

INTERNETGESTÜTZTE ERFASSUNG OFFENER STELLEN

Machbarkeitsstudie im Rahmen eines ESSnet-Projekts
zu Big Data

Dr. Martina Rengers

↳ **Schlüsselwörter:** Big Data – ESSnet – Online-Stellenmarkt – Offene Stellen – Datenqualität

ZUSAMMENFASSUNG

Können Stellenanzeigen aus dem Internet für die amtliche Statistik nutzbar gemacht werden? Um diese Frage drehte sich ein europäisches ESSnet-Big-Data-Projekt, an dem sich das Statistische Bundesamt seit 2016 beteiligte. Der wichtigste Aspekt bei einer Antwort auf diese Fragestellung ist die Beurteilung der Qualität dieser neuen digitalen Datenquelle. Für die Analyse verwendete das Statistische Bundesamt Daten des Jobportals der Bundesagentur für Arbeit und des Europäischen Zentrums für die Förderung der Berufsbildung. Diese wurden der Stellenerhebung des Instituts für Arbeitsmarkt- und Berufsforschung gegenübergestellt. Der Beitrag erläutert das Projekt und gibt die wesentlichen Arbeitsergebnisse für Deutschland wieder.

↳ **Keywords:** big data – ESSnet – online job market – job vacancies – data quality

ABSTRACT

Can internet job offers be utilised for official statistics? This was the focus of a European big data ESSnet project, in which the Federal Statistical Office was involved from 2016 onwards. The main aspect of answering this question refers to the assessment of the quality of this new digital data source. For the analysis, the Federal Statistical Office used data from the job portal of the Federal Employment Agency and of the European Centre for the Development of Vocational Training. They were compared with data from the job vacancy survey of the Institute for Employment Research. This article explains the project and shows the main work results for Germany.



Dr. Martina Rengers

ist Volkswirtin und im Referat „Arbeitsmarkt“ des Statistischen Bundesamtes tätig. Ihre Schwerpunkte liegen in der Analyse ungenutzten Arbeitskräftepotenzials bei Erwerbstätigen, Erwerbslosen und in der Stillen Reserve sowie der konzeptionellen und methodischen Weiterentwicklung der international vergleichbaren Arbeitsmarktstatistik. Zusätzlich befasst sie sich mit dem Thema „Big Data“ und Arbeitsmarktstatistik.

1

Einleitung¹

1.1 Big Data ESSnet 2016 bis 2018

Im Rahmen der Big-Data-Strategie des Europäischen Statistischen Systems (ESS Big Data Action Plan and Roadmap) wurde 2015 ein Big-Data-Projekt des European Statistical System Network (ESSnet) ins Leben gerufen und im November 2015 von einem Konsortium von 22 Partnern aus 20 verschiedenen Ländern vertraglich vereinbart. Das ESSnet-Projekt BIG DATA lief von Februar 2016 bis Mai 2018 und war in zwei Projektphasen² geteilt. Verschiedene Pilotstudien wurden in insgesamt acht Arbeitspaketen durchgeführt, darunter unter anderem Pilotprojekte zur internetgestützten Erfassung (Web Scraping) von offenen Stellen und Unternehmensdaten, zur Nutzung von Mobilfunkdaten, zu Smart-Meter-Daten zum Energieverbrauch sowie zu Navigations- und anderen Schiffsdaten des Automatischen Identifikationssystems (AIS). Das Ziel des Gesamtprojektes war, die generelle Machbarkeit in Bezug auf die Nutzung dieser neuen digitalen Datenquellen durch die amtliche Statistik zu prüfen.³

1.2 Pilotstudie zur internetgestützten Erfassung offener Stellen

An der Pilotstudie zur internetgestützten Erfassung offener Stellen waren neben Deutschland die nationalen statistischen Ämter aus Griechenland, Italien, Schweden, Slowenien und dem Vereinigten Königreich (Vorsitz) beteiligt. In der zweiten Projektphase (ab August 2017) kamen mit Belgien, Dänemark, Frankreich und Portugal weitere Länder hinzu.

Ziel des Arbeitspaketes war es zu untersuchen, inwiefern Online-Stellenanzeigen von Jobportalen, Unternehmenswebseiten oder von privaten Drittanbietern⁴ die Arbeitsmarktberichterstattung unterstützen können. Primäre Technik hierfür ist das sogenannte Web Scraping, das gezielte und automatisierte Herunterladen von Daten auf Internetseiten.⁵

Alle beteiligten Länder standen in ihren Arbeiten zur internetgestützten Erfassung offener Stellen auf Jobportalen vor großen Herausforderungen, insbesondere bei der Dublettenbeseitigung sowie der unvollständigen Abdeckung der offenen Stellen (Coverage-Problem). Die Schnelligkeit in der Branche der Jobportalanbieter und – insbesondere in großen Ländern – die Vielzahl der Anbieter machen eine Bewertung der Jobportale nahezu unmöglich. Auch ist die Datengewinnung von Online-Jobportalen mit rechtlichen und technischen Schwierigkeiten verbunden. Mehrere Länder haben deshalb Kooperationen mit den nationalen Arbeitsagenturen geschlossen, da deren Jobportale oft die umfangreichsten und relevantesten Portale darstellen und der Datentransfer vorzugsweise nicht über Web Scraping, sondern über eine vereinbarte Datenlieferung erfolgt. Auf welche Art und Weise die einzelnen Länder an Daten zu Online-Stellenanzeigen gekommen sind, zeigt [Übersicht 1](#).

Das Statistische Bundesamt hat im Rahmen des Projekts Daten des Jobportals der Bundesagentur für Arbeit und Daten des Europäischen Zentrums für die Förderung der Berufsbildung CEDEFOP (Centre européen pour le développement de la formation professionnelle) verwendet.

Die vorliegende Arbeit stellt die erlangten Projektergebnisse vor und behandelt dabei vorrangig Fragen zur Datenqualität von Online-Stellenanzeigen. Die Ergebnisse sind letztlich Ausgangsbasis der Beurteilung, ob und inwiefern Online-Stellenanzeigen für die amtliche Statistik nutzbar gemacht werden können.

1 Ich danke Herrn Chris-Gabriel Islam für umfangreiche Vorarbeiten zu diesem Aufsatz. Chris-Gabriel Islam studierte Wirtschaftsmathematik und Lehramt für Mathematik und Geschichte an der Technischen Universität Darmstadt. Im Referat „Arbeitsmarkt“ des Statistischen Bundesamtes beschäftigte er sich als wissenschaftlicher Mitarbeiter mit der Etablierung von neuen digitalen Daten in der amtlichen Statistik. Seit August 2018 ist er Referent im Referat „Methoden und Kommunikation in der Preisstatistik“ des Statistischen Bundesamtes.

2 Wegen der zwei Projektphasen wird manchmal auch der Plural „die Big Data ESSnets 2016 bis 2018“ verwendet.

3 Weitere Informationen können auf der Projektseite eingesehen werden: https://ec.europa.eu/eurostat/cros/content/essnet-big-data_en [Zugriff am 5. September 2018].

4 Beispiele für private Drittanbieter sind die niederländische Firma Textkernel mit Hauptsitz in Amsterdam oder die amerikanische Firma Burning Glass mit Hauptsitz in Boston, Massachusetts. Beide Firmen nutzen eine Vielzahl von Daten aus Online-Jobportalen, die unter anderem mithilfe von Web-Scraping-Technologien gewonnen werden. Ihre kommerziellen Marktanalysen und Softwarelösungen haben zum Ziel, Angebot und Nachfrage am Arbeitsmarkt zusammenzuführen und Matchingprozesse zu beschleunigen. Siehe dazu auch Kapitel 4 und die Übersichten 2 und 3 im Abschnitt 5.2.

5 Das Verfahren des Web Scrapings wird vom Statistischen Bundesamt bereits in der Preisstatistik erfolgreich eingesetzt. Siehe dazu beispielsweise Blaudow (2018) und Brunner (2014).

Übersicht 1

Datenzugang der Teilnehmerländer des ESSnet-Projekts Big Data

| | Direct Web scraping | | Agreed Access | | | |
|----------------|---------------------|-------------|---------------|----------------------------|-----------------------------|------------------------------|
| | Enterprise websites | Job Portals | | National Employment Agency | Private Employment Agencies | Other data aggregators |
| Germany | | Yes | | Yes | | Yes (CEDEFOP) |
| Greece | | Yes | | | | |
| Slovenia | Yes | Yes | Yes | Yes | Yes | |
| Sweden | | | Yes | Yes | | |
| United Kingdom | Yes | Yes | Yes | | | Yes (CEDEFOP, Burning Glass) |
| France | | Yes | | Yes | | |
| Belgium | Yes | Yes | | Yes | | |
| Portugal | | Yes | | | | |

Quelle: Swier und andere (2018), Seite 19. Zu Burning Glass siehe Fußnote 4 im Text.

2

Messen der Datenqualität von Online-Stellenanzeigen

2.1 Herausforderungen

Eine wichtige Unterscheidung bei der Analyse des Online-Stellenmarktes ist die zwischen Stellenanzeigen und Stellenangeboten. Eine Stellenanzeige kann mehrere Stellenangebote enthalten, beispielsweise indem ein Arbeitgeber Angestellte für mehrere Standorte sucht, ohne dabei für jedes Angebot eine eigene Anzeige aufzugeben.

Bei den Jobportalen gibt es insgesamt drei Typen:

- › Jobbörsen listen Stellenanzeigen auf, die bei ihnen hochgeladen wurden,
- › Jobsuchmaschinen verfügen über keine eigenen Anzeigen, sondern zählen lediglich Angebote anderer Portale oder von Unternehmenswebseiten auf,
- › hybride Jobportale zeigen sowohl eigene als auch gefundene Stellenanzeigen (Rengers, 2018, hier: Seite 62).

Durch diese Einteilung von Jobportalen wird ersichtlich, dass eine Stellenanzeige nicht zwingend genau einem Portal zugeordnet ist. Möglicherweise stellt ein Arbeit-

geber seine Anzeige bei verschiedenen Portalen ein, um seine Reichweite zu erhöhen, oder aber Portalbetreiber binden automatisch die Anzeigen anderer Anbieter in ihre Trefferliste ein. Daher kann es beim Webscrapen mehrerer Portale vorkommen, dass dieselbe Stellenanzeige mehr als einmal heruntergeladen wird und es zu Dubletten kommt. Eine entsprechende Bereinigung im Datensatz ist nicht trivial und erfordert das Anwenden von intelligenten Algorithmen oder Verfahren des Maschinellen Lernens⁶.

Dubletten können aber nicht nur entstehen, wenn verschiedene Portale gescraped werden, sondern auch, wenn dasselbe Portal mehrmals zu verschiedenen Zeitpunkten gescraped wird und dadurch ein und dieselbe Stellenanzeige mehrmals in den Enddatensatz aufgenommen wird. Eine erste Heuristik, um dieses Dublettenproblem zu lösen, ist es, zwischen Strom- und Bestandsgrößen zu unterscheiden. Dabei sind unter einer Bestandsgröße sämtliche Stellenanzeigen einer gewissen Menge von Jobportalen zu einem bestimmten Zeitpunkt zu verstehen. Stromgrößen hingegen bezeichnen den Zu- und Abgang jener Bestandsgrößen in einem bestimmten Zeitintervall. Ist der Datensatz zu einem bestimmten Zeitpunkt von Dubletten bereinigt und werden im Anschluss daran nur eindeutige Stromgrößen hinzugefügt, kann dieses stromgrößenbedingte Dublettenproblem weitgehend gelöst werden. Technisch sind hier verschiedene Wege denkbar. Beispielsweise kann nach dem Scrapen jeder Stellenanzeige mit einem bestimmten Maß über-

⁶ Dazu siehe Fußnote 20 auf Seite 24.

prüft werden, wie ähnlich eine Anzeige zu den bisherigen Anzeigen ist. Eine weitere Maßnahme ist das Aufrufen der URL in regelmäßigen Zeitabständen. Damit kann geprüft werden, ob eine Anzeige noch geschaltet oder aber abgegangen ist.

Anders sehen die Möglichkeiten bei der jobportalübergreifenden Dublettenbereinigung aus. Hier sind die Erfolgsaussichten weniger gut. Bei den Untersuchungen zur Auswahl geeigneter Datenquellen zur internetbasierten Erfassung offener Stellen wurde deshalb empfohlen, möglichst auf die Einbeziehung hybrider Jobportale zu verzichten, um dieses Dublettenproblem zu minimieren.¹⁷

Neben den genannten Problematiken birgt die Analyse des Online-Stellenmarktes zahlreiche weitere Herausforderungen. Die international vergleichsweise hohe Zahl von 1 563 Jobportalen auf dem deutschen Online-Stellenmarkt¹⁸ verlangt nach einer Reduktion und Beschränkung auf die wichtigsten Jobportale. Der Perspektive von Körner/Rengers und andere (2016) folgend, gehören sowohl die beiden Jobbörsen StepStone und Monster als auch die beiden hybriden Jobportale der Bundesagentur für Arbeit und von Gigajob zu den wichtigsten Jobportalen Deutschlands. Neben Jobportalen lassen sich beispielsweise auch direkt die Webseiten der Unternehmen und die dortigen Stellenanzeigen scrapen. In beiden Fällen ist die legale Basis zu beachten. Sowohl Firmen als auch Portalbetreibende können einem Web Scraping in den allgemeinen Geschäftsbedingungen oder in sogenannten robots.txt-Dateien¹⁹ widersprechen, die genaue Rechtslage ist jedoch unklar. Daher sind eine explizite Einzelfallprüfung oder ein bilateraler Austausch empfehlenswert. Unter den Teilnehmerinnen und Teilnehmern des ESSnet-Projektes ist unumstritten erkannt worden, dass das Web Scraping von Jobportalen sehr mühsam und eine Kooperation mit Portalbetreibern dem vorzuziehen ist (Swier und andere, 2018, hier: Seite 4).

7 Zumindest sollten bei einer Einbeziehung von hybriden Jobportalen nur deren „eigene“ Anzeigen verwendet werden oder aber die Qualität der Jobportal eigenen Dublettenbereinigungsalgorithmen überprüft werden. Siehe dazu auch Rengers (2018), Seite 76 f.

8 Stand 12. Dezember 2017, siehe Crosswater Job Guide (2018).

9 Mithilfe einer Datei robots.txt kann ein Domain-Inhaber das Verhalten von Suchmaschinen-Robots beeinflussen. Wie dies funktioniert, ist im sogenannten Robots-Exclusion-Standard-Protokoll (REP) geregelt. Webmaster können damit festlegen, wie Suchmaschinen auf ihre Site zugreifen.

2.2 Die IAB-Stellenerhebung

Die Qualität eines gescrepten Datensatzes muss anhand bereits bestehender, valider und vergleichbarer Daten gemessen werden. Da die Messung der Arbeitskräftenachfrage auf dem Online-Stellenmarkt ein neues Feld ist, gibt es keine direkt vergleichbaren Daten. Zur Erfassung des gesamtwirtschaftlichen Stellenangebots wird in Deutschland die Stellenerhebung des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) genutzt. Die IAB-Stellenerhebung wird quartalsweise durchgeführt und erfragt auf freiwilliger Basis Informationen zu Anzahl und Struktur offener Stellen.¹⁰

Für die IAB-Stellenerhebung wird seit 1989 im Westen und seit 1992 auch im Osten Deutschlands eine repräsentative Auswahl an Betrieben und Verwaltungen befragt. Es begann mit jährlichen schriftlichen Befragungen im vierten Quartal, die seit dem Jahr 2006 durch telefonische Kurzinterviews im ersten, zweiten und dritten Quartal ergänzt werden. Das vierte Quartal ist dabei das Quartal mit dem größten Stichprobenumfang.¹¹

➤ Grafik 1 zeigt die relevantesten Fragen der IAB-Stellenerhebung. Dabei sind zwei Aspekte besonders zu beachten:

› Derzeit suchend:

Der Fragebogen enthält die Frage, ob die Personalverantwortlichen eines Unternehmens derzeit neue Mitarbeiterinnen und Mitarbeiter suchen. Das hat zur Folge, dass das Ergebnis der Befragung als Bestandsgröße der offenen Stellen innerhalb eines Quartals angesehen werden kann.

› Sofort oder zum nächstmöglichen Termin einstellen ODER zu einem späteren Zeitpunkt:

Die Fragen 6 und 9 in Grafik 1 zeigen, dass nicht nur nach Mitarbeiterinnen und Mitarbeitern gefragt wird, die sofort oder zum nächstmöglichen Termin eingestellt werden können (Frage 6), sondern auch nach

10 Ergebnisse der IAB-Stellenerhebung können unter <http://www.iab.de/de/befragungen/stellenangebot.aspx> abgerufen werden. Ausführliche Informationen zur IAB-Stellenerhebung findet man bei Brenzel und andere, 2016a.

11 In den vergangenen Jahren lag die Rücklaufquote im vierten Quartal zwischen 13 000 und 15 000 auswertbaren Fragebogen. In dem jeweils folgenden ersten, zweiten und dritten Quartal wird eine Teilstichprobe von rund 9 000 Befragungsteilnehmern des vierten Quartals um eine Aktualisierung der Angaben gebeten (Brenzel und andere, 2016a, hier: Seite 9 f.).

Mitarbeiterinnen und Mitarbeitern für einen späteren Zeitpunkt (Frage 9). Die Anzahl der offenen Stellen besteht also aus der Summe der beiden Antworten. Dies ist auch jene Zahl, die an das Statistische Amt der Europäischen Union (Eurostat) berichtet wird. Auf nationaler Ebene wird aber auch die Anzahl der sofort zu besetzenden Stellen veröffentlicht.

Grafik 1

Fragen zur Personalsuche bei der IAB-Stellenerhebung

5. Suchen Sie derzeit neue Mitarbeiter?
Bitte berücksichtigen Sie hier

- KEINE Ausbildungsverhältnisse
- KEINE Entfristungen oder Vertragsverlängerungen
- KEINE Arbeitskräfte, die Sie von Zeit-/Leiharbeitsfirmen entleihen wollen
- KEINE öffentlich geförderte Beschäftigung (z.B. Ein-Euro-Jobs)

Ja Nein ➔ Bitte weiter mit Frage 10

6. Suchen Sie derzeit Mitarbeiter, die sofort oder zum nächstmöglichen Termin eingestellt werden sollen?

Ja Nein ➔ Bitte weiter mit Frage 9

↓

| | | |
|--|--|--|
| Wenn ja, wie viele? <small>ggf. Schätzwerte eintragen, „0“ eintragen, wenn keine</small> | Anzahl sofort einzustellender Mitarbeiter | davon: der Arbeits- agentur gemeldet |
| Insgesamt | <input style="width: 100%;" type="text"/> | <input style="width: 100%;" type="text"/> |

9. Suchen Sie derzeit über die in Frage 6 genannten Personen hinaus Mitarbeiter, die erst zu einem späteren Termin eingestellt werden sollen?

Ja Nein ➔ Bitte weiter mit Frage 10

↓

| | | |
|--|--|--|
| Wenn ja, wie viele? <small>ggf. Schätzwerte eintragen, „0“ eintragen, wenn keine</small> | Anzahl später einzustellender Mitarbeiter | davon: der Arbeits- agentur gemeldet |
| Insgesamt | <input style="width: 100%;" type="text"/> | <input style="width: 100%;" type="text"/> |

Auszug aus dem Fragebogen der IAB-Stellenerhebung aus dem Jahr 2017.

Beim Vergleich der Daten von Online-Stellenanzeigen mit der IAB-Stellenerhebung ist auf die oben beschriebene Interpretation als Bestandsgröße besonderes Augenmerk zu legen. Darüber hinaus ist stets – wie in Abschnitt 2.1 beschrieben – zu beachten, dass auf eine Stellenanzeige mehr als ein Stellenangebot fallen kann.

Weiterhin ist beim Vergleich der Online-Stellenanzeigen mit der IAB-Stellenerhebung zu berücksichtigen, dass letztere den gesamten Stellenmarkt darstellt. Daten von Online-Stellenanzeigen repräsentieren dagegen – günstigstenfalls – nur den Online-Stellenmarkt, wo

bei unklar ist, inwieweit sich diese Ergebnisse auf den Offline-Stellenmarkt übertragen lassen (erstes Coverage-Problem). Aufgrund der Vielzahl an Jobportalen ist es notwendig, eine begrenzte Anzahl an Online-Jobportalen auszuwählen. Dabei ist die Repräsentativität zu gewährleisten. Wegen des unübersichtlichen und schnelllebigen Marktes der Jobportalanbieter ist eine Bewertung der Jobportale nahezu unmöglich. Erschwerend kommt hinzu, dass einzelne Jobportale beispielsweise auf bestimmte Branchen wie das Hotelgewerbe spezialisiert sind. In der Folge kann es zu einer verzerrten Abdeckung kommen (zweites Coverage-Problem; Rengers, 2018, Seite 62 ff.).

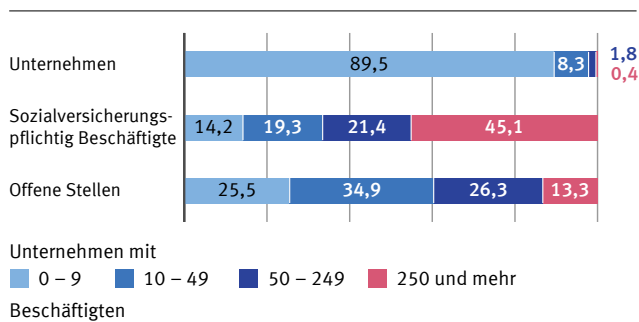
Aus unterschiedlichen Befragungen kann versucht werden, die Bedeutung verschiedener Rekrutierungswege bei der Personalsuche und -gewinnung zu ermitteln. Bei einer Zusatzbefragung der IAB-Stellenerhebung nach der Suchmethode im Jahr 2015 war mit 52 % die Stellenausschreibung auf der eigenen Webseite der am häufigsten genutzte Suchweg der Betriebe. An zweiter Stelle standen persönliche Kontakte und Empfehlungen der eigenen Mitarbeiterinnen und Mitarbeiter (50%), gefolgt vom Kontakt mit der Bundesagentur für Arbeit. Auf Platz 4 der Rekrutierungswege kamen mit 41 % die Online-Jobportale (siehe im Einzelnen Brenzel und andere, 2016b). Befragungen bei den Top-1000-Unternehmen Deutschlands (Weitzel und andere, 2015) kommen dagegen zu einem anderen Ergebnis: Hier steht mit 90 % die eigene Unternehmenswebseite an erster Stelle der Rekrutierungswege. Auf Platz zwei folgt mit 70 % die Personalsuche über Online-Jobportale.¹²

Nach diesen Studien neigen größere Unternehmen also eher dazu, ihre Stellen online auszuschreiben, während kleine Unternehmen eher traditionellere Kanäle, beispielsweise Print, benutzen. Ein Blick auf die Unternehmens- und Beschäftigungslandschaft in Deutschland lohnt: Auswertungen aus dem statistischen Unternehmensregister (URS) zeigen, dass es im Jahr 2016 fast 3,5 Millionen Unternehmen gab. Weniger als zehn Beschäftigte hatten 89,5 % aller Unternehmen. Von den insgesamt 29,5 Millionen sozialversicherungspflichtig Beschäftigten arbeiteten 14,2 % in solchen kleinen Unternehmen, 45,1 % dagegen in Großunternehmen mit 250 und mehr Beschäftigten. Auf der anderen Seite gab es im

¹² Neben den genannten deutschen Untersuchungen ist außerdem auf Carnevale und andere (2014) zu verweisen.

vierten Quartal 2016 laut IAB-Stellenerhebung 1,1 Millionen offene Stellen. Von diesen entfielen 25,5% auf Unternehmen mit weniger als zehn Beschäftigten und 13,3% auf Unternehmen mit 250 und mehr Beschäftigten. Das heißt, dass die Vakanzrate¹³ bei größeren Unternehmen geringer ist als bei kleineren. [↪ Grafik 2](#)

Grafik 2
Ausgewählte Daten nach Unternehmensgröße 2016
in %



Datenbasis: IAB-Stellenerhebung und Unternehmensregister.

2018 - 01 - 0667

3

Online-Stellenanzeigen der Bundesagentur für Arbeit

3.1 Datensatzbeschreibung

Bei den untersuchten Online-Stellenanzeigen der Bundesagentur für Arbeit (BA) handelte es sich um den kompletten Bestand an Stellenanzeigen, die zum 26. Februar 2018 auf dem Jobportal der BA geschaltet waren. Die Datenlieferung der BA bestand aus zwei Teilen. Der erste Teil enthielt die Hauptdatei, die mit insgesamt 75 Variablen die Stellenanzeigen beschreibt. Neben dem Volltext einer Stellenanzeige wurden zahlreiche andere Variablen geliefert. Zu den wichtigsten gehörten dabei die Anzahl der Stellen je Stellenanzeige, die Berufsbezeichnung, der Wirtschaftszweig des Unternehmens, die Berufsklasse und das frühestmögliche Eintrittsdatum der Stellenbesetzung. Im zweiten Teil stand eine

¹³ Diese ist definiert als Quotient aus der Anzahl der offenen Stellen durch die Anzahl der belegten und offenen Stellen.

Datei mit den Ortsangaben der Stellenanzeigen zur Verfügung. Beide Dateien konnten über eine gemeinsame Identifikationsnummer identifiziert und verlinkt werden. Allerdings handelte es sich dabei durchaus um 1-zu-n-Beziehungen, weil zu einer Stellenanzeige mehrere Einsatzorte existieren können.

Da es sich bei dem Online-Jobportal¹⁴ der BA um ein hybrides Jobportal handelt, sind nicht nur originäre Stellenanzeigen der BA enthalten, sondern auch Stellenanzeigen von sogenannten Allianzpartnern. Bei diesen Allianzpartnern handelt es sich entweder um Unternehmen oder andere Jobportale. Die Herkunft jeder Stellenanzeige konnte ermittelt werden, da im Datensatz eine Variable „Allianzpartner“ zur Verfügung stand. Das Jobportal hatte zum Stichtag 26. Februar 2018 einen Bestand von 470 374 Stellenanzeigen, mit insgesamt 748 901 Stellenangeboten. Der Anteil der Stellenanzeigen von Allianzpartnern lag bei 10,4% (49 041 von 470 374 Stellenanzeigen). Durchschnittlich enthielt eine Anzeige 1,6 Stellen. Aufgrund des hohen Anteils an originären BA-Stellenanzeigen lag der Durchschnitt dort ebenfalls bei 1,6 Stellen je Anzeige, während die 49 041 Stellenanzeigen der Allianzpartner durchschnittlich 1,7 Stellen je Anzeige umfassten. [↪ Tabelle 1](#)

Tabelle 1
Das hybride Online-Jobportal der Bundesagentur für Arbeit (BA)

| | Insgesamt | BA originär | BA-Allianzpartner |
|-----------------|-----------|-------------|-------------------|
| Stellenanzeigen | 470 374 | 421 333 | 49 041 |
| Stellenangebote | 748 901 | 663 586 | 85 315 |

Berechnungen basierend auf BA-Daten mit Stand 26. Februar 2018.

¹⁴ Das Online-Jobportal der Bundesagentur für Arbeit ist zurzeit parallel auf zwei Plattformen zu finden: Die neue Plattform ist unter <https://con.arbeitsagentur.de/prod/jobboerse/jobsuche-ui/>, die alte Plattform übergangsweise – bis zu Abschaltung – weiterhin unter <https://jobboerse.arbeitsagentur.de> (Stand: 29. August 2018) erreichbar. Auf der alten Plattform wird die Bezeichnung „Jobbörse“ verwendet, die in diesem Beitrag vermieden wird, um die in Abschnitt 2.1 beschriebene Abgrenzung zwischen Jobbörsen, Jobsuchmaschinen und hybriden Jobportalen stringent anzuwenden.

3.2 Vergleich mit der IAB-Stellenerhebung

Für den Vergleich mit der IAB-Stellenerhebung wurde der Datensatz der Online-Stellenanzeigen der Bundesagentur für Arbeit (BA) zunächst bereinigt. Mit der Bereinigung sollten Karteileichen entfernt werden, denn es wurden Stellenanzeigen gefunden, bei denen das frühestmögliche Eintrittsdatum der Stellenbesetzung mehrere Monate in der Vergangenheit zurück lag (das Maximum war bei gut 20 Monaten). Wie bereits in früheren Untersuchungen wurde als Richtwert von einer „Lebensdauer“ von 30 Tagen für eine Online-Stellenanzeige ausgegangen (Körner/Rengers und andere, 2016, hier: Seite 15).

Da keine Variable für das Veröffentlichungsdatum der Stellenanzeige zur Verfügung stand, wurde das gerade genannte frühestmögliche Eintrittsdatum der Stellenbesetzung als Approximation verwendet. Wie ein Vergleich von Tabelle 2 mit Tabelle 1 zeigt, wird der Datensatz um fast 34 000 Stellenanzeigen (von 470 374 auf 436 499 Stellenanzeigen) reduziert, wenn das frühestmögliche Eintrittsdatum zwischen dem 27. Januar und dem 26. Februar 2018 liegt. Dies entspricht einer Ausschussquote von insgesamt 7,2%. Die Differenzierung nach der Herkunft der Stellenanzeigen zeigt, dass die Ausschussquote bei den Allianzpartnern mit 10,5% deutlich höher ist als bei den originären BA-Stellenanzeigen (6,8%). Durch die Bereinigung verringert sich damit der Anteil der Stellenanzeigen von Allianzpartnern von 10,4 auf 10,1%, da bei den Allianzpartnern mehr alte Stellenanzeigen vorhanden sind als bei den originären BA-Stellenangeboten. Die durchschnittliche Anzahl der Stellen je Anzeige bleibt auch nach der Bereinigung unverändert. [↪ Tabelle 2](#)

Der bereinigte Datensatz der Online-Stellenanzeigen und -angebote kann als (Stichtags-) Ergebnis des ersten Quartals 2018 angesehen werden und wurde hier – wie Tabelle 2 zeigt – mit den Ergebnissen der IAB-Stellenerhebung des ersten Quartals 2017 (!) verglichen. Dafür gab es mehrere Gründe: Zunächst einmal lagen zum Zeitpunkt der Untersuchung die Ergebnisse aus der Stellenerhebung für das erste Quartal 2018 noch nicht vor. Außerdem ist die aus den Online-Stellenanzeigen ermittelte absolute Anzahl an Stellen konzeptionell bedingt nicht mit der absoluten Anzahl an Stellen der IAB-Erhebung vergleichbar: Erstere können nur den Online-Stellenmarkt abbilden, letztere dagegen den gesamten Stellenmarkt. Es geht vielmehr um einen strukturellen Vergleich; mit diesem soll der Frage nachgegangen werden, ob im Online-Stellenmarkt (der BA) bestimmte Wirtschaftszweige, Berufe, Bildungsniveaus oder Betriebsgrößen im Vergleich zum gesamten Stellenmarkt über- oder unterrepräsentiert sind. Da die Quartalsergebnisse der IAB-Stellenerhebung einen Saisonverlauf aufweisen, wurde der bereinigte Datensatz der Online-Stellenanzeigen, der dem ersten Quartal 2018 zugeordnet werden kann, mit dem ersten Quartal der IAB-Stellenerhebung des Vorjahres 2017 verglichen (und nicht mit den aktuelleren Zahlen des vierten Quartals 2017).

Tabelle 2 bestätigt, dass die Anzahl der Stellenangebote der BA nicht an die entsprechende Zahl der IAB-Stellenerhebung heranreicht. Auf Basis der Online-Stellenanzeigen der BA erhält man knapp 695 000 Stellen für das erste Quartal 2018, ein Jahr zuvor wies die IAB-Stellenerhebung allerdings fast 1,1 Millionen Stellen aus. Deutlich geringer ist die Diskrepanz der Online-Stellen des BA-Jobportals zur Zahl der sofort zu besetzenden Stellen, die laut IAB-Stellenerhebung im ersten Quartal 2017 einen Wert von 758 000 annahm. Der Zeitunterschied erklärt diese Differenz nicht, denn tendenziell gab es in den letzten Jahren eine Zunahme

Tabelle 2

BA-Jobportal und IAB-Stellenerhebung: Vergleich der Stellenanzeigen, der Stellenangebote und der sofort zu besetzenden Stellen

| | BA-Jobportal, 1. Quartal 2018 ¹ | | | IAB-Stellenerhebung, 1. Quartal 2017 |
|--|--|-------------|-------------------|--------------------------------------|
| | insgesamt | BA originär | BA-Allianzpartner | |
| Stellenanzeigen | 436 499 | 392 587 | 43 912 | X |
| Stellenangebote | 694 817 | 618 200 | 76 617 | 1 064 000 |
| darunter: sofort zu besetzende Stellen | X | X | X | 758 000 |

¹ Stichtag 26. Februar 2018; bereinigt (30 Tage).
Berechnungen basierend auf BA- und IAB-Daten.

sowohl der offenen Stellen insgesamt als auch der sofort zu besetzenden Stellen. Es ist nicht auszuschließen, dass der Datensatz der Online-Stellenanzeigen der BA aus technischen Gründen möglicherweise unvollständig war. Zumindest wurde auf der Webseite des BA-Jobportals eine deutlich höhere Gesamtzahl an Stellen in Höhe von rund 1,5 Millionen genannt.¹⁵ Ein Teil der Differenz zwischen Online-Stellen und Stellen laut IAB-Stellenerhebung lässt sich allerdings auch dadurch erklären, dass Stellenanzeigen bereits wieder abgeschaltet wurden, obwohl die entsprechenden Stellen noch nicht offiziell besetzt waren. Grund dafür ist, dass der Bewerbungsprozess, das heißt das Auswählen, Einladen und erneute Auswählen der Bewerberinnen und Bewerber, eine gewisse Zeit beansprucht. Würde ein Unternehmen in solch einer Latenzzeit durch das IAB im Rahmen der Stellenerhebung befragt werden, meldete es eine offene Stelle, die aber gleichzeitig nicht mehr im Datensatz der BA auftauchte. Ob Betriebe so handeln oder aber – wie von Brenzel und anderen (2016b) angeführt – eher dazu neigen, Stellenanzeigen noch nach der Besetzung weiterlaufen zu lassen, kann im Rahmen dieser Untersuchung nicht geklärt werden. Es bleibt an dieser Stelle erneut zu berücksichtigen, dass hier der gesamte Stellenmarkt mit dem Online-Stellenmarkt verglichen wird. Eine simple Hochrechnung entsprechend der in Abschnitt 1.2 erwähnten Online-Quote von 41 % hätte sogar eine Überschätzung der Stellenangebote zur Folge.

Wirtschaftszweig

↳ **Grafik 3** zeigt die prozentuale Verteilung von Online-Stellenanzeigen und -angeboten gegliedert nach Wirtschaftszweigen. Im Datensatz der BA gibt es für diese Variable keine fehlenden Werte. Der inkohärente Vergleich zwischen Anzeigen und Stellen zeigt, dass die relative Häufigkeitsverteilung der 436 499 Online-Stellenanzeigen der BA teilweise deutlich von der relativen Häufigkeitsverteilung der knapp 1,1 Millionen Stellen der IAB-Stellenerhebung abweicht. In Prozentpunkten ausgedrückt gibt es hier im Wirtschaftszweig N (Erbringung von sonstigen wirtschaftlichen Dienstleistungen) mit 8,1 Prozentpunkten die maximale Abweichung. Der Durchschnitt der absolut gesetzten Abweichungen liegt bei 1,2 Prozentpunkten, bei einer Streuung von 1,3 Pro-

zentpunkten (mittlere absolute Abweichung). Bei knapp 67 % der Wirtschaftszweige (14 von 21) liegen die absolut gesetzten Abweichungen unterhalb von einem Prozentpunkt. Vergleicht man die Wirtschaftszweigstruktur der IAB-Stellenerhebung dagegen mit der Struktur der 694 817 Stellen des Online-Jobportals, so verschlechtern sich die Maße: Die maximale Abweichung liegt weiterhin im Wirtschaftszweig N, vergrößert sich aber auf 17,7 Prozentpunkte. Der Durchschnitt der Abweichungen erhöht sich auf 1,8 Prozentpunkte, die Streuung auf 1,9 Prozentpunkte. Die absolut gesetzten Abweichungen liegen nur noch bei gut 57 % der Wirtschaftszweige unterhalb von einem Prozentpunkt.

Dieses Ergebnis ist zunächst einmal unerwartet und zeigt, dass die durchschnittliche Anzahl der Stellen je Anzeige in den Wirtschaftszweigen sehr unterschiedlich ist. Das Maximum der durchschnittlichen Anzahl der Stellen je Anzeige liegt im Wirtschaftszweig N und beträgt 2,1 Stellen je Anzeige, gefolgt von Wirtschaftszweig H (Verkehr und Lagerei) mit 2,0 Stellen je Anzeige.

Wird an dieser Stelle zwischen originären BA-Stellenangeboten und solchen von Allianzpartnern unterschieden, fällt auf, dass die im Wirtschaftszweig N beobachtete Diskrepanz zur relativen Häufigkeitsverteilung der Stellen aus der IAB-Erhebung bei den Allianzpartnern mehr als doppelt so stark ausgeprägt ist. Die oben genannten Abweichungsmaße verringern sich deutlich, wenn ausschließlich die BA-originären Stellen den Berechnungen zugrunde gelegt werden.

Berufe

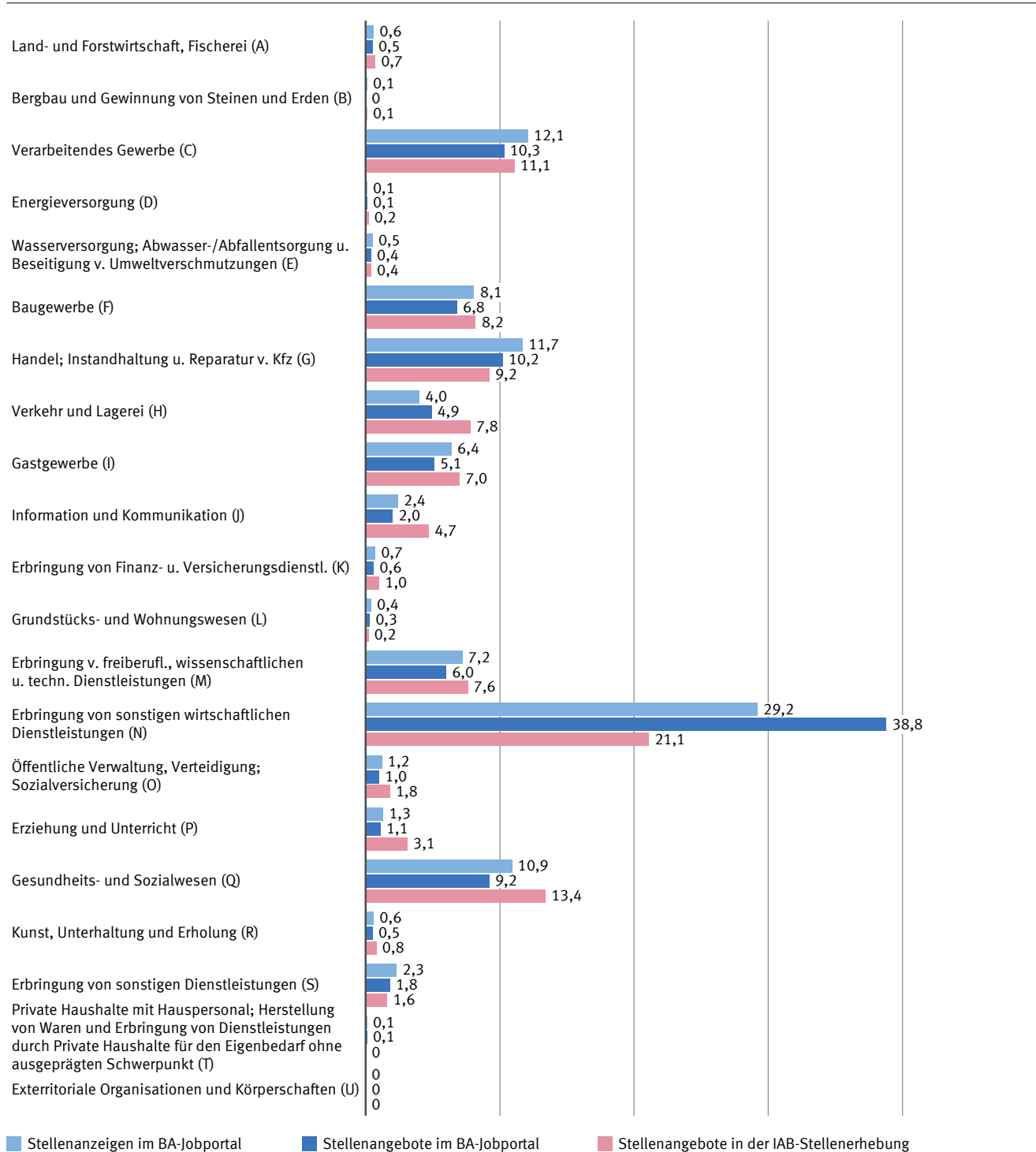
Ein Vergleich der Daten des IAB und der BA auf Berufsebene ist nicht ohne Weiteres möglich, da das IAB der ISCO-08-Klassifikation¹⁶ folgt, im BA-Datensatz die Berufe aber nach einer BA-internen Klassifikation namens Dokumentenkennziffer (DKZ) klassifiziert sind. Zunächst ist es erforderlich, von der DKZ auf die Berufskennziffer (BKZ) umzuschlüsseln. Dazu stehen Tabellen auf dem Download-Portal der BA zur Verfügung (Bundesagentur für Arbeit, 2018a). Aus dieser BKZ kann die Klassifikation der Berufe 2010 (KldB 2010) extrahiert werden, aus der wiederum anhand einer Zuordnungstabelle

15 Angabe laut der alten Plattform (siehe Fußnote 14) zum Stand Februar 2018.

16 Internationale Standardklassifikation der Berufe (ISCO – International Standard Classification of Occupations), die von der Internationalen Arbeitsorganisation (ILO – International Labour Organization) entwickelt wurde. ISCO-08 ist die aktuelle Version aus dem Jahr 2008.

Grafik 3

BA-Jobportal und IAB-Stellenerhebung: Stellenanzeigen und -angebote nach Wirtschaftszweigen¹
in %



BA-Jobportal: Stichtag 26.2.2018; bereinigt (30 Tage). IAB-Stellenerhebung: 1. Quartal 2017, Berechnungen basierend auf Daten der Bundesagentur für Arbeit (BA) und einer Sonderauswertung der Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

¹ Wirtschaftszweige der NACE Rev. 2 (Statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft).

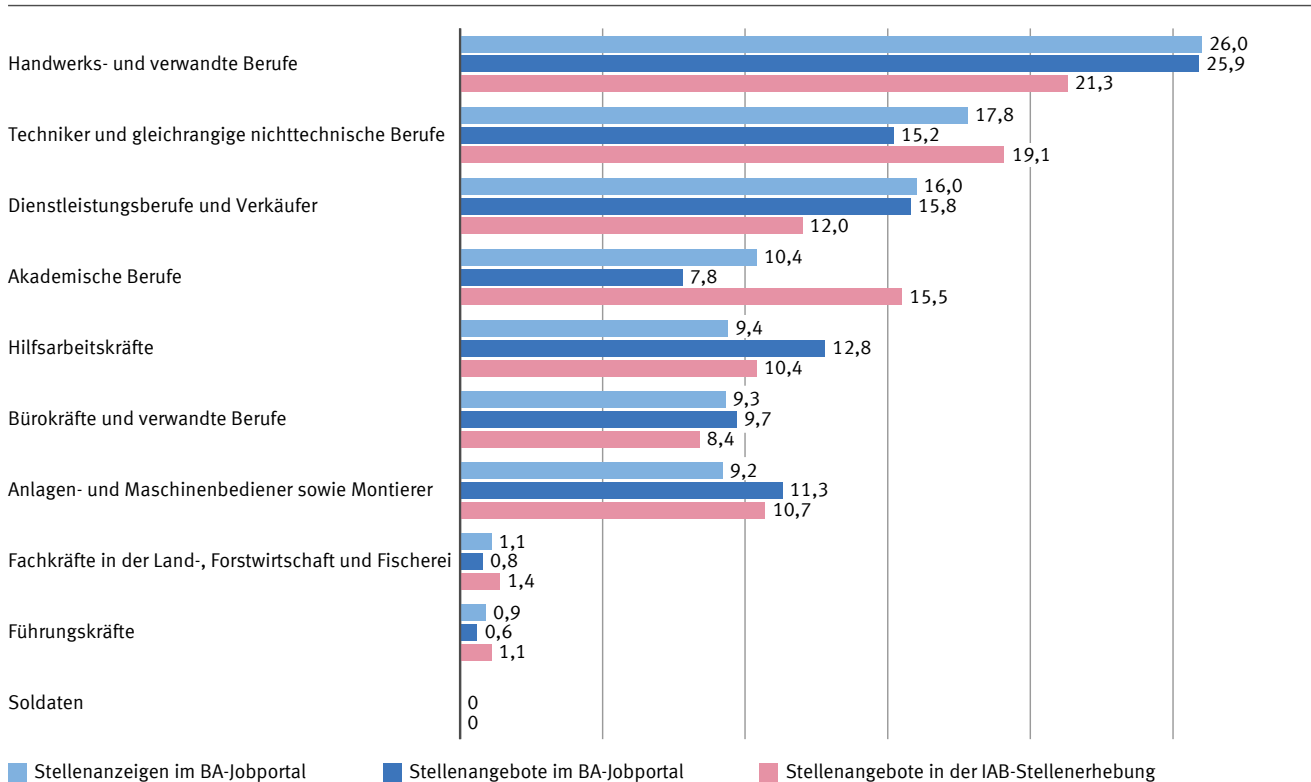
der BA (Bundesagentur für Arbeit, 2018b) auf die ISCO-08-Klassifizierung geschlossen werden kann. Der letzte Schritt ist jedoch nicht eindeutig. Etwa 25 % aller KldB-Klassen können mehreren ISCO-08-Klassen zugeordnet werden. Das Problem der fehlenden Eindeutigkeit von KldB 2010 zu ISCO-08 wurde hier – wie im Mikrozensus beziehungsweise der Arbeitskräfteerhebung – über das sogenannte Schwerpunktprinzip gelöst: Das heißt, dass bei mehreren Alternativen die erstgenannte ISCO-Alternative im BA-Umsteigeschlüssel von der Klassifikation der Berufe 2010 (5-Steller) zur ISCO-08 (4-Steller) gewählt wurde.

↳ **Grafik 4** stellt die Ergebnisse einer auf diese Weise ermittelten ISCO-08-Klassifizierung der Stellenanzeigen und Stellen des BA-Jobportals dar und vergleicht die Struktur mit derjenigen der IAB-Stellenerhebung. Die absolut gesetzten Abweichungen zwischen den relativen Häufigkeitsverteilungen der Stellen des Online-

Jobportals der BA und der Stellen gemäß IAB-Stellenerhebung erreichen mit 7,7 Prozentpunkten ein Maximum in der ISCO-08-Berufsklasse 2 „Akademische Berufe“; hier ist das Online-Jobportal relativ gesehen unterrepräsentiert. Die durchschnittliche absolut gesetzte Abweichung liegt bei 2,5 Prozentpunkten, bei einer Streuung von 2,0 Prozentpunkten. 40 % der dargestellten zehn Berufsklassen haben eine Abweichung von weniger als einem Prozentpunkt. Diese Abweichungsmaße verbessern sich, wenn nur die originären BA-Stellenangebote betrachtet werden: Die maximale absolute Abweichung sinkt auf 7,4 Prozentpunkte, die mittlere Abweichung nimmt einen Wert von 2,3 Prozentpunkten an und bei 50 % der zehn Berufsklassen liegt die Abweichung unterhalb von einem Prozentpunkt (insgesamt mit Allianzpartnern waren dies 40%). Darüber hinaus verbessern sich außerdem – genau wie bei der Betrachtung der Wirtschaftszweige – auch hier diese Kennziffern der Abweichungen, wenn die relative

Grafik 4

BA-Jobportal und IAB-Stellenerhebung: Stellenanzeigen und -angebote nach Berufsklassen
in %



BA-Jobportal: Stichtag 26.2.2018; bereinigt (30 Tage). IAB-Stellenerhebung: 1. Quartal 2017, Berechnungen basierend auf Daten der Bundesagentur für Arbeit (BA) und einer Sonderauswertung der Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

Häufigkeitsverteilung der Anzeigen des BA-Jobportals mit derjenigen der Stellen der IAB-Erhebung verglichen wird. Und das, obwohl es sich dabei um einen Vergleich handelt, der inhaltlich inkohärente Größen zueinander in Bezug setzt.

Bildung

Der BA-Datensatz enthält keine eindeutige Variable zu den Anforderungen an das Bildungsniveau der Bewerbenden. Der Datensatz enthält sechs Variablen, die mit dem Thema Bildung in Verbindung stehen. Keine davon ist allerdings über alle Anzeigen hinweg befüllt und lediglich eine Ausnahme folgt einer eindeutigen Klassifikation. Beispielsweise weist die Variable „Hochschulabschlussart“ 93 % fehlende Werte auf. Ebenso wenig ist eine Zusammenlegung der Variablen möglich. Eine Ursache dafür ist, dass die Variablen zum Bildungsniveau für die Stellenausschreibenden nicht obligatorisch zu befüllen sind. Die IAB-Stellenerhebung folgt in den Ergebnisdarstellungen der ISCED-11-Klassifikation¹⁷. Denkbar wäre daher, die Berufsangaben der BA-Daten umzuschlüsseln, weil bestimmte Berufe für gewöhnlich auch mit einem bestimmten Bildungsniveau verbunden sind. Von ISCO-08 zu ISCED-11, also von den Berufsklassen hin zu den Bildungsklassen, existiert eine solche Umschlüsselung. Angewendet an Testdaten, die sowohl Bildungs- als auch Berufsklassifikationsangaben enthielten, führte die ISCO-zu-ISCED-Umschlüsselung jedoch zu deutlich anderen Ergebnissen bei der Bildungsangabe als die originäre ISCED-Variable. Deshalb wurde auf eine solche Vorgehensweise verzichtet.

Betriebsgröße

Ein wichtiges Merkmal, das im Datensatz der Online-Stellenanzeigen der BA zur Verfügung stand, ist die Variable „Betriebsgröße“. Die Klasseneinteilung der Betriebsgrößen weicht von der in Abschnitt 2.2 verwendeten ab und wird hier durch die sieben nachfolgend genannten Variablenausprägungen determiniert:

- › keine Angabe,
- › bis 6 Beschäftigte,
- › 6 bis 50 Beschäftigte,
- › 51 bis 500 Beschäftigte,
- › 501 bis 5 000 Beschäftigte,
- › 5 001 bis 50 000 Beschäftigte,
- › mehr als 50 000 Beschäftigte.

Für einen konsistenten Vergleich der Größenstruktur der Betriebe hinter den Online-Stellenanzeigen der BA mit derjenigen aus der IAB-Stellenerhebung wurden Sonderauswertungen mit genau dieser Klasseneinteilung zur Verfügung gestellt, ebenso aus dem statistischen Unternehmensregister. Bei 6 651 Online-Stellenanzeigen findet man keine Angabe zur Größe des Betriebes. Dies entspricht 1,5 % aller Stellenanzeigen beziehungsweise 1,8 % der zugehörigen Stellen. In der IAB-Stellenerhebung gibt es dagegen keine Missingfälle. Die Verteilung der knapp 1,1 Millionen Stellen nach der Betriebsgröße weist den relativ größten Wert in der Klasse von 6 bis 50 Beschäftigten auf: Hier lassen sich 47,3 % der aus der IAB-Stellenerhebung ermittelten Stellen zuordnen. Bei den Online-Stellenangeboten Betrieben der Größenordnung von 51 bis 500 Beschäftigten zugeordnet werden. Mehr als die Hälfte (54,8 %) der Stellenanzeigen der Online-Stellenangebote des BA-Jobportals kommen von Kleinst- und Kleinunternehmen (bis 50 Beschäftigte), bei den Stellenangeboten sind es knapp die Hälfte (49,6 %). Die IAB-Stellenerhebung weist 56,1 % der Stellen dieser Betriebsgrößenklasse zu. Den Kleinstunternehmen hatte das Online-Jobportal nicht nur relativ (15,8 gegenüber 8,8 %), sondern auch absolut mehr Stellen zugeordnet, als dies in der IAB-Stellenerhebung der Fall war. Auch wenn es sich hier um zwei unterschiedliche, ein Jahr auseinanderliegende Berichtszeiträume handelt, so fällt in Bezug auf die absolute Zahl an Stellen Folgendes auf: Sie ist beim BA-Jobportal einzig in der Größenkategorie bis fünf Beschäftigte höher als diejenige der IAB-Stellenerhebung. In allen anderen Fällen ist das umgekehrt, bei einer Gesamtzahl, die ebenfalls deutlich unterhalb derjenigen der Stellenerhebung liegt. Um die relative Bedeutung dieser Kleinstunternehmen darzustellen, tragen ergänzende Angaben aus dem Unternehmensregister dieser speziellen Klasseneinteilung der Betriebsgrößen Rechnung. Kleinstunternehmen mit bis zu fünf Beschäftigten hatten demnach im Jahr 2016

17 Internationale Standardklassifikation des Bildungswesens (ISCED – International Standard Classification of Education). ISCED-11 ist die aktuelle Fassung der ISCED, die am 10. November 2011 von der Generalkonferenz der Organisation der Vereinten Nationen für Bildung, Wissenschaft und Kultur (UNESCO) verabschiedet wurde.

Tabelle 3

BA-Jobportal und IAB-Stellenerhebung: Ergebnisse nach Betriebsgröße

| | BA-Jobportal, 1. Quartal 2018 ¹ | | | | | IAB-Stellenerhebung, 1. Quartal 2017 | | Statistisches Unternehmensregister 2016 | |
|---------------------------|--|------|--------------------|-----------------|------|--------------------------------------|------|---|---------------------------|
| | Stellenanzeigen | | Stellen je Anzeige | Stellenangebote | | Stellen | | Unternehmen ² | Beschäftigte ³ |
| | Anzahl | % | | Anzahl | % | Anzahl | % | | |
| bis 5 Beschäftigte | 80 197 | 18,4 | 1,4 | 109 946 | 15,8 | 94 000 | 8,8 | 83,2 | 8,8 |
| 6 bis 50 Beschäftigte | 159 014 | 36,4 | 1,5 | 234 859 | 33,8 | 503 000 | 47,3 | 14,6 | 25,0 |
| 51 bis 500 Beschäftigte | 156 233 | 35,8 | 1,8 | 278 649 | 40,1 | 381 000 | 35,8 | 2,0 | 30,8 |
| mehr als 500 Beschäftigte | 34 404 | 7,9 | 1,7 | 59 012 | 8,5 | 87 000 | 8,2 | 0,2 | 35,4 |
| Keine Angabe | 6 651 | 1,5 | 1,9 | 12 351 | 1,8 | - | - | - | - |
| Insgesamt | 436 499 | 100 | 1,6 | 694 817 | 100 | 1 064 000 | 100 | 100 | 100 |

1 Stichtag 26. Februar 2018; bereinigt (30 Tage).

2 Unternehmen mit steuerbarem Umsatz und/oder mit sozialversicherungspflichtig Beschäftigten.

3 Sozialversicherungspflichtig Beschäftigte.

Berechnungen basierend auf Daten des BA-Jobportals und Sonderauswertungen der IAB-Stellenerhebung und des statistischen Unternehmensregisters.

einen Anteil von 83,2% an allen Unternehmen¹⁸ und beschäftigten 8,8% aller sozialversicherungspflichtig Beschäftigten. [↘ Tabelle 3](#)

[↘ Tabelle 4](#) zeigt die Betriebsgrößenzugehörigkeit der Online-Stellenanzeigen des BA-Jobportals getrennt nach Herkunft der Stellenanzeige. Originäre Stellenanzeigen der BA stammen mit 19,9% deutlich häufiger von Kleinstunternehmen mit bis zu fünf Beschäftigten als dies bei den Stellenanzeigen der Allianzpartner der Fall ist. Letztere erhalten nur 5,1% ihrer Anzeigen von Kleinstunternehmen, allerdings ist die durchschnittliche Anzahl an Stellen bei diesen Anzeigen mit 2,3 deutlich höher als bei den BA-originären Anzeigen von Kleinstunternehmen (1,3 Stellen).

[↘ Grafik 5](#) zeigt die prozentuale Verteilung von Online-Stellenanzeigen und -angeboten nach der Betriebsgröße. Wie bereits oben erwähnt, gibt es in der Betriebsgröße bis fünf Beschäftigte, aber auch in der Größenklasse von 6 bis 50 Beschäftigten deutliche Abweichungen zwischen den Online-Stellenangeboten und der IAB-Stellenerhebung. Das Maximum der absolut gesetzten Abweichungen der Prozentpunkte wird mit 13,5 Prozentpunkten in der letztgenannten Größenklasse erreicht. Der Durchschnitt dieser Abweichungen in Bezug auf die relative Häufigkeitsverteilung der 694 817 Online-Stellen verglichen mit den knapp 1,1 Millionen Stellen der IAB-Stellenerhebung liegt bei 5,4 Prozentpunkten, bei einer mittleren absoluten Abweichung von 3,9 Prozentpunkten. Nur in einer der vier Größenklassen liegt die Abweichung unterhalb von einem Prozentpunkt. Vergleicht man dagegen die Struktur der Stellenanzeigen des BA-Jobportals mit derjenigen der Stellenangebote

18 Hier: Unternehmen mit steuerbarem Umsatz und/oder mit sozialversicherungspflichtig Beschäftigten.

Tabelle 4

Stellenanzeigen im BA-Jobportal¹ nach Betriebsgröße und Datenherkunft

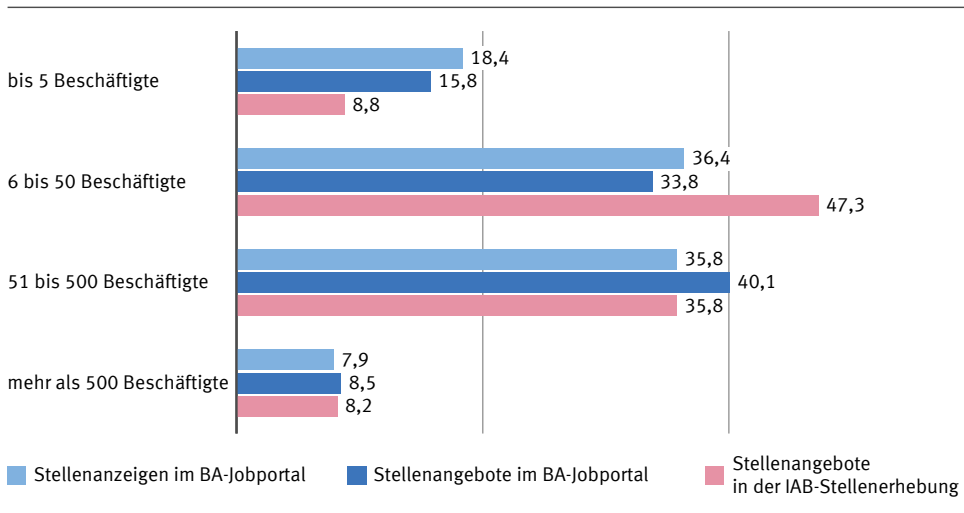
| | Bundesagentur für Arbeit (BA) originär | | | BA-Allianzpartner | | |
|---------------------------|--|------|--------------------|-------------------|------|--------------------|
| | Stellenanzeigen | | Stellen je Anzeige | Stellenanzeigen | | Stellen je Anzeige |
| | Anzahl | % | | Anzahl | % | |
| bis 5 Beschäftigte | 77 964 | 19,9 | 1,3 | 2 233 | 5,1 | 2,3 |
| 6 bis 50 Beschäftigte | 151 678 | 38,6 | 1,5 | 7 336 | 16,7 | 1,8 |
| 51 bis 500 Beschäftigte | 129 680 | 33,0 | 1,8 | 26 553 | 60,5 | 1,8 |
| mehr als 500 Beschäftigte | 27 291 | 6,9 | 1,8 | 7 113 | 16,2 | 1,3 |
| Keine Angabe | 5 974 | 1,5 | 1,9 | 677 | 1,5 | 1,7 |
| Insgesamt | 392 587 | 100 | 1,6 | 43 912 | 100 | 1,7 |

1 1. Quartal 2018; Stichtag: 26. Februar 2018; bereinigt (30 Tage).

Berechnungen basierend auf Daten des BA-Jobportals.

Grafik 5

BA-Jobportal und IAB-Stellenerhebung: Stellenanzeigen und -angebote nach Betriebsgrößen¹ in %²



Darstellung basierend auf Daten der Bundesagentur für Arbeit (BA) und einer Sonderauswertung des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

1 BA-Jobportal: Stichtag 26.2.2018; bereinigt (30 Tage). IAB-Stellenerhebung: 1. Quartal 2017.

2 Das Prozentuieren erfolgte hier ohne Berücksichtigung der Keine-Angabe-Fälle.

2018 - 01 - 0670

aus der IAB-Stellenerhebung, so werden diese Abweichungskennzahlen kleiner. Werden ausschließlich Stellenangebote der BA-Allianzpartner betrachtet, vergrößern sich diese Abweichungen dagegen deutlich. Das Maximum liegt in diesem Fall bei 30 Prozentpunkten, die die Betriebsgrößenklasse von 6 bis 50 Beschäftigten im BA-Jobportal im Vergleich zur IAB-Stellenerhebung unterrepräsentiert ist. An zweiter Stelle steht bei den Stellenangeboten der Allianzpartner die Betriebsgröße von 51 bis 500 Beschäftigten, die – gemessen an der relativen Häufigkeitsverteilung der IAB-Stellenerhebung – mit einer Abweichung von 26,9 Prozentpunkten überrepräsentiert ist.

4

Online-Stellenanzeigen von CEDEFOP

4.1 Datensatzbeschreibung

Das Europäische Zentrum für die Förderung der Berufsbildung CEDEFOP (Centre européen pour le développement de la formation professionnelle) analysiert Job-Bewerber-Matchingprozesse des Arbeitsmarktes insbesondere im Hinblick auf Optimierungspotenziale durch einen

gezielten Abgleich der geforderten und angebotenen berufsspezifischen Fachkompetenzen. Für die Analyse wurden in einem EU-Projekt in Zusammenarbeit mit dem italienischen interuniversitären Forschungszentrum für öffentliche Dienstleistungen CRISP (mit Sitz an der Universität Milano-Bicocca) umfangreiche Daten mit Stellenbeschreibungen und Bewerberprofilen aus Online-Jobportalen automatisiert herausgelesen. Die Forscherinnen und Forscher von CEDEFOP/CRISP haben bei insgesamt 16 Online-Jobportalen aus den fünf Staaten Deutschland, Italien, Vereinigtes Königreich, Tschechische Republik und Irland Beschreibungen offener Stellen gescrept (CEDEFOP, 2016).¹⁹

Für den deutschen Online-Stellenmarkt wurden von CEDEFOP/CRISP die drei hybriden Jobportale (1) Gigajob, (2) meinestadt und (3) Online-Stellenmarkt gescrept. Hybride Jobportale haben eigene Stellenanzeigen und Trefferlistenpartner und zeigen somit zusätzlich die Stellenanzeigen ihrer Partner an, um damit ihre eigene Trefferliste der Stellenanzeigen zu erhöhen. So zeigen die verwendeten Portale „Gigajob“ und „meinestadt“ unter anderem einen Teil der Stellenanzeigen der Bundesagentur für Arbeit an, deren (ebenfalls hybrides) Jobportal mit über einer Million Stellenanzeigen

¹⁹ Weitere Informationen unter <http://www.cedefop.europa.eu/de/events-and-projects/events/feasibility-study-real-time-lmi>.

gen das größte Jobportal Deutschlands ist. Die Datenqualität in Bezug auf die statistische Verwertbarkeit der Daten hybrider Portale hängt entscheidend davon ab, inwieweit das resultierende Dublettenproblem – durch die Portalbetreiber selbst oder nachträglich – zufriedenstellend gelöst werden kann. Die im Zuge des ESS-net-Projektes Big Data durchgeführten Untersuchungen haben diesbezüglich unter anderem bei Gigajob gravierende Mängel aufgezeigt. Hinsichtlich der Auswahl der gescrapten Jobportale bleibt ebenfalls zu bemängeln, dass das BA-Jobportal wegen einer ungünstigen – technisch bedingten – zeitlichen Performanz nicht einbezogen wurde.

Im Zeitraum Juni 2015 bis Februar 2016 wurden insgesamt mehr als 2,4 Millionen Stellenanzeigen gescrappt. Den größten Anteil nahmen mit 42,2% die Stellenanzeigen von meinestadt ein, gefolgt von denen des Portals „Online-Stellenmarkt“ (40,9%). Von den drei hybriden Jobportalen waren Stellenanzeigen von Gigajob mit 17,0% am wenigsten vertreten. Da – neben dem Dublettenproblem innerhalb eines jeden hybriden Jobportals – auch zwischen den drei Portalen Dubletten bestehen können, wurde von CEDEFOP/CRISP eine entsprechende Prüfung vorgenommen. Insgesamt konnten gut 320 000 Stellenanzeigen als Dubletten identifiziert werden. Der bereinigte Datensatz hatte somit eine Größenordnung von 2,1 Millionen Stellenanzeigen und war Ausgangsbasis der weiteren Analysen. [↘ Tabelle 5](#)

Tabelle 5

CEDEFOP-Online-Stellenanzeigen nach Jobportalen Juni 2015 bis Februar 2016

| | Anzahl | % |
|--|-----------|------|
| Gigajob | 417 862 | 17,0 |
| meinestadt | 1 036 645 | 42,2 |
| Online-Stellenmarkt | 1 006 886 | 40,9 |
| Insgesamt | 2 461 393 | 100 |
| Von CEDEFOP/CRISP identifizierte Dubletten | 320 846 | 13,0 |
| Insgesamt (bereinigt) | 2 140 547 | 87,0 |

Berechnungen basierend auf CEDEFOP-Daten.

Der Datensatz von CEDEFOP bestand aus einer Hauptdatei mit 17 Variablen und weiteren Dateien mit zusätzlichen Variablen, die mithilfe einer Identifikationsnummer zugespielt werden konnten. Während die Daten der Hauptdatei über Web Scraping erhoben wurden, sind die Werte aus den übrigen Dokumenten durch

Maschinelles Lernen ermittelt oder durch Text Mining²⁰ extrahiert worden. Dazu zählen beispielsweise der Wirtschaftszweig des Unternehmens, das Bildungsniveau oder die Einsatzorte.

Der bei der Bewertung von Online-Jobportalen erstellten Liste wichtiger Merkmale von Online-Stellenanzeigen (Körner/Rengers, 2016, hier: Seite 8)²¹ folgend, besitzt der CEDEFOP-Datensatz nur zwei wichtige Variablen: die Berufsbezeichnung und den unstrukturierten Volltext der Stellenanzeige. Übrige Variablen sind entweder redundant²² oder besitzen zu viele fehlende Werte (bis zu 98,3%). Trotz ihrer Bedeutung fehlte die Anzahl von Stellen je Anzeige als Variable vollkommen und hätte – soweit überhaupt vorhanden – aus dem Volltext der Stellenanzeige herausgelesen werden müssen.

Eine weitere Problematik, ähnlich zum BA-Datensatz, eröffnet die Variable des Veröffentlichungsdatums. War in der Stellenanzeige ein genaues Datum genannt, nahm die Variable diesen Wert an. Falls nicht, wurde als Variablenwert das Datum gesetzt, an dem die Anzeige das erste Mal gescrappt wurde. Im Datensatz konnte allerdings nicht nachvollzogen werden, auf welche Art die Variable bei den einzelnen Stellenanzeigen gebildet wurde, was nur eine unpräzise Bereinigung von veralteten Stellenanzeigen ermöglichte.

Eine weitere Auswahl aus dem bereits reduzierten Datensatz der oben genannten 2,1 Millionen Stellenanzeigen wurde nicht nur wegen veralteter Stellenanzeigen getroffen, sondern auch, weil es über den Zeitraum von Juni 2015 bis Februar 2016 eine deutliche Ungleichverteilung der Häufigkeiten der gescrapten Stellenanzeigen gab. Am größten war die Anzahl im Juni 2015, aber auch das vierte Quartal 2015 kam zu ähnlich hohen Werten. Im gesamten Juni 2015 wurden 288 000 Stellenanzeigen erfasst, wobei sie in Bezug auf ihre Herkunft mit 55,8% überwiegend dem Portal „meinestadt“ zuzuord-

20 Mithilfe des Text Minings wird versucht, aus textbasierten Quellen die Kerninformationen herauszufiltern. Zum Einsatz kommen dabei sowohl statistische als auch linguistische Analysemethoden. Ebenso können die Techniken des Maschinellen Lernens beim Text Mining Anwendung finden. Das Maschinelle Lernen erfolgt auf Basis von Musterdatenbeständen, bei denen die Strukturen bereits bekannt sind. Die in diesen Lerndaten (Trainingsdaten) erkannten Muster und Gesetzmäßigkeiten werden verallgemeinert und auf neue, unbekannte Datenbestände übertragen.

21 Siehe dazu auch Abschnitt 5.1.

22 Beispielsweise existiert neben der Variablen „Veröffentlichungsdatum“ im Format „DD.MM.YYYY“ die Variable „Veröffentlichungsjahr“ im Format „YYYY“, die keine zusätzliche Information liefert.

Tabelle 6

CEDEFOP-Online-Stellenanzeigen nach Jobportalen

| | Juni 2015 | | 4. Quartal 2015 | | | |
|---------------------|-----------|------|-----------------|------|-----------------------|------|
| | kumuliert | | | | Stichtag ¹ | |
| | Anzahl | % | Anzahl | % | Anzahl | % |
| Gigajob | 33 203 | 11,5 | 99 983 | 16,3 | 42 149 | 18,2 |
| meinestadt | 161 099 | 55,8 | 305 553 | 49,9 | 117 544 | 50,7 |
| Online-Stellenmarkt | 94 283 | 32,7 | 206 317 | 33,7 | 72 339 | 31,2 |
| Insgesamt | 288 585 | 100 | 611 853 | 100 | 232 032 | 100 |

¹ Durchschnittlicher Bestand des Quartals, basierend auf dem Mittelwert der drei Monatsendbestände (Stichtag) eines Quartals. Berechnungen basierend auf CEDEFOP-Daten.

nen waren. Im gesamten vierten Quartal 2015 kumulierte sich ihre Anzahl auf fast 612 000. Um den durchschnittlichen Bestand dieses Quartals zu ermitteln, wurde darüber hinaus der Mittelwert der drei Monatsendbestände (Stichtag) des Quartals bestimmt. Dem vierten Quartal 2015 kann ein Durchschnittsbestand von 232 000 Stellenanzeigen zugeordnet werden. Gut 50 % davon kamen vom Jobportal „meinestadt“, 31,2 % vom Online-Stellenmarkt und 18,2 % von Gigajob.

↘ **Tabelle 6**

4.2 Vergleich mit der IAB-Stellenerhebung

Die in Tabelle 6 dargestellte stichtagsbezogene Auswertung der gescrapten Stellenanzeigen kann mit den Ergebnissen der IAB-Stellenerhebung verglichen werden, die jeweils als durchschnittliche Bestandsgröße eines Quartals interpretiert werden können. Eine Gegenüberstellung dieser beiden Datenquellen bezogen auf das vierte Quartal 2015 zeigt ↘ **Tabelle 7**.

Tabelle 7

CEDEFOP-Online-Stellenanzeigen und IAB-Stellenerhebung: Vergleich der Stellenanzeigen, der Stellenangebote und der sofort zu besetzenden Stellen

| | CEDEFOP-Online-Stellenanzeigen, 4. Quartal 2015, Stichtag ¹ | IAB-Stellenerhebung, 4. Quartal 2015 |
|--|--|--------------------------------------|
| Stellenanzeigen | 232 032 | – |
| Stellenangebote | – | 1 047 079 |
| darunter: sofort zu besetzende Stellen | – | 822 800 |

¹ Durchschnittlicher Bestand des Quartals, basierend auf dem Mittelwert der drei Monatsendbestände (Stichtag) eines Quartals. Berechnungen basierend auf CEDEFOP-Daten.

Da im CEDEFOP-Datensatz keine Angabe zur Anzahl der Stellen je Stellenanzeige vorhanden war, konnte kein kohärenter Vergleich mit den Ergebnissen der IAB-Stellenerhebung vorgenommen werden. Im vierten Quartal 2015 kam diese zu einem Stellenangebot von insgesamt knapp 1,1

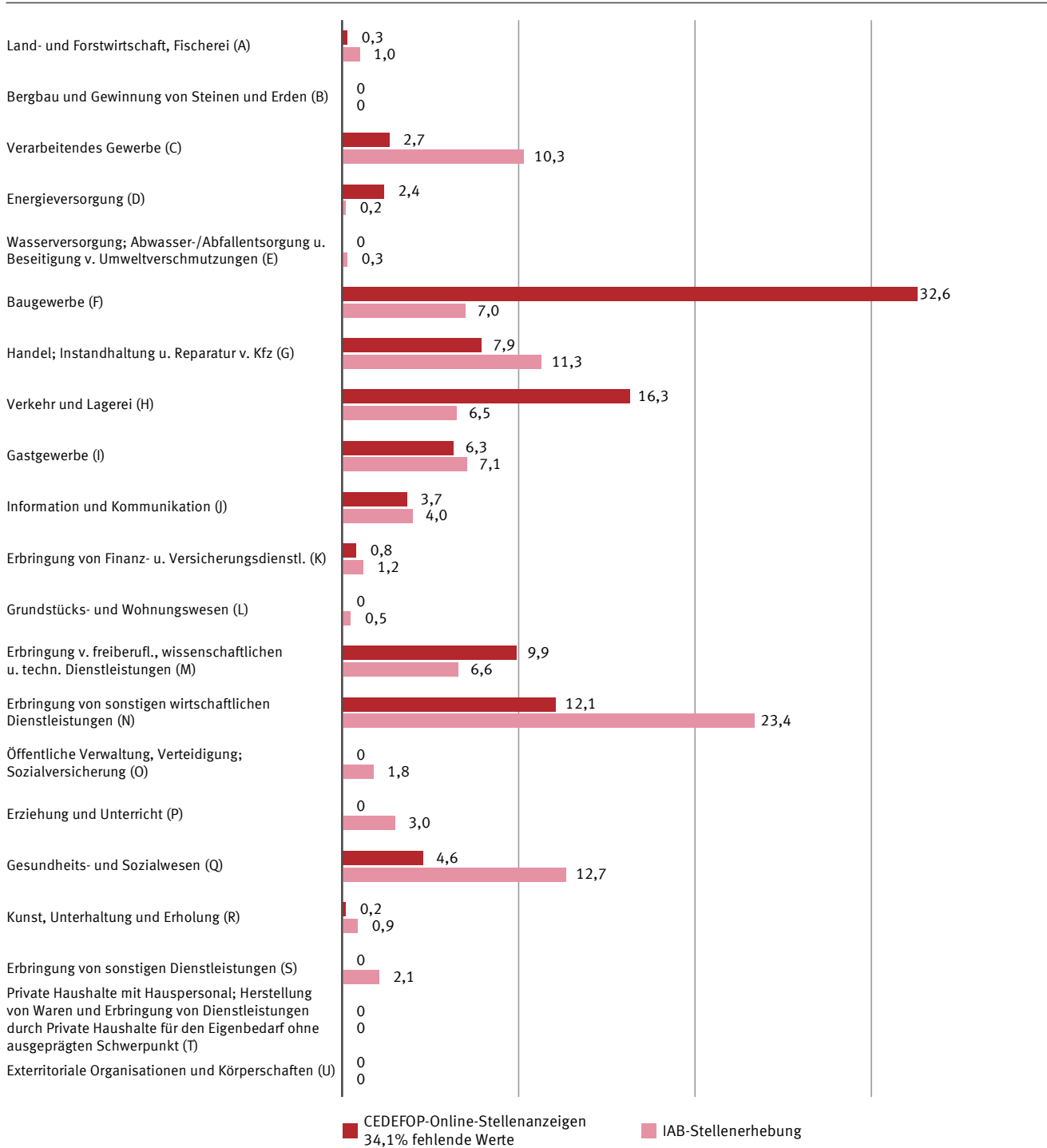
Millionen Stellen, darunter 823 000 sofort zu besetzende Stellen. Demgegenüber stehen die 232 000 Stellenangebote des CEDEFOP-Datensatzes. Legt man die durchschnittliche Zahl von 1,7 Stellen je Anzeige, die bei den Daten des BA-Jobportals beobachtet wurden, zugrunde, dann käme man rein rechnerisch bei den CEDEFOP-Daten auf ein Stellenangebot von knapp 400 000. Bei zusätzlicher Berücksichtigung einer Online-Rekrutierungsquote von 41 % (siehe Abschnitt 2.2) ergäbe sich ein gesamtwirtschaftliches Stellenangebot von 962 000 und damit annähernd in die Größenordnung der IAB-Stellenerhebung. Es gibt jedoch keine gesicherten Erkenntnisse darüber, ob ein solcher Zusammenhang dauerhaft unterstellt werden kann.

Die Strukturvergleiche der Daten des BA-Jobportals mit denen der IAB-Stellenerhebung haben gezeigt (siehe Abschnitt 3.2), dass sich die relativen Häufigkeitsverteilungen von Online-Stellenanzeigen von denen der Online-Stellenangebote teilweise deutlich unterscheiden (siehe auch die Grafiken 3 bis 5). Ein Strukturvergleich der Online-Stellenanzeigen des CEDEFOP-Datensatzes mit den Stellenangeboten der IAB-Stellenerhebung ist deshalb wenig aussagekräftig. Nachfolgend werden die Strukturen dennoch beispielhaft für die Differenzierung nach Wirtschaftszweig verglichen, um zu zeigen, dass sich die Erkenntnisse der obigen Analysen zum BA-Jobportal nicht einfach auf die CEDEFOP-Daten übertragen lassen.

Bezüglich der Zuordnungen zum Wirtschaftszweig im CEDEFOP-Datensatz ist festzuhalten, dass diese mithilfe des Maschinellen Lernens vorgenommen wurden und nicht mit den etablierten Wirtschaftszweigklassifikationen, die bei der IAB-Stellenerhebung eingesetzt werden. Der CEDEFOP-Datensatz hat bei den Angaben zum Wirtschaftszweig darüber hinaus 34,1 % fehlende Werte (Missings). Dabei fällt auf, dass Stellenanzeigen des

Grafik 6

CEDEFOP-Online-Stellenanzeigen und IAB-Stellenerhebung: Stellenanzeigen und -angebote nach Wirtschaftszweigen im 4. Quartal 2015 in %



Wirtschaftszweige der NACE Rev. 2 (Statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft). Berechnungen basierend auf CEDEFOP-Daten und einer Sonderauswertung der IAB-Daten.

Portals „Online-Stellenmarkt“ bei der Wirtschaftszweigangabe vollständig Missings aufwiesen.

↳ **Grafik 6** zeigt die relativen Häufigkeitsverteilungen, wobei die Prozentuierung bei den CEDEFOP-Daten ohne die Missingfälle erfolgte. Im Vergleich zur relativen Häufigkeitsverteilung der Stellenangebote der IAB-Stellenerhebung zeigen die CEDEFOP-Daten eine deutliche Überrepräsentation des Sektors F „Baugewerbe“, dem 32,6% der 232 000 Stellenanzeigen zugeordnet werden. Eine genauere Analyse zeigte, dass diese Stellenanzeigen vollständig dem Portal „meinestadt“ entnommen waren. Der Vergleich der Grafik 6 mit Grafik 3 zeigt keine analoge Systematik zu den obigen Unterschieden zwischen Online-Stellenanzeigen der BA und IAB-Stellenerhebung. Es zeigt sich auch keine einheitliche Systematik struktureller Unterschiede zwischen Online-Stellenanzeigen und Stellen der IAB-Erhebung im Vergleich zu früheren Analysen des Statistischen Bundesamtes. Bei letztgenannten wurde bezüglich der Online-Stellenanzeigen der Jobbörse „Stepstone“ beispielsweise festgestellt, dass der Sektor J „Information und Kommunikation“ deutlich überrepräsentiert war und dreimal häufiger auftrat als bei den Stellenangeboten der IAB-Stellenerhebung (Swier, 2016, hier: Seite 13).

Die beschriebenen Problematiken sind der mangelnden Datenqualität des CEDEFOP-Datensatzes geschuldet und führen dazu, dass eine sinnvolle Gegenüberstellung mit den Ergebnissen der IAB-Stellenerhebung nicht möglich ist. Daher wird hier im Folgenden auf weitere Vergleichsanalysen verzichtet. Bezüglich der ISCED-Bildungsklassifikation und der ISCO-Berufsklassifikation seien jedoch noch ein paar Auffälligkeiten beschrieben: Die Variable zur Kennzeichnung der ISCED-Bildungsklassifikation einer Online-Stellenanzeige wies im CEDEFOP-Datensatz 46,2% Missings auf. Eine Prozentuierung ohne Berücksichtigung dieser Missingfälle zeigte, dass 96% der berücksichtigten Stellenanzeigen einen Bachelorabschluss forderten. Dieser hohe Anteil wurde fast vollständig durch das Portal „meinestadt“ verursacht. Im Gegensatz dazu gab es bei den Angaben zur ISCO-Berufsklassifikation nur 3% fehlende Werte. Die Befüllung dieser Variablen erfolgte auf Basis des Jobtitels. Die Qualität des Klassifikationsverfahrens kann nur schwer überprüft werden und entspricht nicht dem Verfahren der IAB-Stellenerhebung.

5

Fazit und Ausblick

5.1 Big Data ESSnet 2016 bis 2018

Die größten Herausforderungen zum Thema internetgestützte Erfassung offener Stellen waren rechtliche Aspekte, Sicherheitsaspekte, die Bewertung von Jobs, die technische Umsetzung bei der Datengewinnung und nicht zuletzt die Frage der Datenqualität von Online-Stellenanzeigen. Dieser Beitrag hat auf den letztgenannten Aspekt den Schwerpunkt gelegt und die Ergebnisse der Überprüfung der Daten des BA-Jobportals und der Daten des CEDEFOP präsentiert. Beide Datensätze beruhten auf einer vereinbarten Datenlieferung, womit rechtliche und technische Schwierigkeiten des Web Scrapings umgangen werden konnten.

Bei der Messung der Datenqualität von Online-Stellenanzeigen stellt sich zunächst die Frage nach dem „Wie?“. Eine systematische Vorgehensweise führt in einem ersten Schritt zu Überlegungen bezüglich der Marktabdeckung von Online-Stellenanzeigen. Unterschiedliche Befragungen zur Bedeutung verschiedener Rekrutierungswege bei der Personalsuche zeigen, dass größere Unternehmen stärker als kleine und mittlere Unternehmen dazu neigen, Stellen online auszuscriben. Im Ergebnis nutzten zwischen 41 und 70% der Unternehmen Online-Jobportale. Man kann also davon ausgehen, dass der Online-Stellenmarkt nicht den gesamten Stellenmarkt abbildet (erstes Coverage-Problem). Dies hat Auswirkungen auf mögliche Ergebnisabgleiche mit anderen Datenquellen; so sollen die aus der stichprobenbasierten IAB-Stellenerhebung ermittelten offenen Stellen beispielsweise den gesamten Stellenmarkt repräsentieren. Durch die notwendige Auswahl von Jobportalen kann es bei der Abbildung des Online-Stellenmarktes zu einer weiteren fehlenden Repräsentativität kommen, sodass auch der Online-Stellenmarkt nicht vollständig abgebildet wird (zweites Coverage-Problem).

Die Auswahl von Jobportalen ist darüber hinaus mit weiteren Problemen verbunden. So hat die Einbeziehung mehrerer Portale notwendigerweise eine Dublettenbereinigung zur Folge. Dubletten können sowohl innerhalb eines jeden Portals als auch portalübergreifend

vorhanden sein. Der Verzicht auf die Einbeziehung hybrider Jobportale kann das Dublettenproblem minimieren. Es gibt jedoch auch noch andere Besonderheiten, die verstärkt zu Dubletten führen können. Bei der Einschaltung von Personalvermittlern kann es vorkommen, dass ein Arbeitgeber nicht nur ausschließlich die Personalvermittlung beauftragt, sondern – parallel zur Stellenanzeige des Personalvermittlers – auch in Eigenregie mit einer eigenen Stellenanzeige sucht.

Darüber hinaus besteht bei Anzeigen von Personalvermittlern das Problem, dass der eigentliche Arbeitgeber nicht genannt wird, was wiederum zu falschen oder fehlenden Zuordnungen bei der Angabe des Arbeitgebers führen kann. Das Fehlen wichtiger Merkmale ist ein Problem, das die Überprüfung der Datenqualität von Online-Stellenanzeigen erschwert. Doch was sind eigentlich wichtige Merkmale?

Im ersten Bericht zum Big Data ESSnet 2016 bis 2018 wurden wichtige Merkmale zur Beschreibung von Online-Stellenanzeigen identifiziert, die eine Antwort auf die Fragen „Was?“ – „Wann?“ – „Wo?“ – „Wer?“ geben. Konkret verbergen sich dahinter zehn Merkmale: (1) Jobtitel, (2) Beruf, (3) Bildung und (4) Art des Vertrags, differenziert nach dauerhaft/befristet und Vollzeit/Teilzeit, geben eine Antwort auf die Frage „Um was für einen Job handelt es sich?“. Die Merkmale (5) Datum der Veröffentlichung des Stellenangebots, (6) Datum des Endes der Bewerbungsfrist und (7) Datum der Stellenbesetzung beantworten das „Wann?“. Mit (8) Standort des Jobs sollte „Wo?“ geklärt werden und die Merkmale (9) direkter Arbeitgeber oder Zeitarbeitsfirma sowie (10) Wirtschaftszweig NACE des Arbeitgebers können die Frage „Wer?“ beantworten (Körner/Rengers, 2016, hier: Seite 8 f.).

Zum Zeitpunkt der Erstellung dieses ersten ESSnet-Berichts war die Bedeutung eines Merkmals zur Beantwortung der Frage „Wie viele?“ noch nicht im Fokus der Betrachtungen. Das Merkmal (11) Anzahl der Stellenanzeigen gehört allerdings mit zu den wichtigsten überhaupt. Dies haben insbesondere auch die obigen Analysen der beiden Datensätze von Online-Stellenanzeigen gezeigt. Die Stellenanzeigen des BA-Jobportals enthielten durchschnittlich 1,6 Stellen je Anzeige. Den betrachteten knapp 437 000 Online-Stellenanzeigen lagen somit insgesamt fast 695 000 Stellenangebote zugrunde. Doch auch dieser Wert reicht nicht an die Größenordnung der in der IAB-Stellenerhebung gemessenen

offenen Stellen insgesamt heran, da hier der Online-Stellenmarkt mit dem gesamten Stellenmarkt verglichen wird. Eine simple Hochrechnung mit der niedrigsten gemessenen Rekrutierungsquote von Online-Jobportalen (41 %) würde allerdings zu einer Überschätzung des gesamtwirtschaftlichen Stellenangebots führen.

Die Strukturanalysen zu den Online-Stellenanzeigen der BA beim Vergleich mit der IAB-Stellenerhebung haben gezeigt, wie stark sich die Struktur verändert, je nachdem, ob Stellenangebote oder Stellenanzeigen betrachtet werden. Erstaunlicherweise gab es bei den relativen Häufigkeitsverteilungen der Stellenanzeigen des BA-Jobportals teilweise größere Ähnlichkeiten mit derjenigen der Stellen der IAB-Erhebung, als bei Verwendung der Stellenangebote des BA-Jobportals. Auch diese Analysen haben die immense Bedeutung der Angabe der Stellenanzahl je Online-Stellenanzeige bestätigt [oben genanntes Merkmal (11)]. Aufgrund des Fehlens dieses Merkmals bei den Online-Stellenanzeigen des CEDEFOP-Datensatzes waren derartige Vergleichsanalysen nicht möglich. Neben diesem Mangel wiesen die Daten von CEDEFOP weitere große Schwächen auf: Durch die ausschließliche Einbeziehung von hybriden Jobportalen wurde zunächst die Dublettenproblematik stark ausgeweitet. Mit der Auswahl der gescrapten Jobportale, die sich auf drei hybride Jobportale beschränkte und dabei das größte Jobportal Deutschlands nicht mitberücksichtigte, bestanden darüber hinaus erhebliche Zweifel an einer repräsentativen Abbildung des Online-Stellenmarktes. Eine Überprüfung, inwiefern beispielsweise Online-Stellenanzeigen von kleineren Unternehmen vorlagen, konnte nicht durchgeführt werden, da das Merkmal „Betriebsgröße“ im Datensatz fehlte. Ebenfalls einen wesentlichen Beitrag zur Beurteilung der Datenqualität von Online-Stellenanzeigen könnte das obige Merkmal (9) leisten, nämlich die Angabe, ob bei einer Online-Stellenanzeige ein direkter Arbeitgeber oder eine Zeitarbeitsfirma angegeben ist.¹²³

23 Typisch für Zeitarbeitsfirmen beziehungsweise Betriebe der Arbeitnehmerüberlassung ist beispielsweise das Vorhalten eines sogenannten Portfolios, einer Adresskartei mit potenziellen Kandidatinnen und Kandidaten, aus der insbesondere der Bedarf zur kurzfristigen Bedienung von Auftragsspitzen gedeckt wird. „Wenn Betriebe Personal für ihr Portfolio suchen, ohne dass dahinter tatsächlich zu besetzende offene Stellen stehen, kommt es zu einer Überschätzung der Zahl der Stellenangebote, die direkt in ein Beschäftigungsverhältnis münden können. Auch im Falle von Personalsuchen für die Vermittlung an Dritte (Betriebe in anderen Branchen) kommt es zu einer Überschätzung der tatsächlich existierenden Stellenangebote im Bereich der Arbeitnehmerüberlassung“ (Brenzel und andere, 2016a, hier: Seite 48).

Sowohl beim Datensatz der Online-Stellenanzeigen des BA-Jobportals als auch insbesondere bei demjenigen des CEDEFOP fehlten bedeutende Merkmale vollständig oder aber waren von einer hohen Anzahl von Missing-Fällen gekennzeichnet. In Bezug auf die Merkmalsausprägungen gibt es jedoch noch ein weiteres Problem, denn bei den einzelnen Merkmalen (Variablen) muss zwischen originären und generierten Variablen unterschieden werden. Originäre Variablen sind diejenigen, die direkt aus strukturierten Informationen gewonnen werden, während generierte Variablen durch Verfahren des Text Mining oder mithilfe des Maschinellen Lernens gebildet werden. Im CEDEFOP-Datensatz war dies beispielsweise bei der Variablen zur Kennzeichnung des Wirtschaftszweigs der Fall. Hier wurde eine Zuordnung über das Maschinelle Lernen vorgenommen, was allerdings nicht mit den etablierten Wirtschaftszweigklassifikationen, wie sie unter anderem bei der IAB-Stellenerhebung eingesetzt werden, vergleichbar ist. An dieser Stelle ist ein stärkerer interdisziplinärer Austausch notwendig, sodass die Ergebnisse neuer digitaler Datenquellen zu denen traditioneller Datenquellen kohärent sind.

5.2 Big Data ESSnet 2018 bis 2020

Im Mai 2018 wurde ein weiteres Big Data-Projekt des European Statistical Network ausgeschrieben, welches von 2018 bis 2020 andauern soll. Das zukünftige Big Data ESSnet 2018 bis 2020 setzt sich aus drei komplexen Themenschwerpunkten zusammen. Zunächst geht es (I) um eine erste Implementierung neuer „Big Data“-Datenquellen in den statistischen Produktionsprozess, darüber hinaus sollen (II) neue Pilotstudien zu weiteren alternativen Datenquellen durchgeführt werden und schließlich (III) sollen die Arbeiten und Untersuchungen zu innovativen Anwendungen im Bereich der vertrauenswürdigen intelligenten Statistiken (Trusted Smart Statistics) ausgeweitet werden.

Im Zusammenhang mit den Überlegungen zur Fortsetzung der Arbeiten am Thema „Internetgestützte Erfassung offener Stellen“ ist der erste der drei oben genannten Themenschwerpunkte von Bedeutung, denn die Ergebnisse des ESSnets 2016 bis 2018 sind Grundlage des Ziels der ersten Implementierung neuer „Big Data“-Datenquellen in den statistischen Produktionsprozess. Neue „Big Data“-Datenquellen sollen bei vier Statistiken zum Einsatz kommen, darunter sind auch Online-Stellenangebote.

Bezüglich der Datengewinnung und Datenqualität von Online-Jobportalen tauchten im Rahmen des ESSnets 2016 bis 2018 Probleme auf, die bisher nicht alle zufriedenstellend gelöst werden konnten. Vor diesem Hintergrund ist ein Arbeitspaket des zukünftigen Big Data ESSnets 2018 bis 2020 auf die Implementierung der Big-Data-Datenquelle „Online-Jobportale“ in den Statistikproduktionsprozess ausgerichtet. Ziel des neuen Projektes sind Veröffentlichungen von Indikatoren zum Online-Stellenmarkt. Wegen der zu erwartenden eingeschränkten Datenqualität werden derartige Veröffentlichungen wahrscheinlich unter dem Label „Experimentelle Statistiken“ erfolgen. Dieses ist für die Veröffentlichung von Ergebnissen gedacht, die „hinsichtlich Harmonisierung, Erfassung oder Methodik noch nicht vollständig ausgereift sind“ (Eurostat, 2018)²⁴. Unter Berücksichtigung landesspezifischer Gegebenheiten müssen zunächst sinnvolle Indikatoren für den Online-Stellenmarkt entwickelt werden. Grundvoraussetzung ist jedoch ein gesicherter Datenzugang zu Online-Stellenanzeigen. Für Deutschland bieten sich unverändert dieselben Quellen an wie im Pilotprojekt. Dies sind zum einen weiterhin die Daten des Jobportals der Bundesagentur für Arbeit, da deren Analyse aufgrund der Größe und der Vielzahl an Informationen unabdingbar ist. Zum anderen sind dies weiterhin die Daten des Europäischen Zentrums für die Förderung der Berufsbildung. CEDEFOP wird auch künftig dauerhaft Web Scraping von Online-Jobportalen des deutschen Marktes vornehmen, zudem wird sich die Auswahl der Jobportale – nach den CEDEFOP-eigenen Erfahrungen und Beratungen mit den (deutschen) am ESSnet 2016 bis 2018 Beteiligten – deutlich erweitern. Es ist dabei davon auszugehen, dass die Datenqualität der CEDEFOP-Datensätze besser werden wird. Ein dauerhafter Datenzugang ist wegen der gewünschten Data-Sharing-Möglichkeiten innerhalb des ESS von hoher Wahrscheinlichkeit.

Dennoch stellt sich die Frage, ob diese Vorhaben für ein derartiges Implementierungsvorhaben ausreichend sind und man nicht von bereits bestehenden weltweiten Erfahrungen profitieren kann und sollte. [↘ Übersicht 2](#) gibt deshalb einen Ein- und Ausblick auf ausgewählte Beispiele von öffentlich-rechtlichen Institutionen, die bereits Online-Stellenanzeigen in der Arbeitsmarkt-

24 <http://ec.europa.eu/eurostat/de/web/experimental-statistics>; Zugriff am 10. August 2018.

Übersicht 2

Öffentlich-rechtliche Institutionen, die Online-Stellenanzeigen bei der Arbeitsmarktberichterstattung verwenden (Auswahl)

| Land | Institution | Indikatoren | Datengrundlage |
|--------------------------------|---|--|--|
| Australien | Australische Regierung (www.ga.gov.au/) | Internet Vacancy Index (IVI) | 3 größte Jobportale (SEEK, CareerOne and Australian JobSearch); Datenbereitstellung durch das Department of Jobs and Small Business |
| Österreich | AMS Arbeitsmarktservice Österreich, Dienstleistungsunternehmen des öffentlichen Rechts (www.ams.at/) | Qualifikationsbarometer | Inseratsdaten bis zum Erhebungsjahr 2015 aus Printmedien, keine Stellenanzeigen aus Online-Jobbörsen; Inseratsdaten ab dem Erhebungsjahr 2016 basieren auf einer Analyse des nationalen Online-Inserateaufkommens, aus der Big-Data-Plattform Jobfeed von der Firma Textkernel |
| Schweiz | Stellenmarktmonitor Universität Zürich (www.stellenmarktmonitor.uzh.ch) | Kombinierter Print-Online-Index ¹ | Web Scraping von 12 Jobportalen und 1 350 Unternehmenswebseiten |
| Vereinigte Staaten von Amerika | Wirtschaftsforschungsinstitut The Conference Board (www.conference-board.org) | The Conference Board Help Wanted OnLine (HWOL) | Help Wanted Advertising Index (HWI) of print ads wurde 2008 nach 55 Jahren eingestellt; ab 2005 Online-Stellenmarkt Help Wanted OnLine (HWOL); über 16 000 Onlinequellen; Datenlieferungen von der Firma Haver Analytics ; Zusammenarbeit mit Wanted Analytics , die seit November 2015 zur amerikanischen Firma CEB gehören, welche wiederum im April 2017 von Gartner, Inc. übernommen wurden. |
| | US-Bundesstaat Kalifornien: EDD Employment Development Department (https://edd.ca.gov) | HWOL ² -Statistiken | siehe Fußnote 2 |

In der Online-Ausgabe dieses Artikels sind auch die Internet-Verlinkungen auf die Indikatoren sowie die Datengrundlagen hinterlegt.

1 Nähere Informationen bei Sacchi (2014).

2 Siehe dazu die Ausführungen in der Zeile zum Wirtschaftsforschungsinstitut The Conference Board.


Übersicht 3

Firmen, die Online-Stellenanzeigen kommerziell erfassen und auswerten (Auswahl)

| Firma | Firmensitz | Beobachtetes Land | Produktenamen |
|-------------------|---|---|--|
| Burning Glass | Boston, Vereinigte Staaten (www.burning-glass.com) Edinburgh, Vereinigtes Königreich (www.burning-glass.com/uk) Christchurch, Neuseeland (www.burning-glass.com/au) | Analysetool kann prinzipiell in jedem Land eingesetzt werden. Zurzeit gibt es unter anderem folgende Case Studies: <ul style="list-style-type: none"> › Australien und Asien › Vereinigte Staaten: Missouri Economic Research Information Center › Vereinigte Staaten: Northeastern University › Vereinigte Staaten: University of Maryland, Baltimore County › Vereinigtes Königreich: Firma Active Informatics als offizieller Vertriebspartner von Burning Glass | Labour Insight™ |
| Haver Analytics | New York, Vereinigte Staaten (www.haver.com) | Australien, China, Europa, Japan, Kanada, Neuseeland, Vereinigte Staaten, Vereinigtes Königreich | |
| Gartner, Inc. | Hauptsitz in Stamford (Connecticut), Vereinigte Staaten (www.gartner.com) deutsche Standorte in München, Düsseldorf und Frankfurt am Main | weltweit tätig | Gartner TalentNeuron™ (früher: Wanted Analytics) |
| Indeed hiring lab | Indeed, Inc. Austin, Vereinigte Staaten (www.hiringlab.org/) | Australien, Frankreich, Deutschland, Kanada, Vereinigte Staaten, Vereinigtes Königreich | Auswertungen des eigenen Jobportals von indeed |
| Textkernel | Amsterdam, Niederlande (www.textkernel.com) | Daten für Europa (Belgien, Deutschland, Frankreich, Italien, Niederlande, Österreich, Spanien, Vereinigtes Königreich) und Nordamerika (Kanada, Vereinigte Staaten) | Jobfeed |

berichterstattung verwenden. Von besonderem Interesse sind dabei zwei Aspekte: zunächst die Frage des Datenzugangs und dann die Frage der Veröffentlichung geeigneter Indikatoren.

Übersicht 2 ist weiterhin zu entnehmen, dass private Drittanbieter²⁵ in Bezug auf die Datengewinnung eine große Rolle spielen. Der österreichische Arbeitsmarktservice AMS benutzt beispielsweise Online-Stellenanzeigen, die von der niederländischen Firma Textkernel kommerziell gesammelt und aufbereitet wurden. Das amerikanische Wirtschaftsforschungsinstitut The Conference Board veröffentlicht seit 2005 einen Index zum Online-Stellenmarkt, der auf über 16 000 Onlinequellen basiert. Die Datenlieferung kommt von der Firma Gartner, Inc.

↳ Übersicht 3 zeigt, welche weiteren Firmen Online-Stellenanzeigen kommerziell erfassen und auswerten. Es fällt auf, dass viele dieser Firmen weltweit tätig sind. Anbieter, die keinen Sitz in Deutschland haben, wie die Firma Textkernel, bieten dennoch Daten und Analysen des deutschen Online-Stellenmarktes an. Auch für Deutschland bleibt deshalb die Frage, ob ein Rückgriff auf die Datenzusammenstellungen kommerzieller Anbieter eine Option für die Gewinnung von Online-Stellenanzeigen sein kann. Zwar würde man damit rechtliche und technische Schwierigkeiten des Web Scrapings umgehen, die Beurteilung der Datenqualität ist jedoch bedeutend schwieriger. Dazu wäre zum einen die Aneignung eines tiefergehenden Wissens über Big-Data-Technologien generell erforderlich, das die Funktionsweisen von Algorithmen und Technologien wie Web Scraping, Machine Learning oder Text Mining umfasst. Zum anderen wäre erforderlich, dass kommerzielle Firmen der amtlichen Statistik gegenüber die „Black Box“ öffnen und unter anderem folgende Angaben offenlegen: die Auswahl der Jobportale, die eingesetzten Algorithmen zur Dublettenbereinigung, das genaue Vorgehen beim Einsatz von Techniken des Maschinellen Lernens oder Text Minings zur Erzeugung generierter Variablen. 

25 Siehe dazu auch die Ausführungen in Abschnitt 1.2.

LITERATURVERZEICHNIS

Bundesagentur für Arbeit (BA). *Download-Portal*. 2018a. [Zugriff am 18. April 2018].
Verfügbar unter: <https://download-portal.arbeitsagentur.de/files/>

Bundesagentur für Arbeit (BA). *Tabellarische Umsteigeschlüssel zur KldB 2010*. 2018b.
[Zugriff am 18. April 2018]. Verfügbar unter: <https://statistik.arbeitsagentur.de>

Blaudow, Christian. *Fortschritte und Herausforderungen beim Web Scraping – Automatisierung von Preiserhebungen im Internet*. In: METHODEN – VERFAHREN – ENTWICKLUNGEN. Nachrichten aus dem Statistischen Bundesamt. Ausgabe 1/2018, Seite 3 ff. Verfügbar unter: www.destatis.de

Brenzel, Hanna/Czepek, Judith/Kiesl, Hans/Kriechel, Ben/Kubis, Alexander/Moczall, Andreas/Rebien, Martina/Röttger, Christof/Szameitat, Jörg/Warning, Anja/Weber, Enzo. *Revision der IAB-Stellenerhebung – Hintergründe, Methode und Ergebnisse*. IAB-Forschungsbericht 4/2016. Nürnberg 2016a. [Zugriff am 29. August 2018].
Verfügbar unter: <http://doku.iab.de>

Brenzel, Hanna/Czepek, Judith/Kubis, Alexander/Moczall, Andreas/Rebien, Martina/Röttger, Christof/Szameitat, Jörg/Warning, Anja/Weber, Enzo. *Neueinstellungen im Jahr 2015 – Stellen werden häufig über persönliche Kontakte besetzt*. IAB-Kurzbericht 4/2016, Nürnberg 2016b. [Zugriff am 29. August 2018]. Verfügbar unter: <http://doku.iab.de>

Brunner, Karola. *Automatisierte Preiserhebung im Internet*. In: Wirtschaft und Statistik. Ausgabe 4/2014, Seite 258 ff.

Carnevale, Anthony P./Jayasundera, Tamara/Repnikov, Dmitri. *Understanding online job ads data. A Technical Report*. Washington DC, 2014. [Zugriff am 5. September 2018]. Verfügbar unter: <https://cew.georgetown.edu>

CEDEFOP. *Final Report. Project “Real-time labour market information on skill requirements: feasibility study and working prototype”*. 2016. [Zugriff am 3. September 2018]. Verfügbar unter <https://webgate.ec.europa.eu>

Crosswater Job Guide. *Jobbörsen von A – Z*. 2018. [Zugriff am 5. September 2018].
Verfügbar unter: <https://crosswater-job-guide.com>

Eurostat. *Experimentelle Statistiken – Übersicht*. [Zugriff am 5. September 2018].
<https://ec.europa.eu/eurostat>

Körner, Thomas/Rengers, Martina/Swier, Nigel/Metcalf, Liz/Jansson, Ingegerd/Wu, Dan/Nikic, Boro/Pierrakou, Christina. *Deliverable 1.1. Inventory and qualitative assessment of job portals (SGA-1)*. ESSnet Big Data, Work Package 1: Web scraping / Job vacancies 2016.

Rengers, Martina. *Internetbasierte Erfassung offener Stellen im Statistischen Bundesamt*. In: König, Christian und andere. (Herausgeber). Big Data – Chancen, Risiken, Entwicklungstendenzen. Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute. Wiesbaden 2018, Seite 61 ff.

LITERATURVERZEICHNIS

Sacchi, Stefan. *Lange Messreihen zur Entwicklung des Stellenangebots der Schweizer Wirtschaft: Kombiniertes Presse-Online-Index*. SMM Working Paper 2014-1. Zürich 2014. [Zugriff am 5. September 2018]. Verfügbar unter: www.stellenmarktmonitor.uzh.ch

Swier, Nigel/Hajnovic, Frantisek/Declite, Thomas/Rengers, Martina/Islam, Chris-Gabriel/Jansson, Ingegerd/Wu, Dan/Elezovic, Suad/Grahonja, Crt/Pierrakou, Christina/Bisotti, Eleni/Bergat, Maxime/Eidelman, Alexis/Alves, Rui/Fernandes, Maria-Jose. *Deliverable 2.2. Final Technical Report (SGA-2)*. ESSnet Big Data, Work Package 1: Web scraping / Job vacancies 2018.

Swier, Nigel/Hajnovic, Frantisek/Jansson, Ingegerd/Wu, Dan/Nikic, Boro/Pierrakou, Christina/Rengers, Martina. *Deliverable 1.3 Final Technical Report (SGA-1)*. ESSnet Big Data, Work Package 1: Web scraping / Job vacancies 2017.

Swier, Nigel/Jansson, Ingegerd/Wu, Dan/Nikic, Boro/Pierrakou, Christina/Körner, Thomas/Rengers, Martina. *Deliverable 1.2. Interim Technical Report (SGA-1)*. ESSnet Big Data, Work Package 1: Web scraping / Job vacancies 2016.

Weitzel, Tim/Eckhardt, Andreas/Laumer, Sven/Maier, Christian/von Stetten, Alexander/Weinert, Christoph/Wirth, Jakob. *Recruiting Trends 2015 – Eine empirische Untersuchung mit den Top-1.000-Unternehmen aus Deutschland sowie den Top-300-Unternehmen aus den Branchen Finanzdienstleistung, Health Care und IT*. Bamberg 2015. [Zugriff am 5. September 2018]. Verfügbar unter: www.uni-bamberg.de

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Sabine Bechtold

Redaktionsleitung: Juliane Gude

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im Oktober 2018

Das Archiv aller Ausgaben ab Januar 2001 finden Sie unter www.destatis.de/publikationen

Print

Einzelpreis: EUR 18,- (zzgl. Versand)

Jahresbezugspreis: EUR 108,- (zzgl. Versand)

Bestellnummer: 1010200-18005-1

ISSN 0043-6143

ISBN 978-3-8246-1072-3

Download (PDF)

Artikelnummer: 1010200-18005-4, ISSN 1619-2907

Vertriebspartner

IBRo Versandservice GmbH

Bereich Statistisches Bundesamt

Kastanienweg 1

D-18184 Roggentin

Telefon: +49 (0) 382 04 / 6 65 43

Telefax: +49 (0) 382 04 / 6 69 19

destatis@ibro.de

Papier: Metapaper Smooth, FSC-zertifiziert, klimaneutral, zu 61% aus regenerativen Energien

© Statistisches Bundesamt (Destatis), 2018

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.