# **DAGStat-Bulletin**

Neues über Statistik und aus den Gesellschaften der Deutschen Arbeitsgemeinschaft Statistik



Ausgabe 19: Juni 2017



Liebe Leserinnen und Leser,

dieses Bulletin enthält neben den Rubriken "Statistik Aktuell" und "Veranstaltungen" erstmalig die Rubrik "Statistik bei DFG und Statistischem Bundesamt". In dieser neuen Rubrik wird insbesondere von zwei Aktivitäten der Deutschen Forschungsgemeinschaft (DFG) berichtet, die auch die DAGStat betreffen. Die eine Aktivität betrifft die Leitlinien im Umgang mit Forschungsdaten, die die DFG im September 2015 herausgegeben hatte. Dazu gibt es hier nun ein Interview mit Dr. Johannes Fournier, dem Programmdirektor der Gruppe für Wissenschaftliche Literaturversorgungs- und Informationssysteme der DFG. In diesem Interview erläutert er, was die DFG mit diesen Leitlinien erreichen will. Der Zugang zu den Forschungsdaten ist dabei ein zentrales Element für die Qualitätssicherung wissenschaftlicher Arbeiten.

Ein weiteres zentrales Element von wissenschaftlichen Arbeiten ist der Zu-

gang zu den entsprechenden Veröffentlichungen über diese Daten. Diese Veröffentlichungen erscheinen oft in Fachzeitschriften, die sich wegen hoher Gebühren nur wenige Forschungsinstitute leisten können. Ergebnisse, die von der öffentlichen Hand wie der DFG gefördert wurden, sind somit nicht allgemein öffentlich zugänglich. Damit mehr von den öffentlich geförderten Ergebnissen frei zugänglich werden, hat die DFG eine Initiative zu "Open Access" gestartet, in deren Rahmen sie am 19.01.2017 zu einem Rundgespräch mit Vertretern verschiedener Gesellschaften einlud, an dem ich als Vertreterin der DAGStat teilnahm. Einen Bericht über dieses Rundgespräch finden Sie in diesem Bulletin.

Die DFG hatte außerdem im letzten halben Jahr zu Stellungnahmen der Gesellschaften zur Fächerstruktur bei der DFG aufgerufen. Da dieser Aufruf recht unerwartet kam, konnte in der kurzen zur Verfügung stehenden Zeit keine Stellungnahme erarbeitet werden, die von allen DAGStat-Gesellschaften getragen werden konnte. Ich habe aber in einem Brief darauf hingewiesen, dass insbesondere die Statistiker, die sich weder der Ökonometrie, der Biometrie noch der Mathematischen Statistik zuordnen lassen, nicht mit der Fächerstruktur zufrieden sind, und ein Rundgespräch angeregt. Da-

# Aus dem Inhalt:

### Statistik Aktuell

IAB-BAMF-SOEP-Befragung von Geflüchteten

Verliert Deutschland auf dem Zukunftsmarkt den Anschluss?

# Statistik bei DFG und Statistischem Bundesamt

Wozu dienen die DFG-Leitlinien im Umgang mit Forschungsdaten?

DFG-Initiative zu Open Access

Revision der Fächerklassifikation beim Statistischen Bundesamt

### Veranstaltungen

DAGStat-Symposium "Ist Bildung messbar?"

DOTS – 6. Dortmunder Tag der Statistik 2017

Big Data Challenge in Bremen

<u>Personalia</u>













Sektion Methoden der Empirischen Sozialforschung der DGS













gegen wissen wir seit längerem, dass eine Revision der Fächersystematiken des Statistischen Bunin einem fünfjährigen durchgeführt wird. Für den nächsten Revisionsprozess, der dieses Jahr beginnen soll, konnten wir uns diesmal besser vorbereiten und insbesondere auch besser mit der Deutschen Mathematiker-Vereinigung (DMV) absprechen. Eine Arbeitsgruppe, bestehend aus den DMV-Vertretern Jens-Peter Kreiß, Hajo Holzmann, Ralf Korn und den DAGStat-Vertretern Katja Ickstadt, Wolfgang Schmid, Peter Schmidt, hat den Antrag erarbeitet, der am 19.04.2017 an das Statistische Bundesamt versendet wurde und in diesem Bulletin kurz vorgestellt wird. An dieser Stelle möchte ich mich bei diesen Vertretern und insbesondere bei Jens-Peter Kreiß. der den Vorsitz dieser Arbeitsgruppe übernommen hatte, ganz herzlich für ihre Arbeit bedanken.

Was gibt es noch in diesem Bulletin? Am 04.04.2017 fand das DAGStat-Symposium mit dem Thema "Ist Bildung messbar?" statt, das erstmalig in der Berliner Urania durchgeführt wurde. Lesen Sie dazu den Bericht von Christin Jahns, die im letzten halben Jahr für die Öffentlichkeitsarbeit der DAGStat tätig war. Weitere Veranstaltungen mit Statistik-Bezug im letzten Halbjahr waren der Dortmunder Tag der Statistik und die Big Data Challenge in Bremen. Die im letzten Bulletin vorgestellte Befragung von Geflüchteten ist mittlerweile ausgewertet worden, und einige dieser Ergebnisse werden in diesem Bulletin nun vorgestellt. Ferner befindet sich in der Rubrik "Statistik Aktuell" unter dem Titel "Verliert Deutschland auf dem Zukunftsmarkt den Anschluss?" eine Beschreibung einer Studie, die Studierende der TU Dortmund 2016 zu Statistik-Publikationen in hochrangigen internationalen Statistik-Fachzeitschriften durchgeführt haben. Dabei stellte sich heraus, dass Deutschland im Vergleich von 27 Ländern nur mittelmäßig abschneidet.

Ich wünsche Ihnen nun viel Spaß beim Lesen.

Ihre Christine Müller Vorsitzende der DAGStat

# Die IAB-BAMF-SOEP-Befragung von Geflüchteten

von Martin Kroh, Simon Kühne, Manuel Siegert

#### Studiendesign

Die IAB-BAMF-SOEP-Befragung von Geflüchteten ist als jährliche Längsschnittstudie angelegt, in der 2016 erstmals über 4.800 erwachsene Befragte in über 3.500 Haushalten im Rahmen von persönlichen Interviews zu ihren Fluchtwegen und -motiven, Qualifikationen, ihrer Lebenssituation in Deutschland, Gesundheit sowie auch ihren Werteorientierungen befragt wurden. Darüber hinaus werden ab der 3. Welle, die im Jahr 2018 erhoben wird, Informationen zu über 5.700 in den Haushalten bzw. Familien lebenden Kindern zur Verfügung stehen.

Stichprobenbasis bildete das Ausländerzentralregister (AZR), und die Zielpopulation sind Personen, die vom 1. Januar 2013 bis zum 31. Januar 2016 in Deutschland eingereist sind und einen Asylantrag gestellt haben. Das Stichprobendesign basiert auf einer mehrstufig disproportional geschichteten Klumpenstichprobe.

Interviews wurden in Privathaushalten und Sammelunterkünften, aber auch in Erstaufnahmeeinrichtungen durchgeführt. Dabei standen den Befragten sieben Sprachversionen sowohl schriftlich sowie auch als Audiodateien (jeweils über den in den Interviews eingesetzten Laptop) zur Verfügung. Die Ausschöpfungsquote liegt bei knapp 50%. Die Befragungsdaten werden bei Zustimmung der TeilnehmerInnen mit den registerbasierten Daten der Integrierten Erwerbsbiografien (IEB) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) verknüpft.

Tabelle 1 stellt die Zusammensetzung nach Herkunft, Alter und Geschlecht sowohl im Ausländerzentralregister als auch in der realisierten Stichprobe dar. Mit knapp 49% bilden Syrer die größte Herkunftsgruppe in der Stichprobe. Knapp 58% der befragten Erwachsenen war zum Zeitpunkt der Stichprobenziehung bis unter 30 Jahre alt. 61% der erwachsenen Geflüchteten in der Stichprobe sind männlich.

#### **Datenzugang**

Interessierten Forscherinnen und Forschern werden die Daten, unter Berücksichtigung des Datenschutzes, durch die Forschungsdatenzentren (FDZ) des SOEP sowie des IAB ab Herbst 2017 zur Verfügung gestellt. Eine Nutzung der mit den IEB verknüpften Daten für Forschungszwecke wird aus Datenschutzgründen nur am FDZ der Bundes-

## Statistik Aktuell

agentur für Arbeit im IAB im Rahmen eines Gastaufenthalts mit anschließender Datenfernverarbeitung voraussichtlich ab 2018 möglich sein. Auch besteht für grundlagenorientierte ForscherInnen die Möglichkeit, Einfluss auf die Befragungsinhalte der 3. Befragungswelle im Jahr 2018 zu nehmen. Entsprechende Vorschläge sind dem SOEP Survey Team bis zum 1. August 2017 kurz vorzustellen.

#### **Erste Ergebnisse**

Die folgende Liste fasst erste vorläufige Ergebnisse zusammen, die in detaillierter Form in Brücker et al. 2016 veröffentlicht wurden:

- Eine Mehrheit der Geflüchteten gibt Angst vor Krieg und Gewalt (70%), Verfolgung (44%) und Zwangsrekrutierung (36%) als wichtigste Fluchtursachen an.
- 58 Prozent der erwachsenen Geflüchteten haben zehn Schuljahre und mehr in Schulen und Hochschulen sowie beruflicher Bildung verbracht, im Vergleich zu 88 Prozent der deutschen Wohnbevölkerung.
- 90 Prozent der Geflüchteten hatten beim Zuzug keine Deutschkenntnisse. Unter den Befragten,

die sich zwei Jahre und weniger in Deutschland aufhielten, gaben 18 Prozent an, zum Zeitpunkt der Befragung über gute oder sehr gute Deutschkenntnisse zu verfügen, 35 Prozent über mittlere und 47 Prozent über schlechte.

- Rund 90 Prozent der Befragten wollen dauerhaft in Deutschland bleiben.
- 73 Prozent der befragten Geflüchteten im Alter von 18 bis 65 Jahren haben vor dem Zuzug nach Deutschland Berufserfahrungen gesammelt, im Durchschnitt 6,4 Jahre.
- Die Geflüchteten gaben an, dass sie im Mittel seit ihrer Ankunft drei neue Kontakte zu Deutschen und fünf neue Kontakte zu Personen aus den Herkunftsländern geknüpft haben.

#### Publikationen:

Brücker, Herbert; Rother, Nina; Schupp, Jürgen (Hrsg.), 2016: IAB-BAMF-SOEP-Befragung von Geflüchteten: Überblick und erste Ergebnisse. Politikberatung kompakt 116. DIW Berlin.

Tabelle 1
Anzahl erwachsener und minderjähriger Geflüchteter nach Herkunftsland, Geschlecht und Alter

	Erwachsene					Mindoriährigo	
	Männer		Frauen			Minderjährige	
	Bis 30 Jahre	Über 30 Jahre	Bis 30 Jahre	Über 30 Jahre	Sub-Total		Total (in %)
Herkunftsland							
Syrien	96.608	66.638	26.282	30.145	219.673	81 076	40,9
	(591)	(871)	(332)	(551)	(2 345)	(2 853)	(49,3)
Afghanistan	28.600	9.482	7.607	6.020	51.709	31 065	11,3
	(182)	(183)	(128)	(124)	(617)	(847)	(13,9)
Irak	20.776	10.569	6.866	5.927	44.138	19 882	8,7
	(187)	(200)	(93)	(142)	(622)	(789)	(13,4)
Albanien,	8.084	9.234	6.264	7.522	31.104	21 827	7,2
Serbien, Kosovo	(35)	(70)	(42)	(55)	(202)	(342)	(5,2)
Eritrea, Somalia	21.918	4.862	6.585	1.669	35.034	5 614	5,5
	(160)	(47)	(75)	(35)	(317)	(134)	(4,3)
Iran, Pakistan	11.252	7.923	1.859	3.031	24.065	9 143	4,5
	(44)	(57)	(11)	(29)	(141)	(65)	(2,0)
Sonstige	50.548	38.223	14.943	19.641	123.355	37 325	21,9
	(129)	(182)	(83)	(179)	(573)	(687)	(12,0)
Total	237.786	146.931	70.406	73.955	529.078	205 932	735 010
	(1 328)	(1 610)	(764)	(1 115)	(4 817)	(5 717)	(10 534)

Quelle: AZR-Sonderauszählung zum Stand 30. Juni 2016 und in der IAB-BAMF-SOEP-Befragung von Geflüchteten zum Stand September 2016 (Fallzahlen in der Befragung in Klammern)

## Verliert Deutschland auf dem Zukunftsmarkt den Anschluss?

# Studierende der TU Dortmund decken Nachholbedarf im Bereich Statistik auf

von Julian Riehl, Carsten Tegethoff, Christine H. Müller

Kompetenz im Umgang mit großen Datenmengen entwickelt sich im Zeitalter der Digitalisierung immer mehr zu einem entscheidenden Wettbewerbsfaktor. Im Rahmen einer Projektarbeit, deren wesentliche Ergebnisse jetzt im Wirtschafts- und Sozialstatistischen Archiv von AStA (link.springer. com/article/10.1007/s11943-017-0201-0) erschienen sind, untersuchten Studierende der Fakultät Statistik an der Technischen Universität Dortmund den Stellenwert von Statistik und Datenanalyse in Deutschland – und kamen dabei zu einem ernüchternden Ergebnis.

Für ihre Analyse werteten die Studierenden insgesamt fast 3700 Artikel hochrangiger internationaler Statistik-Fachzeitschriften der Jahre 2010 bis 2016 aus. Die Annahme: Aus Ländern, in denen Datenwissenschaften eine hohe Bedeutung beigemessen wird, stammen tendenziell mehr Publikationen. Um Vergleichbarkeit herzustellen, wurde nicht die absolute Anzahl an Veröffentlichungen, sondern die Veröffentlichungszahl pro Universität und pro eine Million Einwohner eines Landes betrachtet. Das erstaunliche Ergebnis: Der Stellenwert der Statistik in Deutschland, immerhin Heimat großer Mathematiker wie Gauß und Leibniz, kann bestenfalls als durchschnittlich angesehen werden. Deutschland befindet sich unter den betrachteten 27 Ländern mit einem Wert von 0,33 Publikationen pro Universität auf dem 14. Platz. Gerademal 1,7 Veröffentlichungen pro eine Million Einwohner entfielen in den untersuchten Daten auf deutsche Autoren und das trotz eines hohen Bruttoinlandsprodukts und Human Development Index, die laut der Studie positiv mit der normierten Veröffentlichungshäufigkeit zusammenhängen.

Ein Blick zum kleinen Nachbarn im Süden zeigt, wie es besser geht: So wurde in der Schweiz mit über 7 Veröffentlichungen pro eine Million Einwohner mehr als vier Mal so häufig publiziert. Auch Belgien, die Niederlande und die skandinavischen Länder sind Deutschland in dieser Hinsicht noch ein gutes Stück voraus. Weiterhin auffällig: Im englischen Sprachraum (Großbritannien, Nordamerika, Australien, Neuseeland) wird ebenfalls durchweg deutlich häufiger zu datenwissenschaftlichen Themen publiziert.

Doch warum ist die Kompetenz in den Datenwissenschaften so wichtig? Der Grund liegt auf der Hand: Im Wirtschaftsleben sind große Datenmengen heute bereits allgegenwärtig. "Industrie 4.0" und "Big Data" sind die Schlagworte der Gegenwart und Zukunft. Man betrachte nur beispielsweise die Automobilbranche, einen der wichtigsten Arbeitgeber in Deutschland: Die Datenflut, welche in modernen Autos anfällt, ist gigantisch. Zur Steuerung moderner Fahrassistenzsysteme werden über Sensoren permanent Ultraschall-, Radaroder Bild- signale erfasst und in Echtzeit ausgewertet. So ist es kein Wunder, dass neben etablierten Automobilherstellern nun auch Konzerne wie Google oder Apple auf den Markt dringen, deren Fähigkeiten im Umgang mit Daten hinlänglich bekannt sind.

Um also im internationalen Wettbewerb nicht ins Hintertreffen zu geraten, sollte Deutschland dringend in sein datenanalytisches Know-how investieren. Wie genau, darauf haben die Studierenden ebenfalls eine Antwort: So stellte sich bei den Untersuchungen heraus, dass vor allem in denjenigen Ländern verstärkt zu datenwissenschaftlichen Themen publiziert wurde, in denen es besonders viele Universitäten mit statistischem Fachbereich gibt. In Deutschland können nicht einmal fünf Prozent aller Hochschulen einen solchen Fachbereich vorweisen. Zum Vergleich: In den USA sind es acht, in Großbritannien sogar zwölf Prozent. Die Schaffung statistischer Fachbereiche kann sich also lohnen, damit Deutschland nicht den Anschluss verliert.

# Wozu dienen die DFG-Leitlinien im Umgang mit Forschungsdaten?

# Antworten von Dr. Johannes Fournier von der DFG auf Fragen der DAGStat

Dr. Johannes Fournier ist Programmdirektor der Gruppe für Wissenschaftliche Literaturversorgungs- und Informationssysteme der Deutschen Forschungsgemeinschaft (DFG). Dort beschäftigt er sich seit 2003 mit Informationsinfrastrukturen und Open Access und ist darüber hinaus verantwortlich für Finanzierungsprogramme elektronischer Publikationen. Dr. Fournier beantwortet hier Fragen der DAGStat zu den neu aufgestellten DFGLeitlinien zum Umgang mit Forschungsdaten.

**DAGStat:** Welches Ziel verfolgt die DFG mit der Verabschiedung der Leitlinien?

Fournier: Forschungsdaten sind ein zentrales Element für die Qualitätssicherung wissenschaftlicher Arbeiten; zugleich kann sich von Fachgebiet zu Fachgebiet stark unterscheiden, was unter diesen Begriff gefasst wird. Mit der Verabschiedung der Leitlinien zielte die DFG daher wesentlich darauf ab, dass eine fachlich angemessene Auseinandersetzung darüber stimuliert wird, ob und in welcher Form die entstehenden Forschungsdaten für die Nachnutzung durch andere herangezogen und welche Mindeststandards an das Forschungsdaten-Management gestellt werden können. Dadurch soll eine Sensibilisierung erreicht werden, insbesondere in den Fachgebieten, die sich bislang noch nicht intensiv mit dem Thema befasst haben. Der Erfolg dieses Ansatzes lässt sich daran erkennen, dass verschiedene Fachbereiche den mit der DFG-Leitlinie abgesteckten Rahmen bereits für eine fachspezifische Ausgestaltung genutzt haben. Im unteren Teil der DFG-Informationsseite zu den Leitlinien finden Sie solche fachspezifischen Regelungen: http://www.dfg.de/foerderung/antragstellung\_begutachtung entscheidung/antragstellende/antragstellung/nachnutzung forschungsdaten/index.html

**DAGStat:** Welchen Nutzen erhofft man sich konkret von der neuen Regelung?

Fournier: Ich möchte zunächst darauf hinweisen, dass die DFG-Leitlinien im Grunde keine neue Regelung darstellen. Vielmehr fassen die Leitlinien Bestimmungen und Richtlinien, die an unterschiedlichen Stellen des DFG-Regelwerks – dem Leitfaden für die Antragstellung, den Verwendungsrichtlinien, der Guten wissenschaftlichen Praxis – bereits vorhanden waren, in einem einzigen Dokument in aller Kürze konsistent zusammen. Das erleichtert es Antragstellerinnen und Antragstellern,

die Erwartungen der DFG an ein professionelles Datenmanagement zu kennen – und gibt diesen Personen zugleich rasch zu erkennen, wie die DFG sie beim Aufsetzen eines professionellen Datenmanagements unterstützen kann. Eine DFG-Leitlinie für dieses komplexe Thema kann jedoch nur einen übergeordneten Rahmen abstecken, der naturgemäß fachlich weiter spezifiziert werden muss.

**DAGStat:** Worin soll der wesentliche Unterschied zum bisherigen Umgang mit Forschungsdaten bestehen?

Fournier: Wie gesagt stellen die DFG-Leitlinien keine grundsätzlich neue Regelung dar, aber - und das zeigt auch das Interesse der Deutschen Arbeitsgemeinschaft Statistik am Thema - das Thema des Forschungsdatenmanagements im DFG-Förderkontext gewinnt deutlich an Aufmerksamkeit. Wichtig ist aus unserer Sicht, dass die DFG-Leitlinien zum Umgang mit Forschungsdaten nicht allein auf den Aspekt der Qualitätssicherung und der Überprüfung vorhandener Forschungsergebnisse rekurrieren. Darüber hinaus wird nämlich betont, dass ein professionelles Forschungsdatenmanagement wesentliche Voraussetzung für die anschließende Verwendung der Daten Forschungsprojekten Dritter ist. Dem entsprechend sollte in Anträgen an die DFG ausgeführt werden, welche Art von Forschungsdaten im Projektverlauf entstehen, wie diese langfristig gesichert werden, ob diese Daten in anderen Forschungskontexten relevant sein können und aus diesem Grund anderen Wissenschaftlern und Wissenschaftlerinnen verfügbar gemacht werden sollten.

**DAGStat:** Soll die Datenbereitstellung zentral oder dezentral erfolgen?

Fournier: Diese Frage muss in jedem einzelnen Fall entschieden werden mit Blick auf die Art und das Volumen der Daten, die generiert oder erhoben werden, und ebenso mit Blick auf die vorhandenen und für ein bestimmtes Fachgebiet geeigneten Infrastrukturen. Deshalb ist es richtig, dass die DFG-Leitlinien keine dezidierten Vorgaben zu diesem Punkt machen. Empfehlungen etwa von Fachgesellschaften können Antragstellern und Gutachtern jedoch wertvolle Hinweise dazu bieten, welche Optionen für eine Datenhaltung bestehen.

**DAGStat:** Welche Infrastruktur könnte die DFG zur Verfügung stellen?

**Fournier:** Die DFG selbst stellt zwar keine Infrastruktur zur Verfügung, bietet jedoch mit dem Programm "Informationsinfrastrukturen für Forschungs- daten" ein Förderinstrument an, das genutzt werden kann, um fächerspezifische Infra-

## **DFG** und DeStatis

strukturen wie z.B. geeignete Forschungsdaten-Repositorien aufzubauen oder weiter zu entwickeln. Als ein Beispiel kann hier GFBio genannt werden, welches die zuvor verteilte Infrastrukturlandschaft der Biodiversitätswissenschaften gebündelt hat, ohne jedoch alle Aspekte der Infrastrukturentwicklung in diesem Feld zu zentralisieren. Über die gerade erwähnte Förderschiene wurde auch das Verzeichnis re3data (http://www.re3data.org/) aufgebaut, das bei der Suche nach einschlägigen Datenrepositorien nützliche Dienste leisten sollte.

**DAGStat:** An welches Format ist gedacht?

Fournier: Die Vielfalt der Formate dürfte fast größer sein als die Vielfalt der Typen von Forschungsdaten – von Messdaten und Simulationen über Surveys bis hin zu Texten, Tönen und Bildern. Auch diese Frage muss daher mit Blick auf die Datentypen und disziplinäre Gegebenheiten beantwortet werden. Generell ist es jedoch wichtig, sowohl für die Forschungsdaten als auch für die beschreibenden Metadaten Formate zu wählen, mit denen eine digitale Langzeitverfügbarkeit der Forschungsdaten gewährleistet wird, wie z.B. mit der Verwendung des TIFF-Formats für Bilder oder von XML für Texte.

**DAGStat:** Wie sollten das Datenmanagement und die Schnittstellen geregelt werden?

Fournier: Wir wollen und können auch hierzu keine spezifischen Vorgaben machen. Unsere Erwartung wäre vielmehr, dass in Anträgen an die DFG spezifiziert wird, über welche Prozesse und Verfahren sichergestellt wird, dass Daten qualitätsgesichert erhoben, kuratiert und ggf. auch für Anschlussnutzungen verfügbar gemacht werden. Ob die in einem Antrag skizzierten Planungen und die für das Datenmanagement beantragten Ressourcen überzeugen – nicht zuletzt im Hinblick auf die in einem bestimmten Fachgebiet etablierten Standards –, muss im üblichen Weg von Begutachtung und Bewertung geklärt werden.

**DAGStat:** Wie soll der Besitz der Daten und der Zugang zu den Daten geregelt werden?

Fournier: Diese Frage führt uns mitten hinein in die aktuell intensive Diskussion darüber, ob es überhaupt ein Eigentum an Daten als immaterielle Rechtsgüter gibt; gegen Ende 2016 wurde dazu eine im Auftrag der EU-Kommission verfasste umfassende "Legal Study on Ownership and Access to Data" vorgelegt. Für den DFG-Kontext scheint mir der Hinweis auf übliche und eingebürgerte Forschungspraktiken wichtig zu sein. So ist es in der Grundlagenforschung üblich, dass Wissenschaftler

als Datenproduzenten selbst darüber bestimmen, wem sie Zugang zu den Daten ermöglichen - z.B. auch dadurch, dass sie ihre Daten bewusst einem Forschungsdaten-Repositorium anvertrauen, das von der Community intensiv genutzt wird. Und es bürgert sich allmählich ein, dass wissenschaftliche Zeitschriften eigene "Data Policies" formulieren und ihre Autoren auffordern, die einem Artikel zugrunde liegenden Datensätze zusammen mit dem Artikel in einer Form zur Verfügung stellen, die ein Überprüfen und Nachvollziehen der publizierten Forschungsergebnisse ermöglicht. Derartige Praktiken dürften den Umgang mit Forschungsdaten in einer Community stärker beeinflussen als juristische Befunde darüber, "wem die Daten gehören". Bei personenbezogenen Daten müssen ohnehin die Erfordernisse des Datenschutzes berücksichtigt werden, der die Möglichkeiten einer wissenschaftlichen Nutzung und des Zugangs zu diesen Daten

**DAGStat:** Wie sollten Daten aus Firmen integriert werden?

Fournier: Empfehlungen und gar Förderrichtlinien der DFG können sich naturgemäß nur auf diejenigen Forschungsdaten beziehen, die aus DFG-geförderten Projekten resultieren. Es wäre gewiss wünschenswert, wenn Wissenschaftler und Wissenschaftlerinnen auch auf Daten aus Wirtschaft und Industrie rekurrieren und damit arbeiten könnten. Die Erfüllung dieses Wunsches dürfte jedoch oft daran scheitern, dass kommerzielle Interessen eine Nachnutzung von Daten erschweren oder unmöglich machen. Gleichwohl werden Fachgebiete, die intensive Schnittflächen mit kommerziellen Bereichen aufweisen, diese spezifischen Erfordernisse in ihre Diskussion über den Umgang mit Forschungsdaten bzw. über die gemeinschaftliche Nutzung von Daten integrieren. Das wäre eine grundlegende Voraussetzung dafür, Forschungsdaten aus beiden Sphären ohne Reibungsverluste nutzen zu können.

**DAGStat:** In welchem Kontext steht die DFG-Initiative zu dem RatSWD (GESIS, FDZ) und der BMBF-Initiative zur Datenvernetzung der Universitätsklinika?

Fournier: Zunächst einmal haben wir es hier mit unterschiedlichen Strukturen und Initiativen zu tun, die von unterschiedlichen Akteuren eigenständig verantwortet werden. Doch selbstverständlich befindet sich die DFG im Austausch mit wichtigen Akteuren und ist z.B. ständiger Gast im Rat für Sozialund Wirtschaftsdaten, nimmt ebenfalls als Gast an den vom BMBF organisierten Begutachtungen im Rahmen der Initiative Medizininformatik teil oder berät sich mit den Kolleginnen und Kollegen im

### DFG und DeStatis

BMBF über die Herausforderungen der klinischen Datenvernetzung. Der gemeinsame Kontext nicht nur für die mit dieser Frage direkt angesprochenen, sondern für viele weitere Initiativen zum Forschungsdaten-Management ist die umfassende wissenschaftspolitische Diskussion über eine effektive und effiziente Nutzung von Forschungsdaten. Diese Diskussion beschäftigt Forscher und Forscherinnen, wissenschaftliche Einrichtungen, Infrastruktureinrichtungen und die Politik - sei es auf nationaler Ebene, sei es auf der europäischen oder der internationalen Ebene, was mit den Stichwörtern "Rat für Informationsinfrastrukturen", "Research Data Alliance", "European Open Science Cloud" oder "The Federal Big Data Research and Development Strategic Plan" hier lediglich angedeutet werden soll. Dass sich die DFG in diesem Feld positioniert und auf die Bedeutung von Forschungsdaten für die Qualitätssicherung wissenschaftlicher Ergebnisse hinweist - und nicht die wirtschaftliche Verwertbarkeit der Daten betont, die gerade von der Politik als "der Rohstoff des 21. Jahrhunderts" vermarktet werden -, entspricht der Verantwortung der DFG für die erkenntnisgeleitete Grundlagenforschung.

DAGStat: Gibt es schon Erfahrungen?

Fournier: Wir beobachten, dass manche Fächer die Verabschiedung der DFG-Leitlinien zum Anlass genommen haben, in ihren Gesellschaften über Fragen des professionellen Umgangs mit Forschungsdaten zu diskutieren und sich mit möglichen Regelungen dafür zu befassen. Auch einige Fachkollegien der DFG haben sich intensiv mit der Thematik beschäftigt und für die Antragsteller und Antragstellerinnen aus ihren Bereichen einschlägige Empfehlungen und Hinweise verfasst. Diese werden auch über die Homepage der DFG publiziert, wo sie derzeit Ausführungen zum Umgang mit Forschungsdaten in der Biodiversitätsforschung, in der Bildungswissenschaft sowie in der Literatur- und Sprachwissenschaft finden. Eher anekdotisch kann ich berichten, dass vermehrt Telefonate und E-Mails mit Fragen dazu auflaufen, wie Wissenschaftler sich in ihren Anträgen zu Fragen des Forschungsdaten-Managements äußern sol-(Weitere Informationen finden sich unlen. ter: http://www.dfg.de/foerderung/ antragstellung\_begutachtung\_entscheidung/antragstellende/antragstellung/nachnutzung forschungsdaten/index.html)

**DAGStat:** An welchen Zeitrahmen ist gedacht?

**Fournier:** Die DFG-Leitlinien sind keine "Roadmap" und wir haben keine Zielvorgaben, die bis zu einem bestimmten Datum erfüllt sein müssten. Wir sind uns sehr bewusst, dass sich der Umgang mit

Forschungsdaten entwickelt. Im Förderhandeln der DFG gehört zu diesem Entwicklungsprozess, dass Antragsteller sich ebenso wie Gutachter und Mitglieder der Fachkollegien mit Fragen des Forschungsdaten-Managements auseinandersetzen. Die daraus resultierenden Wechselwirkungen im Prozess von Antragstellung, Begutachtung und Bewertung werden zunehmend dafür sensibilisieren, was wichtig ist und umgesetzt werden sollte.

**DAGStat:** Das Sammeln von Daten ist von sich aus kein Wert, sondern nur der sachgerechte Umgang damit kann wertvolle Erkenntnisse schaffen. Will die DFG in Zukunft auch mehr die statistische Analyse solcher Daten und die Forschung über statistische Analysemöglichkeiten unterstützen?

Fournier: Der sachgerechte Umgang mit Forschungsdaten ist tatsächlich stets voraussetzungsreich. Auch das erklärt die immer öfter vorgetragene Forderung nach einer professionellen Ausbildung und nach geeigneten Karrierewegen für "data scientists" - wobei klar sein sollte, dass die Bezeichnung "data scientists" als Chiffre für eine Vielzahl ganz unterschiedlicher Berufe und Kompetenzen steht, zu denen ganz gewiss auch statisti-Methodenkompetenz gehört. Ausbildung ist jedoch nicht Sache der DFG, hier sind in erster Linie unsere Universitäten gefragt, die sich ohnehin intensiv mit Fragen des Forschungsdaten-Managements befassen. In der Förderung leiten sich die Bedarfe für fundierte und sorgfältige statistische Analyse von den einzelnen Forschungsfragen ab. Wo statistische Expertise für die erfolgreiche Durchführung eines Forschungsvorhabens erforderlich ist, spricht also viel dafür, die nötigen Ressourcen in einem Antrag an die DFG zu begründen. So spielt die Nutzung ebenso wie die Weiterentwicklung statistischer Methoden in Anträgen aus der Medizininformatik, der Bioinformatik oder den empirischen Sozialwissenschaften eine Rolle. Und mit dem Programm e-Research-Technologie ist ein weiteres Förderangebot etabliert, in dem auch Werkzeuge zur statistischen Analyse entwickelt werden können. Die DFG ist in diesem Bereich also bereits aktiv und wird es bleiben

### **DFG-Initiative zu Open Access**

von Christine Müller

Am 19.01.2017 fand ein Rundgespräch über Open Access bei der DFG statt, bei dem 16 Vertreter von Fachgesellschaften bzw. Dachverbänden Impulsreferate von 7 Minuten zum Stand von Open Access in ihrem Fach hielten und dann darüber diskutiert wurde. Ich habe dabei vor allem von den Ergebnissen des Surveys der DGEpi, IBS-DR und Deutschen Gesellschaft für Sozialmedizin und Prävention zu Open Access berichtet, aber auch die neue Open Access Zeitschrift der GfKI "Archives of Data Science" erwähnt und von meinen Erfahrungen als Koordinierende Herausgeberin von Statistical Papers berichtet. Die Auswertung des Surveys finden Sie unter http://doi.org/10.5281/zenodo.250528, und eine Handreichung mit Antworten zu den am häufigsten benannten Fragen ist unter http://doi.org/10.5281/zenodo.250527 hinterlegt.

Bei den Impulsreferaten und Diskussionen zeigte sich deutlich, wie unterschiedlich die Fächerkulturen sind. In den Wirtschaftswissenschaften, der Psychologie und den Politikwissenschaften haben wie in vielen Gebieten der Statistik Zeitschriften das höchste Ansehen, die in den USA oder Großbritannien herausgegeben werden und nicht Open Access sind. Dagegen haben Open-Access-Zeitschriften kein hohes Ansehen, was auch auf Zeitschriften zutrifft, die von deutschen Fachgesellschaften herausgegeben werden. Allerdings ist in diesen Fächern wie in der Mathematik das Vorveröffentlichen über Preprints weit verbreitet.

In den Geistes- und Kulturwissenschaften ist die Orientierung an Zeitschriften aus dem angelsächsischen Raum nicht so ausgeprägt, aber Open Access ist dort auch nicht etabliert. Insbesondere wird dort die Gefahr gesehen, dass die Gesellschaften Mitglieder verlieren könnten, wenn sie ihre Zeitschrift allen frei zugänglich machen würden.

In der Chemie gibt es mehrere in Deutschland herausgegebene Zeitschriften, die international ein hohes Ansehen besitzen. Da diese aber zu einem großen Teil über Subskriptionen von großen Chemiekonzernen finanziert werden, wird Open Access dort sehr kritisch gesehen.

Ganz anders sieht es in der recht jungen Gesellschaft für Sprachtechnologie und Computerlinguistik aus, die wie die ausländischen Gesellschaften auf diesem Gebiet nur Open-Access-Zeitschriften herausgeben. Diese werden über die Mitgliedsbeiträge bezahlt, aber auch von Google und Co unterstützt. Außerdem entstehen für die Zeitschriften nicht so hohe Kosten, da Editorial-Management-

Systeme selber entwickelt werden. Die Mitgliedschaft in diesen Gesellschaften lohnt wegen reduzierter Beiträge für Konferenzen, die in diesem Gebiet von besonderer Bedeutung sind.

Generell wurde beim Übergang von Subskription zu Open Access das Problem der Doppelfinanzierung von Verlagen gesehen, wenn bei Subskriptionszeitschriften Open-Access-Artikel zusätzlich finanziert werden. Hinzu kommt, dass Verlage im Rahmen der Gewinnmaximierung dazu tendieren bei Open-Access-Zeitschriften mit hohen Impact-Faktoren horrende Gebühren zu nehmen, die sich dann nur sehr wenige leisten können. Damit Gesellschaften aber zu geringen Gebühren Open-Access-Zeitschriften in Eigenregie herausgeben können, werden freie Editorial-Management-Systeme gebraucht.

Dieses Rundgespräch ist Teil der DFG-Initiative zu Open Access, zu der es auch eine ad-hoc-AG der DFG gibt. Im Rahmen dieser Initiative erhalten neue Open-Access-Zeitschriften von der DFG auch eine Anschubfinanzierung. Ausführliche Informationen zu dieser DFG-Initiative finden sich unter: http://www.dfg.de/foerderung/programme/infrastruktur/lis/open access/

# Revision der Fächerklassifikation beim Statistischen Bundesamt

von Christine Müller

Die Revision der Fächersystematiken des Statistischen Bundesamtes soll künftig in einem fünfjährigen Turnus stattfinden. Dazu will der Ausschuss für die Hochschulstatistik im November 2017 die Arbeitsgruppe "Revision der Fächersystematiken" (AG-RF) einsetzen. Diese soll Empfehlungen zur erneuten Revision der Fächerklassifikationen erarbeiten und dem Ausschuss 2019 vorlegen. Nachdem der sehr eilig verfasste Antrag der DAG-Stat aus dem Jahr 2014 abgelehnt worden war, hat ietzt in Zusammenarbeit mit der Deutschen Mathematiker-Vereinigung (DMV) eine Arbeitsgruppe, bestehend aus den DMV-Vertretern Jens-Peter Kreiß, Hajo Holzmann, Ralf Korn und den DAGStat-Vertretern Katja Ickstadt, Wolfgang Schmid, Peter Schmidt, einen neuen Antrag erarbeitet. Dieser wurde am 19.04.2017 an das Statistische Bundesamt gesandt; die einzurichtende Arbeitsgruppe zur Revision der Fächersystematik will sich damit dann befassen.

Was beinhaltet dieser Antrag bezüglich der Statistik? Im Wesentlichen beinhaltet er den Unterpunkt

"Statistik"

im Lehr- und Forschungsbereich "330 Mathematik, Naturwissenschaften allgemein" der Personalstatistik und

im Studienbereich "36 Mathematik, Naturwissenschaften allgemein" der Studierendenstatistik sowie den Unterpunkt

"Datenwissenschaft und Statistik"

im Lehr- und Forschungsbereich "670 Ingenieurwissenschaften allgemein" der Personalstatistik und im Studienbereich "61 Ingenieurwissenschaften allgemein" der Studierendenstatistik. Daneben wird ein neues Fachgebiet

"Statistik und Erhebungsmethoden"

im Lehr- und Forschungsbereich "220 Rechts-, Wirtschafts- und Sozialwissenschaften allgemein" und im Studienbereich "23 Rechts-, Wirtschaftsund Sozialwissenschaften allgemein" sowie ein neues Fachgebiet

"Methoden und Evaluation"

im Lehr- und Forschungsbereich "315 Psychologie" beantragt.

Das Fachgebiet "Ökonometrie" soll im Lehr- und Forschungsbereich "290 Wirtschaftswissenschaften" in

"Statistik und Ökonometrie"

umbenannt werden und auch als neues Fach zum Studienbereich "30 Wirtschaftswissenschaften" hinzugefügt werden.

Für den Lehr- und Forschungsbereich "470 Klinisch-Theoretische Humanmedizin (einschl. Zahnmedizin)" werden mehrere Änderungen beantragt. So sollen das Fachgebiet "Medizinische Statistik und Dokumentation" in "Medizinische Biometrie / Statistik" umbenannt und das Fach "Biomathematik (für Mediziner)" gestrichen werden. Im Lehr- und Forschungsbereich "070 Bibliothekswissenschaft, Dokumentation" soll dagegen das neue Fachgebiet "Medizinisches Informationsmanagement / Medizinische Dokumentation" aufgenommen werden. Im Studienbereich "49 Humanmedizin" sollen die Studienfächer

"Medizinische Biometrie / Statistik" sowie "Epidemiologie"

hinzukommen.

Weiterhin soll der Lehr- und Forschungsbereich "340 Mathematik" mehr untergliedert werden und insbesondere die Gebiete

"Finanz- und Versicherungsmathematik" und "Mathematische Stochastik"

enthalten. Das Gebiet "Mathematische Statistik / Wahrscheinlichkeitsrechnung" im Studienbereich "37 Mathematik" soll dagegen gestrichen werden, da es so einen Studiengang nicht gibt.

Es bleibt abzuwarten, was davon umgesetzt wird.

# DAGStat-Symposium 2017 "Ist Bildung messbar?"

von Christin Jahns

Daten und Statistiken aus nationalen und internationalen Bildungsstudien bestimmen nicht zuletzt seit dem PISA-Schock die Entwicklung unseres Bildungssystems. Doch wie sinnvoll sind solche Bildungsstudien und welche Schlüsse lassen sie zu? Um diese Frage drehte sich das 9. Symposium der DAGStat, das am 4. April in den Räumlichkeiten der Urania Berlin stattfand.

Nach der Begrüßung der Gäste und der Einführung in die Thematik des Symposiums durch die Organisatoren Prof. Dr. Christine Müller, Prof. Dr. Heinz Holling und Prof. Dr. Steffi Pohl begann Prof. Dr. Olaf Köller mit seinem Vortrag zur Rolle der Statistik bei Ländervergleichen. Köller ist Professor für Empirische Bildungsforschung an der Christian-Albrechts-Universität zu Kiel, Gründungsdirektor des Instituts zur Qualitätsentwicklung im Bildungswesen in Berlin und als solcher maßgeblich an der Umsetzung relevanter Bildungsstudien beteiligt. In seinem Vortrag widmete sich Köller unter anderem der Frage, ob die bayerischen Schülerinnen und Schüler wirklich besser sind als die Berliner Schülerinnen und Schüler. Das Hamburger Abendblatt hatte ein paar Tage zuvor berichtet, dass Hamburg Bildungsspitzenreiter wäre, wenn man die Schülerinnen und Schüler mit Migrationshintergrund heraus rechnen würde. Dem stimmte Köller zu. Er gab jedoch zu bedenken, dass jedes Bundesland eigene Defizite habe, weshalb man Bildung zwar messen könne, aber immer auch auf die individuelle Entwicklung der Länder schauen sollte anstatt nur auf die Platzierung im Ranking.

Diese These fand auch in dem Artikel "Eine Frage des Rankings", der im Tagesspiegel über das Symposium am 7. April erschien und unter http://www.tagesspiegel.de/wissen/schuelerleistungen-eine-frage-des-rankings/19624170.html einsehbar ist, große Beachtung. Sie blieb jedoch nicht der einzige interessante Ansatz, der während des Symposiums diskutiert wurde.

Es folgte Dr. Matthias von Davier, Distinguished Research Scientist im National Board of Medical Examiners, zuständig für die Auswertung der Daten aus der PISA 2015-Erhebung. In seinem Vortrag erklärte er, wie neue Datenquellen genutzt werden können, um Schülerleistungen besser zu verstehen.

Mit der Verwendung neuer Datenquellen beschäftigt sich auch Prof. Dr. Susanne Rässler, Leiterin

der Methodengruppe im Nationalen Bildungspanel. Sie zeigte auf, welche Chancen und Stolperfallen beim Zusammenfügen von Datensätzen aus unterschiedlichen Studien bestehen.

Doch nicht nur Schülerleistungen, sondern auch Schulqualität werden gemessen. "Die Zeiten, in denen Schulqualität als "gefühlte" oder "behauptete" Qualität der Arbeit von Lehrkräften und Schulleitung definiert werden konnte, sind seit PISA vorbei", sagte Prof. Dr. Hans Anand Pant. Der Geschäftsführer von "Die Deutsche Schulakademie" und ehemaliger wissenschaftlicher Leiter des Instituts zur Qualitätsentwicklung im Bildungswesen sprach über Vergleichsarbeiten, Schulinspektionen, zentrale Abschlussprüfungen und interne Evaluationen und verglich sie im Hinblick auf ihre Aussagekraft für Schulqualität.

Bei der anschließenden Gesprächsrunde kamen Experten und Besucher miteinander ins Gespräch und diskutierten lebhaft. Somit diente das Symposium auch in diesem Jahr wieder als Forum für den offenen und produktiven Austausch zwischen Wissenschaft und Gesellschaft.

# Ein Ausflug in die Welt der Zahlen und Daten

# Beim Tag der Statistik wird das Fach auch von der praktischen Seite beleuchtet

von Ann-Christin Herbe, TU Dortmund

Der Geschmacksunterschied von Cola-Sorten und Wahlkampfanalyse haben auf den ersten Blick nicht viel gemeinsam, aber hinter beiden verbirgt sich Statistik. Das haben Schülerinnen und Schüler der Oberstufe am Dienstag, 14.02.2017, beim Tag der Statistik an der TU Dortmund erfahren. Bei Informationsveranstaltungen und Mit-Mach-Aktionen konnten die Schülerinnen und Schüler sich über das Fach informieren. "Statistik wird in der Schule ja leider kaum behandelt. Genau deshalb richten wir diesen Tag aus, vor allem um die zahlreichen Anwendungsmöglichkeiten, zum Beispiel in der Wirtschaft oder Medizin, aufzuzeigen", sagte Prof. Dr. Christine Müller von der Fakultät Statistik.



Mitmach-Aktion "Wie viele Fische sind im Teich" (Foto: Nikolas Golsch, TU Dortmund)

Besonders gut kam der Vortrag von Prof. Philipp Doebler bei den Gästen an. Doebler erklärte, dass man Statistik auch im Bereich der Psychologie einsetzt. Als Beispiel analysierte er, wie bei der US-Wahl die Vorlieben der potenziellen Wähler gesammelt wurden. Statistik wird aber auch im Bereich der Sozialwissenschaften eingesetzt. "Statistiker findet man dort, wo es Daten gibt", so Privatdozent Dr. Guido Knapp. "Die Bereiche, in denen das der Fall ist, sind allerdings sehr vielfältig und reichen von ökonomischen Themen bis hin zur Medizin."

Prof. Joachim Kunert versuchte, das Interesse der Schülerinnen und Schüler auf praktische Weise für die Statistik zu wecken. Mit einem sogenannten Dreieckstest stellte er Nachforschungen zum Thema Geschmacksunterschied bei Cola-Sorten an.

Bei diesem Test werden Becher in einer Dreiecksform vor den Teilnehmern aufgestellt und mit den zwei Cola-Sorten gefüllt. Was sich in welchem Becher befindet, wird zufällig ausgewürfelt. Dann gilt es zu probieren und die richtige Sorte herauszuschmecken. Das Besondere an dem Experiment ist, dass sich viele Teilnehmer durch die Anordnung der Becher verwirren lassen. So sind einige beispielsweise fest überzeugt, der mittlere Becher müsse Cola pur enthalten. "Wenn es mehr als der Hälfte gelingt, den Unterschied zu schmecken, dann kann man von einem richtigen Ergebnis sprechen, das nicht zufällig entstanden ist", erklärte Ku-



Mitmach-Aktion "Von der Stichprobe zur Population: Eine Mitmach-Aktion für Naschkatzen" (Foto: Nikolas Golsch, TU Dortmund)

nert

"Ich habe mich vorher schon für Mathe interessiert und der Vortrag am Morgen hat mich bestätigt", sagte der 16-jährige Schüler Oliver Keller. "Ich bin echt überrascht, in wie vielen Bereichen man etwas mit Statistik anfangen kann und könnte mir gut vorstellen, in Zukunft Statistik zu studieren". Aber nicht alle Schülerinnen und Schüler konnten überzeugt werden. "Mathe und alles was mit Zahlen zu tun hat, war schon immer mein Hassfach", sagte die 17-jährige Anna Beilfuß, die mit ihrem Mathe-LK die Veranstaltung besuchte, "das konnte auch der Tag heute nicht ändern."

## Bremen Big Data Challenge (BBDC)

von Tanja Schultz, Jochen Weiner, Felix Putze @ Cognitive Systems Lab, Fachbereich Mathematik/Informatik der Universität Bremen (http://csl.uni-bremen.de)

Die Bremen Big Data Challenge (BBDC, bbdc.csl.uni-bremen.de) soll Interesse an der Disziplin Data Science wecken. Dieser Wettbewerb für alle Studierenden im Land Bremen wurde im Jahr 2016 vom Cognitive Systems Lab (CSL) an der Universität Bremen initiiert. Jedes Jahr stellt das CSL eine neue Aufgabe und einen umfangreichen Datensatz zur Verfügung, aus dem relevante und nutzbare Informationen gezogen werden müssen, um die Aufgabe zu lösen. Am CSL erforschen und entwickeln wir menschzentrierte Technologien und intelligente Anwendungen auf Basis der Erfassung, Erkennung und Interpretation von Biosignalen aus Sprache, Muskel- und Hirnaktivitäten. Maschinelles Lernen, Data Mining und statistische Modellierung spielen dabei eine große Rolle. Als Forschende und Lehrende an der Universität ist es uns ein besonderes Anliegen, Studierende für dieses faszinierende Thema zu begeistern und sie auf zukunftsträchtige Berufsfelder der Informatik wie "Big Data Science" vorzubereiten. Daher sind die Aufgaben der BBDC so gestellt, dass eine erfolgreiche Teilnahme zwar Programmierkenntnisse aber keine Erfahrung im Bereich Data Science erfordert. Außerdem wird Praxisbezug großgeschrieben: Im Jahr 2016 bestand die Aufgabe beispielsweise darin, die Leistung von Online-Spielern auf Basis vergangener Spiele vorherzusagen. In diesem Jahr sollte die Auslastung der Mensa des Studentenwerks Bremen auf dem Campus der Universität prädiziert werden. Der Datensatz wurde mit Unterstützung des Studentenwerks Bremen zusammengestellt.



Das Foto zeigt alle BBDC 2017 Preisträger (sitzend) und Organisatoren des CSL (stehend).
Foto © Harald Rehling, Uni Bremen

Für den Wettbewerb erhalten die teilnehmenden Teams von bis zu drei Studierenden Zugang zum BBDC-Portal, in dem sie ihre Ergebnisse einreichen, bewerten lassen und die aktuelle Rangliste einsehen können. Am ersten März jeden Jahres werden die Aufgabe und der Datensatz vom CSL ausgegeben. Zur Lösung haben die Teams dann 31 Tage Zeit. Danach werden die besten fünf Teams ermittelt und mit Geldpreisen prämiert. In diesem Jahr wurde das Preisgeld von der Firma "neuland Bremen GmbH" gestiftet.

Die BBDC 2017 hat Simon Stelter als Team "The Great Regression" gewonnen. Ihm gelang es, die Zahl der Mensa-Essensbelege in 5-Minuten-Abschnitten mit einer Abweichung von nur 8,6 Belegen zu bestimmen. Dazu betrachtete er das Verhalten der Mensabesucher: Während des Semesterbetriebes gehen Studierende in der Regel zwischen den Vorlesungen zum Essen, was zu mehrfachen Spitzen in der Belege-Verteilung führt. In den Semesterferien dagegen sind die Verteilungen flacher. Um diese Unterschiede zu modellieren, entwickelte er Musterverteilungen für das Sommer- und Wintersemester sowie für die vorlesungsfreie Zeit. Er normalisierte die Standardabweichungen dieser Verteilungen und berechnete mittels eines Regressionsbaumes einen Skalierungsfaktor für jeden Tag in Abhängigkeit der drei Modelle.

Bislang haben 78 Studierende der Universität und der Fachhochschule Bremen in 46 Teams an der BBDC teilgenommen. Das waren allerdings nicht nur Informatik- und Mathematik-Studierende, auch aus anderen Fachrichtungen gab es rege Beteiligung. Die erfolgreichen Lösungen basierten auf unterschiedlichsten Ansätzen. Während einige Teams komplexe Verfahren des maschinellen Lernens und viel Kontextinformationen einsetzten, konnten andere mit geschickten Vorüberlegungen und Analysemethoden punkten. Als Programmiersprache dominiert derzeit Python, was angesichts verfügbarer Bibliotheken und guter Zugänglichkeit nicht überrascht.

Auch in Zukunft sollen spannende Aufgaben in der BBDC gestellt werden. Interessierte können sich jederzeit unter https://bbdc.uni-bremen.de für den Newsletter registrieren, um keine Neuigkeiten zu verpassen.

Personalia **Impressum** 

#### Personalia

### Aus der Fachgruppe Stochastik der DMV

Erst kürzlich ist Prof. Dr. Hans-Otto Georgii, geb. 15.09.1944, am 16. Mai 2017 verstorben.

### Auszeichnungen, Preisträger, Persönlichkeiten

#### Aus der IBS-DR

Auch in diesem Jahr wird die Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR) wieder hervorragende Nachwuchswissenschaftler auszeichnen. Die Preisverleihungen werden im Rahmen der gemeinsamen Tagung des Central European Networks (CEN) der International Biometric Society und der International Society for Biopharmaceutical Statistics (ISBS) am 30. August 2017 in Wien stattfinden.

Mit dem Bernd-Streitberg-Preis 2017, der für herausragende biometrische Abschlussarbeiten vergeben wird, werden Frau Alexandra Bühler für ihre Bachelorarbeit "A Causal Analysis of Ventilator-Associated Pneumonia in Intensive Care", Herr Frank Weber für seine Bachelorarbeit "Modellierung von rekurrenten Ereignissen in der Ereigniszeitanalyse am Beispiel einer Studie zum rezidivfreien Überleben von Harnblasenkrebspatienten" und Frau Regina Stegherr für ihre Masterarbeit "Joint modelling for left-truncated observational studies: Unmeasured baseline covariates as a consequence of delayed study entry" geehrt.

Zudem werden drei Gustav-Adolf-Lienert Preise für Publikationen junger Nachwuchswissenschaftler vergeben. Der 1. Gustav-Adolf-Lienert Preis 2017 geht an Frau Dr. Hong Sun für ihre Arbeit "Comparing a stratified treatment strategy with the standard treatment in randomized clinical trials" in Statistics in Medicine. Der 2. Gustav-Adolf-Lienert Preis 2017 geht an Frau Dr. Annika Hoyer für ihre Arbeit "Meta-analysis for the comparison of two diagnostic tests to a common gold standard: a generalized linear mixed model approach" in Statistical Methods in Medical Research. Frau Dr. Clara Happ erhält den 3. Gustav-Adolf-Lienert Preis 2017 für ihre Veröffentlichung im Journal of the American Statistical Association (JASA) mit dem Titel "Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains".

### **Impressum**

#### DAGStat - Deutsche Arbeitsgemeinschaft Statistik

Prof. Dr. Christine Müller. Vorsitzende Prof. Dr. Wolfgang Schmid, stellv. Vorsitzender apl. Prof. Dr. Hans Peter Wolf, Schatzmeister

#### Geschäftsstelle:

Deutsche Arbeitsgemeinschaft Statistik Universität Bielefeld | Fakultät für Wirtschaftswissenschaften Lehrstuhl für Statistik und Datenanalyse Postfach 10 01 31 33501 Bielefeld

E-Mail: kontakt@dagstat.de | www.dagstat.de

#### Vertreter der Gesellschaften:

Prof. Dr. Wolfgang Schmid, Deutsche Statistische Gesellschaft

Prof. Dr. Yarema Okhrin, Deutsche Statistische Gesellschaft

Prof. Dr. Tim Friede, Internationale Biometrische Gesellschaft Prof. Dr. Katja Ickstadt, Internationale Biometrische Gesellschaft

Prof. Dr. Michael Neumann, Fachgruppe Stochastik der DMV

Prof. Dr. Hajo Holzmann, Fachgruppe Stochastik der DMV

Prof. Dr. Adalbert F. X. Wilhelm, Gesellschaft für Klassifikation e.V.

Prof. Dr. Hans A. Kestler, Gesellschaft für Klassifikation e.V.

Michael Haußmann, Verband Deutscher Städtestatistiker Hartmut Bömermann, Verband Deutscher Städtestatistiker

Prof. Dr. Harald Binder, Fachbereich Biometrie der Deutschen Gesellschaft für

Medizinische Informatik, Biometrie und Epidemiologie e.V.

Prof. Dr. Rolf Biehler, Verein zur Förderung des schul. Stochastikunterrichts e.V.

Dr. Sigrid Behr, Deutsche Gesellschaft für Epidemiologie e.V.

Prof. Dr. Jörg Breitung, Ökonometrischer Ausschuss des Vereins für Socialpolitik

Prof. Dr. Heinz Holling, Fachgruppe Methoden und Evaluation der DGPs

Prof. Dr. Stefanie Eifler, Sektion Methoden der empirischen Sozialforschung der

Dr. Marco Giesselmann, Sektion Methoden der empirischen Sozialforschung

DGS

Bertram Schäfer, Deutsche Sektion der ENBIS Peter Schmidt, Statistisches Bundesamt Dr. Kilian Seng, Sektion Methoden der DVPW Prof. Dr. Joachim Behnke, Sektion Methoden der DVPW