

Entwicklungen in der amtlichen Statistik

Der FDZ-Standort im AfS – Angebot und Nachfrage

von **Ramona Voshage, Daniela Janke** und **Anja Malchin**

Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (FDZ) – als Serviceeinrichtung für die Wissenschaft – sind eine Arbeitsgemeinschaft der Statistischen Ämter des Bundes und der Länder in Deutschland. In jedem Statistischen Amt befindet sich ein regionaler Standort des FDZ. Im Amt für Statistik Berlin-Brandenburg (AfS) gibt es seit dem Jahr 2003 einen regionalen FDZ-Standort für die beiden Länder Berlin und Brandenburg. Für qualitativ hochwertige, unabhängige Forschung und Politikberatung bietet das FDZ eine Infrastruktur, die einen sicheren, schnellen und einfachen Zugang zu einem breiten Angebot amtlicher Mikrodaten gewährleistet. Damit bildet es ein entscheidendes Bindeglied zwischen amtlicher Statistik und unabhängiger Wissenschaft.

Das FDZ bietet Mikrodaten aus allen Bereichen der amtlichen Statistik (Haushalte, Bevölkerung, Bildung, Gesundheit und Soziales, Wirtschaft, Finanzen und Steuern, Rechtspflege, Landwirtschaft, Energie und Umwelt) zur wissenschaftlichen Nutzung an.¹ Damit ist es ein unverzichtbarer Bestandteil der wissenschaftlichen Forschung bei gleichzeitiger Umsetzung der gesetzlich festgeschriebenen Geheimhaltungspflicht. Hochschulen oder sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung können gemäß § 16 Abs. 6 BStatG² ein Forschungsprojekt im FDZ beantragen.

Es stehen verschiedene Zugangswege zur Auswahl: Faktisch anonymisierte Daten können als Scientific-Use-Files in den Räumen der beantragenden wissenschaftlichen Institution ausgewertet werden, während Mikrodaten mit höherem Informationspotenzial (formal anonymisiert) innerhalb der geschützten Räume der amtlichen Statistik verbleiben. Die Auswertung und Erstellung publikationsreifer Ergebnisse der formal anonymisierten Mikrodaten erfolgt entweder über die kontrollierte Datenfernverarbeitung (KDFV) oder an einem Gastwissenschaftlerarbeitsplatz (GWAP).

Zu den Aufgaben des FDZ gehören neben der Datenaufbereitung und -bereitstellung auch die fachliche, teilweise englischsprachige Beratung und Betreuung der Wissenschaftlerinnen und Wissen-

schaftler sowie insbesondere die Geheimhaltungsprüfung der produzierten Ergebnisse während der gesamten Projektlaufzeit.

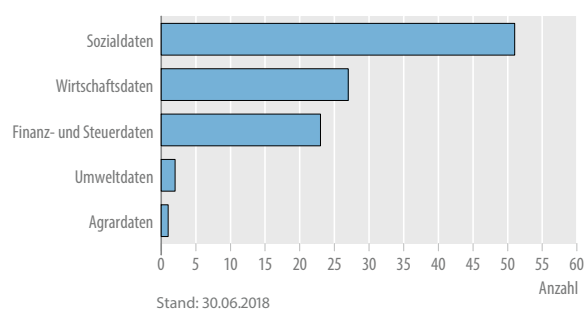
Aufgrund der Vielzahl an Hochschulen und wissenschaftlichen Instituten in der Region Berlin-Brandenburg werden sehr viele und teilweise sehr komplexe Forschungsprojekte durch den FDZ-Standort im AfS betreut. In den letzten Jahren hat sich die starke Nachfrage am Standort manifestiert.

Zahl der Projekte und regionale Zuordnung der Nutzenden

Im AfS werden aktuell 104 aktive FDZ-Projekte betreut, in denen die Wissenschaft hauptsächlich Sozial- und Wirtschaftsdaten nachfragt (Abbildung a). In etwa 75 % der Projekte werden entsprechende Mikrodaten des Mikrozensus, der Todesursachenstatistik oder der Amtlichen Firmendaten für Deutschland (AFiD)³ genutzt.

Wie auch in den vergangenen Jahren werden die Leistungen des FDZ-Standorts im AfS hauptsächlich von der regionalen Wissenschaft in Anspruch genommen, wenngleich deren Anteil etwas gesunken ist: Während 2015 noch 84 % der aktiv Nutzenden einer der lokalen Hochschulen oder einem regionalen wissenschaftlichen Institut angehörten, betrug deren Anteil im ersten Halbjahr 2018 noch etwa 67%.⁴ Knapp über ein Viertel der Projekte wurde von Hochschulen und Forschungseinrichtungen

a | Aktive Projekte am FDZ-Standort im AfS nach Art der verwendeten Statistiken



1 Nähere Informationen zu den FDZ der Statistischen Ämter des Bundes und der Länder unter: www.forschungsdatenzentrum.de

2 Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I S. 2394), das zuletzt durch Artikel 10 Absatz 5 des Gesetzes vom 30. Oktober 2017 (BGBl. I S. 3618) geändert worden ist.

3 Für die AFiD-Produkte wurden Unternehmens- und Betriebsdaten aus verschiedenen Erhebungen über die Unternehmensnummer verknüpft. Sie werden als Panels mit ergänzenden Modulen zur Nutzung angeboten.

4 Für die Entwicklung in früheren Jahren vgl. Zeitschrift für amtliche Statistik Berlin Brandenburg (2015): Datensätze werden gehoben – Die Entwicklung der Nachfrage am regionalen FDZ-Standort Berlin, Ausgabe 4/2015, S. 32 ff.

beantragt, die nicht aus der Region Berlin-Brandenburg stammten (2015: 16 %). Seit 2016 sind auch Datennutzerinnen und -nutzer von Institutionen aus dem Ausland aktiv: Zur Jahresmitte 2018 wurden in Berlin sechs Projekte aus England, Österreich, Italien und den USA betreut (Abbildung b).

Nutzung von Mikrodaten über die kontrollierte Datenfernverarbeitung

Eine Möglichkeit für die Auswertung formal anonymisierter Mikrodaten ist die Nutzung der kontrollierten Datenfernverarbeitung. Dabei verbleiben die Mikrodaten in den geschützten Räumen der amtlichen Statistik. Den Nutzenden wird ein projektspezifisches Datenstrukturfile zur Verfügung gestellt, in dem die enthaltenen Merkmale so verändert sind, dass einerseits keine inhaltlichen Bezüge der Merkmalsträger zu den Originaldaten mehr bestehen, andererseits aber die technische Struktur der Daten erhalten bleibt. Mittels dieser absolut anonymen Datenstrukturfiles werden Auswertungsprogramme von den Wissenschaftlerinnen und Wissenschaftlern erstellt, die von Mitarbeiterinnen und Mitarbeitern des FDZ auf die Originaldaten angewendet werden. Die Nutzenden selber haben keinen Zugriff auf die formal anonymen Einzeldaten, sondern erhalten nach der notwendigen im FDZ durchgeführten Geheimhaltungsprüfung die Ergebnisse ihrer Auswertung per E-Mail.

Die Anzahl eingegangener KDFV-Auswertungsanfragen ist am FDZ-Standort im AfS seit 2014 stark gestiegen: Sie hat sich von 364 Anfragen auf 979 Anfragen im Jahr 2017 nahezu verdreifacht. Im Jahr 2017 ging fast vier Mal pro Arbeitstag eine Auswertungsanfrage ein, die jeweils protokolliert, gestartet und überwacht wurde und deren Ergebnisse entweder zur Ansicht am GWAP bereitgestellt oder auf statistische Geheimhaltung geprüft und der Wissenschaft freigegeben wurden (2014: 1,44 KDFV-Anfragen pro Arbeitstag). Dieser Wert

wurde im ersten Halbjahr 2018 mit durchschnittlich fünf Anfragen pro Arbeitstag noch übertroffen; die KDFV wurde in diesem Zeitraum bereits 662 Mal in Anspruch genommen (Abbildung c).

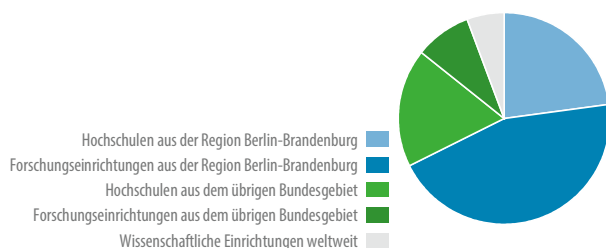
Im FDZ werden KDFV-Projekte üblicherweise durch die regionalen Standorte betreut, die für die genutzten Statistiken fachlich zuständig sind. In der Region Berlin-Brandenburg sind allerdings viele Hochschulen und wissenschaftliche Institutionen ansässig, von denen oftmals eine kombinierte Nutzung von KDFV und GWAP in Berlin beantragt wird. Um die Betreuung der Wissenschaft „aus einer Hand“ zu gewährleisten, werden vom FDZ-Standort im AfS deshalb auch Projekte übernommen, in denen Statistiken außerhalb der eigenen fachlichen Zuständigkeit⁵ ausgewertet werden. Dennoch ist deutlich erkennbar, dass insbesondere die Forschungseinrichtungen aus der Region Berlin-Brandenburg die Hauptnutzenden der KDFV sind.

Nutzung von Mikrodaten am GWAP

Der FDZ-Standort im AfS verfügt über fünf Arbeitsplätze für Gastwissenschaftlerinnen und -wissenschaftler im Gebäude des DIW Berlin (Deutsches Institut für Wirtschaftsforschung e. V.). Dort können faktisch⁶ und formal anonymisierte Einzeldaten in den geschützten Räumen der amtlichen Statistik analysiert werden. Die erforderliche Anonymität der Mikrodaten wird durch eine strikte Regulierung des Datenzugangs erreicht. Die Nutzenden müssen unter anderem zur Geheimhaltung verpflichtet sein (§ 16 Abs. 7 BStatG). Weiterhin verfügen die PC-Arbeitsplätze über keinerlei Kommunikationsmöglichkeiten (beispielsweise einen Zugang zum Internet), über die externes Zusatzwissen generiert werden könnte.

Die Zahl der GWAP-Nutzungen ist seit 2014 deutlich gestiegen. Waren 2014 noch 374 Gastwissenschaftlerinnen und -wissenschaftler vor Ort, stieg diese Zahl im Jahr 2017 bereits auf 749. Zu den

b | Aufteilung der aktiven Projekte am FDZ-Standort im AfS nach regionaler Zuordnung



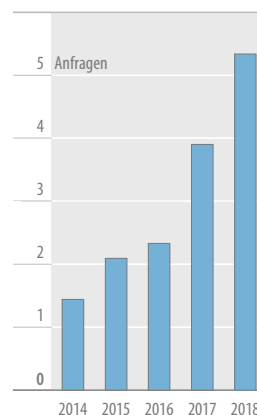
Stand: 30.06.2018

⁵ Der FDZ-Standort im AfS ist für folgende Statistiken fachlich zuständig: IKT Unternehmen, IKT Haushalte, Asylbewerberleistungen, Öffentliche Finanzen und Strukturerhebung im

Dienstleistungsbereich. Weiterhin obliegt dem Standort die Federführung im Projekt „Amtliche Firmendaten für Deutschland“ (AFID).

⁶ Faktisch anonymisierte Mikrodaten sind Einzelangaben, welche den Auskunftgebenden bzw. Merkmalsträgern nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können (vgl. § 16 Abs. 6 S. 1 BStatG).

c | KDFV-Auswertungsanfragen 2014 bis 2018 am FDZ-Standort im AfS pro Arbeitstag



Stand: 30.06.2018

Öffnungszeiten (regulär von Montag bis Donnerstag) kamen im Durchschnitt zwischen drei und vier Datennutzerinnen und -nutzer pro Arbeitstag an die GWAP (Abbildung d). Wie bei den KDFV-Anfragen ist auch hier im ersten Halbjahr 2018 eine Zunahme auf fast fünf GWAP-Besuche pro Arbeitstag zu verzeichnen.

Durch die hohe Nachfrage sowohl nach „klassischer“ Projektbearbeitung am GWAP als auch nach der Ansicht von KDFV-Ergebnissen am GWAP⁷ kommt es teilweise zu Auslastungen über die Kapazitätsgrenzen hinaus. Im ersten Halbjahr 2018 waren bereits an 48 Arbeitstagen mehr als fünf Personen an den verfügbaren PC-Arbeitsplätzen aktiv. Zum Vergleich: 2014 gab es lediglich einen Arbeitstag, an dem mehr als sechs Nutzende am GWAP arbeiteten und somit eine Auslastung von mehr als 100 % vorlag (Abbildung e). Die Häufigkeit von Mehrfachnutzungen⁸ der GWAP-Arbeitsplätze stieg in den letzten Jahren kontinuierlich an (2015: 21 Tage, 2016: 37 Tage,

2017: 44 Tage). Eine solche Auslastung der GWAP über 100 % erfordert von den Mitarbeiterinnen und Mitarbeitern des FDZ eine besondere Sorgfalt in der Terminvergabe und Koordinierung und verursacht einen hohen Betreuungsaufwand für die Nutzerinnen und Nutzer an den GWAP.

Kooperation mit dem IAB

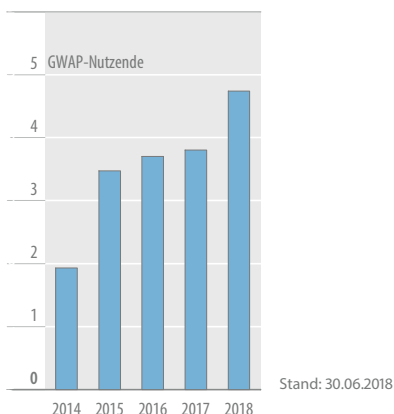
Zusätzlich zu Gastwissenschaftlerinnen und -wissenschaftlern der FDZ der Statistischen Ämter des Bundes und der Länder betreut der FDZ-Standort des AfS im Gebäude des DIW Berlin seit 2012 auch Gastwissenschaftlerarbeitsplätze⁹, die einen Remote-Zugriff zu den Mikrodaten des FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ der BA im IAB) ermöglichen.¹⁰ Die Aufgaben der Mitarbeiterinnen und Mitarbeiter des AfS beschränken sich dabei auf die organisatorische Betreuung des Datenzugriffs; inhaltliche Projektunterstützung wird nicht geleistet.

Diese Kooperation hat sich in den letzten Jahren ebenfalls sehr positiv entwickelt: Die Zahl der Nutzungen stieg von 319 im Jahr 2014 kontinuierlich an. Die letzte Auswertung des Nutzungsverhaltens in den mittlerweile 15 internationalen Standorten zeigt, dass Berlin im Jahr 2017 weltweit der am häufigsten frequentierte Außenstandort des FDZ der BA im IAB war (Abbildung f). Mit 613 dokumentierten Nutzungen der Arbeitsmarktdaten des FDZ der BA im IAB war der Berliner Standort in drei Viertel der regulären Öffnungszeiten zu 100 % ausgelastet.

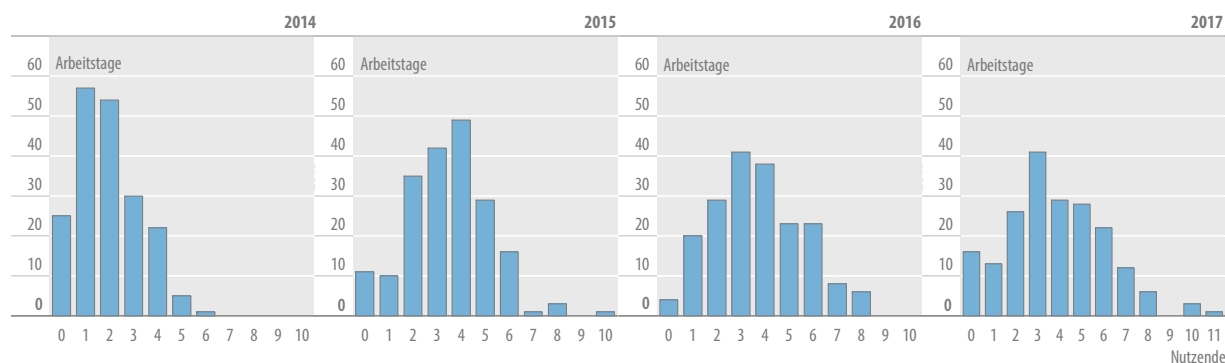
Neuerungen und Herausforderungen im Zuge der BStatG-Novellierung

Das volle Analysepotenzial formal anonymisierter Daten konnte in den FDZ bisher nur über die KDFV ausgeschöpft werden, weshalb dieser Zugangsweg besonders häufig nachgefragt wurde. Im Rahmen der Anpassung des geltenden deutschen Rechts an europäische Anforderungen wurde im Jahr 2016

d | Datennutzende 2014 bis 2018 an den GWAP des FDZ-Standorts im AfS pro Arbeitstag



e | Tatsächliche Auslastung der GWAP des FDZ-Standorts im AfS 2014 bis 2017



Lesebeispiel:

Im Jahr 2017 wurden an drei Arbeitstagen zehn datennutzende Personen pro Tag dokumentiert.

⁷ Hier können Wissenschaftlerinnen und Wissenschaftler die per KDFV erzeugten Ergebnisse eigenständig kontrollieren und ⁸ entscheiden, welche Ergebnisse sie tatsächlich benötigen bzw. die Auswertungen per KDFV fortführen, bis publikationsreifer Output entsteht, was

zu einer Verringerung des Bearbeitungsaufwands im FDZ führt. Wenn eine Person mehrfach am Tag für kürzere Zeiträume am GWAP arbeitet, wird dies als eine einzelne Nutzung pro Arbeitstag erfasst. Der tatsächliche Aufwand für die Koordi-

nierung der vorhandenen Arbeitsplätze ist also nicht vollständig abgebildet.

⁹ Seit Mitte 2018 sind es fünf Arbeitsplätze, zuvor waren es vier.

¹⁰ Nähere Informationen zum FDZ der BA im IAB unter: <https://fdz.iab.de/>.

das Bundesstatistikgesetz zahlreichen Änderungen unterworfen, von denen die FDZ hauptsächlich durch die Novellierung des § 16 Abs. 6 Nr. 2 profitieren. So darf nun „innerhalb speziell abgesicherter Bereiche des Statistischen Bundesamtes und der statistischen Ämter der Länder Zugang zu formal anonymisierten Einzelangaben [gewährt werden], wenn wirksame Vorkehrungen zur Wahrung der Geheimhaltung getroffen“ sind. Als formal anonymisierte Einzelangaben werden dabei gemäß § 5 Abs. 3 BStatG „Einzelangaben ohne Name und Anschrift“ bezeichnet. Zur gesetzlich geforderten Sicherstellung der Geheimhaltung wurden unter allen Statistischen Ämtern des Bundes und der Länder rechtliche, technische und organisatorische Maßnahmen festgelegt und abgestimmt. Der Maßnahmenkatalog für die Bereitstellung formal anonymisierter Angaben an den GWAP der FDZ enthält folgende Punkte:

- Räumliche Abschottung und IT-Abschottung,
- Protokollierung der Arbeitsschritte,
- Sicherstellung der Geheimhaltung,
- Verpflichtung der Empfängerinnen und Empfänger nach § 16 Abs. 7 BStatG,
- Abschluss eines Datennutzungsvertrages.

Die bisherigen Regelungen zur Sicherstellung der statistischen Geheimhaltung und des Datenzugangs gelten weiterhin.

Die Wissenschaftlerinnen und Wissenschaftler können nun auch am GWAP in voller Merkmaltiefe mit den Mikrodaten arbeiten, ohne dass vorab im FDZ individuell anonymisierte Datensätze erstellt werden müssen, was zu einer deutlichen Arbeitsentlastung der FDZ-Standorte führen sollte. Durch diese neu durch den Gesetzgeber eröffnete Möglichkeit könnte zukünftig eine Vielzahl der bestehenden KDFV-Projekte an die GWAP wechseln. Die Wissenschaft kann die formal anonymisierten Mikrodaten selbstständig nutzen, die erzeugten Ergebnisse eigenständig kontrollieren und die Auswertungen so lange fortführen, bis ein publikationsreifer Output entsteht. Das Volumen der final zu prüfenden Ergebnisse durch die FDZ wird durch diese Vorgehensweise voraussichtlich deutlich reduziert.

Die Novellierung des BStatG trat am 20. Oktober 2016 in Kraft. Für die Umsetzung der Gesetzesänderung wurde am 8. November 2017 durch alle Leiterinnen und Leiter der Statistischen Ämter des Bundes und der Länder eine entsprechende Kooperationsvereinbarung unterzeichnet. Im Nachgang äußerte das Bayerische Landesamt für Statistik allerdings bei einigen Statistiken Bedenken, diese tatsächlich in formal anonymisierter Form zur wissenschaftlichen Nutzung am GWAP zur Verfügung zu stellen. Dies führte dazu, dass aktuell vor der Bereitstellung am GWAP für 13 Statistiken spezielle Anonymisierungsmaßnahmen für die bayerischen Daten vorgenommen werden müssen. Dies gilt

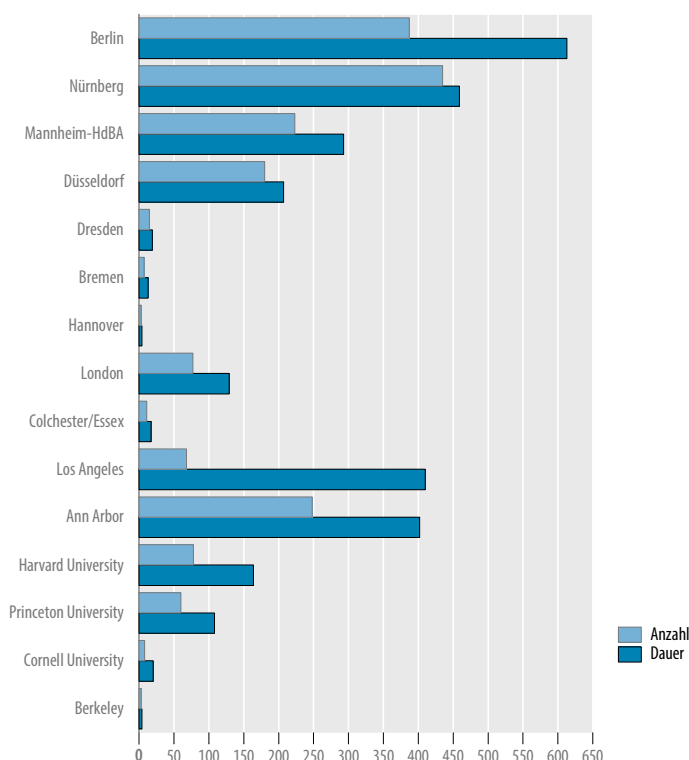
unter anderem für folgende sehr stark nachgefragte Statistiken:

- Mikrozensus,
- Todesursachenstatistik,
- Steuerstatistiken,
- Verdienststrukturerhebung/Gehalts- und Lohnstrukturerhebung.

Die Nutzenden erhalten am GWAP also formal anonymisierte Einzeldaten für 15 Bundesländer und stärker anonymisierte bayerische Einzeldaten. Für wissenschaftliche Analysen, für die auch das uneingeschränkte bayerische Material benötigt wird, muss eine zusätzliche KDFV durchgeführt werden, um konsistente Ergebnisse zu erlangen, was wiederum zu doppelter Datenhaltung und Mehraufwänden in der Bearbeitung führt.

Nicht am GWAP bereitgestellt werden dürfen bayerische Einzeldaten für 24 Statistiken aus den Bereichen Wirtschaft, Energie und Umwelt. Am GWAP sind hier nur die Einzeldaten für 15 Bundesländer verfügbar. Hiervon sind insbesondere die Amtlichen Firmendaten für Deutschland (AFiD) betroffen, für die in den letzten Jahren eine steigende Nachfrage verzeichnet wurde. Wirtschaftsstatistiken werden demnach aktuell kritischer eingestuft als die nicht weniger sensiblen Steuerstatistiken. Sollen wissenschaftliche Analysen für alle Bundesländer durchgeführt werden, muss zwangsläufig zusätzlich die KDFV durchgeführt werden. Diese wird durch die FDZ kostenfrei angeboten, da die Wissenschaft nicht

f | Anzahl und Dauer (in Tagen) der Gastaufenthalte 2017 an den Standorten des FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung



aufgrund von ämter-spezifischen Differenzen mit zusätzlichen Kosten belastet werden soll.

Da gerade die wissenschaftlichen Einrichtungen in Berlin und Brandenburg verstärkt auf die Wirtschaftsdaten zurückgreifen, ist der FDZ-Standort im AfS besonders von dem Vorgehen aufgrund der abweichenden bayerischen Rechtsauffassung betroffen.

Ausblick

Der regionale FDZ-Standort im AfS erfreut sich seit vielen Jahren großer Nachfrage, nicht nur bei der regionalen Wissenschaft. Mehr als 100 aktive Projekte, eine konstant hohe Auslastung der verfügbaren GWAP sowie die Bearbeitung von fast 1000 KDFV-Auswertungsanfragen im Jahr 2017 veranschaulichen den immensen wissenschaftlichen Bedarf an amtlichen Mikrodaten.

Da die Forschungsprojekte aufgrund von Verknüpfungen und Panelbildungen immer komplexer und die auszuwertenden Mikrodaten immer größer und speicherintensiver werden, ist eine zeitgemäße technische Basis für die effiziente Bearbeitung der Anfragen zwingend notwendig. Durch das AfS wurden im Jahr 2018 zwei Hochleistungsrechner beschafft, die eine zügige und parallele Bearbeitung von KDFV-Anfragen erlauben. Da die Auslastung der GWAP am FDZ-Standort des AfS im DIW Berlin bereits jetzt oftmals die 100% übersteigt und mit einer weiter steigenden Nachfrage gerechnet wird, wird eine weitere Ausweitung der Kapazitäten angestrebt, um eine permanente Überbelegung der GWAP zu vermeiden. Die damit verbundene Möglichkeit, dort noch mehr Arbeitsplätze für Gastwissenschaftlerinnen und -wissenschaftler vorzuhalten, ist eine weitere wichtige Voraussetzung für die Wissenschaft, ihren rechtlichen Anspruch auf die Bearbeitung formal anonymisierter Mikrodaten am GWAP nutzen zu können.

In den kommenden Jahren wird es vor allem darauf ankommen, die Gesetzesänderungen des § 16 Abs. 6 Nr. 2 einheitlich für alle Statistiken umzusetzen. Durch die abweichende bayerische Rechtsauffassung über die Zulässigkeit der Bereitstellung formal anonymisierter Mikrodaten und den daraus folgenden Mehraufwänden (doppelte Datenhaltung, Anonymisierungen, Nutzung und Koordinierung mehrerer Zugangswege) kommt es derzeit zu Wartezeiten, insbesondere bei Projekten zu Wirtschaftsstatistiken sowie zu Energie- und Umweltstatistiken. Hier wird eine rechtzeitige Kontaktaufnahme mit den jeweiligen FDZ-Standorten empfohlen. Die FDZ der Statistischen Ämter des Bundes und der Länder haben weiterhin das Ziel, auf eine einheitliche Vorgehensweise bei der Bereitstellung formal anonymisierter Daten am GWAP hinzuwirken.

Ramona Voshage leitet das Referat *Mikrodaten, Forschungsdatenzentrum* des Amtes für Statistik Berlin-Brandenburg.

Daniela Janke und **Anja Malchin** sind Referentinnen im Referat *Mikrodaten, Forschungsdatenzentrum* des Amtes für Statistik Berlin-Brandenburg.