

Rengers, Martina; Hornik, Andrea

Article

Daten zu Online-Stellenanzeigen im ESSnet WIN

WISTA - Wirtschaft und Statistik

Provided in Cooperation with:

Statistisches Bundesamt (Destatis), Wiesbaden

Suggested Citation: Rengers, Martina; Hornik, Andrea (2024) : Daten zu Online-Stellenanzeigen im ESSnet WIN, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 76, Iss. 5, pp. 49-61

This Version is available at:

<https://hdl.handle.net/10419/305182>

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.

DATEN ZU ONLINE-STELLENANZEIGEN IM ESSnet WIN

Martina Rengers, Andrea Hornik

↳ **Schlüsselwörter:** Online-Jobportale – Datenqualität – Arbeitsmarktberichterstattung – Big Data – neue digitale Daten

ZUSAMMENFASSUNG

Der Bedarf an schnell verfügbaren Daten zur kurzfristigen Entwicklung des Arbeitsmarktes steigt. Das Statistische Bundesamt analysiert diesbezüglich bereits seit Jahren das Potenzial von Online-Stellenanzeigen. Insbesondere in Projekten des „ESSnet Big Data“ wurden verschiedene Machbarkeitsstudien zur Nutzung dieser neuen digitalen Daten durchgeführt. Ein neues Projekt verfolgt nun das Ziel, ein sogenanntes Web Intelligence Network (WIN) aufzubauen, welches die Sammlung, Verwaltung und den Transfer von Wissen zur Nutzung und Integration von Webdaten in nationale Statistikproduktionssysteme standardisieren soll. Das kurz als „ESSnet WIN“ bezeichnete Projekt konzentriert sich unter anderem auf Online-Stellenanzeigen.

↳ **Keywords:** online job portals – data quality – labour market reporting – big data – new digital data

ABSTRACT

The demand for readily available data on short-term developments in the labour market is growing. The Federal Statistical Office has been analysing the potential of online job advertisements in this regard for several years. Various feasibility studies on the use of these new digital data have been conducted, particularly as part of the 'ESSnet Big Data' projects. The goal of a new project currently underway is to develop a 'Web Intelligence Network' (WIN), which seeks to standardise the collection, management and transfer of knowledge on the use and integration of web data in national statistics production systems. Known as 'ESSnet WIN' for short, the project focuses, inter alia, on online job advertisements.



Dr. Martina Rengers

ist Volkswirtin und Referentin im Referat „Arbeitsmarkt“ des Statistischen Bundesamtes. Ihre Schwerpunkte liegen in der Analyse des 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“. Im Fokus sind dabei insbesondere Daten zu Online-Stellenanzeigen, ihre Qualitätsbewertung und Eignung für die Arbeitsmarktberichterstattung.



Andrea Hornik

ist Soziologin und Wissenschaftliche Mitarbeiterin im Referat „Arbeitsmarkt“ des Statistischen Bundesamtes. Hauptsächlich befasst sie sich mit der Analyse von Online-Stellenanzeigen aus Big Data im Projekt „Trusted Smart Statistics – Web Intelligence Network“. Im Vordergrund steht dabei neben der Einschätzung der Datenqualität die Nutzbarkeit als Arbeitsmarktindikator.

1

Einleitung

Das Europäische Statistische System (ESS) hat bereits 2015 ein erstes Big-Data-Projekt des European Statistical System Network (ESSnet) initiiert, das die Nutzung neuer digitaler Daten in der amtlichen Statistik analysiert. Das erste Big-Data-ESSnet-Projekt lief von Februar 2016 bis Mai 2018 und hat neben Online-Stellenanzeigen (online job advertisements – OJA) auch andere Anwendungsfälle von neuen digitalen Daten betrachtet.¹

Ein Nachfolgeprojekt ESSnet Big Data von 2018 bis 2020 befasste sich ebenfalls mit Online-Jobportalen als Datenquelle für die Statistikproduktion. Schließlich verfolgt seit März 2021 das ESSnet WIN (Web Intelligence Network) das Ziel, die Wissenssammlung, -verwaltung und den -transfer von Webdaten für die Nutzung und Integration in nationale Statistikproduktionssysteme zu standardisieren. Unter den zu nutzenden Webdaten bilden Online-Stellenanzeigen weiterhin einen Schwerpunkt.

Die Arbeitsmarktberichterstattung des Statistischen Bundesamtes hatte sich am ESSnet Big Data 2016 bis 2018 beteiligt und schon früh für die Frage interessiert, ob und inwieweit Stellenanzeigen von Online-Jobportalen für die amtliche Statistik nutzbar sind. Dieses Interesse wurde in der COVID-19-Pandemie durch den zunehmenden Bedarf an hochfrequenten und hochaktuellen Daten verstärkt.

Die internetbasierte Erfassung von Daten zu Online-Stellenanzeigen bietet dabei das Potenzial, Indikatoren in „Echtzeit“ zu generieren. Mit einem monatlichen Index zur (konjunkturellen) Entwicklung von Online-Stellenanzeigen kann das Potenzial der echtzeitbezogenen Informationen zum Online-Stellenmarkt zeitnah analysiert und aufbereitet werden. Dies erkannten auch Unternehmen, die entweder als Jobportalbetreiber selbst oder als kommerzielle Drittanbieter solche Daten erzeugen. So

1 Zu den anderen Anwendungsfällen gehörten Unternehmensdaten von Webseiten, Mobilfunkdaten, Smart-Meter-Daten zum Energieverbrauch sowie Navigations- und andere Schiffsdaten des Automatischen Identifikationssystems (AIS).

veröffentlichten erste Unternehmen² während der Pandemie – unabhängig von ihrem originären Geschäftsfeld – frei zugängliche, sehr zeitnahe Informationen zur kurzfristigen Entwicklung des Online-Stellenmarktes. Der Trend zu solchen Veröffentlichungen ist bis heute ungebrochen und wird sogar um zusätzliche private Datenanbieter erweitert.³

Das Statistische Bundesamt erhielt durch Data-Sharing bei verschiedenen ESSnet-Projekten Daten zu Online-Stellenanzeigen, ohne selbst ein technisch anspruchsvolles und zeitaufwendiges Datenerfassungssystem aufbauen zu müssen. In den ESSnet-Projekten wurde frühzeitig festgestellt, dass innerhalb des ESS oder der Europäischen Union (EU) bereits Datenerfassungssysteme für Daten zu Online-Stellenanzeigen von anderen Institutionen existieren.

Der vorliegende Beitrag beschäftigt sich mit den Ergebnissen der Analyse, inwieweit die amtliche Statistik Daten zu Online-Stellenanzeigen nutzen kann. Im Einzelnen fasst Kapitel 2 die Ergebnisse der früheren ESSnet-Big-Data-Projekte zusammen, bevor Kapitel 3 das aktuelle ESSnet WIN detaillierter vorstellt. Schließlich untersucht Kapitel 4 die Datenqualität der Online-Stellenanzeigen des ESSnet WIN. Fazit und Ausblick im letzten Kapitel stellen die Kernaussagen zur Nutzbarkeit dieser Daten zusammen und geben einen Ausblick auf die daraus resultierenden Konsequenzen für die Arbeitsmarktberichterstattung des Statistischen Bundesamtes. Darüber hinaus wird die generelle Zukunftsperspektive neuer digitaler Daten im ESS aufgezeigt.

2

ESSnet Big Data

Die Projekte ESSnet Big Data 2016-2018 und ESSnet Big Data 2018-2020 untersuchten die Nutzungsmöglichkeiten neuer digitaler Daten für die amtliche Statistik. Bereits im zweiten Projekt ging es unter anderem um eine erste Implementierung von Big-Data-Quellen in den statistischen Produktionsprozess. Von Beginn an waren

2 Dazu zählen Unternehmen wie Lightcast (ehemals Burning Glass) und Textkernel, aber auch die Jobportalanbieter Indeed und LinkedIn (de Lazzar/Rengers, 2021).

3 Seit September 2023 stellt auch die index Gruppe einen hochaktuellen Stellenmarktindikator zur Verfügung.

Online-Stellenanzeigen eine intensiv betrachtete neue digitale Datenquelle von besonderem Interesse.

Aufgrund des bereits erwähnten technisch anspruchsvollen und zeitaufwendigen Datengewinnungsprozesses sowie ungeklärter rechtlicher Fragen in Bezug auf die erforderlichen Scraping-Prozeduren, wurde nach Möglichkeiten des Data-Sharing gesucht. In diesem Zusammenhang kam das Europäische Zentrum für die Förderung der Berufsbildung CEDEFOP (Centre européen pour le développement de la formation professionnelle) ins Spiel, das bereits für das erste ESSnet Big Data 2016-2018 Daten zu Online-Stellenanzeigen für mehrere europäische Länder durch Web-Scraping zur Verfügung gestellt hatte. Durch eine kooperative Zusammenarbeit konnten die CEDEFOP-Daten dem ESSnet zur Verfügung gestellt werden (Rengers, 2018, hier: Seite 23 ff.).⁴

Auch beim zweiten ESSnet Big Data 2018-2020 lief das Data-Sharing erfolgreich (de Lazzar/Rengers, 2021, hier: Seite 74, 76 ff.). Die 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 begrüßt und weiterverfolgt (Descy und andere, 2018 und 2019).

Beide Projekte haben dabei eine Vielzahl an Herausforderungen offengelegt. Zu nennen sind zunächst die Dublettenbeseitigung und die bei Rengers (2018) und de Lazzar/Rengers (2021) beschriebenen Phänomene der zwei Coverage-Probleme.

Das erste Coverage-Problem ist die unvollständige Abdeckung der Gesamtheit der offenen Stellen durch Online-Stellenanzeigen. So hatten frühere Untersuchungen ergeben, dass zwischen 41 und 70% der Unternehmen ihre offenen Stellen auf Online-Jobportalen bekanntgeben, Online-Stellenanzeigen demnach nicht das gesamte Stellenangebot des Marktes repräsentieren. Außerdem führt die durch ihre hohe Anzahl erforderliche Auswahl von Jobportalen möglicherweise ebenfalls zu einer fehlenden Repräsentativität des Online-Stellenmarktes (zweites Coverage-Problem). Auch die in Deutschland zu beobachtende Schnellebigkeit in der Branche der Jobportalanbieter macht eine Auswahl von

⁴ Der bei Rengers (2018) erwähnte Bericht CEDEFOPs zur Machbarkeitsstudie über die Sammlung von „real-time labour market information“ (CEDEFOP, 2016) ist zum aktuellen Zeitpunkt allerdings nicht mehr im Internet verfügbar.

Jobportalen nötig, sodass der Online-Stellenmarkt nicht vollständig abgebildet werden kann.⁵

Die Beurteilung der Datenqualität von Online-Stellenanzeigen stellt eine der größten Herausforderungen dar. Eine Voraussetzung der Qualitätsprüfung ist, dass Datennutzende einen Einblick in den Datengewinnungs- und Datenaufbereitungsprozess bekommen und gegebenenfalls sogar Einfluss auf den Produktionsprozess haben. Das Data-Sharing oder der Datenankauf von Drittanbietern erfordert deshalb ein Minimum an Transparenz der Black Box der Datenproduktion. Metadaten wie Zeitpunkt der Erfassung, URL der Quelle, Art der Datengewinnung (API, Scraping, Crawling⁶) sind ebenso erforderlich wie die wichtigsten Strukturmerkmale sowie der Volltext der Anzeige. Ein technischer und methodisch-fachlicher Austausch zwischen Datennutzenden und Datenproduzent sollte dabei dauerhaft etabliert und institutionalisiert werden.

Eine intensive Struktur- und Outputanalyse der Daten und ihrer Metadaten ermöglichen den Datennutzenden, Qualitätsmängel und Coverageprobleme zu identifizieren. Das Statistische Bundesamt hat vier Methoden zur Beurteilung der Datenqualität genutzt (Rengers, 2018; de Lazzar/Rengers, 2021):⁷

1. Monitoring der Struktur der Daten aus Online-Stellenanzeigen (OJA) im Zeitverlauf

Das Monitoring erfolgt zunächst in einer einfachen deskriptiven Erfassung der wichtigsten Kenngrößen der Datensätze. Dazu gehört unter anderem eine Variablenliste mit Angabe der Ausprägungen, Missings und der technischen Art ihrer Erfassung (direkt

⁵ „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.“ (de Lazzar/Rengers, 2021, hier: Seite 74).

⁶ Bei Application Programming Interface (API) handelt es sich um Schnittstellen, die vom Portalbetreiber bereitgestellt werden, um gezielt Daten abzufragen. Die Methode (Web-)Scraping bezieht sich auf die gezielte Erhebung von spezifischen Inhalten einer Webseite durch maßgeschneiderte Software. Im Gegensatz dazu ist (Web-)Crawling ein automatisierter Prozess, bei dem Software ohne spezifische Vorgaben nach Inhalten auf verschiedenen Webseiten sucht.

⁷ In einem Video-Beitrag „Qualitätssicherung durch Struktur- und Outputanalyse“ hat Rengers diese Erkenntnisse im November 2021 auf einem von der Bertelsmann Stiftung und dem Bundesinstitut für Berufsbildung ausgerichteten Forum zum Potenzial von Daten aus Online-Stellenanzeigen dargelegt. Verfügbar unter: blog.aus-und-weiterbildung.eu [Zugriff am 11. September 2024].

aus dem Web gescrapt oder über Taxonomien generiert). Ebenso sind Angaben über die Art und Anzahl der verwendeten Quellen erforderlich. Von zentraler Bedeutung ist nicht zuletzt, den Anteil der Dubletten zu beobachten, um die Folgen des (unbekannten) Deduplizierungsprozesses auf die Ergebnisse abzuschätzen.

2. Vergleich der Ergebnisse verschiedener OJA-Daten

Verschiedene Jobportalanbieter machen Angaben zur Anzahl ihrer Online-Stellenanzeigen direkt auf ihren Portalen oder werten sogar ihre eigenen Jobportale aus und veröffentlichen diese Auswertungen. Die eigenen Daten lassen sich deshalb mit diesen Ergebnissen insgesamt oder auf der Ebene einzelner Jobportale vergleichen.

3. Vergleich der Ergebnisse mit anderen Datenquellen

Bei der Überprüfung der Datenqualität mithilfe von anderen Datenquellen ist die Vergleichbarkeit sicherzustellen. Bei Stellenanzeigen ist es naheliegend, die OJA-Daten den Ergebnissen der regelmäßig vom Institut für Arbeitsmarkt- und Berufsforschung (IAB) durchgeführten Betriebsbefragungen zur Entwicklung des gesamtwirtschaftlichen Arbeitskräftebedarfs (IAB-Stellenerhebung) gegenüberzustellen.

4. Bereinigung der OJA-Daten durch gezielte Stichprobenziehung

Eine im Monitoring beobachtete Instabilität der bei der Datenerfassung verwendeten (Jobportal-)Webseiten kann möglicherweise durch eine nachträgliche gezielte Stichprobenziehung aus dem Datensatz behoben werden. Als Kriterien für die Stichprobenziehung können externe Ranglisten⁸ dienen, die regelmäßig den Stellenwert von Jobportalen für den Stellenmarkt bewerten. Auch kann eine Stichprobe anhand des Scrapingdatums gezogen werden, um einen synthetischen Bestand (Pseudostocks) an Online-Stellenanzeigen zu ermitteln und technisch bedingte Brüche im Datengewinnungsprozess nachträglich zu glätten.

Die Erkenntnisse aus der Anwendung der beschriebenen Methoden lassen sich wie folgt zusammenfassen (Rengers, 2018; de Lazzer/Rengers, 2021):

- › Die in den CEDEFOP-OJA-Daten für den deutschen Stellenmarkt verwendeten 159 Jobportale wurden nie alle gleichzeitig genutzt. Im Mittel wurden zwischen 76 und 92 Jobportale je Monat gescrapt, davon waren durchschnittlich 3 bis 5 Portale numerisch bedeutend mit mehr als 100 000 Online-Stellenanzeigen je Monat.
- › Alle untersuchten Datensätze veränderten ihre Struktur im Zeitverlauf deutlich. Dies galt sowohl für die Anzahl der Stellenanzeigen als auch für die relative Bedeutung einzelner Jobportale im Hinblick auf die Anzahl der gewonnenen Stellenanzeigen. Revisionen veränderten die Strukturen sogar rückwirkend.
- › Der Anteil der Dubletten bei den Online-Stellenanzeigen zeigte starke Schwankungen im Zeitverlauf und veränderte sich durch die Revisionen. Die Unterschiede betrugen bis zu 70 Prozentpunkte – ein Hinweis darauf, dass die Datenaufbereitungsprozesse zur Identifikation von Dubletten bei den CEDEFOP-OJA-Daten äußerst volatil waren und häufigen, systematischen Veränderungen unterlagen.
- › Eine Vergleichbarkeit der Ergebnisse von CEDEFOP mit der IAB-Stellenerhebung ist nur eingeschränkt gegeben. Ein Grund ist, dass auf der einen Seite der Online-Stellenmarkt und auf der anderen der Gesamtstellenmarkt betrachtet wird. Auch besteht das Problem, dass die Daten in der Regel keine Angaben zur Anzahl der Stellen je Stellenanzeige enthalten.
- › Vergleiche der Ergebnisse verschiedener OJA-Daten zeigten deutliche Unterschiede. So wurde aus den CEDEFOP-Daten ein monatlicher Online-Stellenindex den vergleichbar gemachten Ergebnissen von Indeed und der Firma Textkernel gegenübergestellt. Hier zeigten die CEDEFOP-Daten eine höchst volatile Zeitreihe, die im Gegensatz zu den anderen OJA-Daten konjunkturelle Einbrüche wie die in der Pandemie zum einen schwer erkennbar machte, diese jedoch zum anderen im Ausmaß deutlich überzeichneten.

⁸ Zum Beispiel von deutschlandsbestejobportale.de oder crosswater-job-guide.com

3

ESSnet WIN

3.1 Entstehung und Bedeutung

Hauptbestandteil des ESSnet-Projekts 2021-2025 „Trusted Smart Statistics – Web Intelligence Network“ (ESSnet WIN) ist das Web Intelligence Hub (WIH), in welchem Eurostat, das Statistische Amt der EU, Webdaten sammelt und für bestimmte Anwendungsfälle vorbereitet. Das WIH soll eine qualitativ hochwertige Quelle für Webdaten, Methoden und Algorithmen etablieren, die zur Erstellung europäischer und nationaler amtlicher Statistiken verwendet werden kann. In einem der Anwendungsfälle werden Online-Stellenanzeigen gesammelt und bezüglich ihrer Verwendung für eine kurzfristige zusätzliche oder eine ergänzende Arbeitsmarktberichterstattung (wie die Arbeitsmarktnachfrage spezieller Qualifikationen, Skills) bewertet. Die Daten zu den Online-Stellenanzeigen werden weiterhin zentral für mehrere Länder in Europa erfasst und unter dem Label „WIH-OJA-Daten“ von Eurostat verwaltet.

↳ Konzept der „Trusted Smart Statistics“

Im Vergleich zu den beiden vorherigen ESSnet-Big-Data-Projekten stellt sich zunächst die Frage, wie es zu der neuen Namensgebung „WIN“ und der Begrifflichkeit „Trusted Smart Statistics“ (TSS) kam.

Dies hat mit den Erfahrungen zu tun, die während und nach der Durchführung der Projekte des ESSnet Big Data gesammelt und welche Schlüsse auf europäischer Ebene daraus gezogen wurden. Es besteht ein großer Druck auf die amtliche Statistik, die neuen Datenquellen zu nutzen. Dies ist einerseits durch die Konkurrenz aus der Privatwirtschaft, andererseits auch durch eine veränderte gesellschaftliche Erwartungshaltung gegenüber der Verwendung digitaler Daten bedingt (Ricciato und andere, 2019, hier: Seite 590 f.; Ricciato/Wirthmann, 2019).

Gleichzeitig reifte die Erkenntnis, dass die Datengewinnung, der -zugriff, die -verarbeitung, die -haltung und die -auswertung neuer digitaler Daten ein vollständig neues System benötigen, welches sich grundlegend von den etablierten Datenquellen unterscheidet. Ein Übergang vom Konzept „Big Data in der amtlichen Statistik“ zum Konzept der „Trusted Smart Statistics“, kurz TSS-Konzept, wurde nötig und inhaltlich konkretisiert. Ein solches neues System soll das Altsystem nicht ersetzen,

sondern neben ihm bestehen. Es steht für einen Paradigmenwechsel, eine neue interinstitutionelle Zusammenarbeit und ein neues Governance-System in Bezug auf die amtliche Statistikproduktion.⁹

Die Qualität amtlicher Statistiken muss auch bei der Verwendung neuer digitaler Datenquellen weiterhin auf einem hohen Niveau gewährleistet sein. Entsprechende offizielle Dokumente, die Qualitätsstandards festlegen, wurden deshalb überarbeitet und diesbezüglich erweitert. Zu nennen sind hier der Verhaltenskodex für europäische Statistiken (Europäische Union, 2018) und der UN-Statistik-Qualitätssicherungsrahmen (United Nations, 2019), die beide Big-Data-Datenquellen explizit mit aufgenommen haben.¹⁰

Der Wechsel der Begrifflichkeit von CEDEFOP-OJA-Daten zu WIH-OJA-Daten von Eurostat hängt mit der Idee des Data-Sharings und der interinstitutionellen Zusammenarbeit im Anwendungsfall der Online-Stellenanzeigen zusammen. So wurde ein speziell zugeschnittenes EU-Projekt 2020-2024 namens „Towards the European Web Intelligence Hub – European System for Collection and Analysis of Online Job Advertisement Data (WIH-OJA)“ aufgesetzt, das gemeinschaftlich von Eurostat und CEDEFOP durchgeführt wird. Es verfolgt das Ziel, Daten in den European Trusted Smart Statistics Centre' Web Intelligence Hub WIH zu implementieren. Bezüglich der Produktion von Daten zu Online-Stellenanzeigen bedeutet das weiterhin ein Outsourcing des gesamten Produktionsprozesses, unter Zusammenlegung von Ressourcen und Fachwissen beider Institutionen. Dabei wird das Erfassungssystem der CEDEFOP-OJA-Daten weiterentwickelt und für die nationalen Statistikämter unter dem Label WIH-OJA-Daten von Eurostat dauerhaft etabliert und institutionalisiert zur Verfügung gestellt (CEDEFOP, 2020; Descy und andere, 2018 und 2019; Eurostat, 2023).

⁹ Es wird der Vergleich mit dem Motor eines Autos angestellt, der nun mit einem neuen Treibstoff der neuen digitalen Daten betrieben werden soll, aber dafür nicht geeignet ist und auch nicht entsprechend umgerüstet werden kann. Es soll deshalb ein zweiter Motor für den neuen Kraftstoff gebaut werden, wobei der alte Motor mit dem alten Kraftstoff der traditionellen Datenquellen weiterhin bestehen bleibt (Ricciato und andere, 2019, hier: Seite 592; Ricciato und andere, 2020, hier: Seite e7-2).

¹⁰ Bei United Nations (2019) sind gesonderte Listen mit Referenzen zu verschiedenen Arten von Datenquellen enthalten. Für Big Data wird unter anderem auf das [Webportal](#) der „Global Working Group on Big Data for Official Statistics“ verwiesen [Zugriff am 11. September 2024]. Einen weiteren Überblick über die Literatur zum Qualitätsrahmen für Big Data findet man beispielsweise bei Kiss-Nagy und andere (2022).

3.2 Öffentlichkeitsarbeit

Frei zugängliche Informationen zum ESSnet WIN stellt Eurostat zum einen klassisch über die [Webseite](#) der Europäischen Kommission zur Verfügung. Zum anderen werden aber auch Social-Media-Kanäle wie LinkedIn, X und YouTube genutzt. [↘ Übersicht 1](#) zeigt die wichtigsten Links.

Übersicht 1

Verbreitungskanäle des ESSnet WIN

	https://cros.ec.europa.eu/landing-page/web-intelligence-network
	Unternehmensseite „ESSnet – Web Intelligence Network“ linkedin.com/company/essnet-project-web-intelligence-network
	X.com/EssnetWIN
	https://www.youtube.com/@ESSnetWIN mit 14 Videos zu Webinaren, Tutorials

Darüber hinaus werden die Themen des ESSnet WIN auf vielen bedeutenden Konferenzen verbreitet (siehe dazu den folgenden Exkurs). Außerhalb turnusmäßiger Konferenzen fand im Rahmen des ESS im Juni 2023 eine sogenannte Web Intelligence Hub Conference WIH-CON¹¹ mit folgenden Themenschwerpunkten statt: Aufbau und Funktionsweise des WIH, neue Anwendungsfälle des WIH außerhalb des Bereichs Online-Stellenanzeigen OJA, innerhalb des OJA-Bereichs die Themen Auswahl von Jobportalen (Landscaping), Qualität, Nutzung von künstlicher Intelligenz zur Datenaufbereitung und Klassifikation. Darüber hinaus wurden auch fachliche Ergebnisse aus den erhobenen OJA-Daten thematisiert, wie sie beispielsweise von CEDEFOP oder Eurostat veröffentlicht werden.¹²

11 Siehe dazu www.wih-con.eu und www.wih-con.eu/en/reference-documents. Bei der letztgenannten URL erhält man nach Angabe einer Mailadresse einen temporären Code und die Video-Aufzeichnungen und Präsentationen der Konferenz lassen sich abrufen (Stand: 5. August 2024).

12 Im Frühjahr 2021 veröffentlichte CEDEFOP die OJA-Daten aus dem WIH in Zusammenarbeit mit Eurostat als Dashboards unter dem Namen [Skills-OVATE](#). Im Dezember 2023 hat Eurostat im Rahmen des „European Year of Skills“ ein [Tool](#) zur Visualisierung von Skills veröffentlicht.

↘ ESSnet WIN-Konferenzbeiträge

Einen umfangreichen Informationsaustausch zum ESSnet WIN gab es beispielweise bei den sogenannten Q-Konferenzen, die die Sicherung der Qualität in der amtlichen Statistik thematisieren und innerhalb des ESS im zweijährlichen Rhythmus stattfinden. Die Q2024 im portugiesischen Estoril hatte jeweils eigene Sessions zur Qualität webbasierter Daten, daraus entwickelten experimentellen Statistiken und zum gesamten Themenkomplex von Online-Stellenanzeigen. Hier sind insbesondere die Konferenzbeiträge von Kvetan/Branka (2024), EUROSTAT/CEDEFOP/Sogeti (2024), Čiginas/Šlevinskas/Burakauskaitė (2024) und Georgiev und anderen (2024) von Relevanz.

Bereits zwei Jahre zuvor hatte die Q2022 im litauischen Vilnius Erkenntnisse zur Nutzung von OJA-Daten thematisiert (siehe dazu insbesondere Kiss-Nagy und andere, 2022, Maślankowski/Nowak, 2022, sowie Branka und andere, 2022).

Eine weitere internationale wissenschaftliche Konferenzreihe ist die NTTS – New Techniques and Technologies for Statistics. Sie wird alle zwei Jahre von Eurostat organisiert und befasst sich mit neuen Techniken und Methoden für die amtliche Statistik sowie mit den Auswirkungen neuer Technologien auf die Systeme zur Erhebung, Erstellung und Verbreitung von Statistiken. Die NTTS 2023 setzte ebenfalls einen Schwerpunkt auf die Themen „Web Data“, „Online Data“ und „Online Job Advertisement“. Die Eurostat-Webseite zur Konferenz (https://ec.europa.eu/eurostat/cros/content/NTTS2023_en) ist jedoch nicht mehr erreichbar, sodass hier das in Papers zusammengetragene Wissen für extern Nutzende nicht mehr zur Verfügung steht.

4

Online-Stellenanzeigen im ESSnet WIN

4.1 Jobportale und Stellenanzeigen

Datenzugang

Seit Anfang 2024 stellt Eurostat die WIH-OJA-Daten über die neue WIH-Datenplattform zur Verfügung. Das sogenannte Datalab ist über eine HTTPS-Verbindung erreichbar und baut auf der Big Data Test Infrastructure (BDTI) der Europäischen Kommission auf. Diese stellt für Einrichtungen der öffentlichen Verwaltung eine Sammlung

von Cloud-basierten Diensten mit Komponenten zur Datenspeicherung, -verarbeitung und -analyse sowie zu Visualisierung und Management unter Berücksichtigung von Datensicherheit bereit. So ist der Datenabruf aus dem Datalab über angehängte Dienste, wie eine RStudio-Umgebung, möglich. Der Zugang erfolgte im Rahmen des in Abschnitt 3.1 beschriebenen Projekts ESSnet WIN.

Verfügbar sind die Mikrodaten für den Zeitraum ab Juli 2018, die jeweils quartalsweise mit einer Verzögerung von ein bis zwei Monaten veröffentlicht werden. Der Gesamtdatensatz enthält 59 Variablen, wobei weitere Variablen in separaten Tabellen unter zusätzlicher Einschränkung vorhanden sind. Die Variablen umfassen unter anderem Informationen zu Einstell- und Ablaufdatum, Einsatzort, Portal und Vertragseigenschaften sowie Klassifikationen der Beschäftigung, des Wirtschaftszweiges, des geforderten Bildungsniveaus und der Fähigkeiten (Skills).

Methode

Zur Untersuchung der Datenqualität werden Stellenangebote von privaten Personalvermittlungen ausgeschlossen. Zur Identifizierung wurden die Namen der Unternehmen mit Angaben aus öffentlichen Branchenbüchern verglichen. Dieses Vorgehen ermöglichte die Unterscheidung zur Zeitarbeit, die in den Daten enthalten bleibt. Etwa 1,2% der verfügbaren Anzeigen einschließlich Dubletten haben private Personalver-

mittlungen veröffentlicht und weitere 3,6% Zeitarbeitsfirmen.¹³ Stellenanzeigen für Praktika verbleiben in der Stichprobe, da die im Datensatz enthaltene Variable aufgrund einer großen Zahl an Fehlwerten nicht verwendet werden kann.

Landscaping

Um geeignete Portale zu erfassen, hat Eurostat in Zusammenarbeit mit internationalen Fachleuten eine Reihe relevanter Portale identifiziert und im Hinblick auf technische Aspekte bewertet (Eurostat, 2023).

Für Deutschland stehen so für den Zeitraum Juli 2018 bis Juni 2023 Stellenanzeigen aus insgesamt 160 Jobportalen zur Verfügung, die jedoch nie alle gleichzeitig erfasst wurden. [↘ Tabelle 1](#) enthält eine Übersicht der Stellenanzeigen und der erfassten Portale. Die Anzahl der Stellenanzeigen sinkt zwischen 2019 und 2022 insgesamt und durchschnittlich je Monat. Der größte Teil der Jobportale liefert im Schnitt weniger als 1 000 Jobanzeigen je Monat, während von zwei oder drei Portalen mehr als 100 000 Stellenanzeigen im Monatsdurchschnitt anfallen. Mit insgesamt 10 Millionen Anzeigen von insgesamt 64 tatsächlich verwendeten Jobportalen je Monat liegen die Daten für das Jahr 2019 über denen der anderen Jahre.

13 Eine Abgrenzung, welche Stellen von privaten Personalvermittlungen vermittelt und welche von ihnen selbst als Arbeitgebende angeboten werden, ist mit den vorliegenden Daten nicht möglich.

Tabelle 1

Online-Stellenanzeigen im ESSnet WIN: realisierte Quellen/Websites (einschließlich Dubletten)

	Monate	Stellenanzeigen		Tatsächlich verwendete Portale	Darunter: durchschnittlich je Monat tatsächlich verwendete Portale, von denen mindestens ...		
		insgesamt	je Monat	Durchschnitt je Monat	1 000	50 000	100 000
					Stellenanzeigen gewonnen wurden		
	Anzahl						
2018	6	4 394 033	732 339	58	19	4	3
2019	12	10 009 271	834 106	64	20	4	2
2020	12	7 863 183	655 265	44	17	4	2
2021	12	7 863 860	655 322	31	17	4	3
2022	12	7 376 241	614 687	24	16	4	3
2023	6	3 894 370	649 062	24	17	4	3

Quelle: WIH-OJA-Daten von Eurostat, eigene Berechnungen

Anmerkung: Daten im Zeitraum von Juli 2018 bis Juni 2023 (Stand: August 2023); unter Einschluss von Dubletten, Praktika und Zeitarbeit; ohne private Personalvermittlung.

Da die Relevanz von Portalen im Zeitverlauf variieren kann, neue Portale hinzukommen und alte wegfallen, ist es nötig, das Landscaping periodisch zu wiederholen.

4.2 Variablen, Missings und Klassifikation

Das Scraping der Portale erfolgt täglich und hängt im ersten Schritt von technischen Aspekten ab. Durch aktive Verhinderung von Scraping, temporäre Nicht-Erreichbarkeit der Webseiten oder Änderungen im Quelltext werden Daten nicht kontinuierlich oder teils gar nicht mehr gesammelt. Neben wegfallenden Portalen können auch systematisch unvollständige Jobbeschreibungen den späteren Datensatz verzerren.

Um die Daten von Doppelerfassungen zu bereinigen, werden in diesem Prozessschritt zudem wiederholte URL entfernt sowie Anzeigen mit gleichen oder ähnlichen Inhalten und Metadaten verschmolzen (Eurostat, 2023).

Nach dem Scraping wird der Textinhalt bereinigt und unter anderem mit Ontologien¹⁴ und Machine-Learning-Modellen in Variablen überführt.

Nutzende können kaum abschätzen, wie valide der Prozess der Klassifikation oder des Scrapings verlief. Ein Indiz können hier die Fehlwerte liefern, die in den Variablen in unterschiedlicher Intensität vorhanden sind. Angaben zum Veröffentlichungsdatum oder der Portalstabilität¹⁵ haben keine Fehlwerte, während etwa die Wochenarbeitszeit oder die Einsatzstadt überproportional viele Missings aufweisen. Exemplarisch lässt sich dies auch anhand der Variablen zu Vertragsarten zeigen. Diese unterscheidet, ob in einer Jobanzeige ein Praktikum, eine Selbstständigkeit, eine befristete oder unbefristete Stelle angeboten wird. [↘ Tabelle 2](#) zeigt einen Auszug dieser Variable für den Zeitraum von Juli 2018 bis Juni 2023 für Deutschland. Der Anteil der Praktika nimmt im Zeitablauf stark ab, aber eine Aussage über mögliche strukturelle Veränderungen bei der Vertragsart von angebotenen Stellen ist aufgrund des gestiegenen

Anteils der Missings nicht möglich. Ab dem zweiten Quartal 2020 gab es bei über 50 % der Stellenanzeigen keine Angabe zur Vertragsart, im vierten Quartal 2020 erreichte der Missinganteil sogar ein Maximum von 83,7 %.

Tabelle 2

Online-Stellenanzeigen im ESSnet WIN: Missings bei Vertragsart (einschließlich Dubletten)

		Stellenanzeigen nach Vertragsart		
		insgesamt	darunter:	
			Praktika	nicht zugeordnet
		Anzahl	%	
2018	3. Quartal	1 941 826	18,5	22,2
	4. Quartal	2 452 207	17,4	21,3
2019	1. Quartal	3 313 191	16,4	25,1
	2. Quartal	2 035 668	14,5	27,7
	3. Quartal	2 536 743	13,7	36,3
	4. Quartal	2 123 669	12,3	43,8
2020	1. Quartal	2 728 636	11,3	48,5
	2. Quartal	1 348 251	7,6	65,1
	3. Quartal	1 901 421	2,3	82,3
	4. Quartal	1 884 875	1,7	83,7
2021	1. Quartal	2 192 789	1,8	83,6
	2. Quartal	1 590 764	2,7	76,2
	3. Quartal	2 001 745	2,9	75,9
	4. Quartal	2 078 562	4,2	68,8
2022	1. Quartal	2 123 346	6,4	59,4
	2. Quartal	1 394 679	6,0	64,0
	3. Quartal	2 074 726	5,7	64,0
	4. Quartal	1 783 490	5,8	65,6
2023	1. Quartal	2 137 435	6,2	59,7
	2. Quartal	1 756 935	6,1	62,4

Quelle: WIH-OJA-Daten von Eurostat, eigene Berechnungen

Anmerkung: Daten im Zeitraum von Juli 2018 bis Juni 2023 (Stand: August 2023); unter Einschluss von Dubletten, Praktika und Zeitarbeit; ohne private Personalvermittlung.

14 Im Kontext von Datenmanagement sind Ontologien eine Menge von Begriffen und ihre Beziehung zueinander. In Matching-Algorithmen werden diese für Klassifizierungszwecke verwendet.

15 Die Variable `source_stability` weist Portalen die Attribute stabil, instabil und neu zu. Für den gesamten Datensatz im Zeitraum von Juli 2018 bis Juni 2023 besitzen Portale insgesamt nur eines dieser Attribute, sodass temporäre Effekte nicht abgebildet werden.

4.3 Dubletten und Revisionen

Nach der Klassifikation werden die Ausprägungen der Online-Stellenanzeigen miteinander verglichen. Dabei werden Stellenanzeigen als Dubletten erkannt, wenn sie identische Ausprägungen in den Variablen zum Veröffentlichungsdatum, der Beschäftigung, dem Einsatzort, den Fähigkeiten und dem Wirtschaftszweig haben. Doppelte Anzeigen werden verschmolzen, sodass im Datensatz nur noch Dubletten verbleiben, die auf unterschiedlichen Portalen veröffentlicht worden sind.¹⁶

↘ **Tabelle 3** zeigt die in den Jahren 2018 bis 2023 erfassten Online-Stellenanzeigen und die identifizierten Dubletten. Der Anteil von Dubletten variiert zwischen 1,6% und 7,5%. Für das Jahr 2021 sind im Vergleich zum Vor- und Folgejahr weniger Dubletten vorhanden, obwohl jeweils gleich viele Stellenanzeigen erfasst worden sind.¹⁷

Neben den Schwankungen im Zeitverlauf sind die WIH-OJA-Daten von Revisionen betroffen. Grund hierfür sind

- 16 Im Dezember 2022 startete Eurostat (Europäische Kommission, 2024) im Rahmen der European Statistics Awards eine Deduplication Challenge mit OJA-Daten aus dem WIH. Ziel war es, einen neuen Ansatz für den Deduplizierungsprozess von Online-Daten zu finden. Es ist möglich, dass sich der bisherige Prozess in Zukunft ändert und zu Revisionen führt.
- 17 Bei früheren Analysen wurde mit Werten zwischen 17,6 und 89,1% ein deutlich höherer Anteil an Dubletten ermittelt. In den damaligen Berechnungen konnte allerdings nicht zwischen aufbereitungsbedingten Vervielfältigungen und „echten“ Dubletten unterschieden werden, weshalb die Angaben bei de Lazzer/Rengers (2021, hier: Seite 80) überschätzt wurden. Vervielfältigungen sind durch den Datenaufbereitungsprozess von Variablen mit Mehrfachnennungen bedingt, während bei „echten“ Dubletten dieselbe Stellenanzeige auf mehreren Jobportalen erscheint.

Fehlerkorrekturen, Änderungen im Klassifikationssystem oder in der methodischen Aufbereitung (Eurostat, 2023).

↘ **Grafik 1** auf Seite 58 enthält die Verläufe der Stellenanzeigen ohne Dubletten insgesamt anhand dreier verschiedener Datensätze, die jeweils ab Juli 2018 beginnen und bis Dezember 2021 (2021Q4), Dezember 2022 (2022Q4) und Juni 2023 (2023Q2) ermittelt worden sind. Für den Datensatz bis Juni 2023 wurden, mit Ausnahme der Monate Mai und August 2021, weniger Stellenanzeigen berücksichtigt als in früheren Versionen. Am stärksten fällt die Differenz im Zeitraum von September 2021 bis November 2022 auf, in dem fast durchgehend 300 000 Stellenanzeigen je Monat weniger ermittelt worden sind.

Frühere Analysen von de Lazzer/Rengers (2021) zeigten, dass eine Begrenzung auf fünf relevante Portale die Volatilität senken kann. Dafür wurden Portale unter anderem anhand externer Ranglisten beurteilt und eine Substichprobe gezogen. Als besonders bedeutend stellten sich die fünf Jobportale Adzuna, Bundesagentur für Arbeit, Jobbörse, XING und StepStone heraus. Bei den aktuellen WIH-OJA-Datensätzen wurden für diese fünf Portale allerdings teilweise keine Daten mehr erfasst. Spätestens mit dem Berichtsmonat Dezember 2021 entfielen vollständig alle Daten der drei Jobportale Adzuna, StepStone und Jobbörse, bei Adzuna und StepStone trat dies bereits Monate früher ein. Der Datenproduzent gab für diesen Datenwegfall keine Gründe an.

Tabelle 3
Online-Stellenanzeigen im ESSnet WIN: Dubletten im Zeitverlauf

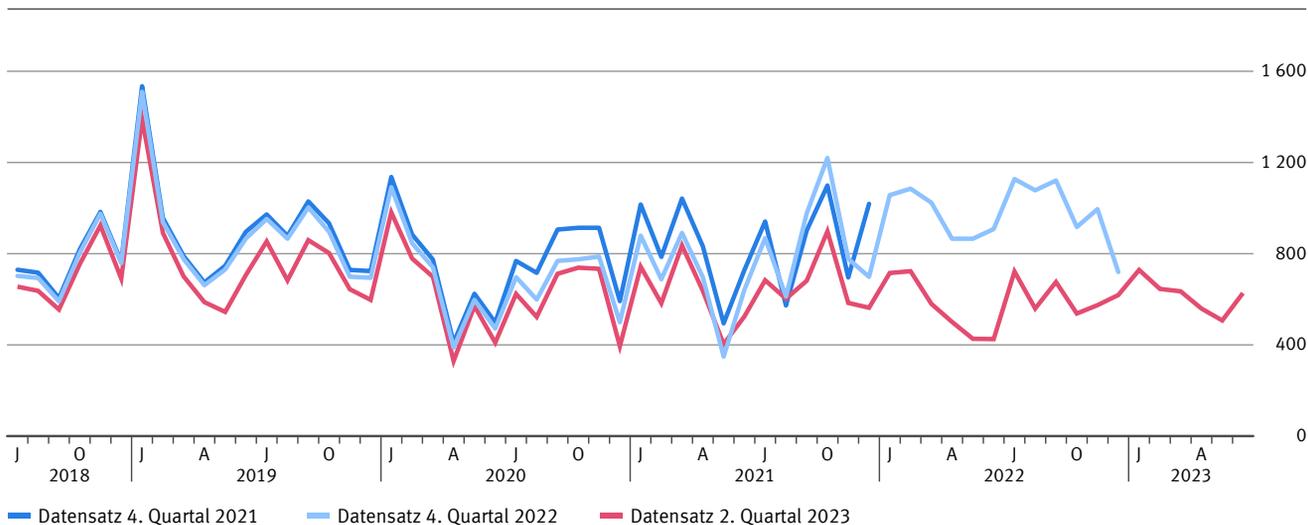
	Monate	Stellenanzeigen einschließlich Dubletten		Stellenanzeigen ohne Dubletten		Anteil der Dubletten
		insgesamt	je Monat	insgesamt	je Monat	
	Anzahl					%
2018	6	4 394 033	732 339	4 215 284	702 547	4,1
2019	12	10 009 271	834 106	9 259 344	771 612	7,5
2020	12	7 863 183	655 265	7 490 570	624 214	4,7
2021	12	7 863 860	655 322	7 737 092	644 758	1,6
2022	12	7 376 241	614 687	7 055 622	587 969	4,3
2023	6	3 894 370	649 062	3 702 565	617 094	4,9

Quelle: WIH-OJA-Daten von Eurostat, eigene Berechnungen

Anmerkung: Daten im Zeitraum von Juli 2018 bis Juni 2023 (Stand: August 2023); unter Einschluss von Dubletten, Praktika und Zeitarbeit; ohne private Personalvermittlung.

Grafik 1

Monatliche Stellenanzeigen im Vergleich: Volatilität und Revisionen (ohne Dubletten)
in 1 000



Daten im Zeitraum von Juli 2018 bis Dezember 2021 (2021Q4), Dezember 2022 (2022Q4) und Juni 2023 (2023Q2); unter Einschluss von Praktika und Zeitarbeit; ohne Dubletten und private Personalvermittlung (Stand: August 2023).

Quelle: WIH-OJA-Daten von Eurostat, eigene Berechnungen

5

Fazit und Ausblick

Die internetbasierte Erfassung von Daten zu Online-Stellenanzeigen hat eine große Bedeutung in der Arbeitsmarktberichterstattung, da sie hochfrequente und hochaktuelle Daten über die Arbeitskräftenachfrage liefern kann. Dies ermöglicht neben der Analyse gefragter spezieller Qualifikationen und Fähigkeiten (Skills) insbesondere auch, die wirtschaftliche Lage durch Nutzung von Echtzeitdaten schnellstmöglich abzubilden.

Unabhängig vom Anwendungsfall der Online-Stellenanzeigen zeigte sich auf europäischer Ebene, dass die Gewinnung, der Zugriff, die Verarbeitung, die Haltung und die Auswertung neuer digitaler Daten ein vollständig neues System benötigen, welches sich grundlegend von dem der etablierten Datenquellen unterscheidet. Als Konsequenz erfolgte der Übergang vom Konzept „Big Data in der amtlichen Statistik“ zum Konzept der „Trusted Smart Statistics“.

Welche Dimension und Bedeutung der Aufbau eines solchen neuen TSS-Konzeptes für die amtliche Statistikproduktion beinhaltet, zeigen die mittlerweile zahlrei-

chen Veröffentlichungen zum Umgang und zur Nutzung neuer digitaler Datenquellen sowie die Öffentlichkeitsarbeit auf Social-Media-Kanälen wie LinkedIn, Twitter, X und YouTube.

Um neue digitale Daten erfassen, aufbereiten und verwenden zu können, sind jedoch eine Vielzahl an Herausforderungen zu bewältigen. Innerhalb des ESS wurde deshalb die Strategie der gemeinsamen Nutzung von europäischen Daten zu Online-Stellenanzeigen entwickelt, um Synergieeffekte nutzbar zu machen und das interinstitutionelle Wissen und technische Ressourcen zu bündeln.

So konnte in den früheren ESSnet-Big-Data-Projekten das bereits existierende CEDEFOP-OJA-Datenerfassungssystem des Europäischen Zentrums für die Förderung der Berufsbildung CEDEFOP als Data-Sharing genutzt werden. Im aktuellen ESSnet WIN erfolgte mit dem Aufbau eines Web Intelligence Network (WIN) eine weitere Standardisierung bei der Nutzung und Integration von Webdaten und insbesondere von Daten zu Online-Stellenanzeigen in nationale Statistikproduktionssysteme. Dazu wurde ein Web Intelligence Hub (WIH) errichtet, mit dem das Erfassungssystem der CEDEFOP-OJA-Daten weiterentwickelt wird und Eurostat die Daten für die nationalen Statistikämter dauerhaft etabliert und institutionalisiert zur Verfügung stellt.

Dieser Beitrag vergleicht die neue Datenbasis mit den Vorgängern des ESSnet Big Data hinsichtlich der Datenqualität. Dabei konnte auf umfangreiche Untersuchungen von Rengers (2018) und de Lazzer/Rengers (2021) zurückgegriffen werden. Bereits das Monitoring der Struktur der WIH-OJA-Daten im Zeitablauf zeigt weiterhin deutliche Qualitätsmängel und bringt keine grundlegend anderen Resultate hervor als für diejenigen der CEDEFOP-OJA-Daten des ESSnet Big Data.

Es bleibt festzuhalten, dass sich die zeitnahe Bereitstellung der Daten sowie der Datenzugang über die neue WIH-Datenplattform verbessert haben. Auch haben Datennutzende mehr Einblick in die Black Box des Datenproduktionsprozesses, indem Eurostat eine standardisierte Veröffentlichung zu WIH-OJA-Metadaten bereitgestellt hat. Die (bisher) dokumentierten Metadaten sind jedoch noch ausbaufähig und nicht detailliert genug. Auch der im ESS gewünschte technische und insbesondere methodisch-fachliche Austausch zwischen Datennutzenden und Datenproduzenten sollte intensiviert werden.

Das Statistische Bundesamt verfolgt dieses Ziel weiterhin, da das Potenzial, das mit der Nutzung von Online-Stellenanzeigen für die (kurzfristige) Arbeitsmarktberichterstattung verbunden ist und die durch Eurostat dauerhaft bereitgestellte Datenquelle an umfangreichen WIH-OJA-Daten nicht ungenutzt bleiben sollten.

Deshalb sollen in künftigen Beiträgen die WIH-OJA-Daten detailliert analysiert und die Erkenntnisse im Rahmen des ESS mit den Datenerzeugern diskutiert werden. Die Diskussion zwischen Datenerzeuger und dem Statistischen Bundesamt als Datennutzer soll insbesondere im Hinblick auf mögliche Anpassungen im Datenproduktionsprozess zur Verbesserung der Datenqualität geführt werden. 

LITERATURVERZEICHNIS

Branka, Jiri/Kvetan, Vladimir/Napierala, Joanna. *From the online job advertisements to official statistics – the aspects of quality assurance*. 2022. Q2022 Conference, Vilnius. Paper Session 21.3. [Zugriff am 12. September 2024]. Verfügbar unter: q2022.stat.gov.lt

CEDEFOP. *Cedefop and Eurostat formalise joint approach to online job advertisement data*. Pressemeldung vom 28. Mai 2020. [Zugriff am 12. September 2024]. Verfügbar unter: www.cedefop.europa.eu

Čiginas, Andrius/Šlevinskas, Donatas/Burakauskaitė, Ieva. *Combining online job advertisements with probability sample data for enhanced small area estimation of job vacancies*. 2024. Q2024 Conference, Estoril. Paper Session 12.1. [Zugriff am 12. September 2024]. Verfügbar unter: airdrive.eventsair.com

De Lazzer, Jakob/Rengers, Martina. *Auswirkungen der Coronakrise auf den Arbeitsmarkt: experimentelle Statistiken aus Daten von Online-Jobportalen*. In: WISTA Wirtschaft und Statistik. Ausgabe 3/2021, Seite 71 ff.

Descy, Pascaline/Kvetan, Vladimir/Wirthmann, Albrecht/Reis, Fernando. *Towards a shared infrastructure for online job advertisement data*. In: Statistical Journal of the IAOS. Band 35, Ausgabe 4, Seite 669 ff. DOI: [10.3233/SJI-190547](https://doi.org/10.3233/SJI-190547)

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 12. September 2024]. Verfügbar unter: dgins2018.statisticsevents.ro

Europäische Union. *European Statistics Code of Practice – revised edition 2017*. 2018. [Zugriff am 12. September 2024]. Verfügbar unter: ec.europa.eu

Europäische Kommission. *Lessons learned from Eurostat's deduplication challenge*. 2024. [Zugriff am 12. September 2024]. Verfügbar unter: cros.ec.europa.eu

Eurostat. *WIH-OJA – Reference Metadata*. 2023. [Zugriff am 12. September 2024]. Verfügbar unter: ec.europa.eu

Georgiev, Kostadin/Inglese, Francesca/Lucarelli, Annalisa/Magistro, Renato Massacci, Giulio/Ruocco, Giuseppina/Stateva, Galya. *Experimental OJA based indicators on labour demand changes: opportunities and challenges*. 2024. Q2024 Conference, Estoril. Paper Session 12.3. [Zugriff am 12. September 2024]. Verfügbar unter: airdrive.eventsair.com

Kiss-Nagy, Anca Maria/Marconi, Gabriele/Paulino, Raquel/Bitoulas, Alexandors/Reis, Fernando. *Path to a quality framework for Online Job Advertisement data*. 2022. Q2022 Conference Vilnius. Paper Session 20.3. [Zugriff am 12. September 2024]. Verfügbar unter: q2022.stat.gov.lt

LITERATURVERZEICHNIS

- Kvetan, Vladimir/Branka, Jiri. *Increasing quality of web-based data: Human Role in production of Consistent Labour Market intelligence*. 2024. Q2024 Conference, Estoril. Paper Session 8.1. [Zugriff am 12. September 2024]. Verfügbar unter: airdrive.eventsair.com
- Nagy, Anca Maria/Gotuzzo, Eliane/Reis, Fernando. *Innovative Approaches to Enhance Data Quality in Official Statistics: A Case Study on Online Job Advertisement Data*. 2024. Q2024 Conference, Estoril. Paper Session 12.2. [Zugriff am 12. September 2024]. Verfügbar unter: airdrive.eventsair.com
- Maślankowski, Jacek/Nowak, Dominika. *Emerging Success Stories in the Production of Experimental and Official Statistics with Web Data*. 2022. Q2022 Conference, Vilnius. Paper Session 21.2. [Zugriff am 12. September 2024]. Verfügbar unter: q2022.stat.gov.lt
- Reis, Fernando/Faria Paulino, Raquel/ Kvetan, Vladimir. *From Web Content to Quality Data: Rules, Roles, and Reliability in the Web Intelligence Hub*. 2024. Q2024 Conference, Estoril. Presentation Session 8.3. [Zugriff am 12. September 2024]. Verfügbar unter: airdrive.eventsair.com
- Rengers, Martina. *Internetgestützte Erfassung offener Stellen – Machbarkeitsstudie im Rahmen eines ESSnet-Projekts zu Big Data*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2018, Seite 11 ff.
- Ricciato, Fabio/Wirthmann, Albrecht. *Trusted Smart Statistics: how new data will change official statistics*. 2019. Data4Policy conference, London. DOI: [10.5281/zenodo.3066061](https://doi.org/10.5281/zenodo.3066061)
- Ricciato, Fabio/Wirthmann, Albrecht/Hahn, Martina. *Trusted Smart Statistics: How new data will change official statistics*. 2020. In: Data & Policy. Band 2, Ausgabe e7. DOI: [10.1017/dap.2020.7](https://doi.org/10.1017/dap.2020.7)
- Ricciato, Fabio/Wirthmann, Albrecht/Giannakouris, Konstantinos/Reis, Fernando/Skaliotis, Michail. *Trusted smart statistics: Motivations and principles*. 2019. In: Statistical Journal of the IAOS. Band 35, Ausgabe 4, Seite 589 ff. DOI: [10.3233/SJI-190584](https://doi.org/10.3233/SJI-190584)
- United Nations. *United Nations National Quality Assurance Frameworks Manual for Official Statistics – Including recommendations, the framework and implementation guidance*. New York, 2019. [Zugriff am 12. September 2024]. Verfügbar unter: unstats.un.org

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im Oktober 2024
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-24005-4, ISSN 1619-2907

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