

de Lazzer, Jakob; Rengers, Martina

**Article**

## Auswirkungen der Coronakrise auf den Arbeitsmarkt: experimentelle Statistiken aus Daten von Online- Jobportalen

WISTA – Wirtschaft und Statistik

**Provided in Cooperation with:**

Statistisches Bundesamt (Destatis), Wiesbaden

*Suggested Citation:* de Lazzer, Jakob; Rengers, Martina (2021) : Auswirkungen der Coronakrise auf den Arbeitsmarkt: experimentelle Statistiken aus Daten von Online-Jobportalen, WISTA – Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 73, Iss. 3, pp. 71-88

This Version is available at:

<https://hdl.handle.net/10419/235267>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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

# AUSWIRKUNGEN DER CORONAKRISE AUF DEN ARBEITSMARKT: EXPERIMENTELLE STATISTIKEN AUS DATEN VON ONLINE-JOBPORTALEN

Dr. Jakob de Lazzer, Dr. Martina Rengers

📌 **Schlüsselwörter:** EXDAT – Online-Stellenmarkt – Online-Jobportal – Datenqualität – Arbeitsmarktberichterstattung

## ZUSAMMENFASSUNG

Die Coronakrise hat den Bedarf an schnell verfügbaren Daten zur Beschreibung der aktuellen wirtschaftlichen Lage stark ansteigen lassen. Dazu zählen auch Informationen zu kurzfristigen Entwicklungen auf dem Arbeitsmarkt. Das Statistische Bundesamt befasst sich bereits seit einiger Zeit damit, aus Online-Stellenanzeigen Indikatoren zur kurzfristigen Arbeitsmarktberichterstattung zu entwickeln und veröffentlicht diese zunächst als experimentelle Statistiken (EXDAT). Der Artikel beschreibt diesen Analyseprozess, prüft die Qualität der so erhobenen Daten und stellt einen monatlichen Index zur (konjunkturellen) Entwicklung von Online-Stellenanzeigen vor.

📌 **Keywords:** EXDAT – online job market – online job portal – data quality – labour market reports

## ABSTRACT

*Due to the coronavirus crisis, the demand for data describing the current economic situation that are early available has sharply increased. This includes information on short-term developments in the labour market. For some time already, the Federal Statistical Office has developed indicators, based on online job advertisements, that can be used for short-term labour market reports and has published them, at first in the form of experimental statistics (EXDAT). The article describes that process of analysis, examines the quality of data collected in this way and presents a monthly index of the development of online job advertisements (in the context of short-term economic development).*



**Dr. Jakob de Lazzer**

ist Volkswirt und Wissenschaftlicher Mitarbeiter im Referat „Arbeitsmarkt“ des Statistischen Bundesamtes. Hauptsächlich befasst er sich mit der Analyse von Online-Stellenmärkten im Rahmen des Projekts ESSnet Big Data: Online-Job-Vacancies. Dabei stehen das Webscraping von Online-Stellenbörsen sowie die Aufbereitung und Qualitätseinstufung der gewonnenen Daten im Vordergrund.



**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

### Die Coronakrise und ihre Folgen für die Arbeitsmarktstatistik

Ende Januar 2020 gab es die erste offizielle Bestätigung, dass das SARS-CoV-2-Virus Deutschland erreicht hatte. Eineinhalb Monate später wurde der erste Lockdown mit Abstands- und Hygieneregeln, Kontakt- und Veranstaltungsverbieten sowie Schließungen von Restaurants und Dienstleistungsbetrieben verhängt. Damit verbunden stieg der Bedarf an aktuellen Informationen stark an – nicht nur zu Infektions- und Sterbefallzahlen, sondern in den folgenden Monaten auch zu Wirtschafts- und Arbeitsmarktdaten. Schnell erfolgte der Ruf nach zusätzlichen, über die traditionellen Statistiken hinausgehenden Daten. Außerdem erschien es dringend notwendig, die Aktualität bisheriger Statistiken zu erhöhen.

Die Einschränkungen durch die COVID-19-Pandemie erschwerten jedoch die Datengewinnung für einige etablierte Statistiken, beispielsweise den Mikrozensus (und die hierin integrierte Arbeitskräfteerhebung). Vor der Coronakrise wurden für den Mikrozensus vor allem persönliche Interviews vor Ort in den Haushalten durchgeführt, im Berichtsjahr 2019 betraf dies 65 % der Interviews (Statistisches Bundesamt, 2020). Genau diese waren aber aufgrund des Lockdowns nicht mehr möglich; die Befragungen wurden nun überwiegend telefonisch, mittels Selbstausfüller-Fragebogen oder mithilfe des neu eingeführten Online-Fragebogens durchgeführt. Welche Effekte dies auf die Qualität der Statistik hat, ist verstärkt zu analysieren.

Diese Auswirkungen der Coronakrise auf die Statistikproduktion gab es nicht nur in Deutschland. Die oben erwähnte Arbeitskräfteerhebung (LFS – Labour Force Survey) basiert auf EU-Verordnungen und wird in den Mitgliedstaaten der Europäischen Union (EU), vier EU-Beitrittskandidatenländern und drei Ländern der Europäischen Freihandelszone (EFTA) durchgeführt. Das Statistische Amt der Europäischen Union (Eurostat) hat die pandemiebedingten Problematiken aufgegriffen und auf einer [Corona-Sonderseite](#) zu Folgen für diese und andere Statistiken zur Verfügung gestellt (Eurostat, 2020a; Eurostat, 2020b). Arbeitskräfteerhebungen werden nicht nur europaweit, sondern in den meisten Ländern der Welt durchgeführt. Ihre statistischen Konzepte

und Definitionen orientieren sich an dem Labour-Force-Konzept der Internationalen Arbeitsorganisation (International Labour Organization – ILO). Für die Produktion von Arbeitsmarktstatistiken hat deshalb auch die ILO eine spezielle Webseite „[COVID-19 and labour statistics](#)“ zusammengestellt. Die dort zur Verfügung stehenden Dokumente beschreiben zum einen die durch COVID-19 entstandenen Schwierigkeiten bei der Datengewinnung und geben zum anderen Hilfestellungen und Empfehlungen zur bestmöglichen Lösung dieser Schwierigkeiten.<sup>1</sup>

Die COVID-19-Pandemie hat die Entwicklung, alternative neue digitale Datenquellen für Statistikzwecke nutzbar zu machen, weiter beschleunigt. So wurde in den Vereinigten Staaten bereits Mitte des Jahres 2020 ein sogenannter Real Time Economic Tracker entwickelt. Opportunity Insights, eine amerikanische Organisation mit Sitz an der Harvard University, hat dazu verschiedene Indikatoren zur schnellstmöglichen Abbildung der wirtschaftlichen Entwicklung in „Echtzeit“ erarbeitet. Für den Bereich des Arbeitsmarkts wurde aus Daten von Online-Jobportalen wöchentlich die Veränderung der Stellenausschreibungen, indiziert auf den Zeitraum vom 4. bis 31. Januar 2020 mit einer zeitlichen Verzögerung (Timelag) von drei bis fünf Tagen nach dem Ende einer Arbeitswoche veröffentlicht.<sup>2</sup>

Für die Indikatoren zum Online-Stellenmarkt verwendete die amerikanische Forschergruppe der Harvard University Daten von Burning Glass, einem privaten Unternehmen mit Hauptsitz in Boston (Massachusetts), das Online-Stellenanzeigen kommerziell erfasst und ausgewertet. Zu den bekanntesten Drittanbietern von Online-Stellenanzeigen gehört neben den amerikanischen Firmen Burning Glass, Haver Analytics und Gartner Inc. die niederländische Firma Textkernel mit Hauptsitz in Amsterdam.

Diese und andere Firmen nutzen eine Vielzahl von Daten aus Online-Jobportalen, die sie unter anderem mithilfe von Web-Scraping-Technologien gewinnen. Normalerweise haben ihre kommerziellen Marktanalysen und Softwarelösungen zum Ziel, Angebot und Nachfrage am Arbeitsmarkt zusammenzuführen, Matching-Prozesse

1 Bereits im April 2020 sind Hinweise für die Arbeitskräfteerhebung erschienen (ILO, 2020a; ILO, 2020b).

2 Eine ausführliche Beschreibung zum Echtzeit-Tracker mit allen Indikatoren zu wirtschaftlichen Folgenabschätzungen der Coronakrise in den Vereinigten Staaten ist bei Chetty und anderen (2020) zu finden.

zu beschleunigen und dies als Dienstleistung anzubieten. Während der Coronakrise reagierten Unternehmen wie Burning Glass und Textkernel, aber auch die Jobportalanbieter Indeed und LinkedIn, jedoch zusätzlich mit sehr zeitnahen Veröffentlichungen und Informationen auf speziellen Unterseiten ihrer eigenen Homepages. Sie betonen damit die Bedeutung von Echtzeitinformationen aus Online-Jobportalen. Am frühesten reagierte Burning Glass, das bereits vier Tage nach Ende einer Berichtswoche kalenderwochenspezifische Änderungen von Online-Stellenanzeigen in einer Excel-Datei zum Download auf einer [COVID-19-Sonderseite](#) für alle Bundesstaaten der Vereinigten Staaten bereitstellte. Burning Glass bewirbt dies mit folgenden Worten: «*The coronavirus pandemic has had an unprecedented impact on the world economy. Burning Glass Technologies is using its database of more than a billion current and historical job postings to measure this economic shock and how it affects the job market. We'll be updating this page frequently with the latest data.*» (Burning Glass Technologies, 2021). Entsprechende COVID-19-Sonderseiten bieten auch die Firmen Indeed und LinkedIn an, die mit „Hiring Lab“ beziehungsweise „Economic Graph“ Daten ihrer eigenen Jobportale auswerten.<sup>3</sup>

Der folgende Beitrag beschreibt die Arbeiten im Statistischen Bundesamt, aus Online-Stellenanzeigen Indikatoren zur kurzfristigen Arbeitsmarktberichterstattung zu entwickeln und zu veröffentlichen. Kapitel 2 stellt Untersuchungen zum Online-Stellenmarkt vor, Kapitel 3 beschreibt die Daten, die für die Untersuchungen genutzt wurden. Wie im Zuge experimenteller Statistiken ein Online-Stellenindex zur Entwicklung des Online-Stellenmarktes konzipiert wurde, ist Thema von Kapitel 4, ebenso ein Vergleich mit externen Daten. Das letzte Kapitel zieht ein Fazit zur Nutzung neuer digitaler Datenquellen für die Berichterstattung zum Arbeitsmarkt und gibt einen Ausblick auf weitere Vorhaben auf europäischer Ebene in diesem Zusammenhang.

---

<sup>3</sup> Siehe dazu auch Rengers (2020) sowie Špeh und andere (2020, hier: Seite 16 ff.).

## 2

---

### Der Online-Stellenmarkt

---

#### 2.1 EU-Projekte zu Online-Stellenanzeigen

---

Die Frage, ob und inwieweit Stellenanzeigen von Online-Jobportalen für die amtliche Statistik nutzbar gemacht werden können, ist nicht erst als Folge der COVID-19-Pandemie entstanden. Das Europäische Statistische System (ESS) hat bereits 2015 mit der Big-Data-Strategie (ESS Big Data Action Plan and Roadmap) ein erstes Big-Data-Projekt des European Statistical System Network (ESSnet) initiiert und mit mittlerweile mehreren Folgeprojekten forciert.

Das Potenzial von Online-Stellenanzeigen für die Abbildung der Arbeitskräftenachfrage wurde zuerst im Big Data ESSnet 2016-2018<sup>4</sup> in einer Pilotstudie zur internetgestützten Erfassung offener Stellen untersucht. Die größten Herausforderungen waren dabei insbesondere die Dublettenbeseitigung, die unvollständige Abdeckung der Gesamtheit der offenen Stellen durch Online-Stellenausschreibungen (erstes Coverage-Problem) und die Schnelllebigkeit in der Branche der Jobportalanbieter. Zudem machte die Vielzahl der Anbieter ein vollständiges Assessment aller Jobportale nahezu unmöglich und zwang zu einer Auswahl davon (zweites Coverage-Problem).

Auch die Datengewinnung von Online-Jobportalen war und ist mit rechtlichen und technischen Schwierigkeiten verbunden. Die statistischen Ämter mehrerer Länder hatten deshalb versucht, Datenlieferungen von Jobportalanbietern vertraglich zu vereinbaren. Gleichzeitig wurde geprüft, ob innerhalb des ESS oder der EU bereits vergleichbare Projekte von anderen Institutionen bestehen und Synergieeffekte genutzt werden können. In diesem Kontext richtete sich der Fokus auf die Arbeiten des

---

<sup>4</sup> 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 einer zweiten Projektphase (ab August 2017) kamen mit Belgien, Dänemark, Frankreich und Portugal weitere Länder hinzu. Eine ausführliche Beschreibung dieses Projekts und seiner Ergebnisse für Deutschland findet sich bei Rengers (2018b).

Europäischen Zentrums für die Förderung der Berufsbildung (Centre européen pour le développement de la formation professionnelle – CEDEFOP).

CEDEFOP analysiert Job-Bewerber-Matchingprozesse des Arbeitsmarkts insbesondere im Hinblick auf Optimierungspotenziale durch einen gezielten Abgleich der geforderten und angebotenen berufsspezifischen Fachkompetenzen. Für ihre Analysen nutzt die Agentur der EU umfangreiche Daten mit Stellenbeschreibungen und Bewerberprofilen aus Online-Jobportalen verschiedener europäischer Länder, die automatisiert ausgelesen werden.<sup>15</sup>

Durch kooperative Zusammenarbeit konnten die CEDEFOP-Daten dem ESSnet Pilotprojekt 2016-2018 als Prototyp zu Forschungszwecken zur Verfügung gestellt werden.

In einem zweiten Big-Data-ESSnet-Projekt 2018-2020 war ein Arbeitspaket auf die Implementierung der Big-Data-Datenquelle „Online-Jobportale“ in den Statistikproduktionsprozess ausgerichtet.<sup>16</sup> Von Beginn an wurden Data-Sharing-Möglichkeiten berücksichtigt und die Arbeiten von CEDEFOP einbezogen. Mit der Nutzung der CEDEFOP-Daten durch das ESSnet Big Data sollten die gegenseitigen technischen und methodisch-fachlichen Austausch intensiviert, Synergien genutzt und gleichzeitig ein dauerhafter und gesicherter Datenzugang zu Online-Stellenanzeigen gewährleistet werden.<sup>17</sup>

---

5 Der Datengewinnungsprozess von CEDEFOP erfolgt in Zusammenarbeit mit dem interuniversitären Forschungszentrum für öffentliche Dienstleistungen CRISP (mit Sitz an der Universität Milano-Bicocca) und der Firma Tabulaex. Tabulaex ist ein Spin-off-Unternehmen der Universität Milano-Bicocca, das im November 2018 von der amerikanischen Firma Burning Glass erworben wurde, aber weiterhin eigenständig handelt.

6 An dem Arbeitspaket der Implementierung von Online-Stellenanzeigen beteiligten sich neben dem Statistischen Bundesamt die nationalen statistischen Ämter aus Bulgarien, Frankreich (DARES), Irland, Italien, Litauen, Polen, Rumänien, Schweden, Slowenien, der Schweiz und dem Vereinigten Königreich.

7 Eine solche Strategie der gemeinsamen Nutzung von europäischen Daten zu Online-Stellenanzeigen wurde 2018 auf der DGINS-Konferenz (Konferenz der Generaldirektorinnen und Generaldirektoren der nationalen statistischen Institute) in Bukarest vorgestellt (Descy und andere, 2018).

## 2.2 Assessment der Jobportale

---

In Deutschland gibt es mehr als 1 000 Jobportale. Die inhaltlich vollständige Erfassung der Online-Stellenanzeigen all dieser Portale wäre – trotz aller Automatisierungen – nicht nur mit einem hohen (Rechner-) Zeitaufwand verbunden, sondern würde auch immense Speicherkapazitäten (Daten-Storage-Problem) benötigen. Daher muss eine Auswahl getroffen werden.

Basis der Auswahlentscheidung kann eine Beurteilung des Stellenwerts eines Jobportals sein.<sup>18</sup> [↘ Übersicht 1](#) zeigt für die Jahre 2018 bis 2020 die Ranglisten zweier externer Anbieter, deutschlandsbestejobportale.de (kurz: DBJ) und crosswater-job-guide.com (kurz: Crosswater). Sie nehmen (regelmäßig) Bewertungen und Rankings der wichtigsten Jobportale Deutschlands vor und machen diese öffentlich zugänglich. Das Ranking von DBJ basiert auf Bewertungen durch Nutzerinnen und Nutzer sowie von Arbeitgebern, das von Crosswater auf dem Verhältnis der Besucherzahlen zur Zahl der Stellenanzeigen.

Auch wenn StepStone Deutschland bei den allgemeinen Jobportalen in den betrachteten Jahren bei beiden Ranglisten stabil den ersten Platz belegt, zeigt sich dennoch eine Veränderung anderer Portale über den Zeitverlauf. FAZjob belegte beispielsweise 2018 bei DBJ den zehnten Platz, tauchte aber ansonsten nie wieder unter den Top Ten auf. Gleiches gilt für ebay-kleinanzeigen, die 2020 erstmalig von DBJ auf Platz acht notiert wurden. Ebenso gibt es neben vielen Übereinstimmungen deutliche Unterschiede zwischen den beiden Ranglisten. Das Jobportal der Bundesagentur für Arbeit, kurz häufig nur Arbeitsagentur genannt, belegt bei Crosswater in den Jahren 2018 und 2019 den siebten Platz, kommt bei DBJ aber nicht unter die zehn Besten. Indeed nimmt bei Crosswater 2019 den dritten Platz ein, erscheint ansonsten jedoch nicht bei den zehn besten allgemeinen Jobportalen. Stattdessen wurde Indeed in allen drei Jahren von DBJ auf den ersten Platz unter den Jobsuchmaschinen gesetzt. Darin zeigt sich die Schwierigkeit, Jobportale eindeutig bestimmten Typen zuzuweisen. So ist Indeed ein Anbieter, der sowohl eigene als auch Stellenanzeigen anderer auflistet, also ein hybrides Jobportal, das nicht nur als reine Jobsuchmaschine fungiert.

---

8 Ausführliche Informationen zum Assessment von Jobportalen bieten Körner und andere (2016) sowie Rengers (2018a, hier: Seite 62 ff.).

# Auswirkungen der Coronakrise auf den Arbeitsmarkt: experimentelle Statistiken aus Daten von Online-Jobportalen

## Übersicht 1

### Externe Ranglisten für allgemeine Jobportale und Jobsuchmaschinen<sup>1</sup>

Rang	2020 <sup>1,2</sup>	2019		2018	
	DBJ	Crosswater	DBJ	Crosswater	DBJ
<b>Allgemeine Jobportale</b>					
1	StepStone DE	StepStone DE	StepStone DE	StepStone DE	StepStone DE
2	stellenanzeigen.de	meinestadt.de	Jobware (geteilter Platz 2)	Indeed.de	regio-Jobanzeiger
3	jobware (geteilter Platz 3)	Indeed.de	regio-jobanzeiger (geteilter Platz 2)	XING	XING (geteilter Platz 3)
4	regio-jobanzeiger (geteilter Platz 3)	XING	XING (geteilter Platz 3)	Monster Deutschland	Stellenanzeigen.de (geteilter Platz 3)
5	meinestadt.de (geteilter Platz 3)	Monster Deutschland	stellenanzeigen.de (geteilter Platz 3)	LinkedIn	Jobstairs (geteilter Platz 3)
6	XING	LinkedIn	N/A	meinestadt.de	Jobware (geteilter Platz 3)
7	LinkedIn	Arbeitsagentur	N/A	Arbeitsagentur	linkedin
8	ebay-kleinanzeigen	mobileJob	N/A	Jobstairs.de	kalaydo
9	kalaydo	Jobware	N/A	Stellenanzeigen.de	meinestadt
10	careerbuilder, jobs, jobscout24	Stellenanzeigen.de	N/A	Jobware	fazjob
<b>Jobsuchmaschinen</b>					
1	Indeed	Kimeta	Indeed	Kimeta	Indeed
2	stellenonline	Jobbörse.de	Kimeta (geteilter Platz 2)	Jobbörse.de	Gigajob
3	kimeta	Jobrapido.de	stellenonline (geteilter Platz 2)	Jobrapido.de	stellenonline
4	jobbörse.de	Jobs.de	gigajob	Jobboerse.de	Kimeta
5	N/A	Jobboerse.de	N/A	Gigajob	jobbörse.de
6	N/A	Jobworld	N/A	Jobturbo	google
7	N/A	Gigajob	N/A	Jobworld	jobboerse.com
8	N/A	Jobturbo	N/A	Jobsuma.de	jobrapido
9	N/A	Opportuno	N/A	Adzuna	jobsuma
10	N/A	CESAR	N/A	N/A	backinjob

1 Ohne Ranglisten für Spezialjobbörsen, die fachspezifische Stellenanzeigen für naturwissenschaftliche, medizinische, Ingenieurs- und ähnliche Berufe anbieten.

2 Für 2020 keine Angaben bei Crosswater vorhanden (Stand: 4. März 2021).

Crosswater = Crosswater Job Guide (Stand: 15. Februar 2021).

2019: <https://crosswater-job-guide.com/archives/75244/die-besten-jobboersen-2019-stepstone-dominiert-staufenbiel-institut-und-jobvector-glaenzen-kimeta-bleibt-beste-jobsuchmaschine/>

2018: <https://crosswater-job-guide.com/archives/68925/die-besten-jobboersen-2018-stepstone-und-indeed-behauptet-jobvector-glaentz-social-media-recruiting-schwaechelt/>

DBJ = Deutschlands Beste Jobportale (Stand: 15. Februar 2021).

2020: <http://deutschlandsbestejobportale.de/index.html>

2019: <http://deutschlandsbestejobportale.de/13.html>

2018: <http://deutschlandsbestejobportale.de/Deutschlandsbeste-Joportale-Jobboersen-2018.html>

Anmerkung: jobbörse.de und jobboerse.de sind zwei unterschiedliche Anbieter.

Dasselbe gilt für viele der im unteren Teil der Übersicht gelisteten Jobsuchmaschinen, von denen ein Großteil Arbeitgebern die Möglichkeit bietet, Stellenanzeigen zu schalten.

Aus Sicht der Datenqualität für die Arbeitsmarktberichterstattung ist bezüglich dieser Ranglisten zu berücksichtigen,

dass bei der Beurteilung der Jobportale für statistische Zwecke weitere Kriterien eine Rolle spielen. So ist das Potenzial für eine Vielzahl von Dubletten bei der Einbeziehung von Jobsuchmaschinen und hybriden Portalen (zum Beispiel meinestadt und Gigajob) deutlich höher als bei reinen Jobbörsen (zum Beispiel StepStone und Monster). Reine Jobbörsen enthalten

nur Stellenanzeigen, die direkt bei ihnen aufgegeben wurden. Zudem können Aufbau und Struktur einer Portal-Webseite nicht nur in technischer Hinsicht bei der Datengewinnung, sondern auch für den Informationsgehalt entscheidend sein.

### 3

## Die CEDEFOP-Daten

### 3.1 Datenbeschreibung

Die nationalen Statistikämter erhielten die CEDEFOP-Daten mit Informationen zu Online-Stellenanzeigen in aufbereiteter, automatisch klassifizierter Form. Der Datentransfer erfolgte über eine eigens eingerichtete EU-Dataplattform (DataLab), auf der bis zum Dezember 2020 vier Datensätze mit einem Timelag von jeweils drei Monaten bereitgestellt wurden.

Allerdings ist kein Zugriff auf die Rohdaten oder den Volltext der Stellenanzeigen möglich, zudem ist der Prozess der Datenaufbereitung nur unvollständig dokumentiert. Darüber hinaus enthalten die Daten keine Angaben zur Anzahl der Stellen je Stellenausschreibung.<sup>9</sup>

Für Deutschland wurden, laut CEDEFOP, Stellenanzeigen von 159 Jobportalen erhoben. Die Datenerfassung

9 Frühere Untersuchungen mit Daten des Jobportals der Bundesagentur für Arbeit haben beispielsweise durchschnittlich 1,6 Stellen je Anzeige gezeigt. So suchen Arbeitgeber möglicherweise Mitarbeiterinnen und Mitarbeiter für mehrere Standorte oder mehrere Beschäftigte mit ähnlichem Stellen- und Anforderungsprofil für einen Standort (Rengers, 2018b, hier: Seite 28).

erfolgt über API, Scraping oder Crawling.<sup>10</sup> Datenqualität und -stabilität nehmen in der Regel mit der Reihenfolge der genannten Erfassungsmethoden ab.

### 3.2 Datenqualität

#### Revisionen, verwendete Jobportale und Anzahl an Stellenanzeigen

Bei jeder neuen Bereitstellung von CEDEFOP-Daten wurden drei zusätzliche Berichtsmonate ergänzt und gleichzeitig für die zurückliegenden Monate Daten revidiert. [↘ Tabelle 1](#) zeigt die wichtigsten Merkmale dieser Datensätze, als Versionen V1, V4, V5 und V8 bezeichnet.<sup>11</sup> Alle vier Datensätze beginnen mit Stellenanzeigen des Berichtsmonats Juli 2018. Die Datensätze unterscheiden sich dabei stark in der Anzahl der Stellenanzeigen je Monat und auch in der Anzahl der verwendeten Jobportale. Bei den jeweils neueren Versionen des Datensatzes erfolgten weitreichende Änderungen an bestehenden Beobachtungen. Da es nicht möglich ist, retroaktiv neue Daten von Online-Jobportalen zu erheben, ist dies nur durch Änderungen in der Datenaufbereitung zu erklären.

Die von CEDEFOP genannten 159 Jobportale werden zu keinem Zeitpunkt alle gleichzeitig zur Informationsgewinnung genutzt. Die Anzahl der durchschnittlich je Monat tatsächlich verwendeten Jobportale schwankt bei

10 Bei der Erfassung über Application Programming Interface (API) werden vom Portalbetreiber bereitgestellte Schnittstellen zur Datenabfrage genutzt. (Web-)Scraping bezeichnet die Erhebung klar umrissener Inhalte durch speziell an eine Webseite angepasste Erfassungssoftware. (Web-)Crawling dagegen ist die Erfassung von Webinhalten mit einer Software, die automatisch und unspezifisch nach passenden Inhalten sucht.

11 Über die Datensätze V2, V3, V6 und V7 ist nichts bekannt. Möglicherweise handelt es sich dabei um interne Zwischenstände des CEDEFOP.

**Tabelle 1**  
Merkmale der verschiedenen Datensätze (einschließlich Dubletten)

	Berichtsmonate	Monate	Stellenanzeigen		Tatsächlich verwendete Jobportale <sup>1</sup>	Darunter: Jobportale <sup>1</sup> , von denen mindestens ... Stellenanzeigen gewonnen wurden		
			insgesamt	je Monat		1 000	50 000	100 000
			Anzahl	Mill.		Anzahl		
Datensatz V1	Juli 2018 bis Dezember 2019	18	17,0	0,9	92	20	5	3
Datensatz V4	Juli 2018 bis März 2020	21	44,0	2,1	88	21	6	5
Datensatz V5	Juli 2018 bis Juni 2020	24	156,0	6,5	76	16	4	3
Datensatz V8	Juli 2018 bis September 2020	27	41,5	1,5	85	20	5	4

1 Durchschnitt je Monat.  
Berechnungen basierend auf CEDEFOP-Daten.

# Auswirkungen der Coronakrise auf den Arbeitsmarkt: experimentelle Statistiken aus Daten von Online-Jobportalen

den einzelnen Revisionsdatensätzen zwischen 76 und 92, wobei die Anzahl an numerisch bedeutenden Portalen, von denen mehr als 100 000 Anzeigen je Monat erhoben wurden, im Durchschnitt zwischen drei und fünf Portalen liegt (siehe Tabelle 1).

Über den gesamten Zeitraum betrachtet, stellten sich für statistische Zwecke die fünf Jobportale Adzuna, Bundesagentur für Arbeit, Jobbörse, XING und StepStone als besonders bedeutend heraus. Laut Angaben von CEDEFOP wurden die Daten bei Adzuna mittels API und bei den anderen vier Jobportalen durch Scraping erfasst. Alle

fünf ausgewählten Jobportale befinden sich mindestens einmal auf den externen Ranglisten der Übersicht 1.<sup>12</sup>

- 12 Weitere wichtige, in den externen Ranglisten enthaltenen Jobportale wie Indeed, Monster, Meinestadt.de und Gigajob schafften es aus nachfolgenden Gründen nicht unter die bedeutendsten fünf: eine Erfassung der Daten von Indeed war nach CEDEFOP-Angaben nicht möglich, da das Scraping unterbunden wird. Für Monster gilt dies auch, dort hatte man jedoch einen API-Zugang, mit dem allerdings in allen Berichtsmonaten immer weniger als 50 000 Stellenanzeigen je Monat gewonnen werden konnten. Daten von meinestadt.de wurden von CEDEFOP zwar über Scraping erhoben, sind aber mit einer schlechten Datenqualität verbunden. Ähnliches gilt für Gigajob, dessen Datenqualität zusätzlich noch durch die verwendete Crawling-Methode beeinträchtigt wird.

**Tabelle 2**

**CEDEFOP-Daten: Stellenanzeigen und Art der Datengewinnung (einschließlich Dubletten)**

	Datensatz V1					Datensatz V4					Datensatz V8				
	Stellenanzeigen	Art der Datengewinnung <sup>1</sup>				Stellenanzeigen	Art der Datengewinnung <sup>1</sup>				Stellenanzeigen	Art der Datengewinnung <sup>1</sup>			
		API	Scraping	Crawling	k.A.		API	Scraping	Crawling	k.A.		API	Scraping	Crawling	k.A.
Anzahl	%				Anzahl	%				Anzahl	%				
<b>2018</b>															
Juli	965 492	30,0	61,5	8,5	0,0	2 238 425	32,2	60,4	7,3	0,0	1 409 243	33,2	60,1	6,7	0,0
August	926 704	35,9	62,4	1,7	0,0	2 225 014	41,6	56,8	1,6	0,0	1 431 108	43,9	54,5	1,6	0,0
September	813 447	34,2	49,3	16,5	0,0	1 795 950	38,9	45,4	15,6	0,0	1 118 707	41,6	43,5	14,9	0,0
Oktober	1 057 380	27,6	62,3	10,2	0,0	2 207 900	30,5	59,3	10,1	0,0	1 328 990	32,1	57,8	10,1	0,0
November	1 292 166	25,2	55,3	19,6	0,0	2 589 953	27,0	54,8	18,1	0,0	1 531 141	28,3	54,3	17,4	0,0
Dezember	1 022 489	32,2	54,8	13,0	0,0	2 067 687	31,8	56,3	12,0	0,0	1 224 565	31,9	56,9	11,2	0,0
<b>2019</b>															
Januar	2 211 067	45,2	49,1	5,6	0,0	6 797 433	37,5	53,0	9,5	0,0	4 798 102	35,6	53,5	10,9	0,0
Februar	1 305 098	28,4	60,8	10,7	0,0	3 140 215	24,4	66,2	9,3	0,0	2 019 768	23,3	67,7	9,0	0,0
März	1 048 708	12,8	71,8	15,5	0,0	2 487 508	12,5	74,8	12,7	0,0	1 573 771	12,9	75,4	11,7	0,0
April	892 891	15,2	70,3	14,5	0,0	2 026 248	15,6	72,2	12,2	0,0	1 264 525	16,2	72,2	11,5	0,0
Mai	867 071	26,1	63,4	10,5	0,0	2 036 028	27,4	64,1	8,5	0,0	1 365 673	31,0	60,9	8,1	0,0
Juni	462 587	13,2	79,8	7,0	0,0	1 062 147	11,7	82,7	5,6	0,0	1 380 957	20,6	67,5	11,9	0,0
Juli	1 015 449	24,7	62,6	12,6	0,0	2 164 094	23,9	65,0	11,0	0,0	1 366 255	22,7	67,7	9,6	0,0
August	1 091 839	37,4	51,2	11,4	0,0	2 241 873	37,0	52,9	10,1	0,0	1 345 852	36,8	54,1	9,1	0,0
September	841 871	43,9	52,8	3,4	0,0	1 734 671	40,3	56,6	3,1	0,0	1 045 096	38,3	58,8	2,9	0,0
Oktober	684 915	36,1	61,8	2,2	0,0	1 383 721	32,8	65,2	1,9	0,0	823 525	31,1	67,1	1,8	0,0
November	576 515	10,9	86,0	3,1	0,0	1 132 095	10,6	86,3	3,1	0,0	660 322	10,5	86,5	3,0	0,0
Dezember	481 312	9,9	84,8	5,1	0,2	1 029 919	10,2	85,3	4,3	0,2	692 040	10,8	85,1	3,6	0,5
<b>2020</b>															
Januar	–	–	–	–	–	2 120 598	7,0	81,0	11,7	0,2	1 967 699	7,8	77,4	14,2	0,7
Februar	–	–	–	–	–	1 205 464	8,2	89,2	2,1	0,5	1 030 998	9,4	87,5	2,3	0,8
März	–	–	–	–	–	697 017	7,2	81,4	10,9	0,5	911 193	8,6	81,1	9,5	0,8
April	–	–	–	–	–	–	–	–	–	–	315 582	11,9	79,9	6,8	1,4
Mai	–	–	–	–	–	–	–	–	–	–	687 667	6,6	80,6	12,0	0,7
Juni	–	–	–	–	–	–	–	–	–	–	1 621 732	14,5	60,4	23,9	1,2
Juli	–	–	–	–	–	–	–	–	–	–	4 470 020	16,2	29,4	51,0	3,4
August	–	–	–	–	–	–	–	–	–	–	1 745 537	37,1	47,7	12,3	2,9
September	–	–	–	–	–	–	–	–	–	–	2 461 304	30,5	59,5	7,2	2,8

Berechnungen basierend auf CEDEFOP-Daten.

1 Siehe Fußnote 10 im Text; k.A.: keine Angabe.

Detaillierte monatspezifische Angaben zur Anzahl der insgesamt verwendeten Jobportale, Anzahl der Stellenanzeigen und Art der Datengewinnung enthalten die [Tabellen 2 und 3](#). Einen visuellen Überblick der monatlichen Verläufe gibt [Grafik 1](#). Da sich der Datensatz V5 aufgrund seiner Größe mit den Mitteln der EU-Dataplattform nicht vollständig analysieren ließ, werden in den Tabellen 2 und 3 sowie der Grafik 1 nur die Datensätze V1, V4 und V8 gegenübergestellt.

Die rückwirkenden revisionsbedingten Änderungen in der absoluten Anzahl der erhobenen Stellenanzeigen sind erheblich. Mit den Revisionen ändert sich die relative Bedeutung einzelner Jobportale. Auch verändern alle Datensätze ihre Struktur im Zeitverlauf. Dies gilt sowohl für die Anzahl der Stellenanzeigen als auch für die Anteile der einzelnen Portale. So nimmt beispielsweise der Anteil von Adzuna, unabhängig von der Datensatzversion, über die Zeit hinweg ab, der von Jobbörse.de steigt dagegen an. Datensatz V8 zeigt eine Auffälligkeit im Juli 2020. Der Anteil von Stellenanzeigen, die nicht zu den fünf bedeutendsten gehören, beträgt 71 %,

**Tabelle 3**  
CEDEFOP-Daten: Verwendete Jobportale

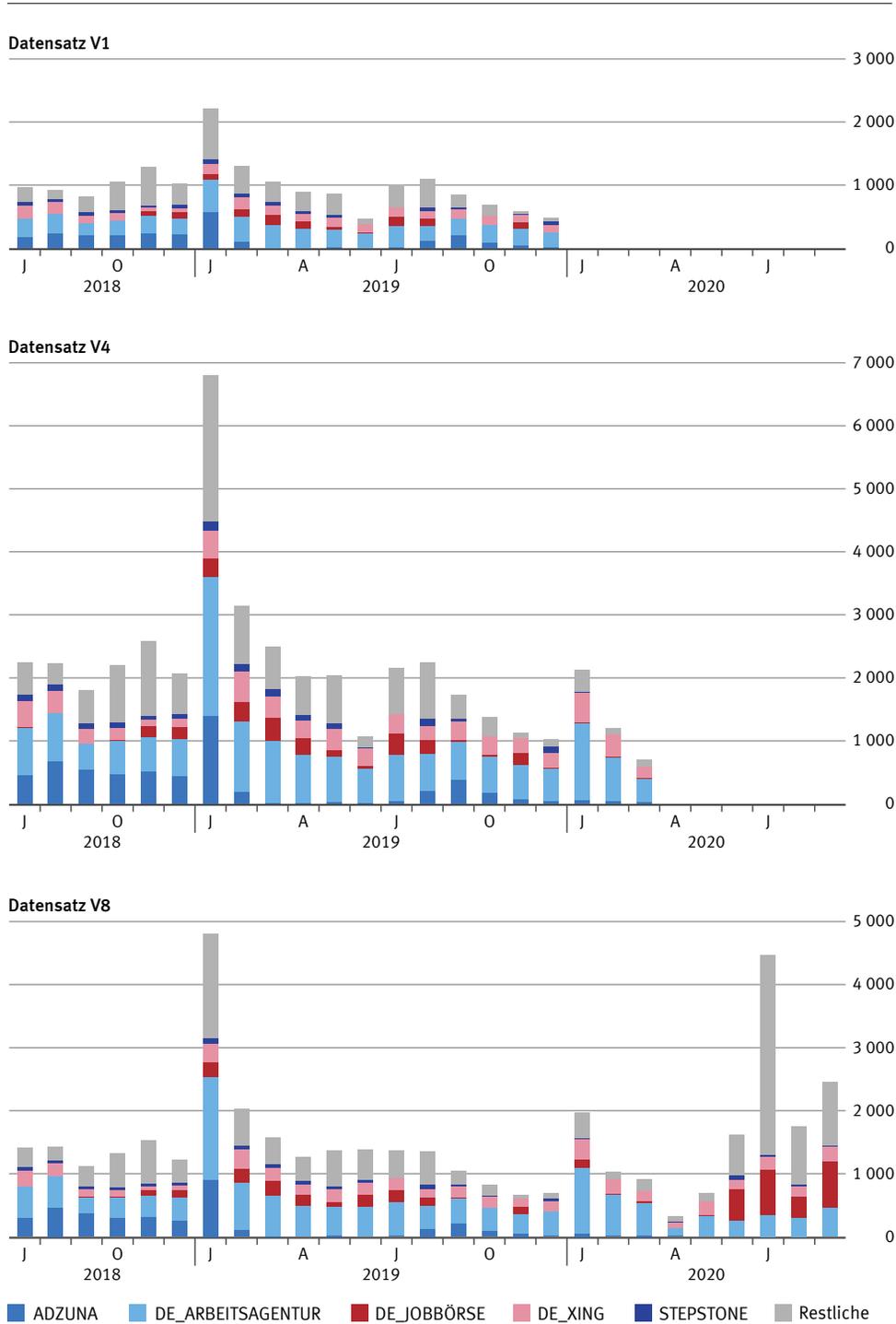
	Datensatz V1				Datensatz V4				Datensatz V8			
	Jobportale insgesamt	Jobportale, von denen mindestens ... Stellenanzeigen gewonnen wurden			Jobportale insgesamt	Jobportale, von denen mindestens ... Stellenanzeigen gewonnen wurden			Jobportale insgesamt	Jobportale, von denen mindestens ... Stellenanzeigen gewonnen wurden		
		1 000	50 000	100 000		1 000	50 000	100 000		1 000	50 000	100 000
<b>2018</b>												
Juli	96	23	5	3	96	94	6	4	94	25	5	4
August	64	15	4	3	64	32	5	4	61	17	5	4
September	108	20	3	3	108	36	6	4	107	21	5	3
Oktober	88	21	6	4	88	35	7	6	83	21	7	4
November	83	25	7	4	83	33	9	6	84	25	7	4
Dezember	116	24	7	3	114	37	9	7	112	25	7	4
<b>2019</b>												
Januar	156	28	8	6	156	49	11	9	151	34	11	8
Februar	128	25	8	5	128	36	8	8	129	25	8	6
März	124	22	6	3	124	34	7	6	126	25	6	4
April	108	22	6	3	108	33	7	5	109	23	6	4
Mai	90	23	3	3	90	28	8	3	89	23	4	3
Juni	58	12	2	2	58	24	2	2	87	24	5	4
Juli	96	19	6	4	96	28	6	6	92	19	6	5
August	82	22	7	5	82	29	9	7	88	22	7	5
September	79	16	5	3	79	29	5	5	82	16	5	4
Oktober	64	13	4	3	64	24	4	4	64	14	4	4
November	63	12	3	3	63	25	4	3	64	12	3	3
Dezember	53	14	3	2	53	23	3	3	55	14	3	2
<b>2020</b>												
Januar	–	–	–	–	80	22	5	3	84	17	5	4
Februar	–	–	–	–	60	22	3	2	69	13	3	2
März	–	–	–	–	53	20	3	2	70	14	4	2
April	–	–	–	–	–	–	–	–	63	10	2	1
Mai	–	–	–	–	–	–	–	–	55	13	2	2
Juni	–	–	–	–	–	–	–	–	68	16	7	6
Juli	–	–	–	–	–	–	–	–	84	24	6	6
August	–	–	–	–	–	–	–	–	69	21	5	4
September	–	–	–	–	–	–	–	–	65	23	6	4

Berechnungen basierend auf CEDEFOP-Daten. Stellenanzeigen einschließlich Dubletten.

# Auswirkungen der Coronakrise auf den Arbeitsmarkt: experimentelle Statistiken aus Daten von Online-Jobportalen

**Grafik 1**

Vergleich der Datensätze V1, V4 und V8: Stellenanzeigen in der monatlichen Häufigkeitsverteilung (einschließlich Dubletten)  
1 000 Stellenanzeigen



Adzuna: API; Arbeitsagentur, Jobbörse, XING, StepStone: Scraping  
Datenbasis: CEDEFOP-Daten.

**Tabelle 4**

**Dubletten in den verschiedenen Datensätzen**

	Berichtsmonate	Monate	Stellenanzeigen einschließlich Dubletten		Stellenanzeigen ohne Dubletten		Dubletten
			insgesamt	je Monat	insgesamt	je Monat	
		Anzahl	Mill.				%
Datensatz V1	Juli 2018 bis Dezember 2019	18	17,0	0,9	14,0	0,78	17,6
Datensatz V4	Juli 2018 bis März 2020	21	44,0	2,1	16,0	0,76	63,6
Datensatz V5	Juli 2018 bis Juni 2020	24	156,0	6,5	17,0	0,70	89,1
Datensatz V8	Juli 2018 bis September 2020	27	41,5	1,5	19,8	0,73	52,3

Berechnungen basierend auf CEDEFOP-Daten.

während er in anderen Monaten maximal 53 % beträgt und im Durchschnitt bei 25 % liegt.<sup>13</sup> Der Juli 2020 zeigt zusammen mit dem Januar 2019 noch eine weitere Auffälligkeit: Die insgesamt hohe monatliche Volatilität bei der Anzahl der absoluten Stellenanzeigen nimmt in diesen beiden Monaten ein extremes Ausmaß ungeklärter Ursache an.

### Dubletten

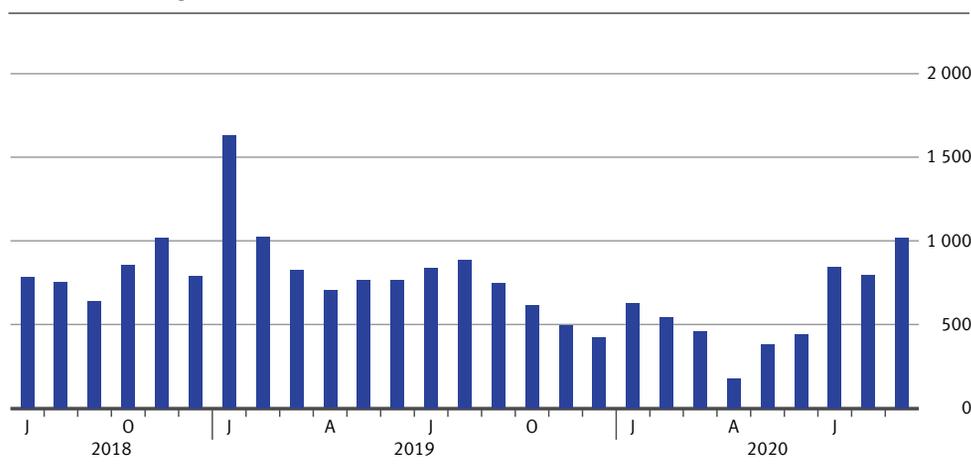
Arbeitgeber annoncieren offene Stellen oft auf mehreren Stellenbörsen gleichzeitig; das erschwert die Auswertung des Stellenbestands. CEDEFOP klassifiziert in der automatischen Datenaufbereitung einen Anteil der Stellenanzeigen als Dubletten, wobei inhaltliche Nähe und zeitnahe Publikation die wichtigsten Kriterien sind. Wie effektiv diese Deduplizierung ist, lässt sich allerdings nicht empirisch überprüfen. Auch CEDEFOP konnte auf Anfrage keine Einschätzung zur Verlässlichkeit der Deduplizierung geben. Der Anteil der Dubletten variiert stark zwischen den verschiedenen Datensätzen, er beträgt beim Datensatz V1 nur 17,6 %, beim Datensatz V5 dagegen 89,1%. [↘ Tabelle 4](#) Dies ist ein Hinweis darauf, dass die Prozesse zur Klassifikation von Dubletten bei CEDEFOP noch äußerst volatil sind und häufigen, systematischen Veränderungen unterliegen.

13 Im Juli 2020 stammten 47 % aller Stellenanzeigen vom Jobportal DE\_Jobangebote. Da dieses Jobportal in den übrigen Monaten aber einen Anteil von weniger als 1 % hatte, war es nicht bei der 5er-Auswahl einbezogen.

Die im Abschnitt „Revisionen, verwendete Jobportale und Anzahl an Stellenanzeigen“ beschriebenen starken Variationen der erfassten Stellenanzeigen über die Zeit bestehen auch nach Eliminierung der Dubletten weiter. [↘ Grafik 2](#)

**Grafik 2**

**Monatliche Häufigkeitsverteilung ohne Dubletten, Datensatz V8**  
1 000 Stellenanzeigen



Berechnungen basierend auf CEDEFOP-Daten.

2021 - 0199

## Pseudostock-Berechnung

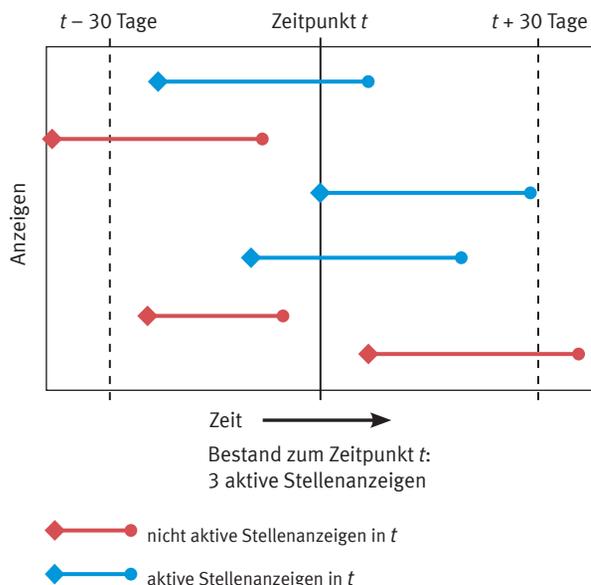
Online-Stellenanzeigen aus den CEDEFOP-Daten sind mit dem Tag ihrer Erfassung datiert. Allerdings entspricht dies nicht notwendigerweise dem Veröffentlichungsdatum, welches unbekannt ist. Das Ablaufdatum der Anzeige ist nur für einen kleinen Teil der Beobachtungen erfasst. Die Anzahl täglich erhobener Anzeigen weist sehr hohe Schwankungen auf, die im Monatsdurchschnitt jedoch etwas geglättet wird.

An der Frequenz der Beobachtungen lässt sich ablesen, dass die CEDEFOP-Stellenanzeigen weder den Bestand (Stocks) zum Zeitpunkt der Datenerfassung, noch die jeweils neu veröffentlichten Stellenanzeigen (Flows) wiedergeben. Deshalb wurde eine Methode entwickelt, um aus den erfassten Anzeigen synthetische Bestände (Pseudostocks) zu errechnen.

Die Berechnung derselben ist in [Grafik 3](#) schematisch dargestellt. Eine Stellenanzeige zählt zum Pseudostock am Zeitpunkt ( $t$ ) wenn sie eine maximale Anzahl von Tagen ( $v$ ) vor  $t$  erfasst wurde, und zum Zeitpunkt  $t$  noch nicht abgelaufen ist. Die Summe dieser Anzeigen ist der Pseudostock in  $t$ . Eine erwartete Gültigkeit ( $v$ ) von 30 Tagen hat sich für Stellenanzeigen als plausibel erwiesen. Die so ermittelten synthetischen Bestände geben also, monatsübergreifend, den Bestand aller Anzeigen wieder, die innerhalb der letzten 30 Tage erfasst wur-

**Grafik 3**

Berechnung von Pseudostocks



2021-0200

den (und noch nicht abgelaufen sind). Der auf diese Weise ermittelte synthetische Bestand approximiert den zugrundeliegenden Bestand der Stellenanzeigen.

## Validierung einzelner Jobportale

Seit Mai 2020 hat das Statistische Bundesamt mithilfe von Webscraping täglich die Anzahl der Stellenanzeigen aus dem Jobportal der Bundesagentur für Arbeit erhoben, ab Ende August 2020 auch die des Jobportals von Xing. Beide Portale zeigen auf einer Übersichtsseite die Summe der Stellenanzeigen an, die von den Nutzenden wählbaren Kriterien entsprechen. Diese Angaben ermöglichen es, die CEDEFOP-Daten für diese beiden Jobportale zu validieren.

### Jobportal der Bundesagentur für Arbeit

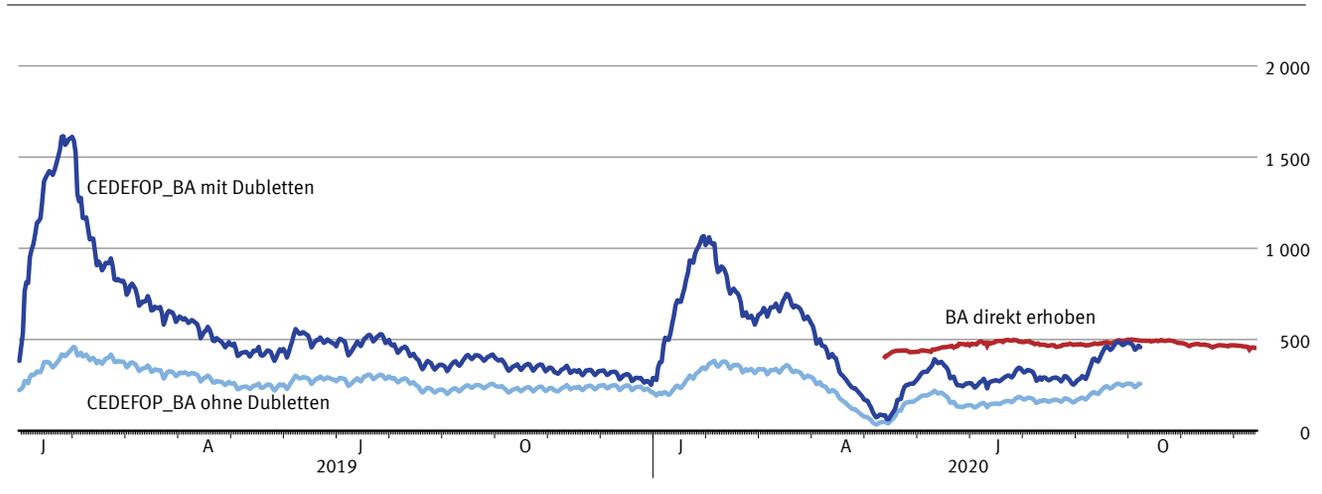
Das Stellenportal der Bundesagentur für Arbeit bietet die Möglichkeit, die Anzahl der Stellen, die innerhalb der letzten vier Wochen veröffentlicht wurden, zu erfassen. Diese werden mit CEDEFOP-Pseudostocks, mit 30 Tagen angenommener Gültigkeit, verglichen. Das Jobportal der Bundesagentur für Arbeit bietet mehrere Filtermöglichkeiten: nach Veröffentlichungszeitraum, nach Angebotsart, nach Befristung, nach Arbeitszeitmodell sowie mit oder ohne Arbeitsvermittler. Es ist davon auszugehen, dass für CEDEFOP die Anzeigen mit den Standardeinstellungen erfasst wurden. Für den folgenden Vergleich wurden diese Parameter angeglichen.

[Grafik 4](#) zeigt die Verläufe der Bestandsangaben auf der Webseite der Bundesagentur für Arbeit im Vergleich zu den aus den CEDEFOP-Daten berechneten Pseudostocks der Bundesagentur für Arbeit. Da das Jobportal der Bundesagentur für Arbeit auch Stellenanzeigen von Allianzpartnern anzeigt und somit doppelte Anzeigen enthalten kann, wird der Vergleich mit und ohne Dubletten gezogen. Die Pseudostocks der für CEDEFOP erhobenen Stellenanzeigen im Jobportal der Bundesagentur für Arbeit sind deutlich volatil als der direkt erhobene Bestand. Während die direkt erhobenen Anzeigen zwischen Mai und September 2020 um maximal 24 % variieren, liegt die Spreizung bei den CEDEFOP-Daten mit Dubletten im gleichen Zeitraum bei 771 %, ohne Dubletten bei 420 %.

Beide aus den CEDEFOP-Daten ermittelten Zeitreihen zeigen nach dem coronabedingten Einbruch einen im Mai 2020 beginnenden Anstieg in der Arbeitskräfte-

**Grafik 4**

**Validierung des Jobportals der Bundesagentur für Arbeit im CEDEFOP-Datensatz**  
Stellenanzeigen in 1 000



CEDEFOP als Pseudostocks mit 30 Tagen Gültigkeit. BA Bestand der letzten 4 Wochen als Summe aller Angebotsarten.  
Quelle: Eigene Erhebungen, Bundesagentur für Arbeit (BA), CEDEFOP

2021 - 0201

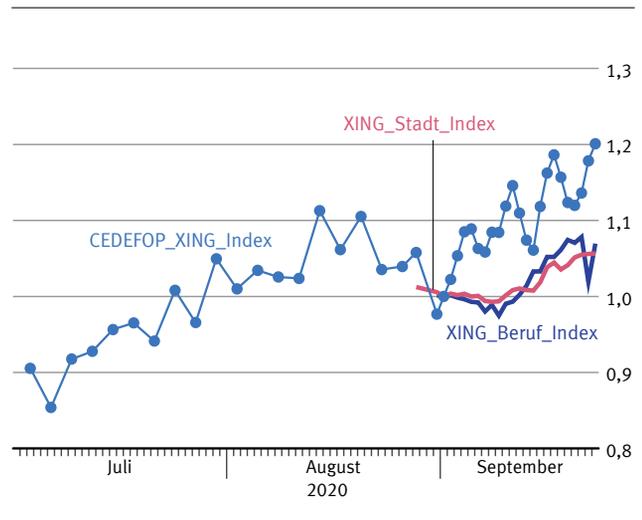
nachfrage des Online-Stellenmarkts. Bei den direkt von der Bundesagentur für Arbeit gescrapten Angaben ist dieser allerdings nur schwach zu erkennen.

**XING-Jobportal**

Auf dem Jobportal von XING ist es nur möglich, die Summe der Stellenanzeigen für die am häufigsten angezeigten Städte und für die häufigsten Berufe zu erfassen. Eine Angabe der Gesamtzahl der Stellenanzeigen ist nicht vorhanden. Die Summenbildung sowohl über die häufigsten Städte als auch über die häufigsten Berufe bildet jeweils nur einen Teil des Gesamtbestands des XING-Jobportals ab. Die auf diese Weise ermittelten Bestände unterscheiden sich stark im absoluten Niveau von den aus den CEDEFOP-Daten für XING berechneten Pseudostocks. Deshalb werden für einen plausiblen Vergleich die Indikatoren normiert und ihre Entwicklung über die Zeit verglichen. Dabei zeigt sich, dass – wie schon beim Jobportal der Bundesagentur für Arbeit – die für CEDEFOP erhobenen Daten stärkeren Schwankungen unterliegen. Gleichzeitig zeigt sich jedoch eine gewisse Übereinstimmung im Anstieg der Stellenanzeigen für alle drei Indizes im Verlauf des Septembers 2020. [↘ Grafik 5](#)

**Grafik 5**

**Validierung des XING-Jobportals im CEDEFOP-Datensatz**  
Indizes, Basis 1.9.2020



CEDEFOP als Pseudostocks, ohne Dubletten.  
Quelle: Eigene Erhebungen, XING, CEDEFOP

2021 - 0202

## 4

### Experimentelle Statistiken zum Online-Stellenmarkt

#### 4.1 EXDAT-Online-Stellenindex zur Entwicklung des Online-Stellenmarkts

Um die Entwicklung der Arbeitskräftenachfrage auf Online-Stellenmärkten zu verfolgen, wurde ein monatlicher Indikator entwickelt. Wie die obigen Ausführungen gezeigt haben, ist die Höhe des absoluten Bestands an Online-Stellenanzeigen mit großen Unsicherheiten behaftet. Deshalb zielt der Indikator darauf ab, Veränderungen im Zeitverlauf darzustellen. Wegen der eingeschränkten Datenqualität werden Ergebnisse dieses Index zunächst unter dem Label „EXDAT – Experimentelle Daten“ veröffentlicht und der Indikator im Folgenden EXDAT-Online-Stellenindex genannt. Statistiken in dieser Rubrik unterscheiden sich im Reifegrad von amtlichen Statistiken. Ähnliche Indikatoren werden in anderen Ländern teilweise seit Jahren veröffentlicht. Zu nennen sind hier insbesondere Australien, die Schweiz, und die Vereinigten Staaten.<sup>14</sup>

Als Basis der Berechnung des EXDAT-Online-Stellenindex wurden die Daten der fünf im Abschnitt „Revisionen, verwendete Jobportale und Anzahl an Stellenanzeigen“ ausgewählten Jobportale des aktuellsten CEDEFOP-Datensatzes V 8 verwendet. Da duplizierte Stellenanzeigen verzerrend wirken, wenn deren Anteil über die Zeit variiert, werden für den EXDAT-Online-Stellenindex die Dubletten wie oben beschrieben herausgefiltert.

Online-Stellenanzeigen werden nicht nur von Arbeitgebern veröffentlicht, ein Teil der erfassten Anzeigen geht auf Personalvermittlungsfirmen zurück. Solche Anzeigen sind jedoch für die Arbeitskräftenachfrage kaum aussagekräftig. Zum einen ist oft unklar, ob hinter der Anzeige eines Personalvermittlers tatsächlich eine

zu besetzende Stelle steht (SWR, 2019). Zum anderen handelt es sich bei Anzeigen von Personalvermittlern regelmäßig um Dubletten von Anzeigen, die auch von Arbeitgebern veröffentlicht wurden. Deshalb werden diese Anzeigen, soweit möglich, vor der Berechnung des EXDAT-Online-Stellenindex herausgefiltert.<sup>15</sup> Diese Anzeigen von Personalvermittlern machen 25,2% der deduplizierten Beobachtungen aus.

Auch Anzeigen, die sich auf Praktika anstatt auf reguläre Stellen beziehen, werden vor der Berechnung des EXDAT-Online-Stellenindex herausgefiltert. Danach wird die Pseudostock-Methode angewandt, um für jeden Beobachtungstag einen synthetischen Bestand zu erhalten (siehe Abschnitt „Pseudostock Berechnung“).<sup>16</sup> Der monatliche Indexwert ist der Durchschnitt der Pseudostocks aller Tage eines Monats. Dies bedeutet, dass der Indexwert eines Monats auch Stellenanzeigen abbildet, die im Verlauf des letzten Monats veröffentlicht wurden. Durch Verwendung von Pseudostocks werden Brüche in der Datengewinnung zwischen den Monaten geglättet und stellenweise hohe Ausschläge in den erfassten Anzeigen in den Folgemonat übertragen. [↘ Grafik 6](#) zeigt diesen Effekt des Pseudostocks-Index im Vergleich mit einem Index, der aus einfachen Monatsmittelwerten gebildet wurde.

Auch durch die Beschränkung auf fünf Portale ergibt sich für den untersuchten Zeitraum ein geglätteter Index, die Volatilität wird reduziert. [↘ Grafik 7](#) kontrastiert den Verlauf dieses Stellenindex mit einer Version des Index, der alle Stellenportale mit einbezieht. Dabei zeigen sich Hinweise darauf, dass die extremen Ausschläge des Index im Februar und Juni 2019 teilweise Effekte der Datenerhebung aus multiplen Quellen sind. Daraus lässt sich auch schließen, dass in den Daten noch substantielle Anteile an Stellenanzeigen enthalten sein müssen, die auf mehreren Stellenbörsen gleichzeitig ausgeschrieben sind. Dadurch werden Veränderungen im Bestand an Online-Stellenanzeigen artifiziell verstärkt, was zu einem volatileren Index führt.

14 Die australische Regierung veröffentlicht einen monatlichen Internet Vacancy Index IVI (<https://data.gov.au>), für die Schweiz existieren der Stellenmarktmonitor Schweiz von der Universität Zürich, ein kombinierter Presse-Online-Index (Sacchi, 2008; Sacchi, 2014) und für die Vereinigten Staaten der Index Help Wanted OnLine HWOL (Conference Board, 2018). Eine Liste dieser und weiterer Indikatoren, die auf Basis von Online-Stellenanzeigen von öffentlich-rechtlichen Institutionen veröffentlicht werden, findet man bei Rengers (2020, hier: Seite 4 f.).

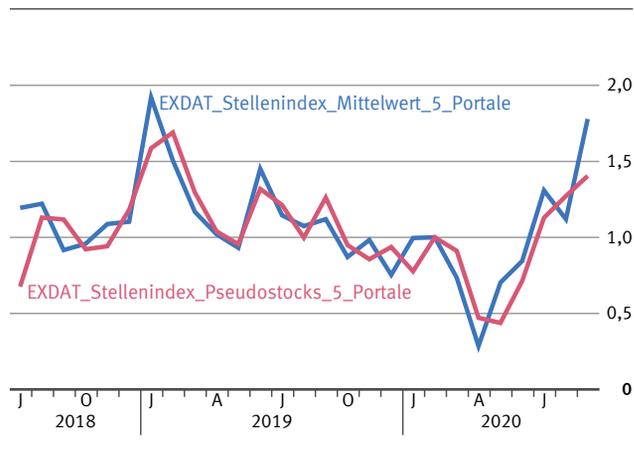
15 Da kein Identifikator für Personalvermittler im CEDEFOP-Datensatz enthalten ist, wird dieser auf Basis des bereinigten Firmennamens generiert. Dabei kommt eine manuell erstellte Filterliste zum Einsatz, die die Firmennamen der, laut Branchenberichten, größten Personalvermittlungsfirmen in Deutschland enthält. Diese Liste wird ergänzt um Firmennamen mit typischen Elementen von Personalvermittlern, beispielsweise „personnel“, „recruitment“ oder „staffing“.

16 Es wird eine durchschnittliche Gültigkeit von 30 Tagen angenommen. Berechnungen mit 20 und 40 Tagen Gültigkeit kommen zu sehr ähnlichen Ergebnissen, die sich hauptsächlich im Grad der Glättung des Index unterscheiden.

**Grafik 6**

**EXDAT-Online-Stellenindex mit Pseudostocks und Monatsmittelwerten**

Indizes, Basis 1.2.2020



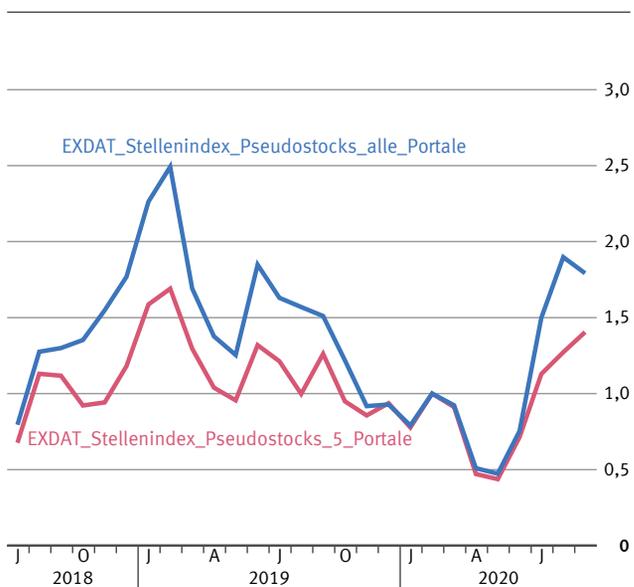
5 Portale: Arbeitsagentur, XING, StepStone, Adzuna und DE\_Jobbörse. Ohne Dubletten, Praktika und Personalvermittler. Berechnungen basierend auf CEDEFOP-Daten.

2021 - 0203

**Grafik 7**

**EXDAT-Online-Stellenindex mit Pseudostocks und unterschiedlicher Anzahl an Jobportalen**

Indizes, Basis 1.2.2020



5 Portale: Arbeitsagentur, XING, StepStone, Adzuna und DE\_Jobbörse. Ohne Dubletten, Praktika und Personalvermittler. Berechnungen basierend auf CEDEFOP-Daten.

2021 - 0204

## 4.2 Externe Vergleichsdaten

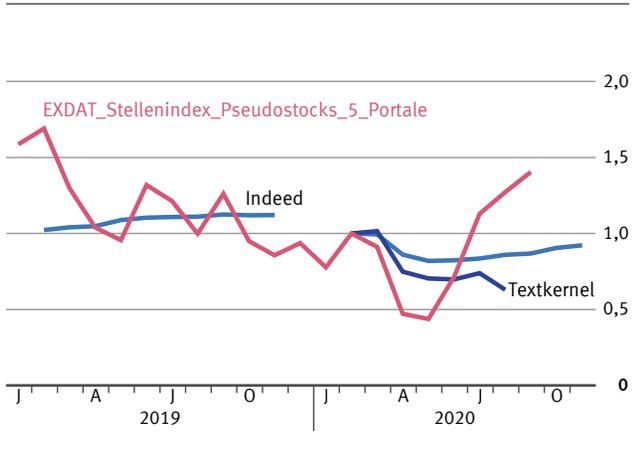
Das Jobportal Indeed hat im Zuge der Coronakrise einen Indikator zum Verlauf von Online-Stellenanzeigen veröffentlicht. Indeed verwendet dafür den über sieben Tage gemittelten Bestand der veröffentlichten Stellen auf dem eigenen Jobportal.<sup>17</sup> Dieser wird mit dem in Abschnitt 4.1 beschriebenen Stellenindex der fünf ausgewählten Portale verglichen, um die Qualität desselben zu evaluieren.<sup>18</sup> Weiterhin betreibt die Firma Textkernel eigenes Webscraping des Online-Stellenmarktes für ihren Analytics-Service Jobfeed. Für Februar bis August 2020 hat Textkernel auf Basis der Jobfeed Daten einen Indikator zum Bestand der Stellenanzeigen veröffentlicht.<sup>19</sup>

In [Grafik 8](#) ist der Verlauf der drei Indizes dargestellt, normiert auf Februar 2020. Alle drei zeigen für das Frühjahr 2020 einen coronabedingten Abschwung am Online-Stellenmarkt. Sowohl bei Indeed als auch beim

**Grafik 8**

**EXDAT-Online-Stellenindex: Vergleich mit Indeed und Textkernel**

Indizes, Basis 1.2.2020



Berechnungen basierend auf CEDEFOP-Daten sowie Auswertungen veröffentlichter Grafiken von Indeed und Textkernel. - Die Lücke in den Indeed-Daten zwischen November 2019 und Februar 2020 ergibt sich daraus, dass Indeed seinen Indikator nur jeweils für den Zeitraum Februar bis November veröffentlicht hat.

2021 - 0205

17 Quelle: [www.hiringlab.org](http://www.hiringlab.org) Die Daten für den Indikator wurden mithilfe der WebPlotDigitizer-Software aus den Grafiken extrahiert.

18 Die Lücke in den Indeed-Daten zwischen November 2019 und Februar 2020 ergibt sich daraus, dass Indeed seinen Indikator nur jeweils für den Zeitraum Februar bis November veröffentlicht hat.

19 Quelle: [www.textkernel.com](http://www.textkernel.com) Die Daten für den Indikator wurden mithilfe der WebPlotDigitizer-Software aus den Grafiken extrahiert.

EXDAT-Online-Stellenindex beginnt der Abschwung im Februar 2020. Bei Textkernel beginnt er erst im März 2020, fällt jedoch deutlicher aus als bei Indeed. Alle drei Indizes erreichen im Mai 2020 ihren Tiefpunkt. Der coronabedingte Abschwung fällt je nach Datenquelle sehr unterschiedlich aus. Zwischen Februar und Mai 2020 verzeichnet der Indeed-Index einen Rückgang um 18 %, der von Textkernel um 29 % und der mit CEDEFOP-Daten ermittelte EXDAT-Online-Stellenindex um 56 %.<sup>20</sup>

Nach dem Mai 2020 zeigen alle drei Indizes eine Erholung am Stellenmarkt an. Im Fall von Textkernel ist diese jedoch kurzlebig, der Index verzeichnet bereits im August 2020 wieder einen Rückgang. Die Erholung, die der EXDAT-Online-Stellenindex anzeigt, ist sehr stark und übertrifft im September 2020 das Niveau von Februar 2020 um 40 %. Es zeigt sich, dass bei der Berechnung von Indizes zum Online-Stellenmarkt die Wahl der Quelle(n) und die Art der Datenerhebung starke Auswirkungen auf den Verlauf haben. Eine große Herausforderung besteht also darin, Indexausschläge durch ökonomische Veränderungen von Ausschlägen durch Veränderungen im Datenerhebungs- und Aufbereitungsprozess zu unterscheiden.

## 5

### Fazit und Ausblick

---

Die Coronakrise hat Entwicklungen in der amtlichen Statistik beschleunigt, alternative neue digitale Datenquellen zu nutzen, um den gestiegenen Bedarf an zeitnahen hochaktuellen Daten zur wirtschaftlichen Entwicklung zu decken. In Bezug auf den Arbeitsmarkt kann die Nutzung von im Internet veröffentlichten Stellenanzeigen dazu einen wichtigen Beitrag liefern. Online-Stellenanzeigen werden auf Jobportalen gebündelt. Da jedoch die Datengewinnung von Online-Jobportalen weiterhin mit rechtlichen und technischen Schwierigkeiten verbunden ist, wurden und werden bei EU-Projekten zu Online-Stellenanzeigen bereits bestehende Datengewinnungsprozesse anderer Institutionen in Europa genutzt. Außerdem stehen die gewonnenen Daten, konkret die CEDEFOP-Daten, in einem Sharingsystem im ESS zur Ver-

fügung. Die CEDEFOP-Daten zu Online-Stellenanzeigen weisen jedoch weiterhin große Qualitätsmängel auf und sind nicht für alle statistischen Zwecke gleichermaßen geeignet.

In einem ersten Schritt hat das Statistische Bundesamt aus den CEDEFOP-Daten einen monatlichen Stellenindex ermittelt, der zwar nicht die absolute Größenordnung von Online-Stellenanzeigen abbildet, wohl aber Veränderungen im Zeitverlauf.

In einem weiteren EU-Projekt 2020-2024 soll der Produktionsprozess für die internetgestützte Erfassung von Online-Stellenanzeigen (online job advertisement – OJA) – bestehend aus Datengewinnung, Datenzugriff, Datenverarbeitung und Datenhaltung – weiterentwickelt werden. Ziel ist, OJA in den European Trusted Smart Statistics Centre' Web Intelligence Hub WIH zu implementieren. Das EU-Projekt trägt den Namen „Towards the European Web Intelligence Hub – European System for Collection and Analysis of Online Job Advertisement Data (WIH-OJA)“. Dieses sogenannte WIH-Projekt wird interinstitutionell, gemeinschaftlich von Eurostat und CEDEFOP, durchgeführt. Bezüglich der Produktion der CEDEFOP-Daten bedeutet das weiterhin ein Outsourcing des gesamten OJA-Produktionsprozesses, unter Zusammenlegung von Ressourcen und Fachwissen beider Institutionen.<sup>21</sup>

Für die nationalen statistischen Ämter als Datennutzer hat gleichzeitig ein weiteres ESSnet-Projekt 2021-2025 begonnen, das zum Ziel hat, ein sogenanntes Web Intelligence Network (WIN) aufzubauen. Damit soll eine Plattform (WIH) entwickelt werden, die die Erstellung von Statistiken mit Internet-Datenquellen auf europäischer und nationaler Ebene ermöglicht. Der Projektname lautet „Trusted Smart Statistics – Web Intelligence Network“ (WIN-Projekt).

Beide EU-Projekte haben zum Ziel, die bereits in den früheren EU-Projekten zu Online-Stellenanzeigen genutzten Data-Sharing-Möglichkeiten bezüglich der CEDEFOP-Daten sowie den gegenseitigen technischen und methodisch-fachlichen Austausch dauerhaft zu etablieren und zu institutionalisieren. Ein starker Fokus liegt dabei auf der Sicherstellung der Qualität und der Entwicklung neuer Methodiken. Damit besteht für die statistischen

---

20 Durch Normierung auf den Februar 2020 sind die Veränderungen zum Februar in Prozent und Prozentpunkten identisch.

21 Siehe dazu auch <https://etendering.ted.europa.eu> und [www.cedefop.europa.eu](http://www.cedefop.europa.eu)

Ämter als Datennutzer die Möglichkeit, Einblick in den Datenerfassungs- und Datenaufbereitungsprozess zu gewinnen. Darüber hinaus ermöglicht der Austausch die Einflussnahme auf den gesamten Produktionsprozess, sodass die Voraussetzungen für eine künftig verbesserte Qualität der CEDEFOP-Daten geschaffen sind.<sup>22</sup> 

---

22 Beim WIN-Projekt liegt der Fokus dabei allerdings nicht nur auf den CEDEFOP-Daten zu Online-Stellenanzeigen. Zweck ist, generell verschiedenste Big-Data-Quellen, einschließlich Webdaten, nutzbar zu machen und sie in die Produktion von amtlichen Statistiken zu integrieren. Für die Zukunft sollen die Dienste des Web Intelligence Hub (WIH) deshalb zusätzlich auch für andere Domänen nutzbar gemacht werden.

## LITERATURVERZEICHNIS

---

Burning Glass Technologies. *COVID-19: The Pandemic's Impact on Jobs*. 2021. [Zugriff am 11. Mai 2021]. Verfügbar unter: [www.burning-glass.com](http://www.burning-glass.com)

Chetty, Raj/Friedman, John N./Hendren, Nathaniel/Stepner, Michael/The Opportunity Insights Team. *How did Covid-19 and Stabilization Policies affect Spending and Employment? – A New Real-Time Economic Tracker Based On Private Sector Data*. In: NBER Working Paper Series – Working Paper 27431. Cambridge, 2020.

Conference Board. *The Conference Board Help Wanted OnLine® Technical Note*. 2018. [Zugriff am 21. April 2021]. Verfügbar unter: <https://conference-board.org>

Eurostat. *Methodological note – Data collection for the EU-labour force survey in the context of the covid-19 crisis*. 2020a. [Zugriff am 21. April 2021]. Verfügbar unter: <https://ec.europa.eu/eurostat>

Eurostat. *Methodological note – EU-labour force survey reaction to the covid-19 crisis under regulation 2019/1700*. 2020b. [Zugriff am 21. April 2021]. Verfügbar unter: <https://ec.europa.eu/eurostat>

Descy, Pascaline/Kvetan, Vladimir/Zuckersteinova, Alena/Wirthmann, Albrecht/Reis, Fernando. *Towards a Common Infrastructure for Online Job Vacancy Data*. Paper der DGINS Conference Bucharest 2018. [Zugriff am 22. April 2021]. Verfügbar unter: <https://dgins2018.statisticsevents.ro>

International Labour Organization (ILO). *COVID-19: Guidance for labour statistics data collection – Guidance to data producers to maintain labour force survey data collection*. 2020a. [Zugriff am 22. April 2021]. Verfügbar unter: <https://ilo.org>

International Labour Organization (ILO). *COVID-19: Guidance for labour statistics data collection – Essential labour force survey content and treatment of special groups (Rev. 1)*. 2020b. [Zugriff am 22. April 2021]. Verfügbar unter <https://www.ilo.org>

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

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

Rengers, Martina. *Internetgestützte Erfassung offener Stellen*. 2018b. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2018, Seite 11 ff.

Rengers, Martina. *Developing meaningful indicators based on OJA [online job vacancies] data*. In: ESSnet Big Data, Work Package B: Implementation – Online Job Vacancies, August 2020. [Zugriff am 22. April 2021]. Verfügbar unter: <https://ec.europa.eu/eurostat>

## LITERATURVERZEICHNIS

---

Sacchi, Stefan. *Lange Messreihen zur Entwicklung des Stellenangebots der Schweizer Wirtschaft: Kombiniertes Presse-Online-Index*. 2014. In: SMM Working Paper 2014-1, Zürich. [Zugriff am 22. April 2021]. Verfügbar unter: [www.stellenmarktmonitor.uzh.ch](http://www.stellenmarktmonitor.uzh.ch)

Sacchi, Stefan. *Adecco Swiss Job Market Index – Stellenmarktmonitor Schweiz, Methodische Grundlagen*. Zürich, 2008. [Zugriff am 22. April 2021]. Verfügbar unter: [www.stellenmarktmonitor.uzh.ch](http://www.stellenmarktmonitor.uzh.ch)

Špeh, Tomaž und andere. *Deliverable B3. Methodological framework for processing online job adverts data for Official Statistics V.2*. In: ESSnet Big Data II, Workpackage WP B: Implementation – Online Job Vacancies 2020. [Zugriff am 22. April 2021]. Verfügbar unter: <https://ec.europa.eu>

Statistisches Bundesamt. *Mikrozensus 2019*. Qualitätsbericht. Wiesbaden 2020.

Südwestrundfunk (SWR). *Bundesagentur für Arbeit: So missbrauchen Datenhändler die Jobbörse*. 2019. [Zugriff am 6. Mai 2021]. Verfügbar unter: [www.swr3.de](http://www.swr3.de)

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

---

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

---

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

---

**Erscheinungsfolge**  
zweimonatlich, erschienen im Juni 2021  
Ältere Ausgaben finden Sie unter [www.destatis.de](http://www.destatis.de) sowie in der [Statistischen Bibliothek](#).

---

Artikelnummer: 1010200-21003-4, ISSN 1619-2907

---

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