

Hochschulstatistikdaten im Angebot des Forschungsdatenzentrums – Analysepotentiale und Zugangsmöglichkeiten

Dipl.-Soz. Patrick Rothe

Bologna-Reform, Exzellenzinitiative, Eliteuniversitäten, Duale Hochschule, Juniorprofessur, Pakt für die Lehre, Bildungsrepublik – das sind nur ein paar der Schlagworte, die die mediale Berichterstattung über, aber auch die wissenschaftliche Beschäftigung mit dem deutschen Hochschulsystem im Laufe der letzten zehn Jahre maßgeblich geprägt haben. Möchte man sich den stattgefundenen strukturellen Umwälzungen aus wissenschaftlicher Perspektive annähern, so ist man als Forscher hierfür auf eine umfassende, valide und belastbare Datengrundlage angewiesen. Die amtliche Statistik trägt mit ihren Erhebungen entscheidend dazu bei, diese Grundlage bereitzustellen. Dabei ist neben den traditionellen Veröffentlichungswegen der amtlichen Statistik (Leitner 2013, Wagner 2012), die sich an eine breite Öffentlichkeit richten, insbesondere die Einzeldatenbereitstellung durch das Forschungsdatenzentrum (FDZ) der Statistischen Ämter der Länder als spezielles Angebot für die nationale und internationale Wissenschaft von Bedeutung.

Das FDZ – ein modernes, nutzerorientiertes Dienstleistungsangebot für die empirische Wissenschaft

Das 2002 als Reaktion auf das sogenannte KVI-Gutachten (Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik 2001) gegründete FDZ der Statistischen Ämter der Länder bietet in enger Zusammenarbeit mit dem FDZ des Statistischen Bundesamtes Einzeldaten¹ aus Erhebungen der amtlichen Statistik für die wissenschaftliche Nutzung an. Im Laufe der vergangenen Jahre konnte sich das FDZ dabei als ein wichtiger Partner von Wissenschaft und Forschung in Deutschland etablieren und das verfügbare Angebot an Daten und Dienstleistungen über die Zeit hinweg stetig ausbauen (Köhler 2012, Nowak 2012, Rothe 2012).

Der entscheidende Vorteil der Bereitstellung statistischer Daten in Form von Einzeldaten gegenüber aggregierten Daten liegt in der Möglichkeit, als Datennutzer flexibel eigene Tabellen variabler Gliede-

rungstiefe erstellen zu können, oder aber die anonymisierten Einzelangaben als Grundlage komplexer statistischer Analyseverfahren zu verwenden. Hierbei sind zudem in vielen Fällen auch regionale Auswertungen möglich (Rothe 2013). Die primäre Zielgruppe des FDZ sind Einrichtungen der unabhängigen wissenschaftlichen Forschung und deren Mitarbeiter, denen gemäß §16 Absatz 6 des Bundesstatistikgesetzes der Zugang zu faktisch anonymisiertem² Datenmaterial gewährt werden kann.

Die insgesamt 13 regionalen Standorte des FDZ – zuzüglich mehrerer Außenstellen in wissenschaftlichen Einrichtungen – sind dabei nicht nur als Ansprechpartner für Wissenschaftler aus dem jeweiligen Bundesland tätig, sondern verfügen jeweils über einen eigenen fachlichen Schwerpunkt. Im Fall des FDZ-Standorts München handelt es sich dabei um das Themenfeld „Bildung und Kultur“, womit insbesondere die bundesweite Verantwortlichkeit für die Einzeldaten der Statistiken aus dem Bereich Hochschule verbunden ist.

1 Einzeldaten – oft auch als Mikro- oder Individualdaten bezeichnet – beinhalten im Gegensatz zu Aggregatdaten die individuellen Angaben zu den Merkmalsausprägungen jeder einzelnen in einer Statistik enthaltenen Einheit, beispielsweise einer Einzelperson, eines Betriebs oder eines Unternehmens.

2 Von faktischer Anonymität kann dann gesprochen werden, wenn der Aufwand, der dafür nötig ist, um einen einzelnen Merkmalsträger anhand seiner individuellen Merkmalskombination innerhalb eines Datenbestands zu identifizieren, aufgrund getroffener Vorkehrungsmaßnahmen so hoch ausfällt, dass dieser in keinerlei rationalem Verhältnis zum durch eine erfolgreiche Reidentifizierung erzielbaren Nutzen steht.

Die Inhalte der angebotenen Bildungsstatistiken

Neben der semesterweise vorliegenden Statistik der Studenten und der Prüfungen (ab dem Wintersemester 1995/1996) befinden sich die jahresweise erhobene Hochschulpersonal- und Stellenstatistik sowie die Statistik der Habilitationen (jeweils ab dem Jahr 1998) im Angebot des FDZ. Enthalten sind dabei grundsätzlich Angaben zu allen Bundesländern.³ Es handelt sich dabei um dasselbe Datenmaterial, das auch als Grundlage für die offiziellen Veröffentlichungen der Statistischen Landesämter sowie des Statistischen Bundesamtes dient.⁴

Die Statistik der Studenten beinhaltet alle relevanten Angaben, die im Verlauf eines Studiums, beispielsweise im Zuge von Einschreibung oder Rückmeldung, über Studierende erhoben werden: neben grundlegenden soziodemographischen Merkmalen wie Geschlecht, Geburtsjahr oder Staatsangehörigkeit beispielweise Informationen zur aktuellen oder früheren besuchten Hochschule, zum Studienfach, dem angestrebten Abschluss oder Art und Ort des Erwerbs der Hochschulzugangsberechtigung. Die Statistik der Prüfungen enthält hingegen Angaben zu denjenigen Personen, die im betreffenden Semester eine Abschlussprüfung abgelegt haben, und beinhaltet für diese neben Hochschule und Studienfach insbesondere Daten zur abgelegten Prüfung wie das dabei erzielte Ergebnis.

In Satzart 1 und 2 der Personal- und Stellenstatistik sind alle Beschäftigten erfasst, die jeweils zum Stichtag 1. Dezember an einer deutschen Hochschule tätig waren. Dies gilt sowohl für Mitarbeiter aus dem wissenschaftlichen oder künstlerischen Bereich als auch für Angehörige der Verwaltung sowie für technische oder sonstige Mitarbeiter. Auch für diese beinhaltet die Statistik Grundinformationen zum Geschlecht oder der Staatsangehörigkeit, daneben aber auch detaillierte Angaben zum Beschäftigungsverhältnis oder der fachlichen Zuordnung. Satzart 3 und 4 ergänzen die statistischen Angaben zum Personal, indem sie Auskunft über die laut den offiziellen Stellenplänen an den Hochschulen vorhandenen, besetzten und nicht-besetzten Stellen und deren wesentliche Merkmale geben. Komplettiert werden die über das FDZ verfügbaren Hochschulstatistikdaten durch die Statistik der Habilitationen,

in der Informationen zu allen innerhalb eines Kalenderjahrs abgeschlossenen Habilitationsverfahren enthalten sind.

Detaillierte Übersichten zu den Merkmalen der einzelnen Statistiken können abgerufen werden unter www.forschungsdatenzentrum.de/datenangebot.asp#p_bildung

Analysepotentiale und Einschränkungen

Die Hochschuldaten der amtlichen Statistik, wie sie im FDZ angeboten werden, ermöglichen es, die Struktur des deutschen Hochschulwesens im Detail zu betrachten und alle über die Zeit hinweg erfolgten maßgeblichen Veränderungen präzise nachzuverfolgen. Regionalisierte Auswertungen sind bis auf Ebene der einzelnen Hochschulen – und noch darüber hinausgehend bis auf Fachbereichsebene – möglich und liefern damit ein exaktes Abbild der strukturellen Gegebenheiten im tertiären Bildungssektor. Das Vorhandensein von Angaben sowohl über die Studierenden als auch zum wissenschaftlichen und nicht-wissenschaftlichen Personal der Hochschulen erlaubt dabei die Betrachtung des Systems Hochschule aus unterschiedlichen Perspektiven.

Anhand der vorliegenden Daten können FDZ-Nutzer unter anderem die langfristige Entwicklung der Studierenden- und Absolventenzahlen sowie der Studienwahl und des Studierverhaltens – ausdifferenziert beispielsweise nach Hochschulart, Fachrichtung oder Geschlecht – verfolgen. Weitere Beispiele für mögliche Analysen sind die Untersuchung der Situation von Migranten und ausländischen Studierenden im deutschen Hochschulsystem oder die Betrachtung innerdeutscher Studentenwanderungen zwischen den Bundesländern. Auch die Berechnung hochschulstatistischer Kennzahlen – beispielsweise im Rahmen von Maßnahmen zum regionalen Bildungsmonitoring – ist ausgehend von den angebotenen Daten möglich. Abseits der klassischen Hochschul- und Bildungsforschung finden die Daten der Hochschulstatistiken aus dem FDZ-Bestand jedoch ebenso Verwendung: So werden sie unter anderem als Hintergrundinformationen für die Modellierung der wirtschaftlichen und infrastrukturellen Leistungsfähigkeit von Regionen herange-

³ Eine Ausnahme stellen die frühen Berichtsjahre (1995–1998) dar, für die für einzelne Bundesländer leider keine Einzeldaten mehr vorlagen, die in den FDZ-Datenbestand hätten übernommen werden können.

⁴ In einigen wenigen Fällen kann es es dennoch zu Abweichungen des FDZ-Datenbestands gegenüber den veröffentlichten Zahlen kommen, da bei der zeitlich nachgelagerten Einzeldatenaufbereitung durch das FDZ Nachmeldungen berücksichtigt werden können, die zum Zeitpunkt der Erstellung der offiziellen Veröffentlichungen noch nicht vorlagen. Es handelt sich hierbei jedoch nur um sehr wenige, sehr gering ausfallende Abweichungen. Diese werden vom FDZ-Standort München entsprechend dokumentiert.

zogen, denn das Vorhandensein von Hochschulen und qualifizierten Absolventen stellt einen wichtigen Standortfaktor im nationalen und internationalen Wettbewerb dar.

Einer der Hauptvorteile der Arbeit mit den Hochschuldaten der amtlichen Statistik liegt darin, dass es sich hierbei um Vollerhebungen mit sehr großen Fallzahlen handelt, womit Probleme, die für gewöhnlich mit der Arbeit mit durch Stichprobenerhebungen gewonnenen Ergebnissen einhergehen – insbesondere Non-Response und damit verbundene Fragen der (Nicht-)Repräsentativität – entfallen. Daher können die Daten nicht nur für eigenständige Untersuchungen, sondern auch immer dann zum Vergleich herangezogen werden, wenn eigene Befragungstichproben hinsichtlich ihrer Repräsentativität und inhaltlicher Aussagekraft anhand offizieller Zahlen in einen umfassenderen Kontext eingeordnet werden sollen. Auch die Durchführung regional tiefgliedriger Analysen sowie die Untersuchung spezifischer Teilpopulationen ist unter diesen Gesichtspunkten problemlos möglich.

Den genannten Vorteilen stehen naturgemäß auch Einschränkungen und Begrenzungen gegenüber: So handelt es sich bei den Hochschulstatistikdaten im FDZ-Bestand – wie bei allen im FDZ angebotenen Daten der amtlichen Statistik – um Sekundärdaten, die ursprünglich nicht explizit für die Nutzung durch Forschung und Wissenschaft, sondern für einen anderen Zweck erhoben wurden. Die Nutzbarkeit für die wissenschaftliche Verwendung, al-

so die Aufbereitung und Dokumentation der Daten, die Bereitstellung für die unterschiedlichen Statistiksoftwarepakete und die Erstellung spezifischer Metadaten,⁵ erfolgt erst im Nachgang zur eigentlichen Produktion der Statistik. Hieraus resultieren eine Reihe inhaltlicher Besonderheiten, auf die sich Forschende bei der Nutzung der Daten über das FDZ einstellen müssen: So sind in den entsprechenden Datenbeständen grundsätzlich keine Angaben zur sozialen Herkunft der enthaltenen Personen vorhanden, was Untersuchungen im Bereich der sozialen Bildungsungleichheit und die Beschäftigung mit Prozessen der Bildungs- und Statusvererbung verhindert. Auch werden durch die amtliche Statistik keinerlei Einstellungsitems erhoben, die Aufschluss über Motivationen und Erwartungen oder die Zufriedenheit von Studierenden und Beschäftigten der Hochschulen geben könnten. Die zur Analyse von Bildungsverläufen im Hochschulbereich wünschenswerte Bereitstellung der Berichtsjahre der Hochschulstatistiken in verknüpfter Form als Längsschnittdatensatz ist ebenso derzeit aufgrund der rechtlichen Rahmenbedingungen nicht möglich.

Möglichkeiten zur Nutzung der Hochschulstatistikdaten über das FDZ

Die Nutzung der Einzeldaten der Hochschulstatistiken im FDZ ist über mehrere unterschiedliche Zugangswege möglich. Diese unterscheiden sich hinsichtlich des Orts der Datennutzung, des Grads der Anonymisierung und des Detaillierungsgrads der enthaltenen Informationen sowie der Zielgruppe (vgl. Abbildung 1).

⁵ Unter dem Begriff Metadaten werden im FDZ-Kontext alle inhaltlich relevanten Informationen zu den angebotenen Daten („Daten über Daten“) zusammengefasst. Hierzu zählen insbesondere Datensatzbeschreibungen, Schlüsselverzeichnisse, Rechtsgrundlagen und Qualitätsberichte sowie aufbereitungsspezifische Dokumentationen.

Unterschiedliche Zugangswege zum FDZ

Abb. 1

Zugangsweg	KDFV	GWAP	CAMPUS-File
Anonymisierung	formal	faktisch	absolut
Informationsgehalt	optimal	sehr hoch	eher gering
Verwendungszweck	detaillierte wissenschaftliche Auswertungen		Lehr- und Übungszwecke
Direkter Datenzugang	nein	ja	ja
Zielgruppe	Allgemeinheit	Wissenschaftler	Studierende, Lehrende

Interessierte Wissenschaftler können den Gastwissenschaftlerarbeitsplatz (GWAP) in München – oder in jedem anderen regionalen FDZ-Standort – besuchen, um dort mittels der Programme STATA, SPSS, SAS oder R in einer abgeschotteten Arbeitsumgebung statistische Auswertungen durchzuführen. Die Daten, die den Nutzern hier zur Verfügung gestellt werden, müssen angesichts der kontrollierbaren Rahmenbedingungen vor Ort weniger stark anonymisiert werden als es der Fall wäre, wenn diese Daten die Räumlichkeiten der amtlichen Statistik verlassen würden. Hierdurch kann der in den Einzeldaten enthaltene Informationsgehalt weitgehend für wissenschaftliche Analysen erhalten werden.

Eine andere Möglichkeit der Datennutzung ist die Kontrollierte Datenfernverarbeitung (KDFV). Bei dieser erstellen die Datennutzer anhand eines Beispieldatensatzes Auswertungsprogramme für das von ihnen bevorzugte Analyseprogramm und lassen dem zuständigen FDZ-Standort – im Fall der Hochschulstatistiken also dem Standort München – ihre Syntax zukommen. Dort wird diese Syntax von den Mitarbeitern auf lediglich formal anonymisierte⁶ Einzeldaten angewandt. Nach einer abschließenden Geheimhaltungsprüfung werden die Ergebnisse den Datennutzern in absolut anonymer⁷ Form zurückübermittelt. Als eine spezielle Form der Sonderauswertung steht die KDFV im Gegensatz zu den anderen FDZ-Datennutzungswegen grundsätzlich allen Interessierten – und nicht nur Angehörigen unabhängiger Forschungseinrichtungen – zur Verfügung. Die Nutzung

der KDFV bietet sich aus Nutzersicht insbesondere dann an, wenn der nächste FDZ-Gastwissenschaftlerarbeitsplatz weit entfernt liegt und somit durch An- und Abreise ein hoher Kosten- und Zeitaufwand anfallen würde, oder wenn bei der Durchführung der faktischen Anonymisierung zu viel an in den Daten enthaltenem Informationsgehalt verloren gehen würde.

Einen Sonderfall stellt das CAMPUS-File der Studenten- beziehungsweise Prüfungsstatistik dar. Bei diesem handelt es sich um einen speziell auf Lehrzwecke zugeschnittenen, absolut anonymen Datensatz, der kostenlos über die FDZ-Internetseite als Download beziehbar ist (Zwick 2008). Dieser ermöglicht es Studierenden, mit einer vereinfachten Form der Studenten- und Prüfungsstatistik erste eigene statistische Auswertungen durchzuführen und dabei Aufbau und Struktur der Daten kennenzulernen. Die Verwendung der CAMPUS-Files im Rahmen von Lehrveranstaltungen ist ausdrücklich erwünscht. Valide inhaltliche Auswertungen sind damit jedoch aufgrund der zahlreichen Maßnahmen, die nötig wurden, um den Grad der absoluten Anonymität zu erreichen, nicht möglich. Hierfür sollte auf die anderen angebotenen Zugangsmöglichkeiten zu den Daten der Studenten- und Prüfungsstatistik zurückgegriffen werden.

Einen Überblick über die verfügbaren Datenbestände und die jeweils möglichen Zugangswege bietet Abbildung 2:

6 Bei formal anonymisierten Datenbeständen werden lediglich die direkten Identifikatoren – beispielsweise im Fall der Studenten- und Prüfungsstatistik die Matrikelnummer – aus dem Datenmaterial entfernt. Da keine darüber hinausgehenden Änderungen an den Daten vorgenommen werden, bleibt das enthaltene Analysepotential unberührt. Formal anonymisierte Daten dürfen nur von Mitarbeitern der amtlichen Statistik innerhalb der Statistischen Ämter verarbeitet werden.

7 Absolute Anonymität bedeutet, dass es unter keinen denkbaren Umständen möglich ist, ausgehend von den betreffenden Daten Rückschlüsse auf die dahinter stehenden Merkmalsträger zu ziehen und diese hierdurch zu reidentifizieren. Gemäß §16 Absatz 1 Satz 4 BStatG dürfen solche Daten einer breiten Öffentlichkeit zugänglich gemacht werden.

Verfügbare FDZ-Datenbestände aus dem Bereich Hochschule
Stand Juli 2013

Abb. 2

Datenbestand	Zugangswege	vorliegend ab
Statistik der Studenten	GWAP, KDFV, CAMPUS-File	WS 1995/96
Statistik der Prüfungen	GWAP, KDFV, CAMPUS-File	WS 1995/96
Personal- und Stellenstatistik Satzart 1: Wissenschaftliches und künstlerisches Personal Satzart 2: Verwaltungs-, technisches und sonstiges Personal Satzart 3: Stellen für wissenschaftliches/künstlerisches Personal Satzart 4: Stellen für Verwaltungs-, technisches und sonstiges Personal	GWAP, KDFV	1998
Statistik der Habilitationen	GWAP, KDFV	1998

Weitere Daten mit Bildungsbezug im Angebot des FDZ

Über die genannten Statistiken aus dem Bereich Hochschule hinaus hält das FDZ-Datenangebot eine ganze Reihe weiterer für die Bildungsforschung potentiell interessanter Daten bereit. Mithilfe von Angaben aus dem Mikrozensus(-Panel), aus der Kinder- und Jugendhilfestatistik, der Berufsbildungsstatistik – die ebenfalls vom FDZ-Standort München betreut wird – oder der Europäischen Erhebung über die betriebliche Weiterbildung (Continuing vocational training survey – CVTS) ist es möglich, auch andere wichtige Bildungsphasen, wie die frühkindliche Bildung, die Berufsausbildung oder den Bereich des lebenslangen Lernens, ausgehend von einer fundierten Datengrundlage, zu untersuchen (Wagner 2009).

Fazit

Die über das FDZ angebotenen Hochschuldaten der amtlichen Statistik sind immer dann eine geeignete Datengrundlage für wissenschaftliche Untersuchungen, wenn es darum geht, ein umfassendes

und zugleich detailliertes Bild der Strukturen des deutschen Hochschulwesens und dessen Entwicklung seit dem Jahr 1995 zu zeichnen. Dabei kommt den Forschenden die Möglichkeit, mit den amtlichen Einzeldaten regionale Vergleiche anzustellen und aufgrund der hohen Fallzahlen auch spezifische Teilpopulationen detailliert untersuchen zu können, zugute. Zudem können anhand der amtlichen Daten die Ergebnisse eigener Befragungen und Studien auf ihre Verallgemeinerbarkeit hin untersucht und verlässlich in den umfassenden institutionellen Kontext eingebettet werden. Prinzipbedingt können mit den über das FDZ angebotenen Daten der amtlichen Statistik jedoch nicht alle wissenschaftlich relevanten Aspekte des Hochschulwesens abgedeckt werden. In Kombination mit anderen im Bereich der Hochschulforschung relevanten Datenquellen – oder auch Daten aus eigenen Erhebungen – ermöglichen sie aber eine multiperspektivische Annäherung an das System Hochschule und ergänzen sich dabei wechselseitig mit den Angeboten anderer Datenproduzenten.

Literatur

Bundesstatistikgesetz (BStatG) (1987), Gesetz über die Statistik für Bundeszwecke.

Köhler, Florian (2012), 10 Jahre Forschungsdatenzentren der amtlichen Statistik – Angebot und Nachfrage nach amtlichen Mikrodaten. Statistische Monatshefte Niedersachsen 6/2012, S. 333–341.

Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Hg.) (2001), Wege zu einer besseren informationellen Infrastruktur. Baden-Baden: Nomos.

Leitner, Cornelia (2013), Daten der Statistischen Ämter des Bundes und der Länder. In: Arbeitsgruppe Regionale Standards (Hg.): Regionale Standards. Ausgabe 2013. Eine gemeinsame Empfehlung des ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V., der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI) und des Statistischen Bundes-

amtes. GESIS-Schriftenreihe Band 12. Mannheim/Köln: GESIS, S. 269–277.

Nowak, Kristin (2012), 10 Jahre Forschungsdatenzentrum der Länder. Statistisches Monatsheft Baden-Württemberg 07/2012, S. 45–49.

Rothe, Patrick (2012), 10 Jahre Forschungsdatenzentrum der Statistischen Ämter der Länder – Ein Blick auf Vergangenheit, Gegenwart und Zukunft der Forschungsdateninfrastruktur der amtlichen Statistik in Deutschland. Bayern in Zahlen 7/2012, S. 492–500.

Rothe, Patrick (2013), Das Angebot der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. In: Arbeitsgruppe Regionale Standards (Hg.): Regionale Standards. Ausgabe 2013. Eine gemeinsame Empfehlung des ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V., der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI) und des

Statistischen Bundesamtes. GESIS-Schriftenreihe Band 12. Mannheim/Köln: GESIS, S. 269–277.

Wagner, Simone (2009), Bildungsdaten der amtlichen Statistik im Forschungsdatenzentrum. Bayern in Zahlen 7/2009, S. 282–298.

Wagner, Simone (2012), Datenangebot und Datenzugangswege des Bayerischen Landesamtes für

Statistik und Datenverarbeitung. Bayern in Zahlen 5/2012, S. 301–316.

Zwick, Markus (2008), CAMPUS-Files – Kostenlose Public Use Files für die Lehre. AStA Wirtschafts- und Sozialstatistisches Archiv, Vol. 2, Nr. 1–2: S. 175–188.