

# IW-Trends

## Gesuchte Datenkompetenzen in Deutschland

Jan Büchel / Jan Felix Engler / Armin Mertens

**IW-Trends 2/2023**

Vierteljahresschrift zur  
empirischen Wirtschaftsforschung  
Jahrgang 50



## Herausgeber

### Institut der deutschen Wirtschaft Köln e.V.

Postfach 10 19 42  
50459 Köln  
[www.iwkoeln.de](http://www.iwkoeln.de)

Das Institut der deutschen Wirtschaft (IW) ist ein privates Wirtschaftsforschungsinstitut, das sich für eine freiheitliche Wirtschafts- und Gesellschaftsordnung einsetzt. Unsere Aufgabe ist es, das Verständnis wirtschaftlicher und gesellschaftlicher Zusammenhänge zu verbessern.

## Das IW in den sozialen Medien

Twitter  
[@iw\\_koeln](https://twitter.com/iw_koeln)

LinkedIn  
[@Institut der deutschen Wirtschaft](https://www.linkedin.com/company/institut-der-deutschen-wirtschaft)

Facebook  
[@IWKoeln](https://www.facebook.com/IWKoeln)

Instagram  
[@IW\\_Koeln](https://www.instagram.com/IW_Koeln)

## Verantwortliche Redakteure

### Prof. Dr. Michael Grömling

Senior Economist  
[groemling@iwkoeln.de](mailto:groemling@iwkoeln.de)  
0221 4981-776

### Holger Schäfer

Senior Economist  
[schaefer.holger@iwkoeln.de](mailto:schaefer.holger@iwkoeln.de)  
030 27877-124

**Alle Studien finden Sie unter  
[www.iwkoeln.de](http://www.iwkoeln.de)**

Rechte für den Nachdruck oder die elektronische Verwertung erhalten Sie über [lizenzen@iwkoeln.de](mailto:lizenzen@iwkoeln.de).

In dieser Publikation wird aus Gründen der besseren Lesbarkeit regelmäßig das grammatische Geschlecht (Genus) verwendet. Damit sind hier ausdrücklich alle Geschlechteridentitäten gemeint.

ISSN 1864-810X (Onlineversion)

© 2023  
Institut der deutschen Wirtschaft Köln Medien GmbH  
Postfach 10 18 63, 50458 Köln  
Konrad-Adenauer-Ufer 21, 50668 Köln  
Telefon: 0221 4981-450  
[iwmedien@iwkoeln.de](mailto:iwmedien@iwkoeln.de)  
[iwmedien.de](http://iwmedien.de)

## **Gesuchte Datenkompetenzen in Deutschland – Ergebnisse einer Big-Data-Analyse**

**Jan Büchel / Jan Felix Engler / Armin Mertens, Mai 2023**

### **Zusammenfassung**

Beschäftigte mit Datenkompetenzen (Datenskills) helfen Unternehmen dabei, Daten effizienter speichern, managen und nutzen zu können. Ihre Bedarfe an Datenskills können Unternehmen decken, indem sie externe Dienstleister beauftragen, bestehende Beschäftigte weiterbilden oder neue Beschäftigte mit Datenskills einstellen. Eine Analyse von Stellenanzeigen in Deutschland aus den Jahren 2019 bis 2023 zeigt, dass Unternehmen in jeder fünften Stellenanzeige Datenskills von neuen Beschäftigten fordern. Das Ergebnis basiert auf einem Machine-Learning-Modell, das mehr als 18 Millionen Stellenanzeigen automatisiert klassifiziert. In 87 Prozent der ausgeschriebenen Datenstellen sollen neue Beschäftigte Unternehmen beim Nutzen ihrer Daten unterstützen, in 84 Prozent beim Speichern und in 67 Prozent beim Managen. Daten zu managen und zu nutzen, wird häufiger von Beschäftigten mit Studienabschluss gefordert, Daten zu speichern eher von Beschäftigten, die eine Berufsausbildung abgeschlossen haben. Dabei möchten Unternehmen Beschäftigte mit Datenskills in 59 Prozent der aktuellen Datenstellen als Allrounder für alle drei Einsatzbereiche einsetzen. Die hohe Schnittmenge zeigt, wie universell Datenskills eingesetzt werden können. Insgesamt sind die wachsenden Bedarfe an Datenskills der Unternehmen ein positives Signal für die Entwicklung der Datenökonomie in Deutschland. Angesichts des sich verschärfenden Fachkräftemangels in Digitalisierungsberufen kann es für Unternehmen jedoch herausfordernd sein, ihre Datenkompetenzbedarfe langfristig zu decken.

Stichwörter: Arbeitsmarkt, Daten, Datenwirtschaft, Machine Learning

JEL-Klassifikation: C45, C55, J24

DOI: 10.2373/1864-810X.23-02-01

## Warum Unternehmen Datenkompetenzen benötigen

Daten stellen für Unternehmen eine wertvolle Ressource dar. Sie können beispielsweise datenbasiert neue Produkte, Dienstleistungen oder Geschäftsmodelle entwickeln (Niebel et al., 2019) oder ihre Produktivität steigern (Brynjolfsson/McElheran, 2016). Umfragen zur Data Economy Readiness zeigen jedoch, dass im Jahr 2022 nur 31 Prozent der Industrieunternehmen und industrienahen Dienstleister in Deutschland bereit für die Datenökonomie waren, im Jahr 2021 waren es 29 Prozent (Büchel/Engels, 2022a). Das bedeutet, dass nur knapp ein Drittel der Unternehmen Daten umfangreich digital speichert, strukturiert verarbeitet und vielseitig nutzt (Büchel/Engels, 2022b). Es bleibt demnach offensichtlich viel Datenpotenzial ungenutzt. Damit sich dies in Zukunft ändern kann, sollten Unternehmen nicht nur die Potenziale erkennen, die Daten für ihr Geschäftsmodell haben können. Zudem benötigen sie Beschäftigte mit datenbezogenen Kompetenzen (Datenskills), denn es sind die Mitarbeiter mit diesen Fähigkeiten, die Datenprojekte in den Unternehmen durchführen.

Dabei haben Unternehmen mehrere Optionen, um Zugang zu Datenkompetenzen oder Datenskills zu erhalten. Sie können erstens bestehende Beschäftigte entsprechend weiterbilden oder zweitens externe Dienstleister beauftragen, um Datenprojekte im Unternehmen durchzuführen. Externe Dienstleister haben den Vorteil, dass Unternehmen kurzfristig auf spezifisches Know-how zugreifen und temporäre Bedarfe decken können. Im Vergleich zu Datenskills bei eigenen Beschäftigten können sie allerdings perspektivisch teuer sein, wenn Unternehmen langfristige Lösungen anvisieren. Das trifft zum Beispiel auf Digitalisierungsvorhaben zu, die meist langfristig angelegt sind und kontinuierlich optimiert und gepflegt werden müssen (KfW Research, 2023). Zudem können stetig neue Datenprojekte hinzukommen. Daher fungieren externe Dienstleister eher als Katalysator, um digitale Transformationsprozesse in den Unternehmen zu initiieren und weniger, um einen etablierten operativen Betrieb zu begleiten und schrittweise zu optimieren (Stettes, 2023). Als dritte Option können Unternehmen neue Beschäftigte einstellen, welche die geforderten Datenskills bereits besitzen und diese unmittelbar in das Unternehmen einbringen können.

Ob Unternehmen externe Dienstleister beauftragen oder bestehende Beschäftigte weiterbilden, ist in der Regel nur schwer beobachtbar. Derartige Aktivitäten laufen

intern ab und Unternehmen haben meist keinen Anreiz, entsprechende Informationen offenzulegen. Neueinstellungen dagegen haben den Vorteil, dass die ausgeschriebenen Stellenanzeigen meist öffentlich einsehbar sind. Sie eignen sich daher dafür, um Bedarfe an Datenskills der Unternehmen in Deutschland zu analysieren. Daten zu Online-Stellenanzeigen werden gerade in der Arbeitsmarktforschung immer häufiger verwendet (Acemoglu et al., 2022; Azar et al., 2020). Dabei umfassen derartige Datensätze typischerweise sehr viele Datenpunkte, gerade wenn sie mehrere Datenquellen wie unterschiedliche Stellenportale zusammenführen. Aufgrund ihres Umfangs ermöglichen sie sehr präzise und robuste Analysen. Beispielsweise kann untersucht werden, in welchen Regionen Deutschlands Unternehmen besonders hohe Bedarfe haben (Büchel/Mertens, 2022) oder welche Muster sich in Anforderungsprofilen abzeichnen (Büchel/Mertens, 2021). In dem vorliegenden Kontext werden Stellenanzeigen verwendet, um den Bedarf an Datenskills in Deutschland zu messen.

Die nachfolgende Analyse ist zweigeteilt: Zunächst wird mithilfe von Stellenanzeigen untersucht, wie häufig Unternehmen Datenskills bei neuen Beschäftigten fordern und wie sich die Bedarfe im Zeitverlauf entwickelt haben. Daraufhin wird für die Stellenanzeigen mit geforderten Datenskills (sogenannte Datenstellen) jeweils ermittelt, wofür neue Beschäftigte ihre Datenskills einsetzen sollen. Konkret wird untersucht, in welchen der drei Bereiche der Data Economy Readiness (Datenspeicherung, Datenmanagement und Datennutzung) die neuen Beschäftigten die ausschreibenden Unternehmen unterstützen sollen. Für beide Analyseschritte werden jeweils Machine-Learning-Modelle (ML-Modelle) trainiert. Das erste Modell identifiziert automatisch pro Stellenanzeige, ob Datenskills gefordert werden. Das zweite Modell ordnet eine Datenstelle basierend auf den geforderten Datenskills automatisch den drei genannten Bereichen der Data Economy Readiness zu.

### **In welchem Umfang Unternehmen Datenskills benötigen**

Als Basis der Analyse dient ein Datensatz des Anbieters Textkernel zu Online-Stellenanzeigen in Deutschland. Der Datensatz aggregiert Stellenanzeigen aus über 60.000 Quellen, wie den großen Stellenportalen Stepstone oder der Bundesagentur für Arbeit für die jeweils ersten Quartale der Jahre 2019 bis 2023 und bereinigt um Duplikate. Ob Datenskills gefordert werden oder nicht, ist besonders aus den Informationen

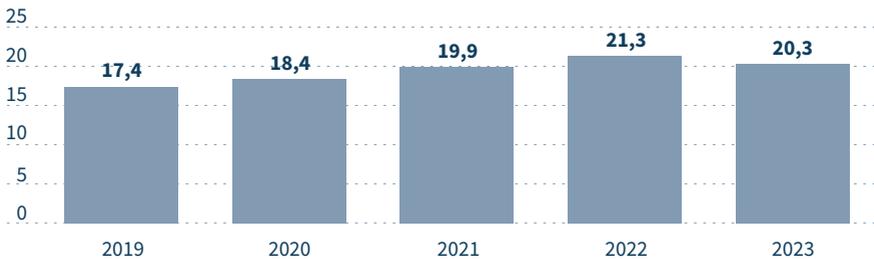
zum Aufgabenbereich der neuen Beschäftigten und zu den geforderten Kompetenzen erkennbar. Daher ist das Ziel der Analyse zunächst, jede einzelne Stellenanzeige danach zu klassifizieren, ob sie Datenskills fordert oder nicht. Da der Datensatz über alle Jahre hinweg insgesamt mehr als 18 Millionen Stellenanzeigen umfasst, wäre er manuell nur mit sehr hohem Aufwand zu verarbeiten. Aus diesem Grund wird ein ML-Modell trainiert, um den gesamten Datensatz mit vertretbarem Aufwand präzise analysieren zu können.

Aufbauend auf Büchel et al. (2023) wird ein ML-Modell verwendet, um Datenskills in deutschen Stellenanzeigen aus den jeweils ersten Quartalen der Jahre 2019 bis 2023 zu identifizieren. Das Modell beruht auf einer domänenspezifischen Transformer-Architektur – also einem Sprachmodell, das auf deutschsprachige Stellenanzeigen trainiert wurde (Gnehm et al., 2022). Der dahinterliegende Algorithmus kann von bekannten Trainingsdaten lernen, Muster erkennen und diese auf neue, bisher unbekannte Daten übertragen. Er muss dafür aber auf spezifische Anwendungsfälle angepasst werden. Dafür werden manuell 1.000 Stellenanzeigen dahingehend annotiert, ob darin Datenskills gefordert werden oder nicht. Dazu wurde die Annotationssoftware doccano verwendet (Nakayama et al., 2018). Um diese Trainingsdaten möglichst objektiv zu annotieren, wurde zunächst jeweils einzeln vorgegangen. Anschließend wurden die Konfliktfälle besprochen, um zu übereinstimmenden Annotationen zu gelangen. Im Zuge dessen wurde iterativ ein Regelwerk entwickelt, das inhaltlich auf den drei Bereichen der Data Economy Readiness (Büchel/Engels, 2022b) aufbaut. Von Beschäftigten werden beispielsweise Datenskills gefordert, wenn sie Kenntnisse zu Cloud-Technologien aufweisen oder Datenbanken pflegen sowie Software entwickeln oder programmieren sollen. Tätigkeiten, bei denen unklar war, ob sie digital oder analog durchgeführt werden, zählten hingegen nicht zu den Datenskills. Beschäftigte, die Kundeninformationen ablegen, können dies auch vollständig analog durchführen. Auch eher grundlegende Kompetenzen wie allgemeine Microsoft-Office-Skills waren nicht ausschlaggebend, um Datenskills in den Stellenanzeigen zu identifizieren. Die 1.000 annotierten Stellenanzeigen dienten dann als Trainingsdaten für das ML-Modell, das darauf basierend die übrigen Stellenanzeigen automatisiert bewertete.

## Datenstellen in Deutschland

Abbildung 1

Anteil der Datenstellen<sup>1)</sup> in Prozent aller Online-Stellenanzeigen im jeweiligen ersten Quartal



1) Datenstellen: Online-Stellenanzeigen mit explizit geforderten Datenkompetenzen (Datenskills).

Quelle: Institut der deutschen Wirtschaft

Abbildung 1: <http://dl.iwkoeln.de/index.php/s/b4FbM78GfYWiTda>

Zunächst wurde analysiert, wie hoch der Anteil der Datenstellen an allen Stellenanzeigen in Deutschland in den jeweils ersten Quartalen der Jahre 2019 bis 2023 war (Abbildung 1). Demnach forderten im ersten Quartal 2023 die Unternehmen in 20,3 Prozent der etwa 4,2 Millionen ausgeschriebenen Stellenanzeigen Datenskills von neuen Beschäftigten. Im Vergleich zum Jahr 2022 war der Anteil an Datenstellen um 1 Prozentpunkt niedriger. Das erste Quartal des Jahres 2022 war nicht nur generell sehr ausschreibungsstark mit insgesamt 4,5 Millionen Stellenanzeigen, sondern es wurden auch relativ betrachtet häufiger Datenskills gefordert. Davor hatte sich in den Jahren 2019 bis 2021 ein kontinuierlicher Wachstumstrend abgezeichnet. Dabei stieg nicht nur der Anteil der Datenstellen, sondern in einzelnen Jobs wurden im Zeitverlauf immer mehr Datenstellen ausgeschrieben. Zum Beispiel wurden im ersten Quartal des laufenden Jahres 28.566 Datenstellen mit der von Textkernel standardisierten Jobbezeichnung SAP-Berater ausgeschrieben. Das ist ein Anstieg um 77 Prozent gegenüber dem Jahr 2019. Seit 2022 werden zudem viele Datenstellen für Controller veröffentlicht, die eher Aufgaben aus den Bereichen Datenmanagement und Datennutzung betreffen. Gleichzeitig werden aber auch seit 2019 konstant viele Datenstellen für kaufmännische Angestellte, Personalsachbearbeiter oder Fachkräfte Lagerlogistik ausgeschrieben, die in hohem Umfang am Datenspeichern beteiligt sind.

Wenngleich in etwa jeder fünften Stellenanzeige in Deutschland Datenskills gefordert werden, kann daraus nicht abgeleitet werden, dass auch ein Fünftel der derzeitigen Beschäftigten in Deutschland Datenskills aufweisen. Stellenanzeigen zeigen ausschließlich die Nachfrageseite des Arbeitsmarkts und sind ein Indikator für den Beschäftigungszuwachs. Somit besagt die Analyse erst einmal nur, dass Unternehmen im Durchschnitt von einem Fünftel der neuen Beschäftigten Datenkompetenzen fordern. Ist die Fluktuation unter diesen Beschäftigten hoch, kann es für Unternehmen schwer sein, die ausgeschriebenen Stellen langfristig zu besetzen. Diese Notlage der Unternehmen kann zudem der vorliegende Fachkräftemangel in Digitalisierungsberufen in Deutschland verstärken (Burstedde, 2023, 13 f.). Insgesamt können demnach aus den Bedarfen an Datenskills nur bedingt Rückschlüsse über den Bestand an Beschäftigten mit Datenskills in den Unternehmen gezogen werden. Dafür wären zusätzliche Informationen zur Fluktuation und zu weiteren Charakteristika der Belegschaft notwendig. Allerdings können Stellenanzeigen zeigen, wofür Unternehmen neue Beschäftigte mit Datenskills primär einsetzen möchten und demnach, in welchen Bereichen die Bedarfe besonders hoch sind.

### **Für welche Einsatzbereiche Unternehmen Datenskills benötigen**

Neue Beschäftigte können ihre Datenskills in verschiedenen Bereichen einsetzen. Sie können Unternehmen darin unterstützen, Daten effizienter zu speichern, zu managen und/oder zu nutzen. Da die drei Bereiche der Data Economy Readiness die inhaltliche Grundlage bilden, um Datenskills zu identifizieren, können die Datenstellen auch entsprechend zugeordnet werden. Die Annahme ist dabei, dass Datenstellen einem der drei Einsatzbereiche zugewiesen werden, wenn mindestens ein identifizierter Datenskill aus dem jeweiligen Bereich stammt. Allerdings schließen die drei Bereiche sich nicht gegenseitig aus. Datenskills können mehreren Bereichen zugeordnet werden. Beispiele hierfür sind neue Beschäftigte, die Key Performance Indicators (KPIs) für ein ausschreibendes Unternehmen identifizieren, notwendige Datenquellen erschließen und darauf basierend Datenanalysen durchführen sollen. Das Tätigkeitsprofil betrifft Elemente des Datenmanagements und der Datennutzung. Die möglichen Überschneidungen erhöhen jedoch die Komplexität. Datenstellen korrekt den drei Bereichen zuzuordnen, ist daher komplexer, als Datenskills in Stellenanzeigen zu identifizieren.

Denn im ersten Analyseschritt gab es nur zwei Optionen, die sich gegenseitig ausschließen – entweder forderten Stellenanzeigen Datenskills oder nicht.

Daher wird ein zweites ML-Modell entwickelt, um Datenstellen automatisch ihrem Einsatzbereich zuzuordnen. Grundlegend folgt das zweite Modell der Methodik des ersten Modells: 1.000 manuelle Annotationen dienen als Trainingsdaten, um alle übrigen Datenstellen automatisch ihrem Einsatzbereich zuzuordnen. In einem iterativen Prozess wurden jeweils auftretende Konfliktfälle besprochen und darauf basierend das Regelwerk der Annotationen verfeinert. Das ML-Modell wurde dabei nicht gänzlich neu entwickelt. Als Basis diente analog zu dem ersten ML-Modell das bereits vortrainierte Modell von Gnehm et al. (2022). Es ist besonders auf deutsche Stellenanzeigen und die deutsche Sprache spezialisiert. Somit lernt das ML-Modell neben den bereits vortrainierten Eigenschaften zur Sprache und zu den Besonderheiten deutscher Stellenanzeigen, zusätzlich aus den 1.000 Trainingsdaten, Datenstellen korrekt ihren Einsatzbereichen zuzuordnen (Devlin et al., 2019).

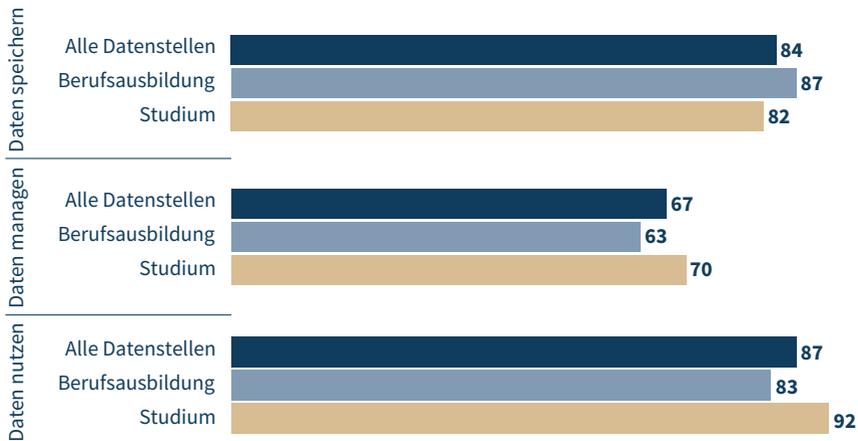
Bei etwa zwei Drittel der insgesamt 855.975 Datenstellen im ersten Quartal des Jahres 2023 gibt Textkernel zusätzlich an, ob Unternehmen ein Studium (36 Prozent) oder eine Berufsausbildung (32 Prozent) als Abschluss fordern. Das relative Verhältnis der Abschlüsse ist im Gegensatz zu eher spezialisierten Stellenanzeigen aus dem KI-Bereich ausgeglichen. Darin machen Studienabschlüsse etwa 80 Prozent der geforderten Abschlüsse aus (Büchel/Mertens, 2021). Nachfolgend wird nicht nur analysiert, ob neue Beschäftigte eher Daten speichern, managen oder nutzen sollen, sondern auch, welche Abschlüsse dabei in den einzelnen Bereichen gefordert werden (Abbildung 2).

Alle drei Bereiche der Data Economy Readiness sind in den Datenstellen hochrelevant. In 87 Prozent aller Datenstellen sollen neue Beschäftigte die suchenden Unternehmen darin unterstützen, Daten besser zu nutzen, in 84 Prozent Daten zu speichern und in 67 Prozent Daten zu managen. Beschäftigte, die eine Berufsausbildung abgeschlossen haben, sollen leicht häufiger Daten speichern und seltener managen und nutzen. Umgekehrt sollen Beschäftigte, die ein Studium abgeschlossen haben, häufiger Daten managen und nutzen, aber seltener Daten speichern. Die Unterschiede in den Abschlussformen sind jeweils signifikant. Es entsteht der Eindruck, als würden Manage-

## Datenstellen nach Einsatzbereichen und geforderten Bildungsabschlüssen

Abbildung 2

Anteil der Datenstellen, die den jeweiligen Einsatzbereichen zuzuordnen sind, und nach Art des geforderten Bildungsabschlusses in Prozent aller Datenstellen im ersten Quartal 2023



Grundgesamtheit: 855.975 Datenstellen; Mehrfachzuordnungen zu den drei Einsatzbereichen sind möglich.  
Quelle: Institut der deutschen Wirtschaft

Abbildung 2: <http://dl.iwkoeln.de/index.php/s/JiJHTTWcA3g6B9>

ment- und Nutzungsaktivitäten vermehrt Kenntnisse voraussetzen, die in einem Studium erlernt werden. In diesen Bereichen werden oft sehr technische Skills wie das Einrichten von Datenschnittstellen (APIs) oder Managementskills verlangt, um etwa die Datenstrategie des Unternehmens zu entwickeln. Dagegen kann das Datenspeichern Aufgaben umfassen, die vorwiegend in einer Berufsausbildung erlernt werden. Beispielsweise können Beschäftigte, die in der Produktion tätig sind und Maschinen bedienen, besser einschätzen, wie Produktionsdaten über Sensoren digital gespeichert und Störungen im Datenfluss behoben werden können. Dennoch sind die Unterschiede in den Abschlussformen hinsichtlich der drei Einsatzbereiche eher gering. Beide Abschlussformen werden häufig in Kombination mit allen drei Bereichen gefordert. So können Beschäftigte in ihrer Berufsausbildung zum Beispiel auch lernen, mit Datenverarbeitungssoftware umzugehen und damit Daten zu speichern, zu managen und zu nutzen.

Die Aufteilung auf die drei Bereiche der Data Economy Readiness verändert sich seit 2019 nur minimal. Seitdem sollen insgesamt nur geringfügig mehr Beschäftigte Daten managen (plus 1 Prozentpunkt), während leicht weniger Beschäftigte Daten speichern sollen (minus 3 Prozentpunkte). Die Konstanz der Einsatzbereiche ist beachtlich, da in jedem Jahr komplett neue Stellenanzeigen analysiert werden. Insgesamt haben Unternehmen demnach höhere Kompetenzbedarfe beim Speichern und Nutzen von Daten als beim Managen. Dies muss allerdings nicht bedeuten, dass Unternehmen dem Datenmanagement eine zu geringe Bedeutung zuschreiben. Ein genauerer Blick in die Umfrageergebnisse der Data Economy Readiness zeigt sogar das Gegenteil: Unternehmen geben an, dass sie im Durchschnitt etwa die Hälfte der relevanten Aspekte einer effizienten Datenspeicherung und eines effizienten Datenmanagements erfüllen, bei den Aspekten einer effizienten Datennutzung sind es jedoch nur 14 Prozent (Büchel/Engels, 2022c, 8 f.). Stattdessen ist es vorstellbar, dass in den Unternehmen generell weniger Datenmanagementaktivitäten anfallen als in den Bereichen Speichern und Nutzen, wonach auch die Kompetenzbedarfe geringer sind. Unternehmen können beispielsweise nur vereinzelt Beschäftigte suchen, die im Unternehmen ein sogenanntes Data Warehouse einrichten, in dem Daten aus verschiedenen Quellen zentral gebündelt und für verschiedene Zwecke weitergegeben werden. Ein gut funktionierendes Data Warehouse besitzt einen hohen Automatisierungsgrad. Im Vergleich dazu kann zum Beispiel die Datenbankpflege oder die Analyse verschiedener Datensätze nach unterschiedlichen Gesichtspunkten einen höheren manuellen Bedarf für die Unternehmen ausmachen. Daraus können höhere Kompetenzbedarfe beim Speichern und Nutzen der Daten entstehen.

### **Welche Kombinationen von Datenskills gesucht werden**

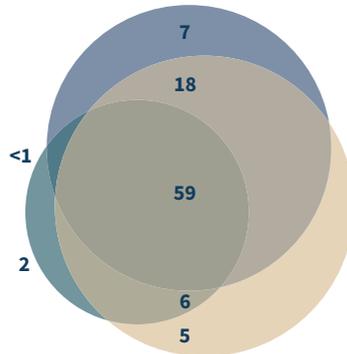
Vorstellbar ist darüber hinaus, dass sich die drei Bereiche in hohem Umfang überschneiden. Gerade die hohen Anteile in allen drei Bereichen deuten darauf hin. Neue Beschäftigte könnten daher auch indirekt am Datenmanagement beteiligt sein, ohne dass die Unternehmen dies explizit in die Datenstellen schreiben. Eine Untersuchung der Schnittmengen zeigt, wie eng die einzelnen Bereiche inhaltlich miteinander verknüpft sind (Abbildung 3).

## Schnittmenge der Einsatzbereiche

Abbildung 3

Anteil der Datenstellen, die den jeweiligen Einsatzbereichen und den Schnittmengen aus mehreren Einsatzbereichen zuzuordnen sind<sup>1)</sup>, im ersten Quartal 2023 in Prozent aller Datenstellen

● Datenspeicherung ● Datenmanagement ● Datennutzung



1) 4 Prozent der Datenstellen können keinem Einsatzbereich zugeordnet werden.  
Quelle: Institut der deutschen Wirtschaft

Abbildung 3: <http://dl.iwkoeln.de/index.php/s/R4zbPCKMwo2nz5p>

Mit 59 Prozent der Datenstellen sollen neue Beschäftigte ihre Datenkompetenzen am häufigsten in allen drei Bereichen (Speichern, Managen und Nutzen von Unternehmensdaten) einsetzen. Beispiele hierfür sind Beschäftigte, von denen Kenntnisse zu Cloud-Technologien gefordert werden. In der Cloud können Unternehmen Daten zentral speichern, unternehmensübergreifend teilen sowie vereinfacht visualisieren oder analysieren. In 18 Prozent der Stellenanzeigen sollen neue Beschäftigte Unternehmensdaten speichern und nutzen. Das betrifft etwa Beschäftigte, die zunächst passende Daten aus verschiedenen Quellen suchen, speichern und daraufhin analysieren. In 7 Prozent der Stellenanzeigen sollen neue Beschäftigte Daten lediglich speichern, etwa indem sie Datenbanken aktualisieren. Nur Daten zu speichern, wird dabei in Datenstellen, die eine Berufsausbildung fordern, mit 12 Prozent deutlich häufiger gefordert als in Datenstellen, die ein Studium voraussetzen (2 Prozent). Anders sieht es bei Datenstellen aus, die das gleichzeitige Managen und Nutzen von Daten betreffen. Insgesamt 6 Prozent aller Datenstellen fordern diesen Mix, 9 Prozent der Datenstellen mit Studium und nur 3 Prozent der Datenstellen mit einer Berufsausbildung. Beispiele hierfür sind neue Beschäftigte, die KPIs (weiter-)entwickeln und

darauf basierend strategische Unternehmensentscheidungen ableiten sollen. Diese Erkenntnis ist konsistent zu der insgesamt höheren Bedeutung des Datenspeicherns bei Beschäftigten, von denen eine Berufsausbildung gefordert wird, sowie zur höheren Bedeutung des Managens und Nutzens von Daten bei Beschäftigten, von denen ein Studium verlangt wird (s. Abbildung 2).

In 5 Prozent der Stellenanzeigen sollen neue Beschäftigte Daten nur nutzen, beispielsweise um Prozessänderungen zu simulieren und daraus Prognosen abzuleiten. Unternehmen suchen selten Beschäftigte, die Daten lediglich managen sollen (2 Prozent). Beispiele hierfür sind Data Engineers, deren primäre Aufgabe es ist, kontinuierliche Datenströme (Datenpipelines) zwischen der Erhebung und Nutzung der Daten bereitzustellen. Die hohe Schnittmenge des Datenmanagements mit anderen Bereichen zeigt, dass Beschäftigte nur in seltenen Fällen ausschließlich Daten managen sollen. Oftmals sind sie auch beim Speichern und Nutzen der Daten involviert. Selten sollen neue Beschäftigte Daten speichern und managen, nicht aber nutzen. Das ist beispielsweise der Fall, wenn neue Beschäftigte Datenbanken nicht nur pflegen, sondern ebenfalls neu entwickeln sollen. Da derartige Beschäftigte sehr vertiefte Kenntnisse der Datenbanken aufweisen, kann es für Unternehmen sinnvoll sein, wenn sie die Datenbanken auch direkt nutzen. Andere Beschäftigte müssten sich in die Datenbankstruktur erst zeitintensiv einarbeiten. In diesen Fällen überschneiden sich daher meist alle drei Einsatzbereiche.

Seit 2019 ändern sich ebenfalls die Anteile der Schnittmengen nur minimal. Einzig die Schnittmenge zwischen Daten speichern und nutzen nahm seit 2019 leicht von 21 auf 18 Prozent ab. Dafür sollen mehr Beschäftigte gleichzeitig Daten managen und nutzen (plus 1 Prozentpunkt). Das kann in der bereits erwähnten insgesamt wachsenden Bedeutung des Datenmanagements begründet sein. Nicht zuletzt zeigt die Bedeutung der Schnittmengen, dass Beschäftigte oftmals in mehreren Einsatzbereichen aktiv sein sollen, zum Beispiel, um neue Daten zu erheben und gleichzeitig zu nutzen. Dass jedoch mehr als die Hälfte der Datenstellen alle drei Bereiche betrifft, verdeutlicht, wie eng die Einsatzfelder miteinander verbunden sind und dass Unternehmen Datenskills bereichsübergreifend einsetzen möchten. Die gleichbleibende Struktur der Schnittmengen im Zeitverlauf kann zudem ein Indiz dafür sein, dass

Unternehmen eine für sie optimale Zusammensetzung der drei Einsatzbereiche für Datenskills gefunden haben.

### **Was die Daten über Datenskills erkennen lassen**

Unternehmen in Deutschland suchen immer mehr Beschäftigte mit Datenskills. Dies ist ein positives Signal für die Datenökonomie in Deutschland. Werden entsprechende Datenstellen langfristig besetzt, können Unternehmen die Potenziale ihrer Daten besser nutzen. Dabei möchten Unternehmen Beschäftigte mit Datenskills meist als Allrounder einsetzen, um Daten sowohl besser speichern, managen und nutzen zu können. Die hohe Schnittmenge zeigt, wie universell Datenskills eingesetzt werden können. Generell gewinnt Datenmanagement zunehmend an Bedeutung. Das kann bedeuten, dass Unternehmen vermehrt begreifen, welche Vorteile eine strukturierte Datenverarbeitung haben kann. Daten effizient zu managen, kann sich zudem positiv auf das Speichern und Nutzen von Daten auswirken, wozu Unternehmen in der Vergangenheit schon viele Beschäftigte suchten. Angesichts des sich verschärfenden Fachkräftemangels in Digitalisierungsberufen (Burstedde, 2023, 13 f.) bleibt abzuwarten, in welchem Umfang die Datenkompetenzbedarfe der Unternehmen in Zukunft gedeckt werden können. Zielführend für Unternehmen kann daher sein, parallel zum Ausschreiben von Datenstellen auch Datenskills bei bestehenden Beschäftigten über entsprechende Weiterbildungen aufzubauen oder kurzfristige Bedarfe über externe Dienstleister zu decken.

*Diese Studie ist Teil des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekts „Incentives and Economics of Data Sharing – IEDS“.*

## Literatur

Acemoglu, Daron / Autor, David / Hazell, Jonathon / Restrepo, Pascual, 2022, Artificial Intelligence and Jobs: Evidence from Online Vacancies, in: Journal of Labor Economics, 40. Jg., Nr. S1, S. 293–340

Azar, José / Marinescu, Ioana / Steinbaum, Marshall / Taska, Bledi, 2020, Concentration in US labor markets: Evidence from online vacancy data, in: Labour Economics, 66. Jg., Nr. 101866

Brynjolfsson, Erik / McElheran, Kristina, 2016, Data in Action: Data-Driven Decision Making in U.S. Manufacturing, US Census Bureau Center for Economic Studies Paper, Nr. CES-WP-16-06, Rotman School of Management, Working Paper, Nr. 2722502, Zürich

Büchel, Jan / Engels, Barbara, 2022a, Viele Unternehmen sind nicht bereit für die Datenwirtschaft, IW-Kurzbericht, Nr. 96, Köln

Büchel, Jan / Engels, Barbara, 2022b, Datenbewirtschaftung von Unternehmen in Deutschland: Eine empirische Bestandsaufnahme, in: IW-Trends, 49. Jg., Nr. 1, S. 73–90

Büchel, Jan / Engels, Barbara, 2022c, Branchentrends beim Data Sharing. Status quo und Use Cases in Deutschland, IW-Report, Nr. 53, Köln

Büchel, Jan / Mertens, Armin, 2021, KI-Bedarfe der Wirtschaft am Standort Deutschland. Eine Analyse von Stellenanzeigen für KI-Berufe, Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi), Köln / Berlin

Büchel, Jan / Mertens, Armin, 2022, KI-Bedarfe in Deutschland. Regionale Analyse und Entwicklung der Anforderungsprofile in KI-Stellenanzeigen, Gutachten im Projekt „Entwicklung und Messung der Digitalisierung der Wirtschaft am Standort Deutschland“ im Auftrag des Bundesministeriums für Wirtschaft und Klimaschutz (BMWK), Berlin / Köln

Büchel, Jan / Engler, Jan / Mertens, Armin, 2023, The Demand for Data Skills in German Companies: Evidence from Online Job Advertisements, in: EconPol Forum, 24. Jg., Nr. 2, S. 56–61

Burstedde, Alexander, 2023, Kompetenzbarometer: Fachkräftesituation in Digitalisierungsberufen – Beschäftigungsaufbau und Fachkräftemangel bis 2026, Gutachten im Rahmen des Projekts „Entwicklung und Messung der Digitalisierung der Wirtschaft am Standort Deutschland“ im Auftrag des Bundesministeriums für Wirtschaft und Klimaschutz (BMWK), Berlin / Köln

Devlin, Jacob / Chang, Ming-Wei / Lee, Kenton / Toutanova, Kristina, 2019, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, S. 4171–4186

Gnehm, Ann-Sophie / Bühlmann, Eva / Clematide, Simon, 2022, Evaluation of transfer learning and domain adaptation for analyzing german-speaking job advertisements, in: Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), Marseille, S. 3892–3901

KfW Research, 2023, KfW-Digitalisierungsbericht Mittelstand 2022, Deutlicher Digitalisierungsschub im zweiten Jahr der Corona Pandemie, Kluft zwischen Vorreitern und Nachzüglern droht sich zu vertiefen, Frankfurt am Main

Nakayama, Hiroki et al., 2018, doccano: Text Annotation Tool for Human, <https://github.com/doccano/doccano> [28.3.2023]

Niebel, Thomas / Rasel, Fabienne / Viète, Steffen, 2019, BIG data – BIG gains? Understanding the link between big data analytics and innovation, in: Economics of Innovation and New Technology, 28. Jg., Nr. 3, S. 296–316

Stettes, Oliver, 2023, Solo-Selbstständige und Werkvertragsbeschäftigte als Katalysatoren des digitalen Wandels in deutschen Unternehmen, in: IW-Trends, 50. Jg., Nr. 1, S. 63–80

## The Demand for Data Skills in Germany – Results of a Big Data Analysis

Employees with data skills help companies to store, manage and use data more efficiently. Firms can meet their need for data skills by engaging external service providers, offering training on the job for existing employees or hiring new employees who already have data skills. An analysis of job advertisements in Germany from 2019 to 2023 shows that companies require data skills from their new employees in every fifth job advertisement. The result is based on a machine learning model that automatically classifies more than 18 million job advertisements. In 87 per cent of the job advertisements with data skills new employees should help their employers to use data, in 84 per cent to store data and in 67 per cent to manage data. Managing and using data is more often required of college graduates, while new employees who have completed vocational training are more likely to be required to store data. However, in 59 per cent of current job advertisements with data skills companies want new employees to function as all-rounders working in all three areas. The high overlap shows how universally data skills can be applied. Overall, the growing demand for data skills among companies suggests that the data economy in Germany is developing positively. However, given the growing shortage of skilled workers in digitalisation professions, it can be challenging for companies to meet their long-term data skills needs.