stichproben für den Mikrozensus (MZ) 2020 umgestellt. Ein zentraler Aspekt dieser Umstellung ist die Integration der bisher freiwilligen und eigenständig durchgeführten europäischen Erhebungen EU-SILC (European Union Statistics on Income and Living Conditions – Erhebung über Einkommen, Armut und Lebensbedingungen der Bevölkerung) und IKT (Informations- und Kommunikationstechnologien). Wie bisher enthalten bleibt die Arbeitskräfteerhebung LFS (Labour Force Survey).³ Während die integrierte MZ-SILC-Erhebung in weiten Teilen mit Auskunftspflicht belegt wird, ist für den Erhebungsteil MZ-IKT weiter eine freiwillige Beantwortung vorgesehen. Beide Erhebungen finden nur in einem begrenzten Zeitraum des Jahres statt. Die Erhebung des Kernprogramms und der MZ-LFS-Substichprobe erfolgt weiterhin über das gesamte Jahr. Für alle Erhebungsteile wurde ab 2020 zudem eine feste Berichtswoche eingeführt, das heißt auch bei eventuellen Verzögerungen der Befragung erfolgen die Antworten der Haushalte zu einer bestimmten Kalenderwoche. Detaillierte Informationen zu den Neuerungen im Mikrozensus sind der Literatur zu entnehmen (Bihler/Zimmermann, 2016; Hundenborn/Enderer, 2019). Dieser Artikel behandelt die Hochrechnung und sich darauf auswirkende Neuerungen im Mikrozensus ab 2020.

Dazu erläutert Kapitel 2 zunächst die hochrechnungsrelevanten Neuerungen im Mikrozensus, die sich ab dem Erhebungsjahr 2020 aus der Neugestaltung ergeben. Kapitel 3 beschreibt die zum Einsatz kommende Methodik der gebundenen Hochrechnung, die Ziehungswahrscheinlichkeiten sind Thema in Kapitel 4. Die Berechnung von Antwortwahrscheinlichkeiten wird in Kapitel 5 dargestellt, während sich die Kapitel 6 und 7 mit den verschiedenen Hochrechnungsfaktoren, der Bereitstel-

lung von Eckwerten und den verwendeten Hochrechnungsmodellen befassen. Methodischer Anpassungsbedarf aufgrund von Erhebungseinschränkungen im Jahr 2020 ist in Kapitel 8 aufgeführt. Ein kurzes Fazit beschließt den Beitrag.

2

Hochrechnungsrelevante Neuerungen im Mikrozensus ab 2020

2.1 Periodizität der Erhebung

Neben den neu integrierten Erhebungen wurde auch die Erfassung zur Arbeitsmarktbeteiligung neu konzipiert. Um unterjährige Veränderungen im Arbeitsmarkt besser abzubilden, erfolgt die Erfassung dieser Substichprobe unterjährig nach dem Rotationschema 2-(2)-2. Das bedeutet, dass die Befragungen in den entsprechenden Auswahlbezirken zunächst in zwei aufeinanderfolgenden Quartalen erfolgt, in den zwei darauffolgenden Quartalen pausiert und dann in den zwei anschließend folgenden Quartalen nochmals durchgeführt wird (Hundenborn/Enderer, 2019, hier: Grafik 2). Um die Befragungsbelastung zu reduzieren und da einige der Merkmale als zeitstabil angenommen werden, unterscheiden sich MZ-LFS-Befragungen in der Zusammenstellung der Erhebungsmerkmale. Es gibt sogenannte Strukturmerkmale, die einmal jährlich abzufragen sind. Diese werden fest in den Befragungswellen eins und drei abgefragt, sowie zusätzlich in der vierten Befragungswelle, sofern die Berichtswoche im ersten Quartal liegt.⁴ In den weiteren Befragungen des MZ-LFS werden nur die Quartalsmerkmale – das sind saisonale Merkmale – erfragt.

Für alle weiteren Substichproben gilt das bisherige Rotationsschema. Ein Viertel der Stichprobe rotiert jedes Jahr heraus und wird durch ein neues, hereinrotierendes Viertel ersetzt. Es liegt daher keine einheitliche Rotation über die gesamte Stichprobe vor (Hundenborn/Enderer, 2019).

1 Erneuerung der Auswahlgrundlage auf Basis des Zensus 2011.

2 Die Rechtsgrundlagen sind am Ende des Beitrags aufgeführt.

3 Um die in den Mikrozensus integrierten Erhebungen von den eigenständigen beziehungsweise den europäischen Erhebungen abzugrenzen, werden im Beitrag die Bezeichnungen MZ-LFS, MZ-SILC und MZ-IKT verwendet, wo sie zutreffen.

4 Diese letzte Bedingung kam hinzu, um den gesetzlichen Anforderungen zu entsprechen: Danach muss jeder Haushalt, der Teil der MZ-LFS-Stichprobe ist, in jedem Kalenderjahr seiner Teilnahme an der Erhebung auch Angaben zu den Strukturmerkmalen geben.

2.2 Stichprobe

Beim integrierten Mikrozensus handelt es sich weiterhin um eine geschichtete Klumpenstichprobe von 1 % der Bevölkerung (Bihler/Zimmermann, 2016). Eine proportional einheitliche Verteilung der Stichprobe über alle Bundesländer und Regierungsbezirke kann für den integrierten Mikrozensus nicht mehr gewährleistet werden. Präzisions⁵- und Feldphasenanforderungen in Verbindung mit einem Gesamtauswahlsatz von maximal 1 % der Bevölkerung erfordern, dass die Auswahlsätze der Substichproben auf Ebene der Regierungsbezirke (NUTS 2) variieren (Hundenborn/Enderer, 2019).

Um den Anforderungen eines Panels für die EU-SILC-Erhebung zu genügen, werden Haushalte der MZ-SILC-Substichprobe, die aus ihrer Wohnung wegziehen, bevor sie aus der Stichprobe herausrotieren, weiterverfolgt; ihre Befragung zum Kernprogramm und MZ-SILC erfolgt dann auf freiwilliger Basis. Eine Ausnahme von dieser Regelung sind sogenannte Einzugshaushalte⁶, diese werden im Falle eines Auszugs nicht mehr weiterverfolgt.

2.3 Ergebnisbereitstellung

Mit dem neuen System sollen die Ergebnisse des Mikrozensus deutlich früher als bisher zur Verfügung stehen. Dennoch ist auch ein ausreichender Zeitraum notwendig zur Berücksichtigung von Auskünften, die gegebenenfalls durch ein Verwaltungsverfahren infolge der Auskunftspflicht gewonnen werden. Aufgrund dieses Spannungsfelds sind ab 2020 zwei Ergebnisarten vorgesehen: Erst- und Endergebnis.

Beide Ergebnisarten beruhen auf vollständig aufbereiteten und hochgerechneten Daten. Sie unterscheiden sich jedoch in der Menge der vorliegenden Beobachtungen und in der Konsistenz der Hochrechnung der Substichproben. Für Letztere werden hochgerechnete Daten zu den jeweiligen Fristen an Eurostat geliefert und zeitnah

⁵ Die Präzisionsanforderungen des Statistischen Amtes der Europäischen Union (Eurostat) an die jeweiligen Substichproben LFS, SILC und IKT sind gegeben als obere Grenze an die geschätzten Standardfehler (siehe Anhang II der Verordnung [EU] 2019/17001700). Diese implizieren wiederum einen vom Designeffekt abhängigen Mindeststichprobenumfang.

⁶ Das sind Haushalte, die in einen in Vorwellen bereits befragten Wahlbezirk der Stichprobe einziehen.

veröffentlicht. Das Endergebnis wird für Kernprogramm und alle Substichproben zum gleichen Zeitpunkt erstellt, sodass Ergebniskonsistenz für genau die in der Hochrechnung verwendeten Anpassungsmerkmale erreicht werden kann. Eine Ausnahme bildet die IKT-Erhebung, für die lediglich ein Endergebnis bereitgestellt wird. Grund dafür ist, dass durch deren Freiwilligkeit die Möglichkeiten einer höheren Ausschöpfung begrenzt sind.

3

Gebundene Hochrechnung

Bei einer Hochrechnung im Sinne einer Vorhersage von unbekanntem Größen einer Grundgesamtheit aus einer Zufallsstichprobe wird allgemein zwischen freier und gebundener Hochrechnung unterschieden. Eine freie Hochrechnung bedingt eine Zufallsauswahl, in der jedes Element mit einer Wahrscheinlichkeit größer 0 unabhängig von anderen Elementen ausgewählt wird. Existieren Einflüsse, bei denen man von Auswirkungen auf die Ziehung beziehungsweise von einer hohen Korrelation mit den Ergebnissen ausgehen kann, und sind diese bekannt, so wird auf eine gebundene Hochrechnung zurückgegriffen. Die gebundene Hochrechnung nutzt Eckwerte aus externen Quellen, um die Schätzung in der freien Hochrechnung zu korrigieren (Deville und andere, 1993). Um die Varianz beziehungsweise Standardabweichung von Ergebnissen erwartungstreu zu schätzen, müssen die externen Eckwerte korrekt sein. Das heißt die Schätzung der Varianz für Ergebnisse der Hochrechnung vernachlässigt jede Unsicherheit in den verwendeten externen Hilfsmerkmalen.

Die gebundene Hochrechnung im Mikrozensus hat grundsätzlich zwei Ziele:

- › Sie soll die systematischen Fehler durch unerkannte Antwortausfälle verringern und ebenso
- › die Zufallsfehler derjenigen Erhebungsmerkmale, die mit den Bezugsmerkmalen hoch korreliert sind.

Durch die Anpassung wird auch eine gewisse Konsistenz zu den in der amtlichen Statistik verfügbaren Bevölkerungsdaten erreicht. Ziel der Hochrechnung ist, den Personen beziehungsweise Haushalten in der Stichprobe Hochrechnungsgewichte zuzuordnen, mit deren Hilfe

Ergebnisse (Totalwerte, Anteilswerte) zu gewissen Merkmalen berechnet werden können.

Wie bereits bisher im Mikrozensus wird der verallgemeinerte Regressionschätzer als Kalibrierungsverfahren zur Erzeugung aller Hochrechnungsfaktoren verwendet.

3.1 Der verallgemeinerte Regressionschätzer

Der verallgemeinerte Regressionschätzer (GREG – Generalized Regression Estimator) fungiert als modellunterstützte Erweiterung des designbasierten Horvitz-Thompson-Schätzers (HT-Schätzer). Der HT-Schätzer kann allgemein als freie Hochrechnung verstanden werden und der GREG-Schätzer korrigiert diesen anhand eines Modells aus externen Hilfsvariablen und einem Vektor von geschätzten Regressionskoeffizienten. In einer gewichteten Regression wird also der HT-Schätzer um den Korrekturfaktor des GREG-Schätzers erweitert. Genauere Ausführungen sowie die Herleitung der genutzten Formeln finden sich bei Afentakis und Bihler (2005; hier: Seite 1042). Um die Handhabung bei der Ergebniserstellung aus dem Material des Mikrozensus zu vereinfachen, wird der zu berechnende Hochrechnungsfaktor w_k wie bisher an das Einzelmateriale angefügt.

Antwortausfälle können nicht vernachlässigt werden (siehe Abschnitt 2.4 in Afentakis/Bihler, 2005). Es wird somit weiterhin auf das zweistufige Hochrechnungsverfahren zurückgegriffen. In der ersten Stufe soll möglichen Verzerrungen vor der Hochrechnung durch Berechnung von Antwortwahrscheinlichkeiten begegnet werden. Die Wahrscheinlichkeit für Person k , in die Stichprobe zu gelangen, wird daher zusammengesetzt aus der Ziehungswahrscheinlichkeit π_k und der geschätzten Antwortwahrscheinlichkeit $\hat{\theta}_k$ (siehe Kapitel 5). Für deren Kehrwert d_k gilt:

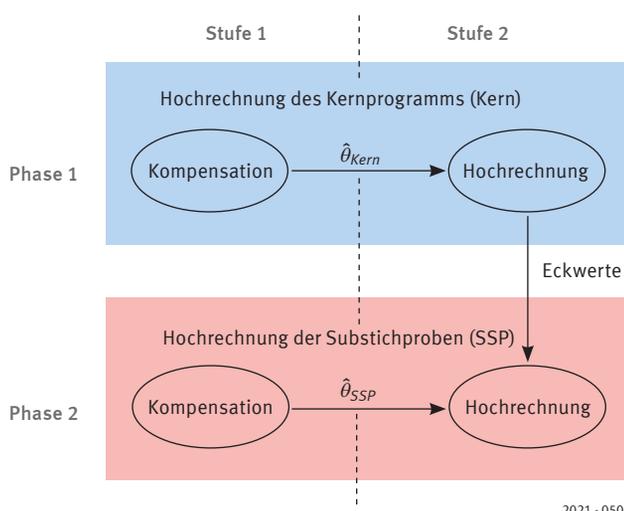
$$(1) \quad d_k = \frac{1}{\pi_k \hat{\theta}_k}$$

Alle Personen eines Haushalts erhalten wie bisher den gleichen Hochrechnungsfaktor (Afentakis/Bihler, 2005).

3.2 Zwei-Phasen-Hochrechnung

Zum Zeitpunkt der Substichprobenhochrechnung (Substichproben: MZ-LFS, MZ-SILC und MZ-IKT) steht der bis dahin jeweils erhobene Teil des Kernprogramms zur Verfügung. Dieser enthält auch einige der für die Substichproben relevanten Merkmale. Für diese Merkmale liegen also zum Zeitpunkt der Hochrechnung validere, für eine größere Stichprobe erfasste Informationen vor. Mit dem Ziel der Kohärenz zwischen Kernprogramm und Substichproben (und da einige Merkmale als hochkorreliert angenommen werden können) wird für den Mikrozensus ab 2020 eine Hochrechnung der Substichproben in zwei Phasen angestrebt. Die erste Phase ist die Hochrechnung des kompletten Kernprogramms in der geforderten zeitlichen Untergliederung, beispielsweise Quartal, Halbjahr oder Jahr. In der zweiten Phase findet eine gebundene Hochrechnung der Substichproben an die aus der ersten Phase gewonnenen Eckwerte statt. Dabei wird für jede Phase das zweistufige Verfahren des bisherigen Mikrozensus angewendet. Als erste Stufe erfolgt die Kompensation von Antwortausfällen und in einer zweiten Stufe die eigentliche Hochrechnung an die Grundgesamtheit. Für die Substichproben wird eine getrennte Kompensation mit jeweils eigenen Modellen implementiert. [↪ Grafik 1](#)

Grafik 1
Zwei-Phasen-Hochrechnung des Mikrozensus



2021 - 0508

4

Ziehungswahrscheinlichkeit

Die Einheit für die Hochrechnungen ist die Person beziehungsweise der Haushalt. Im Mikrozensus sind die Auswahlseinheiten jedoch nicht Personen oder Haushalte, sondern Auswahlbezirke. Die Ziehungswahrscheinlichkeiten der Auswahlbezirke gelten für alle Personen/Haushalte des betreffenden Auswahlbezirks, da alle auf der Fläche des Auswahlbezirks in der Berichtswoche wohnenden Personen in die Erhebung einbezogen werden. Bei der Auswahl steht zusätzlich eine zeitliche Komponente⁷ zur Verfügung, die Auswahlseinheit ist also im Grunde die Kombination von Auswahlbezirk und Kalenderwoche. Bei den Substichproben MZ-SILC und MZ-IKT sind die Merkmale zeitlich eher stabil, daher kann dieser Aspekt zur einfacheren Darstellbarkeit entfallen. Die zeitliche Komponente ist aber zwingend bei allen Hochrechnungen des MZ-LFS und für die Jahreshochrechnung des Kernprogramms zu berücksichtigen.

⁷ Ausgenommen sind diejenigen Merkmale von EU-SILC, die sich nicht wie die demografischen Merkmale oder Haushaltstypen auf die Berichtswoche, sondern auf einen zurückliegenden Stichtag/Zeitraum beziehen.

4.1 Kernhochrechnung

Im Jahresmaterial können die Einheiten der Hochrechnung (Personen beziehungsweise Haushalte) mehrfach vertreten sein, da für die MZ-LFS-Substichprobe in Privathaushalten die Kernmerkmale in sieben der neun am Jahr beteiligten Rotationsgruppen für zwei Wochen erhoben wurden. [↪ Grafik 2](#)

Weil das Jahresergebnis als Durchschnitt über die 52 Kalenderwochen angelegt ist, sind alle Befragungen – egal, ob ein oder zwei Befragungszeitpunkte im Jahr vorliegen – gleichwertig. Die normierte Ziehungswahrscheinlichkeit $\pi_{ij,k}$ einer Einheit k , die zum Regierungsbezirk j des Bundeslandes i gehört und keine Gemeinschaftsunterkunft ist, berechnet sich wie folgt:

$$(2) \pi_{ij,k} = f_{ij}^{gesamt} - f_{ij}^{lfs} + \frac{16}{9} f_{ij}^{lfs} = f_{ij}^{gesamt} + \frac{7}{9} f_{ij}^{lfs}$$

Dabei entsprechen f_{ij}^{gesamt} und f_{ij}^{lfs} jeweils dem Auswahlatz der gesamten beziehungsweise der LFS-Stichprobe im Regierungsbezirk j des Bundeslandes i .

Für Gemeinschaftsunterkünfte ergeben sich andere Ziehungswahrscheinlichkeiten. Für die Gemeinschaftsunterkünfte der Substichprobe MZ-LFS wird nur die erste

Grafik 2

Rotationschema der Auswahlbezirke für die Unterstichprobe zur Arbeitsmarktbeteiligung ab 2020

	I. Quartal	II. Quartal	III. Quartal	IV. Quartal
Rotationsgruppe 1	4. Befragung			
Rotationsgruppe 2	3. Befragung	4. Befragung		
Rotationsgruppe 3		3. Befragung	4. Befragung	
Rotationsgruppe 4			3. Befragung	4. Befragung
Rotationsgruppe 5	2. Befragung			3. Befragung
Rotationsgruppe 6	1. Befragung	2. Befragung		
Rotationsgruppe 7		1. Befragung	2. Befragung	
Rotationsgruppe 8			1. Befragung	2. Befragung
Rotationsgruppe 9				1. Befragung

Hellblau signalisiert Rotationsgruppen mit je zwei Befragungen innerhalb eines Jahres. Dunkelblau sind entsprechend Befragungen mit Strukturmerkmalen (sowie Befragungen von Gemeinschaftsunterkünften).

2021 - 0509

und dritte Befragung durchgeführt, außerdem die vierte Befragung, falls sie in das erste Quartal fällt. Hierdurch ist die zeitliche Gleichverteilung bei den Gemeinschaftsunterkünften des MZ-LFS-Teils nicht mehr gegeben: Im ersten Quartal sind drei Rotationsgruppen beteiligt, in den übrigen Quartalen jeweils zwei Rotationsgruppen. Der Anteil der Ziehungswahrscheinlichkeit aus der Arbeitskräfteerhebung LFS ist somit $\frac{12}{9} f_{ij}^{lfs}$ für Gemeinschaftsunterkünfte im ersten Quartal beziehungsweise $\frac{8}{9} f_{ij}^{lfs}$ für Gemeinschaftsunterkünfte in den anderen Quartalen. Die Ziehungswahrscheinlichkeiten der übrigen Teile – MZ-SILC, MZ-IKT und nur Kernprogramm – unterscheiden sich für Gemeinschaftsunterkünfte und Privathaushalte nicht. Die Ziehungswahrscheinlichkeit des übrigen Teils ist also $f_{ij}^{gesamt} - f_{ij}^{lfs}$. Für die Jahreshochrechnung des Kernprogramms ist die normierte Ziehungswahrscheinlichkeit einer Einheit k einer Gemeinschaftsunterkunft, die zum Regierungsbezirk j des Bundeslandes i gehört, somit

$$(3) \quad \pi_{ijk} = (f_{ij}^{gesamt} - f_{ij}^{lfs}) + \frac{12}{9} f_{ij}^{lfs} = f_{ij}^{gesamt} + \frac{3}{9} f_{ij}^{lfs},$$

falls die Berichtswoche der Einheit k im ersten Quartal liegt, und

$$(4) \quad \pi_{ijk} = (f_{ij}^{gesamt} - f_{ij}^{lfs}) + \frac{8}{9} f_{ij}^{lfs} = f_{ij}^{gesamt} - \frac{1}{9} f_{ij}^{lfs}$$

in den sonstigen Fällen.¹⁸

Unterjährige Hochrechnung des Kernprogramms

Für die unterjährige Hochrechnung des Kernprogramms werden die beschriebenen Ziehungswahrscheinlichkeiten $\pi_{ij,k}$ der Jahreshochrechnung entsprechend der berücksichtigten Zeiträume mit einem Faktor multipliziert. Für die Quartalshochrechnung entspricht der Faktor $\frac{1}{4}$, für das Halbjahr $\frac{1}{2}$.

4.2 Substichprobe MZ-LFS

Je Quartal werden vier der neun Rotationsgruppen befragt, das heißt mehr als ein Viertel des Jahresmaterials. Jedoch wird im Quartal jede Person/jeder Auswahlbezirk nur einmal zur Befragung herangezogen. Die zeitlich stabilen Strukturmerkmale werden fest in den Befragungswellen 1 und 3 abgefragt, sowie zusätzlich in der vierten Befragungswelle, sofern die Berichtswoche im ersten Quartal liegt.¹⁹ Auch hier liegt jede Befragungseinheit nur einmal vor, allerdings besteht keine zeitliche Gleichverteilung, da im ersten Quartal drei statt zwei Rotationsgruppen beteiligt sind. Für die Hochrechnung der MZ-LFS-Stichprobe ergeben sich damit folgende Ziehungswahrscheinlichkeiten:

Quartal

$$(5) \quad \pi_{ijk}^{lfsq} = \frac{4}{9} f_{ij}^{lfs},$$

Jahr

$$(6) \quad \pi_{ijk}^{lfsj} = \frac{12}{9} f_{ij}^{lfs}, \text{ falls die Berichtswoche} \\ \text{der Einheit } k \text{ im ersten Quartal liegt,} \\ = \frac{8}{9} f_{ij}^{lfs} \text{ in den sonstigen Fällen.}$$

4.3 Substichproben MZ-SILC und MZ-IKT

Die Ziehungswahrscheinlichkeiten entsprechen für die Substichproben MZ-SILC und MZ-IKT dem Auswahlatz je Regierungsbezirk der jeweiligen Stichprobe. Entsprechend dem Vorgehen in der Kernhochrechnung wird einem weggezogenen MZ-SILC-Haushalt die Ziehungswahrscheinlichkeit des Auswahlbezirks zugeordnet, zu dem er im Vorjahr gehörte. Es erfolgt somit keine Modifikation ihrer ursprünglichen Eingangswahrscheinlichkeiten, auch gibt es bei einem Ausfall keine Kompensation (Antwortwahrscheinlichkeit = 1). Aufgrund der geringen Fallzahlen wird angenommen, dass eine etwaige Verzerrung zu vernachlässigen ist.

¹⁸ Anmerkung: Im Durchschnitt über die vier Quartale ergibt sich für die Gemeinschaftsunterkünfte gerade der vorgegebene Jahresauswahlatz.

¹⁹ Letzteres wurde hinzugenommen, um den gesetzlichen Anforderungen zu entsprechen, wonach jeder Haushalt, der Teil der MZ-LFS-Stichprobe ist, in jedem Kalenderjahr seiner Teilnahme an der Erhebung auch Angaben zu den Strukturmerkmalen geben muss.

5

Antwortwahrscheinlichkeit – Kompensation von Antwortausfällen

Antwortausfälle können nicht als zufällig angenommen werden. Etwaigen Verzerrungen wird daher durch die Berechnung von Antwortwahrscheinlichkeiten begegnet.

Um die Antwortwahrscheinlichkeit $\hat{\theta}_k$ eines Haushalts k zu bestimmen, werden Hilfsinformationen aus der Bruttostichprobe identifiziert, denen ein erwartbarer Unterschied im Antwortverhalten unterstellt werden kann. Geplant ist, den antwortenden Haushalten mittels eines Logitschätzers eine Antwortwahrscheinlichkeit zuzuweisen (Ekholm/Laakson, 1991). Die Teilnahme wird als binäre abhängige Variable Y der Bruttostichprobe angenommen und die Antwortwahrscheinlichkeit durch ein binäres, logistisches Regressionsmodell zwischen Teilnahme und Hilfsvariablen geschätzt (Körner und andere, 2006):

Teilnahme der Einheit k :

$$(7) \quad Y_k = 1 \quad \text{Prob}(Y_k = 1) = \rho(\beta^t; x_k)$$

Nichtteilnahme der Einheit k :

$$(8) \quad Y_k = 0 \quad \text{Prob}(Y_k = 0) = 1 - \rho(\beta^t; x_k)$$

Antwortwahrscheinlichkeit der Einheit k :

$$(9) \quad \hat{\theta}_k = \text{Prob}(Y_k = 1) = \rho(\hat{\beta}^t; x_k) = \frac{e^{\hat{\beta}^t x_k}}{1 + e^{\hat{\beta}^t x_k}}$$

Dabei entspricht x_k den Ausprägungen der Hilfsvariablen für Einheit k und $\hat{\beta}$ den geschätzten Regressionskoeffizienten.

Neu für den Mikrozensus ab 2020 ist, dass nicht nur finale Ausfälle, sondern auch verzögerte Dateneingänge zu berücksichtigen sind. Aufgrund der festen Berichtswoche sind Befragungen einem festen Berichtszeitraum zugeordnet und können nicht wie bisher dem folgenden Berichtszeitraum zugeschlagen werden. Insbesondere für die frühe Hochrechnung der Erstergebnisse ist dies relevant. Um den zeitlichen Verzug zu berücksichtigen, kann auf Basis der Daten aus vorherigen Erhebungsjahren die Antwortwahrscheinlichkeit bestimmt werden. Hierzu können Materialstände des Erstergebnisses mit den Ständen zum Endergebnis abgeglichen werden.

Unter Verwendung des logistischen Regressionsmodells werden Antwortwahrscheinlichkeiten je Merkmalskombination bestimmt. Vorteilhaft hierbei ist, dass für alle berücksichtigten Datensätze der volle Merkmalskranz vorliegt und somit eine große Auswahl an erklärenden Merkmalen zur Verfügung steht. Berechnete Antwortwahrscheinlichkeiten berücksichtigen allerdings lediglich die Verzögerungen, nicht die Totalausfälle. Aus diesem Grund wird eine weitere Antwortwahrscheinlichkeit, welche die aktuelle Bruttostichprobe berücksichtigt, berechnet. Dies geschieht unter Verwendung von Informationen, die für die gesamte Bruttostichprobe vorliegen. Auch für die Berechnung dieser Wahrscheinlichkeiten ist ein logistisches Regressionsmodell vorgesehen. Der Merkmalskranz hierfür ist begrenzt. Vollständig und qualitativ gut liegen alle Merkmale zur Verwendung in der Stichprobenziehung vor. Allerdings besteht für diese Merkmale eine geringere Korrelation zum Antwortverhalten. Aus Beobachtungen der Interviewenden liegen Daten für weitere soziodemografische Merkmale vor. Umfang und Qualität zu den frühen Hochrechnungszeitpunkten müssen allerdings künftig in einem funktionierenden System untersucht werden.

Die Antwortwahrscheinlichkeit $\hat{\theta}_k$ für eine Einheit k wird ermittelt durch

$$(10) \quad \hat{\theta}_k = \alpha \hat{\theta}_{k1} + (1 - \alpha) \hat{\theta}_{k2}$$

mit

$\hat{\theta}_{k1}$: Antwortwahrscheinlichkeit der Einheit k berechnet auf vorherigem Datenmaterial,

$\hat{\theta}_{k2}$: Antwortwahrscheinlichkeit der Einheit k berechnet aus der aktuellen Stichprobe und

$\alpha \in [0,1]$. Durch die Wahl des Parameters α kann der Fokus stärker auf die aktuelle Stichprobe oder auf das Antwortverhalten der Vorjahre gelegt werden.

Sonderfall Substichprobe MZ-IKT

Für die freiwillige Substichprobe MZ-IKT ergibt sich die Antwortwahrscheinlichkeit als Produkt aus der Antwortwahrscheinlichkeit des Mikrozensus und Teilnahme-wahrscheinlichkeit, das heißt der Wahrscheinlichkeit, an der MZ-IKT-Erhebung teilzunehmen. Auch die Teilnahmewahrscheinlichkeit wird über das logistische Regressionsverfahren bestimmt. Hierfür werden die an der IKT teilnehmenden Haushalte den restlichen Haushalten, die

auskunftspflichtig zum Kernteil geantwortet haben und Teil der IKT-Bruttostichprobe sind, gegenübergestellt.

Für die gesonderte Hochrechnung der Personen wird analog eine Teilnahmewahrscheinlichkeit auf Personenebene erstellt. Die Antwortwahrscheinlichkeit des Haushalts wird für alle Personen übernommen.

6

Hochrechnungsfaktoren

Für jeden Erhebungsteil findet eine getrennte Hochrechnung statt, die die jeweils zugrunde liegende Stichprobe berücksichtigt. Die wichtigsten Hochrechnungsfaktoren des Mikrozensus einschließlich ihrer entsprechenden Bezeichnung bei europäischen Datenlieferungen zeigt [Übersicht 1](#).

Für Auswertungen von Merkmalskombinationen muss immer auf die gemeinsame Stichprobe zurückgegriffen werden. Sollen beispielsweise Kernmerkmale zusammen mit MZ-SILC-Merkmalen ausgewertet werden, ist hierfür die MZ-SILC-Stichprobe mit dem zugehörigen Hochrechnungsfaktor zu verwenden.

Übersicht 1

Hochrechnungsfaktoren im Mikrozensus

Faktoren	Stichprobe	Nationale Bezeichnung	Bezeichnung Eurostat
Monatliche Arbeitsmark-Schnellmeldung	MZ-Kern	HR000MM	–
Nationales Quartalsergebnis	MZ-Kern	HR000QQ	–
Nationales Jahresergebnis	MZ-Kern	HR000JJ	–
Tief regionalisierte Auswertung des Jahresergebnisses	MZ-Kern	HR000RG	–
MZ-LFS-Quartalsergebnis	MZ-LFS	HR100QQ	COEFFQ
MZ-LFS-Quartalsdurchschnitt zur Jahresauswertung	MZ-LFS	HR100JQ	COEFFQ/4
MZ-LFS-Jahresergebnis	MZ-LFS-Struktur	HR100JJ	COEFFY = COEFF2Y = COEFFHH
Ad-Hoc-Modul LFS	MZ-LFS-Ad-Hoc-Modul	HR100MO	COEFFMOD
MZ-SILC-Querschnitt	MZ-SILC	HR200JJ	DB090 = RB050 = PB040
MZ-SILC-Faktor für Kinder bis 12 Jahre	MZ-SILC	HR200KI	RL070
2-, 3-, 4-jähriger Längsschnitt	MZ-SILC	HR200L2, HR200L3, HR200L4	RB062, RB063, RB064
MZ-IKT Personen	MZ-IKT	HR300PN	IND_WGHT
MZ-IKT Haushalte	MZ-IKT	HR300HH	HH_WGHT

7

Bereitstellung der Eckwerte

Für die Substichproben steht zum Zeitpunkt der Hochrechnung der bis dahin jeweils erhobene Teil des Kernprogramms zur Verfügung. Hieraus können Aggregate für die zweite Phase der gebundenen Hochrechnung generiert werden. Für die erste Phase der gebundenen Hochrechnung werden Informationen über die Größe und Struktur der Gesamtbevölkerung verwendet, sogenannte externe Eckwerte. Die Datenquelle hierfür ist die Fortschreibung des Bevölkerungsbestands (Bevölkerungsfortschreibung). Die jeweils aktuellen Bevölkerungszahlen werden durch die Fortschreibung der Ergebnisse des jeweiligen letzten Zensus ermittelt und sind drei bis vier Monate nach dem Ablauf des Berichtsmonats beziehungsweise sechs Monate nach dem Berichtsjahr verfügbar. Zum Zeitpunkt der ersten Hochrechnung liegen noch keine aktuellen Werte vor, daher werden die Bevölkerungseckzahlen quartalsweise für die drei zurückliegenden und die vier kommenden Monate vorausgeschätzt. Da die Mikrozensuserhebung gleichmäßig über alle Kalenderwochen eines Berichtszeitraums durchgeführt wird, stellen Quartalseckwerte den Mittelwert aus den Monatswerten, Jahreseckwerte den Durchschnitt aus den Quartalen dar. Monatliche Eckwerte werden als Durchschnitt aus dem Vormonat und dem Berichtsmonat gebildet.

Eine weitere Neuerung ist, dass die Eckwerte zur Staatsangehörigkeit künftig aus der Bevölkerungsfortschreibung und nicht wie bisher aus den Zahlen des Ausländerzentralregisters generiert werden.¹⁰ Seit dem Zensus 2011 erfolgt die Fortschreibung nach einzelnen Staatsangehörigkeiten und bietet somit eine konsistente Datenquelle für alle Eckwerte zur Hochrechnung der Mikrozensusstichprobe.

Zur Vorausschätzung ab 2020 wurde das bisher verwendete Verfahren weiterentwickelt: Zunächst wird die „Bevölkerung insgesamt“ unter Verwendung des sogenannten Holt-Winters-Prognoseverfahrens je Bundesland vorausgeschätzt. Basis hierfür sind Zeitreihen über die monatliche Entwicklung der Geburten, Sterbefälle und Zu- und Fortzüge über die Grenzen des Bundeslands, die mit der letzten vorhandenen Bevölkerungszahl verrechnet werden. Im nächsten Schritt werden hieraus die untergliederten Eckwerte auf Landesebene abgeleitet. Dafür werden zunächst die Anteile dieser Eckwerte an der Gesamtbevölkerung des Bundeslands in der Vergangenheit nach Monaten berechnet. Die monatlichen Zeitreihen werden anschließend extrapoliert und mit der projizierten Bevölkerung multipliziert. Die verwendeten Zeitreihen basieren auf der monatlichen Entwicklung ab Januar 2016. Lediglich für die Prognose der Anteile nach den drei nichtdeutschen Staatsangehörigkeitsgruppen¹¹ haben die Testrechnungen ergeben, dass ein kür-

zerer Zeitraum von 24 Werten eine bessere Anpassung liefert.

Im letzten Schritt werden die Eckwerte nach Kreisen berechnet. Hierfür wird das bisherige Vorgehen übernommen. Die vorausgeschätzten großen Aggregate auf Landesebene werden nach Kreisen verteilt. Da die Entwicklung der einzelnen Eckwerte nach Kreisen stark volatil ist, wird auf eine Zeitreihenextrapolation verzichtet. Die relativen Verteilungen nach Kreis-Eckwerten werden deshalb im Prognosezeitraum konstant gehalten.

Zum Zeitpunkt der zweiten Hochrechnung für Endergebnisse liegen bereits endgültige Zahlen der Bevölkerungsfortschreibung vor. Diese werden in gleicher Untergliederung verwendet.

Geschlechtsangabe divers/kein Eintrag

Da aus dem Zensus 2011 keine Angaben zu den Merkmalsausprägungen divers/ohne Angabe vorliegen, stehen aktuell für die Hochrechnung keine Eckwerte hierzu aus der Bevölkerungsfortschreibung zur Verfügung. Derzeit werden die Personen ohne Angabe zum Geschlecht in der Bevölkerungsfortschreibung und im Mikrozensus noch zufällig den Ausprägungen männlich oder weiblich zugeordnet.

Hilfsvariablen für die gebundene Hochrechnung

↳ **Übersicht 2** stellt den Anpassungsrahmen für die gebundene Hochrechnung des Kernmaterials an die Bevölkerung dar. Grundlage zur Bestimmung des Anpas-

10 Es gab aufgrund unterschiedlicher Berechnungsmethodik regelmäßig Abweichungen zwischen den Zahlen der Bevölkerungsfortschreibung und denen des Ausländerzentralregisters.

11 Folgende Staatsangehörigkeitsgruppen werden gesondert berücksichtigt: türkisch, EU-25, nicht EU-25.

Übersicht 2

Anpassungsrahmen Kernprogramm

Merkmal	Ausprägung	Regionale Ebene	Berichtszeitraum
Alter (zum 31.12. des Jahres) *Geschlecht	(<15, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+) *(männlich, weiblich)	NUTS- 1	Jahr/ Quartal
Alter (zum 31.12. des Jahres) *Geschlecht *Staatsangehörigkeit	(15-24, 25-54, 55-74) *(männlich, weiblich) *(deutsch, nichtdeutsch)		Jahr/ Quartal
Staatsangehörigkeit *Geschlecht	(deutsch, türkisch, EU, Rest) *(männlich, weiblich)		Jahr/ Quartal
Monatsanpassung	Bevölkerung je Monat		Jahr/ Quartal
Staatsangehörigkeit *Geschlecht	(deutsch, nichtdeutsch) *(männlich, weiblich)	NUTS-2	Jahr/ Quartal
Geschlecht	(männlich, weiblich)	Regionale Anpassungsschicht	Quartal
Staatsangehörigkeit *Geschlecht	(deutsch, nichtdeutsch) *(männlich, weiblich)		Jahr
Geschlecht	(männlich, weiblich)	Regionale Schichtuntergruppe	Jahr

Übersicht 3

Anpassungsrahmen des MZ-LFS für die Quartalshochrechnung

Haushalt/Personen	Merkmal	Ausprägung	Regionale Ebene
Personen „Bevölkerung in Privathaushalten am Hauptwohnsitz“	Monatsanpassung	Bevölkerung je Monat	NUTS-0
	Staatsangehörigkeit *Geschlecht	(deutsch, türkisch, EU, Rest) *(männlich, weiblich)	Ost/ West
	Erwerbsstatus*Alter *Geschlecht	(erwerbstätig, erwerbslos, Nichterwerbsperson) *(15-24, 25-54, 55-74) *(männlich, weiblich)	NUTS-1
	Alter*Geschlecht	(15-24, 25-34, 35-44, 45-54, 55-64, 65-74) *(männlich, weiblich)	NUTS-2
	Bevölkerung insgesamt		Regionale Anpassungsschicht

sungsrahmens waren fachliche Analysen sowie die erwarteten Besetzungszahlen der Anpassungsklassen. Die Untergliederung mit einer gesonderten Ausweisung türkischer Staatsangehöriger wurde beibehalten (Afen-takis/Bihler, 2005). Grund dafür ist, dass auch auf Basis aktueller Daten die Stichprobenverteilung nach freier Hochrechnung für Personen mit türkischer Staatsangehörigkeit weniger von den Eckwerten abweicht als bei der übrigen nicht deutschen Bevölkerung.

Beibehalten wurde auch die automatische Reduktion der Hochrechnungsterme, wenn eine Anpassungszelle weniger als zehn befragte Personen enthält. Ziel ist es, eine instabile Schätzung durch zu geringe Besetzungszahlen in den Anpassungsklassen zu vermeiden. Hierbei wird eine mögliche fachliche gegenüber einer geografischen Reduktion bevorzugt, sodass möglichst alle in

der Anpassung vorgesehenen regionalen Ebenen verwendet werden.

Zur Hochrechnung der Substichproben kann auf die Merkmale des Kernprogramms zugegriffen werden. Hierbei sind die ab 2021 geltenden Vorgaben Eurostats und die gewünschte fachliche Konsistenz zwischen Stichproben zu berücksichtigen. Die Lieferungen an Eurostat umfassen lediglich Personen in privaten Haushalten am Hauptwohnsitz (Hochgürtel/Weinmann, 2020), daher bezieht sich die Hochrechnung für Substichproben auf diesen Personenkreis.

Insbesondere die neue Durchführungsverordnung des EU-LFS enthält deutlich stärkere Anforderungen als bisher. Künftig soll die Verteilung der Referenzpopulation nach Geschlecht, Altersgruppen und NUTS-2-Region

Übersicht 4

Anpassungsrahmen des MZ-LFS für die Jahreshochrechnung

Haushalt/Personen	Merkmal	Ausprägung	Regionale Ebene
Haushalt „Privathaushalte am Hauptwohnsitz“	Anzahl Haushalte		NUTS-0
	Haushaltsgröße	(1, 2, 3, 4, 5 und mehr)	
Personen „Bevölkerung in Privathaushalten am Hauptwohnsitz“	Erwerbsstatus*Alter*Geschlecht	(erwerbstätig, erwerbslos, Nicht- erwerbsperson) *(15-24, 25-34, 35-44, 45-54, 55-64, > 64 nM))*(männlich, weiblich)	NUTS-0
	Monatsanpassung	Bevölkerung je Monat	
	Staatsangehörigkeit* Geschlecht	(deutsch, türkisch, EU, Rest) *(männlich, weiblich)	Ost/West
	Erwerbsstatus*Alter*Geschlecht	(erwerbstätig, erwerbslos, Nichterwerbsperson) *(15-24, 25- 54, 55-74)*(männlich, weiblich)	NUTS-1
	Alter*Geschlecht	(15-24, 25-34, 35-44, 45-54, 55-64, 65-74) *(männlich, weiblich)	NUTS-2
	Bevölkerung insgesamt		Regionale Anpassungsschicht

Übersicht 5

Der Anpassungsrahmen für MZ-SILC

Haushalt/Personen	Merkmal	Ausprägung	Regionale Ebene
Haushalt „Privathaushalte am Hauptwohnsitz“	Haushaltsgröße	(1, 2, 3, 4, 5+)	NUTS-0
	Haushaltstyp	(Einpersonenhaushalt, Paare ohne Kind, Alleinerziehende, Paare mit Kindern, Sonstige)	
	Haushaltseinkommen	(<1 500, 1 500-2 000, 2 000-3 000, 3 000-4 000, >4 000)	
	Haushaltstyp2	(ohne Kind/mit Kindern)	NUTS-1
	Anzahl		NUTS-2
Personen „Bevölkerung in Privathaushalten am Hauptwohnsitz“	Alter*Geschlecht	(25-49, 50-64, 65+) *Geschlecht	NUTS-0
	Staatsangehörigkeit	(deutsch, EU, nicht EU)	
	Alter	(0-15, 16-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+)	
	ISCED	(niedrig, mittel, hoch)	
	Hauptstatus	(arbeitslos, angestellt, selbstständig, Schüler/-in/Studierende, im Ruhestand; nicht erwerbstätig)	NUTS-1
	Geschlecht	(männlich, weiblich)	
	Alter	(0-24, 25-64, 65+)	
	Anzahl		

berücksichtigt werden. Eine Kombination der Merkmale ist gewünscht, aber nicht zwingend notwendig. Zusätzlich soll bei der Gewichtung der Strukturvariablen (Jahresergebnis) die Anzahl der Haushalte sowie die Verteilung nach Haushaltsgröße einbezogen werden. Zudem ist Konsistenz zum Quartalsdurchschnitt hinsichtlich des Erwerbsstatus aufgedgliedert nach Alter und Geschlecht sicherzustellen. Somit müssen bei Hochrechnung der Strukturmerkmale die Eckwerte der LFS-Quartalsstichprobe berücksichtigt werden. Aufgrund der monatlichen Schnellmeldung zum Arbeitsmarkt liegen einige der wichtigsten LFS-Merkmale im gesamten Kernprogramm vor, an Eurostat wird jedoch nur die Substichprobe geliefert. Europäische Konsistenzanforderungen zwischen Stichproben werden gegenüber nationalen Konsistenz-

anforderungen zum Kernjahresergebnis priorisiert. Daher wird die Stichprobe der LFS-Strukturmerkmale an den Quartalsdurchschnitt der Kernstichprobe hochgerechnet und nicht an das Jahresmaterial. Gewählte Anpassungsrahmen für die Quartals- und Jahreshochrechnung enthalten die [Übersichten 3 und 4](#).

[Übersicht 5](#) zeigt den geplanten Hochrechnungsrahmen für MZ-SILC. Es wird wie bisher an das Haushaltsnettoeinkommen angepasst. Die Klassenzusammenfassung wurde entsprechend der geschätzten Quintilsverteilung im Mikrozensus vor 2020 vorgenommen. Aufgrund der Erhöhung des Stichprobenumfangs im System ab 2020 ist ein gesonderter Faktor für Personen ab 16 Jahren, wie er bisher vorlag, nicht mehr notwendig.

Übersicht 6

Anpassungsrahmen MZ-IKT für Haushalte

Haushalt/Personen	Merkmal	Ausprägung	Regionale Ebene
Haushalt „Privathaushalte am Hauptwohnsitz“ mindestens eine Person zwischen 16 und 74 Jahren	Internetzugang	(ja, nein)	NUTS-0
	Haushaltsnettoeinkommen	(<1 500, 1 500-2 500, 2 500-4 000, >4 000)	
	Haushaltstyp	(Einpersonenhaushalt, Haushalt ohne Kind, Haushalt mit Kindern)	
	Anzahl Haushalte		NUTS-1 Zusammenfassung

Übersicht 7

Anpassungsrahmen MZ-IKT für Personen

Haushalt/Personen	Merkmal	Ausprägung	Regionale Ebene
Personen „Bevölkerung in Privathaushalten am Hauptwohnsitz“ zwischen 16 und 74 Jahren	Internetnutzung	(ja, nein)	NUTS-0
	Alter *Geschlecht	(16-24, 25-34, 35-44, 45-54, 55-64, 65-74) *(männlich, weiblich)	
	ISCED	(niedrig, mittel, hoch)	
	Soziale Stellung	(erwerbstätig, nicht erwerbstätig)	
	Gesamtbevölkerung		NUTS-1-Zusammenfassung

Für MZ-IKT werden auch weiterhin zwei Hochrechnungsfaktoren bereitgestellt, um Personen- und Haushaltsauswertungen abzudecken. Die Anpassungsrahmen sind abgebildet in den [Übersichten 6 und 7](#).

8

Erhebungseinschränkungen im Jahr 2020

Im Jahr 2020 war die Durchführung der Mikrozensus-erhebung eingeschränkt. Dies lag zum einen an der Umstellung des IT-Systems und wurde zum anderen durch die Corona-Pandemie zusätzlich verschärft, da bisher überwiegend persönlich vor Ort stattfindende Befragungen nicht mehr durchgeführt werden konnten. Die durchschnittliche Ausfallquote auf Bundesebene lag 2020 hierdurch bedingt bei etwa 38% und damit deutlich höher als in vorherigen Jahren (nähere Informationen hierzu unter www.destatis.de/mikrozensus2020).

Die beobachteten Antwortausfälle waren fachlich, regional und zeitlich über das Jahr 2020 sehr unterschiedlich verteilt. Zusätzlich war aufgrund der Erhebungseinschränkungen eine vollständige Erfassung der Bruttostichprobe mit Kompensationsmerkmalen auf Haushaltsebene nicht möglich. Um dennoch etwaigen Verzerrungen entgegenzuwirken, wurde die Annahme getroffen, dass für zeitstabile Merkmale die Bruttostichprobenverteilung 2020 der Stichprobenverteilung 2019 entspricht. Unter Berücksichtigung der neuen Auswahlsätze wurden hierdurch synthetische Bruttofallzahlen für 2020 bestimmt. Die Antwortwahrscheinlichkeiten für den Mikrozensus 2020 wurden unter Verwendung des bisherigen Verfahrens (Anpassung der Netto- an die Bruttofallzahlen [Afentakis/Bihler, 2005]) je Stich-

probe auf Bundeslandebene berechnet. Das in Kapitel 5 beschriebene Verfahren wurde bisher nicht angewandt.

Weiteren Anpassungsbedarf gab es in den Hochrechnungsmodellen. Es konnten weniger Terme als in den beschriebenen Modellen (Kapitel 7) aufgenommen werden. Eine Anpassung für die Kernstichprobe wurde nur bis zur Ebene der Anpassungsschichten durchgeführt, auf eine monatliche Anpassung wurde verzichtet. Auch für die Substichproben waren Reduktionen der Modelle notwendig. Die Qualität der Jahresergebnisse aus dem Mikrozensus 2020 ist auf Bundesebene weiterhin gewährleistet, jedoch sind die Möglichkeiten zur Auswertung in fachlicher und regionaler Tiefe für dieses außergewöhnlich schwierige Erhebungsjahr eingeschränkt.

9

Fazit

Mit der Umstellung auf den Mikrozensus 2020 ist es möglich, eine Vielzahl von Anforderungen an die amtlichen Haushaltsstatistiken aus einer gemeinsamen Quelle konsistent und flexibel zu bedienen. Da die Komplexität des Systems stark gestiegen ist, wurde die Methodik soweit möglich auf moderne und den Anpassungen gerechte Weise aktualisiert und erweitert. Die wichtigsten methodischen Änderungen wurden in diesem Beitrag vorgestellt. Aufgrund des Umfangs und der Komplexität konnte jedoch nur eine Auswahl der erzeugten Hochrechnungsfaktoren des Mikrozensus beleuchtet werden. Darüber hinaus werden derzeit weitere Entwicklungen bearbeitet und analysiert. 

LITERATURVERZEICHNIS

Afentakis, Anja/Bihler, Wolf. [*Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005*](#). In: *Wirtschaft und Statistik*. Ausgabe 10/2005, Seite 1039 ff.

Bihler, Wolf/Zimmermann, Daniel. [*Die neue Mikrozensusstichprobe ab 2016*](#). In: *WISTA Wirtschaft und Statistik*. Ausgabe 6/2016, Seite 20 ff.

Deville, Jean-Claude/Särndal, Carl-Erik/Sautory, Olivier. *Generalized Raking Procedures in Survey Sampling*. In: *Journal of the American Statistical Association*. Jahrgang 88. Ausgabe 09/1993, Nr. 423, Seite 1013 ff.
[DOI 10.1080/01621459.1993.10476369](https://doi.org/10.1080/01621459.1993.10476369)

Ekholm, Anders/Laaksonen, Seppo. *Weighting via Response Modeling in the Finnish Household Budget Survey*. In: *Journal of Official Statistics*. Jahrgang 1991. Ausgabe 7, Nr. 3, Seite 325 ff.

Hochgürtel, Tim/Weinmann, Julia. [*Haushalte in der Berichterstattung des Mikrozensus ab 2020*](#). In: *WISTA Wirtschaft und Statistik*. Ausgabe 3/2020, Seite 89 ff.

Hundenborn, Janina/Enderer, Jörg. [*Die Neuregelung des Mikrozensus ab 2020*](#). In: *WISTA Wirtschaft und Statistik*. Ausgabe 6/2019, Seite 9 ff.

Körner, Thomas/Nimmergut, Anja/Nökel, Jens/Rohloff, Sandra. [*Die Dauerstichprobe befragungsbereiter Haushalte*](#). In: *Wirtschaft und Statistik*. Ausgabe 5/2006, Seite 451 ff.

Statistisches Bundesamt. *Zu den Auswirkungen der methodischen Neugestaltung sowie der Corona-Krise auf die Ergebnisse des Mikrozensus 2020*. 2021. [Zugriff am 8. November 2021]. Verfügbar unter: www.destatis.de

RECHTSGRUNDLAGEN

Verordnung (EU) 2019/17001700 des Europäischen Parlaments und des Rates vom 10. Oktober 2019 zur Schaffung eines gemeinsamen Rahmens für europäische Statistiken über Personen und Haushalte auf der Grundlage von Einzeldaten aus Stichprobenerhebungen, zur Änderung der Verordnungen (EG) Nr. 808/2004, (EG) Nr. 452/2008 und (EG) Nr. 1338/2008 des Europäischen Parlaments und des Rates sowie zur Aufhebung der Verordnung (EG) Nr. 1177/2003 des Europäischen Parlaments und des Rates und der Verordnung (EG) Nr. 577/98 des Rates (Amtsblatt der EU Nr. L 261, Seite I/1).

Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und die Arbeitsmarktbeteiligung sowie die Wohnsituation der Haushalte (Mikrozensusgesetz – MZG) vom 7. Dezember 2016 (BGBl. I Seite 2826), das durch Artikel 178 der Verordnung vom 19. Juni 2020 (BGBl. I Seite 1328) geändert worden ist.

DER INTERAKTIVE GEHALTS- VERGLEICH – EINE PROFILBASIERTE SCHÄTZUNG VON VERDIENSTEN

Anja Überschaer

↳ **Schlüsselwörter:** Gehaltsrechner – Verdienste nach Berufen – Regression – Verdiensterhebung – Lohnschätzung

ZUSAMMENFASSUNG

Das Statistische Bundesamt entwickelt sich zu einem digitalen, nutzerorientierten und innovativen Informationsdienstleister weiter. Dazu gehören auch neue Produkte, die Informationen nutzungsfreundlich bereitstellen – wie der Gehaltsvergleich. Diese interaktive Webanwendung ermöglicht es den Nutzerinnen und Nutzern, sich auf Basis von individuellen Angaben ein Gehalt schätzen zu lassen. Der Gehaltsrechner ist leicht zu bedienen und legt den Fokus auf individuelle Profile. Dadurch ermöglicht die Anwendung die zielgruppenorientierte und bedarfsgerechte Aufbereitung von Verdienstdaten und steigert somit deren Nutzen.

Der Aufsatz beschreibt den konzeptionellen Aufbau des Gehaltsrechners sowie die Methodik der Schätzung des individuellen Bruttomonatsverdienstes.

↳ **Keywords:** salary calculator – earnings by occupation – regression – earnings survey – wage estimate

ABSTRACT

The Federal Statistical Office is developing into a digital, user-oriented and innovative information provider. This also includes creating new products which provide information in a user-friendly form, for example the salary calculator for salary comparison. That interactive web application allows users to have a salary estimated on the basis of individual information. The salary calculator is easy to use and focuses on individual profiles. It enables earnings data to be prepared in a target-group-oriented and needs-based manner, thus increasing its usefulness.

This paper describes the conceptual design of the salary calculator and the methodology used to estimate individual gross monthly earnings.



Dr. Anja Überschaer

ist im Referat „Verdienste, Umschulung“ des Statistischen Bundesamtes tätig. Sie hat im Rahmen eines Doppelmasterprogramms Betriebswirtschaftslehre und Volkswirtschaftslehre studiert. Ihr Tätigkeitsbereich umfasst methodische Fragestellungen rund um die neue Verdienstatistik und sie hat die interaktive Anwendung „Gehaltsvergleich“ entwickelt.

1

Einleitung

Viele Personen interessieren sich für Angaben zu durchschnittlichen Verdiensten. Dabei geht es oft um die Einordnung und Bewertung des eigenen Verdienstes und um Jobbewerbungen, wobei der Beruf, der Ausbildungsabschluss, die Branche und weitere Aspekte eine Rolle spielen. Um dieser Nachfrage gerecht zu werden, ist im Statistischen Bundesamt der [interaktive Gehaltsvergleich](#)¹ entwickelt worden. Er richtet sich vorwiegend an Privatpersonen, die beispielsweise aufgrund von bevorstehenden Gehaltsverhandlungen, Bewerbungsgesprächen oder rein aus Interesse eine spezifische Bruttomonatslohnschätzung erhalten möchten.

Die Veröffentlichung durchschnittlicher Verdienste in den Fachveröffentlichungen und in der Datenbank GENESIS-Online des Statistischen Bundesamtes sind für Nutzerinnen und Nutzer, die sehr spezifische und individuelle Informationen benötigen, nicht optimal. So stehen unter anderem Geheimhaltungsvorschriften einer spezifischen Nutzung entgegen, da konkrete Angaben zu Verdiensten nach Berufen, zusätzlich gegliedert nach Bundesländern, Wirtschaftszweigen oder Ausbildungsabschlüssen, nicht veröffentlicht werden dürfen. Zudem wäre eine solche spezifische Aufschlüsselung der Verdienste in Tabellenform nicht nutzungsfreundlich darstellbar. Die Methodik der Regression ermöglicht, dass der Gehaltsrechner Verdienstinformationen, die den Bedürfnissen der Zielgruppe entsprechen, individuell und profilbasiert bereitstellt. Der interaktive Gehaltsvergleich baut durch seine leichte Zugänglichkeit potenzielle Hemmschwellen ab und überzeugt durch die spielerische Aufbereitung der Informationen. Das interaktive Onlinetool ist durch diese Kombination an Eigenschaften – individuelle Informationen, leichte Zugänglichkeit und spielerische Handhabung – auch für Laien attraktiv.

Beim interaktiven Gehaltsvergleich sind verschiedene gehaltsbestimmende Merkmale vorgegeben, zum Beispiel Beruf, Ausbildungsabschluss oder Branche, anhand derer sich der Bruttomonatsverdienst schätzen lässt. Dazu gibt eine Person die für sie passende Aus-

prägung dieser Merkmale an und erhält eine individuelle Schätzung des Bruttomonatsverdienstes. Der Gehaltsvergleich ermöglicht somit den Nutzerinnen und Nutzern einen schnellen und unkomplizierten Überblick über den (geschätzten) Verdienst, unter der Berücksichtigung von persönlichen und arbeitsplatzspezifischen Merkmalen. Dabei handelt es sich um regressionsbasierte Schätzungen und nicht um Mittelwerte aus der Erhebung an sich, wie sie die Fachveröffentlichungen enthalten. Mit dem Gehaltsvergleich ist es dabei auch möglich, Veränderungen im (geschätzten) Bruttomonatslohn zu beobachten, die entstehen, wenn einzelne verdienstbestimmende Merkmalsausprägungen modifiziert werden.

Der Artikel beschreibt zunächst die Datengrundlage des interaktiven Gehaltsvergleichs (Kapitel 2) und wie die Aufbereitung der Daten erfolgt (Kapitel 3). Kapitel 4 stellt das entwickelte Regressionsmodell vor, während Kapitel 5 die Variablen der Regression ausführlich erläutert. Die Qualität des Modells wird in Kapitel 6 detailliert untersucht. Der Beitrag schließt mit einem Fazit und einem Ausblick auf eine mögliche jährliche Aktualisierung des Gehaltsrechners mit den jeweiligen Daten des Monats April der neuen digitalen Verdiensterhebung.

2

Datengrundlage

Der interaktive Gehaltsvergleich basiert zunächst auf den Daten der Verdienststrukturerhebung, die alle vier Jahre Daten zu den Verdiensten erfasst hat, zuletzt für das Berichtsjahr 2018. An dieser Erhebung nahmen etwa 60 000 repräsentativ ausgewählte Betriebe teil und lieferten unter anderem Angaben zum Bruttomonatslohn von etwa 1 Million abhängig Beschäftigten.²

Die Verdienstdaten dieser Erhebung sind untergliedert nach Wirtschaftszweigen und Regionen sowie persönlichen Angaben über die Beschäftigten, wie Geschlecht, Geburtsjahr, die Dauer der Unternehmenszugehörigkeit, Beruf und Ausbildungsabschluss. Zudem wurden Merkmale zum Beschäftigungsverhältnis erhoben, beispielsweise Angaben zu Tarifvertrag, Art der Beschäftigung (befristet oder unbefristet) und der Umfang des

1 Eine Beschreibung der Anwendung und ihrer Methodik befindet sich unter www.destatis.de.

2 Zusätzliche Informationen zur Verdienststrukturerhebung 2018 enthält der [Qualitätsbericht](#) (Statistisches Bundesamt, 2018).

Urlaubsanspruchs. Die Verdienststrukturerhebung ermöglicht damit Aussagen über die Verteilung der Verdienste sowie über den Einfluss wichtiger, die individuelle Verdiensthöhe bestimmender Faktoren.

Der Gehaltsrechner nutzt Angaben zu den einzelnen Beschäftigten, zum Beispiel den Verdienst. Diese Angaben stammen aus der Lohnabrechnung der Betriebe, dadurch ist eine hohe Datenqualität gewährleistet. Die auskunftgebenden Betriebe sind gesetzlich verpflichtet, vollständige und korrekte Angaben zu machen. Zudem erfolgt eine Plausibilisierung der Daten durch die Statistischen Ämter der Länder. Auch dies sichert eine sehr genaue und zuverlässige Datenbasis.

Künftig wird der Gehaltsvergleich auf Basis der neuen monatlichen Verdiensterhebung ab 2022 jährlich aktualisiert. Bei der Konzeption des Regressionsmodells wurde daher darauf geachtet, nur Merkmale zu verwenden, die auch in den Daten der neuen Verdiensterhebung vorliegen werden. Dies soll einen möglichst reibungslosen Übergang beim Wechsel der Datengrundlage zur Berechnung des Gehaltsvergleichs sicherstellen.

3

Datenaufbereitung

Aus Gründen der Vergleichbarkeit wurden Teilzeitbeschäftigte bei der Schätzung des Regressionsmodells ausgeschlossen: Deren Bruttomonatsverdienst hängt stark von der Zahl der Arbeitsstunden ab. Branchen mit einem vergleichsweise hohen Anteil an Teilzeitbeschäftigten (die in der Regel einen niedrigeren Bruttomonatsverdienst haben als Vollzeitbeschäftigte) würden die durchschnittlichen Monatsverdienste verzerren. Zudem wäre ein Vergleich der Verdienste über verschiedene Branchen hinweg erschwert, sofern diese einen unterschiedlich hohen Anteil an Teilzeitbeschäftigten haben. Da mehr Frauen als Männer in Teilzeit arbeiten, gilt Gleiches auch für die Verdienste von Männern und Frauen. Auch hier würde die Einbeziehung von Teilzeitbeschäftigten dazu führen, dass die durchschnittlichen Verdienste von Frauen systematisch unterschätzt würden. Das Regressionsmodell basiert daher auf den Verdiensten von etwa 600 000 Personen, die in Vollzeit arbeiten und keine Auszubildenden oder Personen in Altersteilzeit sind.

Im Zuge der Datenaufbereitung wurde auch auf mögliche Ausreißer eingegangen. Nach Aguiñes und anderen (2013) sind Ausreißer Datenpunkte, die sich stark von anderen Datenpunkten unterscheiden. Den Autoren zufolge kann es sich dabei beispielsweise um Fehler in den Daten handeln (wie Tippfehler) oder um „besondere Fälle“. Letztere unterscheiden sich in ihrer Ausprägung zwar von anderen Fällen, sind aber dennoch korrekt. Eine Entfernung von Ausreißern, bei denen es sich nicht um echte Fehler handelt, hat zur Folge, dass Informationen eines (korrekten) Datenpunkts nicht genutzt werden können. Dies kann dazu führen, dass die Stichprobe die Grundgesamtheit nur noch verzerrt darstellt. Da aufgrund der Plausibilisierung der Daten durch die Statistischen Ämter der Länder falsche Angaben (gerade bei Extremfällen) sehr unwahrscheinlich sind, würde dies lediglich die Varianz künstlich reduzieren. Deshalb werden alle Werte zur Schätzung des Regressionsmodells verwendet.

4

Das Regressionsmodell

Um das Regressionsmodell zu entwickeln, wurde zunächst basierend auf theoretischen Modellen (wie dem Lohnmodell von Mincer, 1974) und mithilfe der vorhandenen Daten Folgendes ermittelt: Welche Merkmale beeinflussen den Verdienst einer Person und wie stark ist ihr Einfluss? Dies ist möglich, da in den Daten sowohl die Verdienste als auch die gehaltsbestimmenden Merkmale hinterlegt sind. Mithilfe des Regressionsmodells, das die ausgewählten Einflussfaktoren und die Größe ihres Einflusses umfasst, können anschließend Schätzungen ausgegeben werden. Die ausgegebenen Zahlenwerte des Gehaltsvergleichs sind folglich Prognosewerte einer Regressionsgleichung. Die Regression wird dabei nicht etwa zur Laufzeit durchgeführt, sondern zuvor bei der Programmierung des Gehaltsvergleichs. Im Gehaltsrechner sind allein die ermittelten Regressionskoeffizienten hinterlegt. Für jede Merkmalsausprägung liegt ein merkmalspezifischer Prognosewert (= Regressionskoeffizient) vor. Die finalen Prognosewerte ergeben sich durch eine Aufsummierung der von der Person individuell ausgewählten merkmalspezifischen Regressionskoeffizienten (bei metrischen Variablen erfolgt zunächst die Multiplikation des hinterlegten Prognosewerts mit

der von der Nutzerin oder vom Nutzer angegebenen Zahl) und der Konstanten. Der so ermittelte Wert wird dann der anfragenden Person ausgegeben.

Da es sich bei der Verdienststrukturerhebung um eine Stichprobe mit komplexem Design (zweistufige, geschichtete Stichprobe) handelt, ist dies auch bei der Auswertung zu berücksichtigen. In der vom Statistischen Bundesamt genutzten Statistiksoftware SAS steht hierzu die Prozedur SURVEYREG zur Verfügung (Finke, 2010). Das SURVEYREG-Verfahren führt Regressionsanalysen für Stichprobenerhebungsdaten durch. Das Verfahren kann mit komplexen Stichprobenentwürfen für Umfragen umgehen, einschließlich Entwürfen mit Schichtung, Clustering und ungleicher Gewichtung (SAS Institute Inc., 2013).

5

Variablen der Regression

5.1 Die abhängige Variable Bruttomonatsverdienst

Die abhängige Variable ist der Bruttomonatsverdienst (Gesamtbruttoentgelt gemäß §1 Absatz 2 Nummer 2c Entgeltbescheinigungsverordnung) abzüglich sonstiger Bezüge wie 13. Monatsgehalt, Urlaubsgeld oder Weihnachtsgeld. Die Entscheidung fiel für die Ausgabe von Monatsverdiensten, da die Verdienste in der Lohnabrechnung der Betriebe in Monatsgehältern vorliegen und auch so erhoben werden. Darüber hinaus handelt es sich bei den monatlichen Bruttoverdiensten um sogenannte laufende Bezüge. Anhand dieser lässt sich nach Abzug von Steuern und Abgaben (die individuell unterschiedlich sind, wie die Beiträge zur Krankenversicherung) das monatlich verfügbare Erwerbseinkommen bestimmen. Somit stellt der Bruttomonatsverdienst eine wichtige Entscheidungsgrundlage dar. Außerdem ist den meisten Personen der Bruttomonatsverdienst aufgrund der gesetzlich vorgeschriebenen Entgeltbescheinigung geläufig. Folglich lassen sich Angaben dazu gut einordnen und leicht vergleichen.

Zudem unterliegen Verdienstdaten einer rechtsschiefen Verteilung (Mincer, 1974). Dies kann durch Loga-

rithmierung korrigiert werden (Fields, 2010), was eine Annäherung an die Normalverteilung ermöglicht. Aus diesem Grund wird in dem Regressionsmodell der logarithmierte Bruttomonatsverdienst als abhängige Variable verwendet:

$$\ln Y_i = \beta_0 + \sum_{j=1}^n \beta_j X_{ij} + e_i$$

Dabei sind:

$\ln Y_i$ = logarithmierter Bruttomonatsverdienst einer Person i

β_j = Regressionskoeffizient eines Merkmals j

β_0 = Regressionskonstante

X_{ij} = beobachtetes Merkmal j einer Person i

e = Störterm

5.2 Die unabhängigen Variablen

Die nachfolgend beschriebenen Variablen (beziehungsweise Merkmale) wurden in das Regressionsmodell mit aufgenommen, um das Gehalt zu schätzen. Hierbei handelt es sich sowohl um arbeitsplatzspezifische Merkmale (wie den Beruf) als auch um persönliche Merkmale (wie die Ausbildung). Die Auswahl der Merkmale orientiert sich am (bereinigten) Gender Pay Gap (Finke, 2010).

Interessen von Nutzerinnen und Nutzern können sehr unterschiedlich sein. Beispielsweise sucht eine Person Informationen zu einer konkreten Stelle, eine andere möchte jedoch lediglich eine grobe Orientierung über Verdienstmöglichkeiten in einem Beruf. Daher müssen nicht für alle Merkmale zwingend Angaben vorgenommen werden. Während für Beruf, Branche, Ausbildung und Bundesland eine Auswahl zu treffen ist, gibt es bei anderen Merkmalen auch die Option, „keine Angabe“ auszuwählen. In diesem Fall werden für die betroffenen Merkmale Annahmen getroffen. Damit diese Annahmen für möglichst viele Nutzerinnen und Nutzer „passend“ sind, basieren sie auf Durchschnittswerten oder (im Rahmen der Anwendung) „üblichen“ Werten.

Pflichtangaben

Die Pflichtangaben sind immer anzugeben und sind Merkmale, die einen besonders starken Einfluss auf das Ergebnis haben. Diese Merkmale sind nicht übermäßig spezifisch und können auch von Personen angegeben werden, die sich nur grob orientieren möchten.

Beruf

Anhand der Stichwortliste der Bundesagentur für Arbeit, die auf dem Klassifikationsserver der Statistischen Ämter des Bundes und der Länder (klassifikationsserver.de) zur Verfügung steht, werden einzelne Berufsbezeichnungen den Berufsgattungen der Klassifikation der Berufe 2010 zugeordnet. Nähere Informationen zu den Klassifikationen der Berufe sind in Band 1 der Klassifikation der Berufe 2010 (Bundesagentur für Arbeit, 2011) zu finden.

Die Berufsgattung dient als Ausgangsbasis, um die Merkmale Berufsgruppe, Anforderungsniveau und Aufsichtsbeziehungsweise Führungsfunktion abzuleiten. In der Klassifikation der Berufe 2010 wird jeder Berufsgattung eine fünfstellige Zahl zugeordnet (5-Steller). Die ersten drei Stellen des 5-Stellers bezeichnen die Berufsgruppe. Die vierte Stelle gibt Aufschluss darüber, ob es sich um eine Aufsichts- beziehungsweise Führungsfunktion handelt. Umfasst eine berufliche Tätigkeit vorwiegend Aufsichts- oder Führungsaufgaben, wird sie entsprechend als Aufsichts- beziehungsweise Führungsposition definiert. Die letzte Stelle kann die Werte 1 bis 4 annehmen und bezieht sich auf das Anforderungsniveau des Arbeitsplatzes. Das Anforderungsniveau gibt an, wie komplex eine Tätigkeit ist. Es gibt vier Ausprägungen: (1) Helfer- und Anlernertätigkeiten, (2) fachlich ausgerichtete Tätigkeiten, (3) komplexe Spezialistentätigkeiten und (4) hoch komplexe Tätigkeiten.

Beispielsweise gehört die Berufsbezeichnung „Sekretär/in“ der Berufsgattung „Büro- und Sekretariatskräfte (ohne Spezialisierung) – Fachlich ausgerichtete Tätigkeiten“ an (5-Steller 71402). Diese wiederum ist eine Untergruppe der Berufsgruppe „Büro- und Sekretariat“ (3-Steller 714), wird nicht als Aufsicht- beziehungsweise Führungsfunktion eingeordnet (vorletzte Ziffer ungleich 9) und hat das Anforderungsniveau „fachlich ausgerichtete Tätigkeiten“ (Stufe 2).

Der Vorteil einer Aufsplittung des 5-Stellers liegt darin, dass die Nutzerinnen und Nutzer die Bestimmung des Anforderungsniveaus nicht selbst vornehmen müssen. Stattdessen erfolgt dies automatisch über den ausgewählten Beruf und dessen Zuordnung zu einem Anforderungsniveau. Darüber hinaus lässt sich so auch die Anzahl an Dummyvariablen im Regressionsmodell reduzieren.

Branche

In die Regression gehen auch die einzelnen Abteilungen der Branchen ein. Basierend auf der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008) wird zunächst die Oberkategorie (= Abschnitt der WZ 2008) ausgewählt und dann die Unterkategorie (= Abteilung der WZ 2008) bestimmt. Beispielsweise kann als Oberkategorie „Verarbeitendes Gewerbe (Herstellung)“ und als Unterkategorie „Automobilindustrie (Herstellung von Fahrzeugen und Fahrzeugteilen)“ ausgewählt werden. Die Originalbezeichnungen aus der WZ 2008 wurden zum besseren Verständnis leicht angepasst. So wird zum Beispiel statt der Bezeichnung „Herstellung von Kraftwagen und Kraftwagenteilen“ die Formulierung „Automobilindustrie (Herstellung von Fahrzeugen und Fahrzeugteilen)“ verwendet.

Ausbildung

Die Ausbildung einer Person wird mit dem höchsten beruflichen Ausbildungsabschluss erfasst. Dieser umfasst sechs Kategorien: ohne beruflichen Ausbildungsabschluss, Abschluss einer anerkannten Berufsausbildung, Meister (beziehungsweise Techniker- oder gleichwertiger Fachschulabschluss), Bachelor, Master (beziehungsweise Diplom/Magister/Staatsexamen) und Promotion.

Bundesland

Die Nutzerinnen und Nutzer wählen zudem auch das Bundesland aus. Dies ermöglicht die Berücksichtigung von regionalen Verdienstunterschieden.

Wahlangaben

Für die Wahlmerkmale ist keine Angabe erforderlich. Werden diese nicht spezifiziert, wird ein Durchschnittswert hinterlegt.

Alter

Anhand des Alters wird approximativ die Berufserfahrung ermittelt und in die Analyse mit einbezogen. Dazu erfolgt eine näherungsweise Ermittlung über das Alter und die Ausbildung einer Person (Achatz und andere, 2005). Durch diese Form der Berechnung kann es theoretisch vorkommen, dass die Berufserfahrung einen negativen Wert annimmt (potenzielle Berufserfahrung = Alter – durchschnittliche Ausbildungsdauer). Für den Fall, dass das Alter unter der durchschnittlichen Ausbildungsdauer liegt, wird der Wert der Berufserfahrung jedoch auf null gesetzt. Die potenzielle Berufserfahrung wird zudem auch als quadrierter Wert in das Regressionsmodell mit aufgenommen, um den im Zeitverlauf verminderten Einfluss von Berufserfahrung auf den Verdienst abzubilden (Achatz und andere, 2005).

Befristung

Das Regressionsmodell berücksichtigt auch, ob eine Stelle befristet oder unbefristet ist. Gründe für eine Befristung können beispielsweise der zeitlich begrenzte Einsatz im Rahmen einer Projektstelle oder eine Elternzeitvertretung sein.

Anzahl Beschäftigte im Unternehmen

Durch die Aufnahme der Anzahl der Beschäftigten im Unternehmen in das Regressionsmodell wird die Größe des Unternehmens abgebildet. Um den Nutzerinnen und Nutzern die Angabe zu erleichtern, gibt es sechs Kategorien: weniger als 10 Beschäftigte, 10 bis 49 Beschäftigte, 50 bis 249 Beschäftigte, 250 bis 499 Beschäftigte, 500 bis 999 Beschäftigte sowie 1 000 und mehr Beschäftigte.

Tarifbindung

Das Regressionsmodell berücksichtigt auch, ob in dem Betrieb ein Tarifvertrag (Kollektivvertrag oder Firmentarifvertrag) gilt oder nicht.

Dauer der Unternehmenszugehörigkeit

Wie lange eine Person bereits in einem Unternehmen arbeitet, wird über die Dauer der Unternehmenszugehörigkeit erfasst. Da – ähnlich wie bei der Berufserfahrung – der positive Einfluss auf den Verdienst im Zeitverlauf geringer ausfällt, wird auch in diesem Fall der quadrierte Term in das Regressionsmodell mit aufgenommen.

Geschlecht

Das Geschlecht ist ebenfalls als unabhängige Variable im Regressionsmodell enthalten. Allerdings erfolgt aus Gründen der Transparenz keine Auswahl durch die Nutzerinnen und Nutzer. Stattdessen wird das Ergebnis für beide Gruppen ausgegeben. Die Beschränkung auf die binäre Ausgabe (Mann oder Frau, ohne die Angabe von „divers“) ist durch die unzureichende Datenbasis für die Gruppe „divers“ bedingt.

6

Qualität des Modells

Die Güte des Modells wurde zum einen mit klassischen Qualitätskriterien überprüft, wie dem (korrigierten) R^2 . Zum anderen wurde der mittlere absolute prozentuale Fehler (mean absolute percentage error, MAPE) herangezogen, um eine Kreuzvalidierung durchzuführen. Des Weiteren werden in der Webanwendung Plausibilisierungsmaßnahmen ergriffen, um Falscheingaben zu unterbinden und eine ausreichende Anzahl an Datensätzen zu gewährleisten.

6.1 R^2 und korrigiertes R^2

Das R^2 (auch Bestimmtheitsmaß genannt) ist ein Gütemaß zur Beurteilung der Qualität eines Regressionsmodells und kann Werte zwischen 0 und 1 annehmen. Es gibt an, wie viel Prozent der Varianz der abhängigen Variable durch das Modell erklärt werden kann. Ein R^2 von 0 bedeutet, dass es keinen Zusammenhang zwischen den unabhängigen Variablen und der abhängigen Variablen gibt, während ein Wert von 1 (dies entspricht 100 %) einem perfekten Zusammenhang entspricht. Da das R^2 immer weiter ansteigt, je mehr Variablen in das

Modell aufgenommen werden, wird gerne das korrigierte R^2 zur Beurteilung herangezogen. Dieses „bestraft“ zu komplexe Modelle und ist in der Regel niedriger als das R^2 . Das für den Gehaltsvergleich verwendete Regressionsmodell hat ein R^2 von 0,66 und ein korrigiertes R^2 von ebenfalls 0,66. Folglich erklärt das Modell 66 % der Varianz (beziehungsweise Änderungen) der abhängigen Variablen, was für Modelle dieser Art eine gute Leistung ist. Dass das korrigierte R^2 nicht deutlich unter dem R^2 liegt, ist darüber hinaus ein erster Anhaltspunkt dafür, dass das Modell nicht zu viele (unnötige) Variablen umfasst. Weitere Informationen zur Überanpassung enthält der folgende Abschnitt 6.2.

6.2 Kreuzvalidierung

Da das Modell der Vorhersage von Verdiensten dient, ist seine Qualität auch diesbezüglich zu testen. Eine der größten Gefahren bei der Entwicklung von Vorhersagemodellen ist die Überanpassung (Shmueli, 2010). Bei Erstellung eines Modells mit vielen Variablen kann es passieren, dass das Modell übermäßig genau auf die aktuelle Stichprobe eingestellt wird (mit allen stichprobenbedingten Besonderheiten), aber neue Daten deutlich schlechter schätzt (Putka und andere, 2018). Das führt zu einer zu optimistischen Beurteilung der Güte des Modells. Ein überangepasstes Modell kann folglich zwar sehr gut die Daten erklären, mit denen es erstellt wurde, liefert aber bei neuen Daten deutlich schlechtere Schätzungen. Die Qualität eines Modells zur Vorhersage von Ergebnissen wird daher mittels Kreuzvalidierung bestimmt. Im einfachsten Fall unterteilt man die Stichprobe in zwei Teile und nutzt einen Teil der Daten, um das Modell zu erstellen, und den anderen Teil, um mit diesem Modell die Werte für die abhängige Variable zu schätzen. Ist die Schätzqualität in beiden Teilen der Stichprobe gleich gut, dann ist eine Überanpassung unwahrscheinlich und davon auszugehen, dass das Modell die Grundgesamtheit (und nicht nur die Stichprobe) abbildet (John/Roth, 1999).

Daher wurde das Modell zusätzlich noch einmal mit 90 % der Daten geschätzt (Trainingsdatensatz) und mit 10 % der Daten getestet (Validierungsdatensatz). Die Aufteilung erfolgte dabei zufallsbasiert. Zum Vergleich wurde der mittlere absolute prozentuale Fehler herangezogen. Dieses Abweichungsmaß entspricht der mittleren prozentualen Abweichung der geschätzten Werte von den

erhobenen Werten (hier: die Bruttomonatsverdienste in Euro). Während im Trainingsdatensatz der mittlere absolute prozentuale Fehler 19,27 % beträgt, liegt er im Validierungsdatensatz bei 19,21 %. Dieser Unterschied ist nicht signifikant, eine Überanpassung ist daher unwahrscheinlich. Folglich kann das Modell zur Schätzung neuer Daten verwendet werden. Um für die Schätzung des finalen Modells nicht auf Informationen verzichten zu müssen, werden dafür alle Daten verwendet. Dies ist möglich, da die Modelle, die mit 90 % beziehungsweise 100 % der Daten geschätzt wurden, nahezu gleich sind und Überanpassung bereits ausgeschlossen wurde.

Darüber hinaus sagt der mittlere absolute prozentuale Fehler aus, dass die vom Modell geschätzten Bruttomonatsverdienste (in Euro) im Mittel etwa 20 % von den erhobenen (= wahren) Bruttomonatsverdiensten abweichen. Dabei ist zu bedenken, dass das Modell den mittleren Bruttomonatsverdienst schätzt. So wie die individuellen erhobenen Verdienste um den Mittelwert der erhobenen Verdienste streuen, streuen die individuellen erhobenen Verdienste auch um den geschätzten Mittelwert. Zudem lassen sich bei einer Schätzung nicht alle Einflussfaktoren perfekt abbilden. Beispielsweise wird erfasst, welchen Abschluss eine Person hat, aber nicht deren Notendurchschnitt. Folglich kann es auch innerhalb dieser Gruppe von Personen mit dem gleichen Abschluss zu Unterschieden im Verdienst kommen. Vereinfacht gesagt wird daher der erhobene Mittelwert (und somit auch der geschätzte Mittelwert) des Bruttomonatsverdienstes für Personen mit einer sehr guten Abschlussnote tendenziell zu niedrig sein, für Personen mit einer sehr schlechten Abschlussnote dagegen eher zu hoch.

Zur Veranschaulichung dieser Differenzen soll ein Beispiel mit dem Beruf-5-Steller „Büro- und Sekretariatskräfte (ohne Spezialisierung) – fachlich ausgerichtete Tätigkeiten“ dienen. Für diese Gruppe liegt der Mittelwert der geschätzten Verdienste (3 186 Euro) vergleichsweise nah am Mittelwert der erhobenen Verdienste (3 374 Euro). Die Abweichung beträgt 5,57 %. Der mittlere absolute prozentuale Fehler für diese Gruppe liegt jedoch bei 21,72 %. Da Einzelwerte in der Regel um den Mittelwert streuen, kommt es zu Abweichungen. Folglich weicht die Schätzung des Mittelwerts auch von den individuellen erhobenen Werten ab. Vergleicht man aber die Mittelwerte der Schätzungen mit den Mittelwerten der erhobenen Daten, sind die Unterschiede deutlich

geringer. So liegen die Abweichungen der Mittelwerte (geschätzter und wahrer Bruttomonatsverdienst) für die 500 häufigsten Berufe im Schnitt unter 6,50%. Daher muss bei der Interpretation der Ergebnisse des Gehaltsvergleichs bedacht werden, dass es sich um Mittelwert-schätzungen für das ausgewählte Profil handelt. Die ausgegebenen Werte können und sollen daher nur der Orientierung dienen.

6.3 Plausibilisierung innerhalb der Webanwendung

Maßnahmen zur Qualitätssicherung sind ebenfalls in den interaktiven Gehaltsvergleich integriert. Diese verhindern die Eingabe unplausibler Angaben und die Anzeige von Ergebnissen ohne ausreichende Anzahl an Datensätzen. So darf als Alter nur ein Wert zwischen 16 und 67 Jahren angegeben werden und die Dauer der Unternehmenszugehörigkeit 51 Jahre nicht überschreiten. Darüber hinaus muss die Differenz zwischen Alter und Dauer der Unternehmenszugehörigkeit mindestens 16 Jahre betragen. Des Weiteren werden Ergebnisse nur ausgegeben, wenn in der Kombination Beruf 3-Steller und in der Oberkategorie des gewählten Wirtschaftszweigs mindestens 140 Fälle in den Daten vorliegen. Dies entspricht umgerechnet 30 Fällen für jede Kombination von Wirtschaftszweig 2-Steller (die Gliederungsebene „Abteilungen“ der WZ 2008) und Beruf 3-Steller. Die Einschränkung soll zum einen gewährleisten, dass für die einzelnen Kombinationen dieser sehr stark untergliederten Merkmale genügend Fälle vorliegen. Zum anderen soll sie verhindern, dass versehentlich völlig unsinnige Kombinationen ausgewählt werden können. Sowohl Beruf als auch Wirtschaftszweig haben vergleichsweise hohen Einfluss auf die Höhe des Verdienstes. Doch gerade die Auswahl des korrekten Wirtschaftszweigs ist für Laien nicht ganz einfach. Daher verhindert diese Prüfung auch, dass den Nutzerinnen und Nutzern für falsche Kombinationen von Beruf 3-Steller und Abteilung der WZ 2008 Ergebnisse ausgegeben werden.

7

Fazit und Ausblick

Zielgruppe für die Nutzung des interaktiven Gehaltsvergleichs sind vorwiegend Privatpersonen, die beispielsweise aufgrund von bevorstehenden Gehaltsverhandlungen, Bewerbungsgesprächen oder rein aus Interesse eine spezifische Bruttomonatslohnschätzung erhalten möchten. Den Nutzerinnen und Nutzern soll eine unkomplizierte Schätzung von Bruttomonatsverdiensten für individuell spezifizierte Profile möglich sein. Diese Profile können nach Belieben anhand verschiedener gehaltsbestimmender Merkmale (wie Beruf, Ausbildungsabschluss oder Branche) konfiguriert werden. Verdienstdaten werden somit zielgruppenorientiert aufbereitet und ihr Nutzen dadurch erhöht. Der Gehaltsvergleich trägt folglich zur Ausrichtung des Statistischen Bundesamtes als kundenorientierter und innovativer Informationsdienstleister bei.

Der Anwendung liegt ein Regressionsmodell zugrunde, das eine individuelle Schätzung der Verdienste anhand von Profilen ermöglicht. Dieses Vorgehen verhindert zum einen die Verletzung von Geheimhaltungsvorschriften, die bei einer so spezifischen Darstellung andernfalls unumgänglich wäre. Zum anderen werden Verdienstinformationen auf diese Weise nutzungsfreundlich aufbereitet und leichter zugänglich. Das Verfahren der multiplen Regression erlaubt es, anhand der ermittelten Regressionskoeffizienten neue Profile zu schätzen und den Bruttomonatsverdienst zu prognostizieren. Zur Beurteilung der Qualität des Modells wurde das (korrigierte) R^2 herangezogen. Da es sich um ein Vorhersagemodell handelt, wurde darüber hinaus im Rahmen einer Kreuzvalidierung auch der mittlere absolute prozentuale Fehler (MAPE) betrachtet. Dies stellt sicher, dass keine Überanpassung des Modells auf die zur Schätzung verwendeten Daten vorliegt. Die Qualitätsprüfungen zeigen, dass das Modell geeignet ist, um Schätzungen für neue Daten auszugeben. Dabei ist jedoch zu beachten, dass die geschätzten Verdienste lediglich als Anhaltspunkte zur Orientierung dienen können.

Die Anwendung selbst enthält zudem mehrere Plausibilitätskontrollen, die gravierende Fehleingaben verhindern. Ebenso wird sichergestellt, dass nur dann Verdienstinformationen ausgegeben werden, sofern genügend Daten für das angeforderte Profil vorliegen.

Die Nutzerinnen und Nutzer des interaktiven Gehaltsvergleichs benötigen Verdienstinformationen, um ihre gegenwärtige Situation einzuschätzen. Daher ist eine möglichst aktuelle Datengrundlage erstrebenswert. Die Verwendung der neuen Verdiensterhebung ermöglicht eine solche zeitnahe Aktualisierung der Daten. Im Jahr 2021 wurde einmalig der April erhoben, ab 2022 liefert diese Erhebung monatlich Informationen zu den Verdiensten. Es ist vorgesehen, zur Aktualisierung des Gehaltsrechners jeweils die Aprildaten zu verwenden. Dieser Monat wurde bereits bei der Verdienststrukturerhebung als repräsentativer Berichtsmonat genutzt. Dadurch ist künftig eine Aktualisierung des Rechners bereits im laufenden Jahr möglich. Im Gegensatz zur Verdienststrukturerhebung erfasst die neue Verdiensterhebung darüber hinaus alle Mitarbeiterinnen und Mitarbeiter eines Unternehmens und nicht nur einen Teil davon. Dadurch liegen künftig die Einzeldaten von etwa 7 Millionen Beschäftigten vor. Dieser Zugewinn an Einzeldaten sollte es ermöglichen, für eine größere Anzahl an Beruf-Branchen-Kombinationen Informationen bereitzustellen. 

LITERATURVERZEICHNIS

- Achatz, Juliane/Gartner, Hermann/Glück, Timea. *Bonus oder Bias? Mechanismen geschlechtsspezifischer Entlohnung*. In: KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie. Jahrgang 57. Ausgabe 3/2005, Seite 466 ff.
- Aguines, Herman/Gottfredson, Ryan K./Joo, Harry. *Best practice recommendations for defining, identifying, and handling outliers*. In: Organizational Research Methods. Jahrgang 16. Ausgabe 2/2013, Seite 270 ff. DOI: doi.org/10.1177/1094428112470848
- Bundesagentur für Arbeit. *Klassifikation der Berufe 2010 – Band 1: Systematischer und alphabetischer Teil mit Erläuterungen*. 2011.
- Field, Andy/Miles, Jeremy. *Discovering statistics using SAS*. 2010.
- Finke, Claudia. [Verdienstunterschiede zwischen Männern und Frauen 2006](#). Herausgeber: Statistisches Bundesamt im Auftrag des Bundesministeriums für Familie, Senioren, Frauen und Jugend. Wiesbaden 2010.
- Mincer, Jacob A. *Schooling, experience, and earnings*. New York 1974.
- Putka, Dan J./Beatty, Adam S./Reeder, Matthew C. *Modern prediction methods: New perspectives on a common problem*. In: Organizational Research Methods. Jahrgang 21. Ausgabe 3/2018, Seite 689 ff.
- SAS Institute Inc. *SAS/STAT® 13.1 User's Guide*. Cary 2013.
- Shmueli, Galit. *To explain or to predict?* In: Statistical Science. Jahrgang 25. Ausgabe 3/2010, Seite 289 ff.
- Statistisches Bundesamt (Herausgeber). [Verdienststrukturerhebung 2018](#). Qualitätsbericht. 2020.
- Statistisches Bundesamt (Herausgeber). [Interaktiver Gehaltsvergleich](#). Methodenbericht. 2020.
- St. John, Caron H. /Roth, Philip L. *The Impact of Cross-Validation Adjustments on Estimates of Effect Size in Business Policy and Strategy Research*. In: Organizational Research Methods. Jahrgang 2. Ausgabe 2/1999, Seite 157 ff. DOI: doi.org/10.1177/109442819922003

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im Dezember 2021, Seite 43 korrigiert am 23.02.2022
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-21006-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2021
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.