

# Machbarkeitsstudie: Forschermobilität und Innovation

Daniel Schiller (NIW)

mit fachlicher Beratung durch Stefan Hennemann (Kapitel 5.2)  
unter Mitarbeit von Immo Böke und Miriam Richter-Tokar

---

Studien zum deutschen Innovationssystem

9-2014

---

Niedersächsisches Institut für Wirtschaftsforschung e.V. (NIW), Hannover

Februar 2014

Diese Studie wurde im Auftrag der Expertenkommission Forschung und Innovation (EFI) erstellt. Die Ergebnisse und Interpretationen liegen in der alleinigen Verantwortung der durchführenden Institute. Das BMBF hat auf die Abfassung des Berichts keinen Einfluss genommen.

Studien zum deutschen Innovationssystem

Nr. 9-2014

ISSN 1613-4338

Herausgeber:

Expertenkommission Forschung und Innovation (EFI)

Geschäftsstelle:

c/o Stifterverband für die deutsche Wissenschaft

Pariser Platz 6

10117 Berlin

[www.e-fi.de](http://www.e-fi.de)

Alle Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie die Übersetzung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der EFI oder der Institute reproduziert oder unter Verwendung elektronischer Systeme gespeichert, verarbeitet, vervielfältigt oder verbreitet werden.

#### **Kontakt und weitere Informationen:**

PD Dr. Daniel Schiller

Niedersächsisches Institut für Wirtschaftsforschung (NIW)

Königstraße 53

30175 Hannover

Tel.: +49-511-1233-16-38

Fax: +49-511-1233-16-55

Email: [schiller@niw.de](mailto:schiller@niw.de)

**Inhaltsverzeichnis**

<b>1</b>	<b>Einführung</b> .....	<b>1</b>
<b>2</b>	<b>Methodische und inhaltliche Vorüberlegungen</b> .....	<b>5</b>
2.1	Definitionen.....	5
2.1.1	Forscher.....	5
2.1.2	Mobilität.....	6
2.2	Anforderungen an ein Berichtssystem.....	7
2.2.1	Relevante Inhalte .....	7
2.2.2	Methodische Anforderungen .....	8
<b>3</b>	<b>Deutsche Datensätze und Datenquellen</b> .....	<b>10</b>
3.1	Mikrozensus.....	10
3.2	Sozio-ökonomisches Panel (SOEP).....	12
3.3	Integrierte Erwerbsbiographien (IEB) der Bundesagentur für Arbeit (BA) .....	12
3.4	Ausländerzentralregister des Bundesamtes für Migration und Flüchtlinge (BAMF).....	13
3.5	Panelstudien zu Absolventen und Promovierenden.....	15
3.6	Kerndatensatz Forschung.....	16
3.7	Zwischenfazit .....	18
<b>4</b>	<b>Internationale Datensätzen und Datenquellen</b> .....	<b>19</b>
4.1	Internationale Daten .....	19
4.1.1	Publikationsdaten.....	19
4.1.2	Patentdaten.....	21
4.1.3	Forscherdatenbanken .....	21
4.1.4	Professionelle soziale Netzwerke .....	23
4.1.5	Natural Language Processing (NLP) .....	24
4.2	Nationale Daten der Hauptzielländer deutscher Forschermobilität.....	25
4.2.1	Arbeitskräfteerhebungen anderer Länder .....	25
4.2.2	Daten von Einwanderungsbehörden .....	26
4.2.3	Studien zu Absolventen und Promovierten .....	26
4.2.4	Nationale Forscherregister, Forscher- und Lebenslaufdatenbanken .....	28
4.3	Zwischenfazit .....	29
<b>5</b>	<b>Ausführliche Darstellung von zwei vielversprechenden Forschungsszenarien</b> .....	<b>30</b>
5.1	International vergleichende Auswertung von Zensusdaten .....	30
5.2	Generierung eines Datensatzes durch Natural Language Processing (NLP) .....	35
<b>6</b>	<b>Auswirkungen und Determinanten der Forschermobilität im Innovationssystem</b> .....	<b>40</b>
6.1	Individuelle Karriereentwicklung .....	40
6.2	Organisationale Innovationsleistung und Wissensspillover .....	41
6.3	Determinanten der Forschermobilität.....	42
<b>7</b>	<b>Zusammenfassung</b> .....	<b>44</b>
<b>8</b>	<b>Literaturverzeichnis</b> .....	<b>48</b>

### 1 Einführung

Neues Wissen und neue Technologien sind nicht das Ergebnis eines anonymen Produktionsprozesses, sondern hängen in kritischer Weise von einzelnen herausgehobenen Wissenschaftler- und Erfinderpersönlichkeiten und ihren Interaktionen ab. Diese Personen verfügen in ihren jeweiligen Spezialgebieten über einen erheblichen impliziten Wissensbestand, den sie über einen langen Zeitraum entwickelt haben und der nur zu sehr hohen Kosten auf andere übertragen werden kann. Diese Schlüsselpersonen fungieren als Wissensspillover-Agenten, da sie durch ihre räumliche Mobilität und Präsenz an bestimmten Standorten zur Diffusion impliziten Wissens und der Überführung wissenschaftlichen Wissens in neue Technologien beitragen (Schiller und Revilla Diez 2010).

Hochqualifizierte sind weltweit das mobilste Bevölkerungssegment. In Westeuropa sind sieben Prozent dieser Bevölkerungsgruppe in ein anderes OECD-Land gewandert, während der Wert für andere Qualifikationsgruppen nur zwei Prozent beträgt. Gleichzeitig sind ein Drittel der Zuwanderer in OECD-Staaten Hochqualifizierte, während ihr Anteil an den Erwerbstätigen lediglich zehn Prozent beträgt (Docquier und Marfouk 2004). Wissenschaftler sind innerhalb dieser Gruppe besonders mobil. Ioannidis (2004) hat gezeigt, dass ein Drittel der Spitzenforscher in den USA im Ausland geboren wurde und auch in europäischen Ländern sehr hohe Werte erreicht werden (Deutschland: 19%, Großbritannien: 24%). In der Bilanz können bei den meistzitierten Wissenschaftlern jedoch lediglich die USA einen Nettozugewinn erzielen, während Deutschland auf der Basis des Vergleichs von Geburts- und Arbeitsort dieser Personen einen Nettoverlust von 28% hinnehmen muss (Schiller und Revilla Diez 2008). Wenn man Hochqualifizierte (mit tertiärem Bildungsabschluss) insgesamt betrachtet, ist Deutschland allerdings ein Nettoimporteur (OECD 2008).

#### **Auswirkungen der Mobilität**

Aus der hohen Mobilitätsneigung von Wissenschaftlern und den starken Ungleichgewichten in den Wanderungsbilanzen entstanden Debatten darüber, welche ökonomischen Effekte der Zugewinn oder Verlust von Wissenschaftlern und Erfindern mit sich bringt. Klassische unidirektionale Ansätze (*brain drain* und *brain gain*) wurden dabei in jüngster Zeit durch bi- bzw. multidirektionale Ansätze (*brain circulation*) erweitert (Saxenian 2005), da viele dieser Personen sich nur temporär an Einrichtungen außerhalb ihrer Heimatländer aufhalten bzw. durch grenzüberschreitende Kooperationen in mehrere lokale Umfelder eingebettet sind. Wissensspillover-Agenten besitzen das Potenzial, Transferkanäle zwischen den verschiedenen Standorten ihrer Kooperationspartner zu eröffnen. Diese Wissensflüsse sind auf verschiedenen räumlichen (lokal, national, international) und institutionellen (Wirtschaft, Wissenschaft oder sektorübergreifend) Maßstabsebenen angesiedelt (Schiller und Kiese 2010). Häufig erfüllen mobile Wissenschaftler eine sogenannte „Antennenfunktion“ (Fritsch und Schwirten 1998), da sie die Schnittstelle zwischen der Wissensproduktion in globalen Netzwerken und der Anwendung im lokalen Kontext bilden. Untersuchungen der Auswirkungen von Forschermobilität sollten daher neben den unidirektionalen Effekten, die sich aus der Abwanderung einer Person ergeben, auch die Interaktionen der mobilen Forscher untersuchen. Nur so können die Effekte für Forschungs- und Innovationsaktivitäten bei sendenden und empfangenden Forschungseinrichtungen und Unternehmen vollständig abgebildet werden. Beispielsweise kann der Wechsel eines Forschers an eine andere Einrichtung weitere Kooperationen mit seiner Herkunftsein-

richtung nach sich ziehen und dieser so Zugang zu Wissen der neuen Einrichtung ermöglichen.

Empirische Studien für die Biotechnologie haben gezeigt, dass das Engagement von Spitzenforschern einen positiven Effekt auf Wachstum und Wissenstransfer neugegründeter Biotechnologieunternehmen hat (Zucker und Darby 2006). Frühere Arbeiten haben in allgemeinerer Weise den regionalwirtschaftlichen Nutzen von einem umfangreichen Pool wissenschaftlicher Talente nachgewiesen (Horowitz 1966). Im Rahmen des von der EU-geförderten Forschungsprojekts „Intangible Assets and Regional Economic Growth“ (IAREG) haben Schiller und Revilla Diez (2010, 2012) die räumliche Mobilität und regionale Vernetzung von Spitzenforschern untersucht und gezeigt, dass auch die meistzitierten Forscher eines Wissenschaftsfeldes in vielfältigen wissenschaftlichen und industriellen Kooperationen engagiert sind, die nicht selten in der Region des aktuellen Wirkens eingebettet sind.

Es besteht ein interdependentes Verhältnis zwischen Mobilität und individueller Forschungsleistung der Wissenschaftler und Erfinder. Leistungsmerkmale können einerseits positive Auswirkungen auf die Mobilitätsintensität haben, da sich leistungsstarken Forschern mehr Möglichkeiten bieten, im Zeitverlauf an andere (bessere) Einrichtungen zu wechseln. Andererseits hat die Mobilität Auswirkungen auf die Forschungsleistung, die vor allem dann positiv sein werden, wenn sich durch den Wechsel des Arbeitsumfelds und zusätzliche Kooperationspartner neue Forschungsimpulse ergeben. Für die Beurteilung der Auswirkungen der Forschermobilität ist daher neben der Betrachtung der organisationalen Ebene (Forschungseinrichtung/Unternehmen) der Zusammenhang zwischen Mobilität und Forschungs- und Erfinderleistung auf der individuellen Ebene zu untersuchen.

### **Mobilitätsmotive und Ausmaß lokaler Vernetzung**

Zentrale Determinanten für Art und Ausmaß der Auswirkungen der Wissenschaftler- und Erfindermobilität sind die Motive, die der Mobilität zugrunde liegen und die Dauer der Aufenthalte an den verschiedenen Standorten im In- und Ausland. Schiller und Revilla Diez (2012) haben gezeigt, dass längere Aufenthalte im Ausland die Produktivität von Wissenschaftlern dauerhaft erhöhen, ohne dass es eine geringere Bereitschaft gäbe, sich in lokalen Verbänden am neuen Standort zu engagieren. Ein Informationssystem für Forschermobilität sollte daher in der Lage sein, zwischen der Art und Länge des Aufenthalts (z.B. kurzfristiger Gastaufenthalt, mittelfristiger Aufenthalt im Rahmen von Qualifikationsphasen oder langfristige Tätigkeit) zu differenzieren. Von besonderem Interesse sind hier Faktoren, die dafür verantwortlich sind, dass ein Übergang von temporären Aufenthalten in deutschen Forschungseinrichtungen und Unternehmen hin zu einer dauerhaften Migration erfolgt oder behindert wird.

Als kritisch für die dauerhafte Gewinnung mobiler Forscher haben sich die jeweiligen Rahmenbedingungen erwiesen, da die relevante Personengruppe aufgrund ihrer herausragenden Qualifikation in der Regel zwischen einer Vielzahl potenzieller Standorte wählen kann. Reiner (2010) systematisiert unter dem Begriff ‚brain competition policy‘ die verschiedenen Aspekte des Regulierungsrahmens, die von der Einwanderungsgesetzgebung über Arbeitsmarkt- und sozialpolitische Regeln bis hin zu weichen Faktoren wie der Lebensqualität oder Willkommenskultur reichen. In Deutschland hat es gerade in den letzten Jahren Veränderungen der Einwanderungsregeln gegeben, die zu einer Verbesserung der Zuzugsmöglichkeiten Hochqualifizierter geführt haben (Blue Card, Anerkennung von Abschlüssen, Wegfall der Vorrangprü-

fung etc.). Daneben engagieren sich verschiedene öffentliche Akteure und gemeinnützige Einrichtungen seit vielen Jahren durch explizite Programme zur Mobilitätsförderung im Wettbewerb um die besten Köpfe (Schiller 2011). Mit Hilfe einer umfassenden Datenbasis mit Längsschnittinformationen zu Karriereverläufen können hier Aussagen über die Auswirkungen von Gesetzesänderungen und Mobilitätsprogrammen in Deutschland sowie im europäischen Raum gemacht werden.

### **Ziele und Relevanz der Machbarkeitsstudie**

Ein zentrales methodisches Ergebnis existierender Arbeiten war, dass es an wissenschaftlichen Untersuchungen und Datengrundlagen mangelt, die explizit herausgehobene Wissenschaftler- und Erfinderpersönlichkeiten in den Mittelpunkt stellen und Informationen zur Mobilität und ihren Auswirkungen beinhalten. Trotz existierender Ansätze zur Ermittlung von Wanderungsbilanzen fehlt es an einem laufenden Berichtssystem, das Auskunft über Zu- und Abwanderung von Forschern in Deutschland insgesamt gibt und in regionaler Hinsicht sowie zwischen Wissenschaftsdisziplinen bzw. Branchen differenziert. Vor diesem Hintergrund verfolgt die Machbarkeitsstudie vier zentrale Ziele:

- Definition von inhaltlichen und technischen Anforderungen an ein Berichtssystem zur Erfassung der Forschermobilität (Kapitel 2),
- Bewertung deutscher und internationaler Datensätze/-quellen und empirischer Forschungsansätze hinsichtlich ihrer Einsatzmöglichkeiten als Basis quantitativer Untersuchungen zur Forschermobilität (Kapitel 3 und 4),
- Ausführliche Bewertung von zwei besonders vielversprechenden Ansätzen hinsichtlich Kosten, Ressourceneinsatz und Risiken, die mit dem jeweiligen Ansatz verbunden sind (Kapitel 5) und
- Möglichkeiten der Verknüpfung des Berichtssystems mit anderen Informationen zu Auswirkungen und Determinanten der Forschermobilität (Kapitel 6).

Der Aufbau eines Berichtssystems zur Forschermobilität bezieht seine gesellschaftliche Relevanz aus den inhaltlichen Fragen, die sich durch die verbesserte Verfügbarkeit von Daten zu diesem Thema beantworten lassen. Vier zentrale Fragestellungen der Forschungs- und Innovationspolitik seien an dieser Stelle beispielhaft genannt:

1. Wissensbilanzen: Forscher sind zentrale gesellschaftliche Wissensträger. Daher trägt ihre personelle Mobilität dazu bei, den Wissensbestand in der Zielregion zu erhöhen. Durch ein Berichtssystem können Bilanzen erstellt werden, in welchen Bereichen Deutschland durch die internationale Forschermobilität in der Summe Wissen dazugewinnt oder verliert.
  - Kommt es im Zuge der Forschermobilität für Deutschland zu einem *brain drain* oder einem *brain gain* von Forschern in öffentlichen und privaten Einrichtungen?
  - Welches Bild ergibt sich bei einer nach Fachdisziplinen, Wirtschaftszweigen, Institutionen und Regionen differenzierenden Betrachtung?
  - Welche Rolle spielt die Rückwanderung von Deutschen aus dem Ausland im Rahmen der Forschermobilität?

- Wie viele ausländische Forscher bleiben langfristig oder dauerhaft in Deutschland?
  - Welche Implikationen ergeben sich daraus für den erwarteten Fachkräftemangel?
2. Ökonomische Bilanzen: Ansatzpunkt sind hier die Kosten die mit der Ausbildung von Wissenschaftlern verbunden sind und zumindest in Deutschland auf allen Stufen des Bildungssystems ganz überwiegend von der öffentlichen Hand getragen werden. Die Abwanderung eines in Deutschland ausgebildeten Doktoranden in ein anderes Land hat also neben dem Verlust eines Wissensträgers auch den Verlust der Investitionen in die Ausbildung zur Folge.
- Welche Verluste an Investitionen in die Ausbildung entstehen Deutschland durch die Abwanderung von Forschern?
  - Welche Einsparungen an Investitionen in die Ausbildung erzielt Deutschland durch die Zuwanderung von Forschern aus dem Ausland und die zusätzliche Qualifikation Deutscher, die aus dem Ausland zurückkehren?
3. Auswirkungen der internationalen Forschermobilität:
- Welche Auswirkungen hat die internationale Forschermobilität auf die wissenschaftliche Produktivität und Karriereentwicklung des mobilen Forschers selbst (individuelle Perspektive)?
  - Welche Auswirkungen hat die internationale Forschermobilität für die wissenschaftliche Produktivität und den wissenschaftlichen bzw. wirtschaftlichen Erfolg der abgebenden und aufnehmenden Einrichtungen bzw. Unternehmen (institutionelle Perspektive)?
  - Welche Auswirkungen hat die internationale Forschermobilität für die wissenschaftliche und technologische Leistungsfähigkeit, die Vernetzung von Forschern und Innovatoren, regionale Wissensspillover sowie den wirtschaftlichen Erfolg der betroffenen Innovationssysteme (systemische Perspektive)?
4. Determinanten der Forschermobilität
- Wie attraktiv ist das deutsche Wissenschaftssystem für die Gewinnung ausländischer Forscher sowie den Verbleib und die Rückgewinnung deutscher Forscher?
  - Welchen Einfluss haben öffentliche und stiftungsfinanzierte Mobilitätsförderung der vergangenen Jahre auf die Forschermobilität?
  - Welche Wirkungen haben die veränderten Einwanderungsregelungen in Deutschland (u.a. Blue Card, Abschaffung von Vorrangprüfungen für bestimmte Berufe und verbesserte Anerkennungsregelungen von Abschlüssen)?
  - Welche Auswirkungen haben Veränderungen der Finanzierung und Rahmenbedingungen in den Zielländern auf die Forschermobilität, z. B. Rückgang der staatlichen Ausgaben für Forschung in den USA?

## 2 Methodische und inhaltliche Vorüberlegungen

Vor der Prüfung von Methoden zur Erfassung der Forschermobilität in den Kapiteln 3 und 4 werden in diesem Abschnitt Vorüberlegungen zur Definition der Begriffe „Forscher“ und „Mobilität“ sowie zu den inhaltlichen und technischen Anforderungen an ein Berichtssystem dargestellt. Diese bilden die Basis für die Bewertung von Vor- und Nachteilen der Vielzahl im deutschen und internationalen Raum bisher genutzten Methoden zur Erfassung der Forschermobilität und ermöglichen, Handlungsempfehlungen für die Verbesserung der Erfassung dieses Phänomens abzuleiten.

### 2.1 Definitionen

#### 2.1.1 Forscher

Das geplante Berichtssystem hat das Ziel, möglichst treffsicher die Mobilität von Forschern abzubilden. Dazu bedarf es zunächst einer operationalisierbaren Definition dieser Personengruppe. Eine solche Abgrenzung erfolgt üblicherweise unter Nutzung von Informationen über Qualifikation und Beruf.

Das *Canberra-Manual* der OECD (1995) definiert den Kernbereich der Humanressourcen in Wissenschaft und Technologie (HRST) als Personen mit einem Bildungsabschluss auf dem Level 5 oder 6 (tertiäre Bildungsabschlüsse) der International Standard Classification of Education (ISCED) und anhand einer Liste von wissenschaftlichen und technischen Berufen auf Basis der ISCO-88 (International Standard Classification of Occupations). Laut Eurostat werden EU-weit etwa 20 % der Beschäftigten dem HRST Core zugeordnet. Diese Abgrenzung ist für die Erfassung von Forschern eindeutig zu weit, da in ihr auch alle nicht forschenden Hochqualifizierten in wissenschaftlich-technischen Berufen erfasst sind.

Das *Frascati-Manual* der OECD (2002) bietet Definitionen für Personal in Forschung und Entwicklung (FuE) sowie explizit zu Forschern. FuE-Personal ist als der Personenkreis definiert, der direkt mit FuE befasst ist und direkte Dienstleistungen für FuE-Aktivitäten leistet, z. B. FuE-Manager und Mitarbeiter in Technik und Verwaltung von FuE-Abteilungen. Dieser Personenkreis ist deutlich kleiner und umfasst laut Eurostat nur noch ca. 1,5 % der Beschäftigten in der EU. Durch die Zählung von Technikern und Verwaltungskräften sind hier aber immer noch Personen enthalten, die nicht selbst forschen. Erst die noch engere Abgrenzung der Forscher (ca. 1 % der Beschäftigten in der EU) ist hinreichend präzise, um für die Erfassung von Forschermobilität herangezogen zu werden: „Researchers are professionals engaged in the conception or creation of new knowledge, products, processes, methods and systems and also in the management of the projects concerned“ (OECD 2002: 93).

Zur Identifizierung von Forschern anhand der ISCO-88 empfiehlt das Frascati-Manual die Nutzung der Gruppe 2 „Professionals“ und der Gruppe 1237 „Research and Development Department Managers“. Aus Sicht des Autors dieser Studie ist jedoch davon auszugehen, dass diese Operationalisierung einerseits zu weit ist (in der Gruppe 2 sind auch viele Nicht-Forschende enthalten), andererseits zu eng ist, da nur FuE-Manager, nicht aber FuE-Mitarbeiter erfasst werden, sofern diese nicht unter Professionals eingeordnet sind. Durch eine weitere Eingrenzung der Qualifikation (z. B. nur Promovierte, ISCED-6) und/oder Fokussierung auf bestimmte Branchen (z. B. Hochschulen und Forschungseinrichtungen) kann je nach



Verfügbarkeit der notwendigen Informationen in der genutzten Datenquelle eine differenziertere Identifizierung erfolgen, als es alleine auf Basis der ISCO-88 möglich ist. Die genaue Operationalisierung der Definition von Forschern wird in den Kapiteln 3 und 4 für jeden einzelnen untersuchten Datensatz erläutert und hinsichtlich ihrer Treffsicherheit bewertet.

Die durch das Berichtssystem zur Forschermobilität zu erfassenden Forscher sind in drei Bereichen zu finden: Hochschulen, öffentliche Forschungseinrichtungen, private Unternehmen. Zwischen diesen drei Bereichen gibt es deutliche Unterschiede bezüglich der Verfügbarkeit von Informationen über die relevanten Individuen. Tendenziell werden Forscher an Hochschulen und öffentlichen Forschungseinrichtungen leichter zu identifizieren sein und mehr Informationen zu ihnen verfügbar sein, als im Falle von Forschern in privaten Unternehmen. Dies liegt zum einen im unterschiedlichen Selbstverständnis der beiden Forschergruppen. Forscher im öffentlichen Sektor haben ein großes Interesse an der Selbstvermarktung ihrer individuellen Forschungsleistung und stellen dazu Lebenslaufinformationen, Publikationsdaten und Forschungsprojekte auf persönlichen Homepages dar bzw. tauchen in öffentlich zugänglichen Datenbanken (z. B. Publikationsdatenbanken, soziale Netzwerke) auf. Forscher in FuE-Abteilungen der Unternehmen sind von außen nicht individuell identifizierbar und ihre Tätigkeit und Forschungsleistungen unterliegen häufiger der Geheimhaltung. Zum anderen ist der öffentliche Forschungssektor einfacher anhand von Branchenklassifikationen bzw. institutioneller Zugehörigkeit abgrenzbar, da davon ausgegangen werden kann, dass ein Großteil der Beschäftigten mit einem Hochschulabschluss an Universitäten und öffentlichen Forschungseinrichtungen auch als Forschende in diesen Einrichtungen tätig sind. Für Mitarbeiter mit Hochschulabschluss in privaten Unternehmen gilt dies aufgrund der Vielzahl anderer Tätigkeitsfelder außerhalb der Forschung nicht.

Da Forscher in privaten Unternehmen in Deutschland etwa 57 % und EU-weit etwa 43 % der gesamten Forscherpopulation ausmachen, erscheint ihre Berücksichtigung in einem Berichtssystem zur Forschermobilität trotz erschwelter Erfassbarkeit unerlässlich. Bei der Beurteilung der Methoden in den Kapitel 3 und 4 wird auch immer darauf hingewiesen werden, ob und auf welche Weise Forscher in privaten Unternehmen durch den jeweiligen Ansatz miterfasst werden.

### **2.1.2 Mobilität**

Auch der Begriff der Mobilität bedarf einer differenzierten Definition, um festzulegen, welche Arten der Mobilität durch ein Berichtssystem zur Forschermobilität erfasst werden sollen. Die Machbarkeitsstudie bezieht sich vor allem auf räumliche Mobilität und dabei ausschließlich auf internationale Mobilität. Die Mobilität von Forschern innerhalb Deutschlands, die ebenfalls noch nicht hinreichend erforscht scheint, ist daher nicht Gegenstand dieser Machbarkeitsstudie. Institutionelle Mobilität, d. h. Jobwechsel zwischen öffentlicher Forschung und FuE-Abteilungen privater Unternehmens, wird nur erfasst, wenn sie länderübergreifend erfolgt und steht daher ebenfalls nicht im Mittelpunkt der Erfassung von Forschermobilität. Die betrachtete internationale räumliche Mobilität umfasst drei unterschiedliche Tatbestände:

- Zuwanderung ausländischer Forscher nach Deutschland
- Abwanderung deutscher Forscher ins Ausland
- Rückwanderung deutscher Forscher aus dem Ausland nach Deutschland

Darüber hinaus kann Mobilität in sehr unterschiedlichen zeitlichen Dimensionen erfolgen. Ein Problem liegt darin, dass häufig nicht von Anfang an absehbar ist, wie lange ein Aufenthalt im Ausland andauern wird. Von vornherein langfristig (mehr als drei Jahre) angelegt sind Aufenthalte, bei denen eine dauerhafte Tätigkeit im Gastland angenommen wird. Mittelfristiger Natur (ein bis drei Jahre) sind Aufenthalte für einzelne Karrierephasen, z. B. Promotion oder Postdoc-Phase in der wissenschaftlichen Forschung oder der Auslandseinsatz eines privatwirtschaftlichen FuE-Mitarbeiters für einige Jahre. Diese beiden Arten sind eindeutig im Rahmen des Berichtssystems Forschermobilität zu erfassen und tendenziell leichter zu identifizieren, da sie in der Regel mit der Änderung des Arbeitgebers bzw. der institutionellen Zuordnung einhergehen.

Ein ebenfalls wichtiger Aspekt der internationalen Forschermobilität ist kurzfristig angelegte Mobilität im Bereich von sechs Monaten bis zu einem Jahr, z. B. im Rahmen eines Sabbaticals eines Wissenschaftlers, eines Austauschprogramms oder eines längeren projektbezogenen Einsatzes in einem ausländischen Unternehmensteil. Auch diese Art der Mobilität soll, sofern durch das genutzte Verfahren identifizierbar, im Rahmen des Berichtssystems als Forschermobilität erfasst werden. Allerdings ist hier die Abgrenzung schwieriger, da in der Regel keine Änderung des Arbeitsgebers bzw. der institutionellen Zuordnung erfolgt. Grundsätzlich nicht erfasst werden sollen sehr kurzfristige Aufenthalte von wenigen Wochen oder Monaten im Rahmen von Tagungen, Sommerschulen, Gastaufenthalten oder empirischen Erhebungen im Ausland.

Die Eignung der geprüften Datenbanken für die Erfassung der drei verschiedenen Mobilitätsarten und der zeitlichen Dimension von Mobilität wird ebenfalls in den Kapiteln 3 und 4 erläutert. Wünschenswert ist es, alle drei Mobilitätsarten mit einer Methode abdecken zu können. Allerdings besteht hier die Schwierigkeit, dass in deutschen Quellen nur die Zuwanderung oder Rückwanderung nach Deutschland, nicht aber die Abwanderung aus Deutschland systematisch erfasst werden dürfte, während ausländische Quellen vor allem Abwanderung aus Deutschland abbilden können. Grundsätzlich ist bei diesem Aspekt die Chance am größten, dass in internationalen Quellen, die für alle Länder in gleicher oder vergleichbarer Form vorliegen, alle drei Mobilitätsarten enthalten sind. Ein weiterer Ansatz ist die Konzentration auf Hauptzielländer der deutschen Forschermobilität – dies ist vor allem die USA – und die Identifikation deutscher Forscher anhand nationaler Quellen dieser Länder.

## **2.2 Anforderungen an ein Berichtssystem**

### **2.2.1 Relevante Inhalte**

Zur vollständigen Beantwortung der in der Einleitung aufgeworfenen inhaltlichen Fragen, wäre es notwendig eine im Folgenden zusammengefasste Vielzahl von Inhalten personenspezifisch zu erfassen. Es ist allerdings nicht zu erwarten, dass sich eine Datenquelle bzw. ein Datensatz finden lässt, der alle Informationen in gleichem Umfang abdeckt. Die nachfolgende Übersicht ist daher eher als Wunschliste zu verstehen, die von den zu prüfenden Methoden mehr oder weniger umfassend erfüllt werden wird.

- Identifizierbarkeit von Forschern und von Mobilität im Sinne der obengenannten Definitionen
- Abbildung individueller Karriereverläufe
- Art, Anzahl und Zeiträume der Mobilität
- Arbeitsorte (räumlich und institutionell)
- Individuelle Forschungsleistung und Vernetzung
  - Forschungsleistung: Publikationen, Patente, Einwerbung von Forschungsmitteln, etc.
  - Vernetzung: Ko-Publikationen, Ko-Patente, Mitgliedschaft in Forschungsverbänden, etc.
- Forschungsleistung und Vernetzung der Herkunfts- und Zieleinrichtung des mobilen Forschers
  - Forschungsleistung: Publikationen, Patente, Einwerbung von Forschungsmitteln, etc.
  - Vernetzung: Ko-Publikationen, Ko-Patente, Mitgliedschaft in Forschungsverbänden, etc.
- Mobilitätsmotive und -barrieren
- Weitere strukturelle personenbezogene Charakteristika
  - Alter, Qualifikation Fachbereich, Geschlecht, sozialer und familiärer Hintergrund, Migrationshintergrund etc.

Zur Darstellung von Migrationsbilanzen (wie viele Forscher wandern zu, ab, zurück?) wäre der erste Inhalt ausreichend. Die Analyse von Migrationsmustern erfordert zusätzliche Informationen zu Karriereverläufen. Sofern Interdependenzen von Migration und Forschungsleistung (individuell und/oder institutionell) untersucht werden sollen, sind weitere Informationen zu diesen Aspekten notwendig. Die Erfassung von Mobilitätsmotiven und -barrieren würde schließlich vertiefte Untersuchungen zu Determinanten der internationalen Forschermobilität ermöglichen.

### 2.2.2 Methodische Anforderungen

Studien zur Forschermobilität haben in der Vergangenheit häufig Befragungen genutzt, um Antworten auf spezifische Fragen zum Mobilitätsverhalten zu gewinnen. Dazu wurden Forscher anhand von Datenbanken (z. B. Hochschullehrerverzeichnis oder im Rahmen von Mobilitätsprogrammen geförderte Forscher) identifiziert und um die Teilnahme an einer Befragung gebeten. Der große Vorteil dieser Methode ist, dass tatsächlich die Informationen abgefragt werden können, die für die Beantwortung der obengenannten Fragen zur Forschermobilität notwendig sind. Insbesondere qualitative Einschätzungen zu Motiven und Hemmnissen können mit diesem Ansatz gewonnen werden. Dem steht allerdings eine Vielzahl von Nachteilen gegenüber, die sich auf methodische Aspekte und den notwendigen Ressourceneinsatz beziehen. Zu nennen sind hier vor allem folgende Argumente, die in der Summe gegen eine Weiterverfolgung von Befragungen im Rahmen dieser Machbarkeitsstudie sprechen:

- Relativ hoher Ressourceneinsatz für die Erstellung und Implementierung der Befragungen bzw. sehr niedrige Rücklaufquoten. Das RESCAR-Projekt der EU erreichte 2006 lediglich 13 % Rücklauf und das MORE-Projekt musste sich 2010 sogar mit 11 % begnügen.
- Geringe Rücklaufquoten stellen die Repräsentativität der Methode in Frage. Es ist davon auszugehen, dass Befragungen zur Mobilität überdurchschnittlich häufig von mobilitätsaffinen Forschern beantwortet werden und so die Mobilität überschätzt werden kann.
- Durch die bei Befragungen notwendige Eingrenzung der Grundgesamtheit auf eine bestimmte Forscherpopulation ist die gleichzeitige Erfassung von abgewanderten, zugewanderten und rückgewanderten Forschern nicht leistbar.

In der Vergangenheit wurden häufig Forscher befragt, die an Förderprogrammen zur internationalen Mobilität teilgenommen haben, z. B. DAAD (DAAD und HIS 2012), Alexander von Humboldt-Stiftung (Jöns 2007), oder Mitglieder von Netzwerken zur Rückkehrförderung sind, z. B. GAIN (Engin und Reifenberg 2012). Ein Vorteil dieses Ansatzes ist, dass deutsche Forscher im Ausland erfasst werden. Anhand von geförderten mobilen Forschern Aussagen über alle mobilen Forscher zu treffen ist allerdings aus zwei Gründen problematisch. Zum einen ist nicht bekannt, wie hoch der Anteil der geförderten Mobilität an der gesamten Forschermobilität ist. Zum anderen ist davon auszugehen, dass geförderte Mobilität sich von nicht-geförderter Mobilität unterscheidet. So ist erstere fast immer temporär und auf die Dauer der Förderung begrenzt. Außerdem ist die Rückkehrwilligkeit in das Heimatland häufig eine Auflage des Förderprogramms. Das GAIN-Netzwerk wendet sich daher auch explizit an Rückkehrwillige nach Deutschland, so dass diese Datenbasis für Aussagen über die Rückkehrwahrscheinlichkeit aller deutschen Forscher methodisch ungeeignet erscheint.

Aufgrund der Einschätzung, dass die inhaltlichen Verzerrungen und methodischen Probleme bei gleichzeitig hohem Ressourceneinsatz gegen die Eignung der Befragungsmethodik für ein Berichtssystem zur Forschermobilität sprechen, ist es der methodische Anspruch im Rahmen dieser Machbarkeitsstudie vor allem solche Methoden zu prüfen, die eine automatisierte bzw. automatisierbare Datenbereitstellung ermöglichen oder existierende personenbezogene Mikrodaten zur Mobilität und zur Forschungs- und Innovationsleistung nutzen. Die Voraussetzungen für die Eignung solcher Datenquellen oder Datensätze sind, dass sie eine hinreichend genaue Abgrenzung des Untersuchungsgegenstands Forscher erlauben, Informationen zu den obengenannten relevanten Themenbereichen enthalten, große Fallzahlen für nach verschiedenen Merkmalen differenzierende Analysen umfassen und intertemporal und interregional vergleichbar vorliegen.

## 3 Deutsche Datensätze und Datenquellen

Entsprechend der Fokussierung der Machbarkeitsstudie auf Forschermobilität von und nach Deutschland werden zunächst Datenquellen bewertet, die für Deutschland vorliegen. Diese werden in Kapitel 4 um internationale Datensätze und Quellen aus den Hauptzielländern der deutschen Forschermobilität ergänzt. Die in die Untersuchung einbezogenen Datenquellen, aus denen sich potenziell Informationen zur Forschermobilität gewinnen lassen, umfassen:

- regelmäßig erhobene Bestandsstatistiken mit Individualinformationen (Mikrozensus, Sozio-ökonomisches Panel (SOEP)),
- Prozessdaten aus der Arbeitsverwaltung (Integrierte Erwerbsbiographien (IEB) der Bundesagentur für Arbeit (BA)),
- Zugangsdaten der Einwanderungsbehörden (Ausländerzentralregister),
- Panelstudien zu Verbleib und Karriereentwicklung von Hochschulabsolventen und Promovierenden (z. B. Absolventenpanel des HIS, ProFile-Promovierendenpanel des iFQ) und
- die vom Wissenschaftsrat jüngst vorangetriebenen Bemühungen zur Erstellung eines Kerndatensatzes Forschung, durch den an Hochschulen vorliegende Daten in einem standardisierten Format austauschbar gemacht werden sollen.

### 3.1 Mikrozensus

Der Mikrozensus ist eine seit 1957 von der amtlichen Statistik durchgeführte repräsentative Haushaltsbefragung, in der durch eine 1 %-Stichprobe der Wohnbevölkerung in Deutschland etwa 830.000 Personen in etwa 370.000 privaten Haushalten stellvertretend für die gesamte Bevölkerung zu ihren Lebensbedingungen befragt werden. Die befragten Haushalte werden nach einem statistischen Zufallsverfahren ausgewählt. Die Daten des Mikrozensus gehen in die Arbeitskräfteerhebung der Europäischen Union ein (European Labour Force Survey) und sind daher potenziell für internationale Vergleiche geeignet (s. Kapitel 4.2.1). Für die Analyse der Forschermobilität sind insbesondere folgende im Mikrozensus enthaltene Informationen von Bedeutung:

- Qualifikation, Beschäftigung und Wirtschaftszweig zur Identifikation von Forschern,
- Zuzugsjahr zur Identifikation von mobilen Personen und Dauer des bisherigen Aufenthalts und
- Geburtsort und Nationalität zur Unterscheidung von „echten“ Zu- und deutschen Rückwanderern.

Die Identifikation von Forschern kann im Mikrozensus durch die Abgrenzung von Personen mit Hochschulabschluss (ISCED 5a und 6), die eine potenziell forschende Tätigkeiten nach der International Standard Classification of Occupations (ISCO) ausüben, vorgenommen werden. Auf der im deutschen Mikrozensus enthaltenen Ebene der 3-Steller bieten sich die Gruppen 211-214, 221, 222, 231, 241 und 244 an, um der Definition von Forschern nach dem Frascati-Manual nahe zu kommen. Weitere Spezifizierungen sind durch die Eingrenzung des Personenkreises auf Personen mit einem tertiären Bildungsabschluss und der Einschränkung auf

forschungsaffine Wirtschaftszweige möglich, um z. B. promovierte niedergelassene Ärzte von solchen zu unterscheiden, die an Universitäten oder in Pharmaunternehmen forschen. Neben den Standardvariablen werden zusätzliche in einzelnen Jahren Sonderthemen abgefragt. Zum Thema Mobilität gibt es in ausgewählten Jahren (zuletzt 2008) zusätzliche Informationen. Aus diesen Daten können Informationen zu Mobilitätsmotiven sowie ausländischen Bildungsabschlüssen gewonnen werden. Außerdem wurden seit 1996 alle vier Jahre und ausnahmsweise letztmalig im Jahr 2007 im Mikrozensus die schwerpunktmäßig ausgeübten Tätigkeiten der befragten Personen erfasst. Mit Hilfe des Tätigkeitsschwerpunkts „Forschen, Entwerfen, Konstruieren, Gestalten von Produkten, Plänen, Programmen“ lassen sich Forscher noch treffsicherer abgrenzen als alleine durch das oben geschilderte Vorgehen.

Die Verwendung des Mikrozensus ist aufgrund des amtlichen und repräsentativen Charakters der Erhebung vorteilhaft. Dieser lässt zum einen die Hochrechnung auf die Gesamtzahl der deutschen und ausländischen Forscher zu und hält zum anderen den Ressourcenbedarf relativ gering, da die Daten zu überschaubaren Kosten für wissenschaftliche Zwecke genutzt werden können. Darüber hinaus sind Forscher des öffentlichen und privaten Sektors in der Datenquelle gleichermaßen erfasst. Zentrale Informationen zur Abgrenzung von ausländischen bzw. mobilen Forschern liegen in allen Wellen vor.

Der Mikrozensus erfüllt jedoch bei weitem nicht alle Kriterien, um Antworten auf alle obengenannten Fragenkomplexe zu geben, und ist lediglich für die Darstellung von Strukturen der Forschermobilität geeignet. Das Erstellen vollständiger Migrationsbilanzen ist nicht möglich, da nur Forscher erfasst sind, die sich zum Befragungszeitpunkt in Deutschland aufhalten. Damit sind nur Zu- und Rückwanderer erfasst, nicht jedoch die Auswanderer. Außerdem sind keine Karriereverläufe oder Interdependenzen mit wissenschaftlicher Leistung darstellbar, da keine Informationen zu Erwerbsbiographien erfasst sind und die Anonymisierung keine Verknüpfung mit anderen Datenquellen (z. B. zur wissenschaftlichen Leistung) ermöglicht. Die Abgrenzung von Forschern ist nur annäherungsweise eindeutig möglich, da auf Basis der Berufsklassifikation nicht die tatsächlichen Tätigkeiten abgebildet werden. Letztere werden nur in einigen Jahren abgefragt. Ein Problem der 1 %-Stichprobe ist, dass trotz mehr als 800.000 Fällen die Subpopulation ausländischer Forscher in Deutschland zu gering ist, um nach einer Vielzahl von Merkmalen differenziert auszuwerten. Potenziell kann die Fallzahl durch das Poolen mehrere Wellen des Mikrozensus erhöht werden. Dabei muss allerdings der Panel-Charakter des deutschen Mikrozensus beachtet werden, bei dem jede gezogene Person über vier Jahre befragt wird. Überschneidungsfrei sind daher nur Wellen, die mindestens vier Jahre auseinanderliegen, z. B. die Jahre 2006 und 2010. Die gleichzeitige Verwendung mehrerer Wellen kann jedoch auch ohne Poolen dazu genutzt werden, um die Robustheit des Stichprobenfehlers zu bewerten. Im Vergleich mit den Arbeitserhebungen anderer Länder ist ein weiterer Nachteil des deutschen Mikrozensus die lange Zeitspanne bis zur Veröffentlichung neuer Daten. Während in Großbritannien bereits Daten für das Jahr 2012 verfügbar sind und in Frankreich mit den Daten von 2011 gearbeitet werden kann, liegen für Deutschland im Jahr 2013 lediglich die Daten für das Jahr 2010 vor. Positiv muss für den deutschen Mikrozensus allerdings die im internationalen Vergleich hohe Fallzahl angemerkt werden, die mit 1 % der Wohnbevölkerung und damit ca. 400.000 Erwerbspersonen deutlich über den ca. 80.000 bis 90.000 befragten Personen in Großbritannien und Frankreich liegt.

Hinsichtlich der abschließenden Bewertung der Ansätze kann an dieser Stelle bereits vorweggenommen werden, dass der Mikrozensus die einzige aktuell für Deutschland verfügbare Datenbasis ist, aus der sich mit hinreichender Genauigkeit und Stichprobengröße für die gesamte Forscherpopulation Aussagen treffen lassen. Einen Nachweis für die Nutzbarkeit der Daten für Fragestellungen der Forschermobilität haben Klingert und Block (2013) mit Mikrozensus-Daten des Jahres 2007 erbracht. Durch die relativ gute internationale Vergleichbarkeit der Erhebung ist es möglich, mit Hilfe der Europäischen Arbeitskräfteerhebung oder nationaler Arbeitskräfteerhebungen (Kapitel 4.2.1) länderübergreifende Vergleiche der Struktur der internationalen Forschermobilität durchzuführen und den Umfang der Präsenz deutscher Forscher im Ausland zu quantifizieren. Letzteres findet jedoch in kleinen Ländern bzw. kleinen Stichproben eine Beschränkung durch die geringen Fallzahlen für die Subpopulation der deutschen Forscher.

#### **3.2 Sozio-ökonomisches Panel (SOEP)**

Das Sozio-ökonomische Panel (SOEP) ist eine seit 25 Jahren jährliche durchgeführte repräsentative Wiederholungsbefragung, die im Auftrag des DIW Berlin von TNS Infratest Sozialforschung durchgeführt wird. Das SOEP umfasst 20.000 Personen aus rund 11.000 Haushalten und gibt Auskunft zu Fragen über Einkommen, Erwerbstätigkeit, Bildung und Gesundheit. Forscher können im SOEP anhand der Klassifikation der Berufe (KldB) ähnlich wie im Mikrozensus abgegrenzt werden. Der Panelcharakter ermöglicht es, Karriereverläufe abzubilden, da das SOEP die Ausbildungs- und Erwerbsbiographien erfasst. Prinzipiell sind anhand des SOEP daher die gleichen Untersuchungen möglich, die auch anhand des Mikrozensus durchgeführt werden können. Zusätzlich sind Fragen zu Karriereverläufen enthalten. Die praktische Anwendbarkeit des SOEP für Fragen der Forschermobilität scheitert allerdings an der zu geringen Fallzahl an Forschern. Unter der Annahme, dass sich unter den 20.000 Befragten etwa 10.000 Erwerbstätige befinden, sind in den Daten selbst bei einem Anteil der Forscher an der Gesamtbevölkerung von 2 % nicht mehr als 200 Forscher zu erwarten. Die Zahl der zugewanderten ausländischen und rückgewanderten deutschen Forscher dürfte in diesem Sample selbst unter der Annahme, dass 10 % der Forscher zugewandert sind, für statistische Analysen bei weitem zu gering sein. Auch zur Identifizierung von ausgewanderten Forschern ist die Datengrundlage zu klein. Eine Pilotstudie von Erlinghagen und Stegmann (2009) hat gerade einmal 288 Personen für den Zeitraum 2002 bis 2006 als Auswanderer aus dem SOEP identifiziert, davon 46 mit Hochschulabschluss. Die Zahl der darin enthaltenen Forscher dürfte im einstelligen Bereich liegen. Das SOEP wird daher als Datenquelle für Forschermobilität aufgrund zu geringer Fallzahlen nicht weiter verfolgt.

#### **3.3 Integrierte Erwerbsbiographien (IEB) der Bundesagentur für Arbeit (BA)**

Im Gegensatz zum SOEP sind die Fallzahlen bei den Integrierten Erwerbsbiographien (IEB) der Bundesagentur für Arbeit (BA) kein Ausschlusskriterium. Diese Daten umfassen die Prozessdaten der Arbeitsverwaltung und damit alle mehr als 25 Millionen sozialversicherungspflichtig Beschäftigten in Deutschland. Für diesen Personenkreis sind die Erwerbsbiografien von 1975 bis zum aktuellen Rand abgebildet, wodurch Informationen über Karriereverläufe vorliegen. Die Identifikation von Forschern ist vergleichbar mit dem Mikrozensus anhand der Tätigkeit (nach KldB 1988 bzw. seit dem 01. Januar 2011 nach KldB 2010), dem Bildungsab-

schluss (nur Hochschulabschluss enthalten, keine differenzierte Erfassung von Promovierten) und forschungsaffinen Wirtschaftszweigen möglich.

Trotz der Vorteile aufgrund der Größe des Datensatzes gibt es drei wesentliche Beschränkungen, die in der Summe gegen eine Verwendung der IEB für ein Berichtssystem zur Forschermobilität sprechen. Erstens sind nur die sozialversicherungspflichtig Beschäftigten erfasst. Mit den Selbständigen und Beamten fehlen daher zwei wesentliche Segmente der Erwerbstätigen. Für die Identifizierung von Forschern bedeutet dies, dass alle Professoren und alle übrigen verbeamteten Forscher an Universitäten und öffentlichen Forschungseinrichtungen nicht erfasst sind. Das gleiche gilt für Unternehmer privater Forschungsunternehmen. Der Datensatz wäre alleine aus diesem Grund höchstens für Forscher in der Privatwirtschaft geeignet.

Zweitens sind Migrationsereignisse nicht direkt aus den Daten der IEB ablesbar. Die Abwanderung einer Person würde sich im Verschwinden des Datensatzes aus den IEB äußern. Für dieses Verschwinden wird jedoch kein Grund hinterlegt, so dass Abwanderung einer Person ins Ausland nicht von anderen Gründen für ein Verschwinden getrennt werden kann. Diese anderen Gründe können unter anderem die Pensionierung, die Verbeamtung, die Gründung eines eigenen Unternehmens, das vorübergehende Beenden der Erwerbstätigkeit (z. B. Elternzeit) oder sogar den Tod des Arbeitnehmers umfassen. Sie sind zahlenmäßig höchstwahrscheinlich von viel größerer Bedeutung als die Abwanderung. Ähnliches gilt für die Zuwanderung, die sich ebenso wie viele andere Tatbestände (z. B. erstmaliger Eintritt in ein sozialversicherungspflichtiges Beschäftigungsverhältnis nach Abschluss der Ausbildung), die nichts mit Mobilität zu tun haben, in einem neuen Erscheinen einer Person äußern würde. Die fehlende eindeutige Identifizierung von Wanderungsereignissen ist der schwerwiegendste Grund, der gegen eine Nutzung der IEB-Daten für ein Berichtssystem zur Forschermobilität spricht. Drittens unterliegen die Daten besonderen Anforderungen der Geheimhaltung und sind anders als die Daten der Arbeitskräfteerhebung nicht für internationale Vergleiche geeignet.

An dieser Stelle ist jedoch darauf hinzuweisen, dass die Probleme der eindeutigen Identifizierung von Forschern und der Mobilitätserschaffung durch Zuspielen (Merging bzw. Record Linkage) anderer Datensätzen teilweise umgangen werden können. So werden in einem laufenden Projekt des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) mit Prof. Dr. Dietmar Harhoff (Max-Planck-Institut für Immaterialgüter- und Wettbewerbsrecht, MPI-IP, München) Patentdaten mit den IEB-Daten verknüpft und so ein Forschungsdatensatz erstellt, der eine eindeutige Identifizierung von Erfindern, einer Teilgruppe der Forscher, und ihrer Mobilität (anhand der Adressen der Erfinder) zulässt. Die Eignung von Patentdaten zur Erfassung von Forschermobilität wird detailliert in Kapitel 4.1.2 diskutiert.

#### **3.4 Ausländerzentralregister des Bundesamtes für Migration und Flüchtlinge (BAMF)**

Das Bundesamt für Migration und Flüchtlinge (BAMF) verfügt mit dem Ausländerzentralregister (AZR) über eine Datenbank, in der Informationen über Zuwanderer nach Deutschland erfasst werden. Die Datenbank enthält Informationen zu allen Einwanderern aus Nicht-EU-Staaten, sog. Drittstaaten, die für einen längerfristigen Aufenthalt im Bundesgebiet eine vorübergehende Aufenthalts- oder dauerhafte Niederlassungserlaubnis benötigen. Die im AZR zu speichernden Informationen sind in der Verordnung zur Durchführung des Gesetzes über das Ausländerzentralregister (AZRG-DV) festgelegt. Sie umfassen neben Namen, Geburtsdatum, Geburtsort, Staatsangehörigkeit und letztem Aufenthaltsort im Ausland keine persönlichen



Merkmale, die Aufschluss über Qualifikation oder Tätigkeit der einreisenden Personen geben. Es sind lediglich alle im Aufenthaltsgesetz definierten Aufenthaltszwecke im Detail abgelegt. Diese können indirekte Hinweise auf Personen geben, die eine Hochschulausbildung in Deutschland durchlaufen (§ 16 Abs. 1), eine solche abgeschlossen haben und in Deutschland eine Arbeit suchen (§ 16 Abs. 4), als anerkannte Fachkräfte im IT-Bereich oder in akademischen Berufen einreisen (§ 27 BeschV), als Hochqualifizierte nach § 19 AufenthG gelten oder entsprechend der Umsetzung der „EU-Forscherrichtlinie“ explizit als Forscher einreisen (§ 20 Abs. 1 AufenthG).

Im Sinne der Definition von Forschern in dieser Machbarkeitsstudie beschränkt sich der relevante Personenkreis auf die nach § 20 Abs. 1 AufenthG einreisenden Forscher. Hier sind Forscher, die an Hochschulen und öffentlichen Forschungseinrichtungen tätig sind, genauso erfasst, wie solche in der Privatwirtschaft. Informationen über den Arbeitgeber werden im AZR allerdings nicht gespeichert. Im Jahr 2011 wurden 317 dieser Aufenthaltserlaubnisse erteilt, davon 53 an Chinesen, 45 an Inder und 40 an Amerikaner (BAMF 2013). Im ersten Halbjahr 2012 belief sich die Zahl der Bewilligungen auf nur noch 84. Zum Erteilungszweck nach § 20 Abs. 5 AufenthG an Forscher aus Drittstaaten, die bereits in einem anderen EU-Mitgliedsstaat zugelassen sind, wurden keine Aufenthaltserlaubnisse erteilt (BAMF 2012). Diese Zahlen zeigen aufgrund ihrer geringen Höhe, dass sie für eine Beurteilung der Forschermobilität nach Deutschland weitgehend ungeeignet sind, da einreisende Forscher mit der Staatsangehörigkeit eines EU-Staates nicht erfasst werden. Außerdem wird die Zuwanderung deutscher bzw. die Abwanderung der betreffenden Personen nicht systematisch erfasst.

Neben der Feststellung der fehlenden Eignung als Datenquelle für die Forschermobilität sei an dieser Stelle jedoch auch eine inhaltliche Einordnung erlaubt. Die Größenordnung, in der sich die Zuwanderung von Forschern aus Nicht-EU-Staaten in Deutschland bewegt, ist für die größte Volkswirtschaft Europas äußerst gering. Die Zahl von 84 Aufenthaltsgenehmigungen für Forscher im ersten Halbjahr 2012 (BAMF 2012) deutet auf eine nach wie vor viel zu geringe Attraktivität des Forschungsstandortes Deutschland hin. Ebenfalls verschwindend gering sind die 75 Niederlassungserlaubnisse für Hochqualifizierte nach § 19 AufenthG (BAMF 2012). Diese Zahlen stehen in krassem Widerspruch zu der medialen Aufmerksamkeit, die diesen Paragraphen des Aufenthaltsgesetzes im Rahmen der Bekämpfung des Fachkräftemangels zugekommen sind.

Zusätzlich erfasst die Wanderungsstatistik des Statistischen Bundesamtes grenzüberschreitende Wanderungen anhand der Informationen der örtlichen Meldeämter. Die Abmeldung einer Person aus dem Bundesgebiet wird als Abwanderung gewertet, die erstmalige Anmeldung einer zu diesem Zeitpunkt noch nicht oder nicht mehr in Deutschland gemeldeten Person als Zuwanderung. Diese Daten kommen allerdings von vornherein nicht für ein Berichtssystem zur Forschermobilität in Frage, da es sich um eine fallbezogene Statistik handelt, die nur Mobilitätsereignisse festhält, nicht aber Aussagen über die Zuordnung von mehreren Mobilitätsereignissen der gleichen Person oder personenbezogene Merkmale, die eine Identifizierung von Forschern ermöglichen würden, enthält.

#### 3.5 Panelstudien zu Absolventen und Promovierenden

Panelstudien über Hochschulabsolventen werden in Deutschland bereits seit vielen Jahren durch das HIS-Institut für Hochschulforschung (HIS-HF) durchgeführt. Mit der Kohorte des Prüfungsjahrgangs 1997 wurden zum ersten Mal Absolventen in drei Wellen befragt und zwar 1 bzw. 1,5 Jahre, 3,5 bzw. 4,5 Jahre und 10 Jahre nach dem Abschluss. Die aktuellste Kohorte stützt sich auf die Antworten von ca. 10.000 Absolventen des Prüfungsjahrgangs 2009. Die Rücklaufquote der Befragung beträgt schätzungsweise 25 %. Die Absolventenbefragung des HIS-HF stellt die einzige repräsentative Datenbasis für alle Hochschulabsolventen in Deutschland dar. Potenziell sind diese Daten geeignet, um Aussagen über die Aufnahme einer Tätigkeit in der Forschung (getrennt nach öffentlichem und privatem Sektor) und den räumlichen Verbleib von Hochschulabsolventen zu machen.

Ein Problem ist allerdings die Erreichbarkeit von Absolventen, die Deutschland verlassen, so dass trotz der für viele andere Fragestellungen sehr gut geeigneten Methodik davon ausgegangen werden muss, dass eine vollständige Erfassung der Abwanderung von Hochschulabsolventen aus Deutschland mit Hilfe des Absolventenpanels nicht möglich ist. Enders und Bornmann (2002) merken dazu bereits selbstkritisch an, dass offen bleiben muss, ob mit solchen Befragungen ein repräsentativer Rücklauf im Hinblick auf die internationale Mobilität erreicht werden kann. Auch sie nehmen an, dass durch die schwierige Adressenlage die im Ausland lebenden Erwerbstätigen seltener erreicht werden.

Zuwanderung von Forschern nach Deutschland kann in Absolventenpanels nur für die Personen abgebildet werden, die vor der Aufnahme einer Tätigkeit in der Forschung in Deutschland einen Hochschulabschluss (Bachelor oder Master) erworben haben und dadurch in das Absolventenpanel aufgenommen werden. Trotz des großen Umfangs der Kohorten ist außerdem davon auszugehen, dass die Teilpopulationen der auswandernden Forscher und der als Studierende zugewanderten Absolventen, die später eine Tätigkeit in der Forschung aufnehmen, relativ gering ist. Zur Einschätzung der Größenordnung, in der sich die internationale Mobilität unter Akademikern insgesamt bewegt, bietet sich die Nutzung der Daten allerdings durchaus an. Für entsprechende Analysen zur Auswanderung wurden die Daten in der Vergangenheit u. a. von Arndt et al. (2010) genutzt.

Ein interessanter Ansatz, der allerdings noch in den Anfängen steckt, ist der Aufbau des Profile-Promovierendenpanels durch das Institut für Forschungsinformation und Qualitätssicherung (iFQ). Das Panel umfasst 6.195 Teilnehmer, von denen in der Initialbefragung 2011 lediglich 12 % zum Befragungszeitpunkt bereits promoviert waren. Weitere Befragungen sind bei Abschluss der Promotion und vier Jahre nach Abschluss der Promotion vorgesehen (Haus et al. 2012). Da der Fokus auf den aktuell Promovierenden liegt und die Panelbefragungen bereits vier Jahre nach Abschluss der Promotion enden sollen, werden Aussagen über den weiteren Karriereverlauf und spätere Wanderung ins Ausland auch in Zukunft nicht erfasst. Es sei denn, man würde die Panelpopulation auch zu einem späteren Zeitpunkt noch einmal befragen. Im Panel erfasst sind jedoch ausländische Promovierende in Deutschland, so dass nach weiteren Befragungswellen Aussagen zur Zuwanderung dieser Personengruppe und ihrem Verbleib nach der Promotion getroffen werden können. Die temporäre Mobilität von Promovierenden aus Deutschland ins Ausland im Rahmen der Promotion wird durch das Panel ebenfalls erfasst. Die Datenbasis über Promovierende in Deutschland bietet potenziell die Möglichkeit, das Mobilitätsverhalten von Promovierenden während ihrer Qualifizierungsphase zu erfassen,

sobald die Panelpopulation ihre Promotionsphase beendet hat. Die Probleme der Erreichbarkeit von Personen aus dem Panel, die ins Ausland wandern gelten allerdings für das Promovierendepanel des iFQ genauso wie für das Absolventenpanel des HIS. Repräsentative Untersuchungen zu Promovierenden werden darüber hinaus dadurch erschwert, dass die Grundgesamtheit dieser Personen nicht eindeutig abgrenzbar ist, wie Hornbostel (2012) ausführlich erläutert.

Der wissenschaftliche Nachwuchs wird ebenfalls vom HIS im Rahmen des sogenannten WiNbus-Panels befragt, bei dem es sich allerdings nicht um ein echtes Panel handelt, sondern lediglich um eine Personengruppe, die als wissenschaftlicher Nachwuchs identifiziert wurde und die regelmäßig zu verschiedenen Themen befragt wird. WiNbus ist daher als herkömmliche Befragung einzuordnen, die sich lediglich immer wieder an den gleichen Personenkreis wendet, die aber im Sinne dieser Machbarkeitsstudie nicht als innovative Methode gewertet wird. Das Projekt wird vom Bundesministerium für Bildung und Forschung (BMBF) gefördert und dient vor allem dazu, vertiefte Einblicke in die Wahrnehmung und Beurteilung des Wissenschaftsstandortes Deutschland zu gewinnen. Bisher wurden drei Befragungen durchgeführt, die den Themen (i) Einstellung des wissenschaftlichen Nachwuchses zum Bologna-Prozess, (ii) Karriereorientierung und (iii) internationale Vernetzung, Ausrichtung und Mobilität gewidmet waren. Die Ergebnisse zum letztgenannten Thema sind bei Jaksztat et al. (2011) dokumentiert und bieten spannende Einblicke in die Wahrnehmung und Beurteilung internationaler Mobilität durch den wissenschaftlichen Nachwuchs, sind allerdings nicht geeignet, um die in der Einleitung dieser Machbarkeitsstudie genannten Fragestellungen zum Thema Forschermobilität zu beantworten. Die Ergebnisse können allerdings in qualitativer Hinsicht einen Beitrag zur Einordnung der Bedeutung von Mobilität in der Karriereplanung des wissenschaftlichen Nachwuchses und zu den Motiven liefern.

#### **3.6 Kerndatensatz Forschung**

Im Januar 2013 hat der Wissenschaftsrat (2013) seine Empfehlungen zur Erstellung eines Kerndatensatz Forschung veröffentlicht. Darin wird ein auf 18 Monate angelegtes Vorhaben skizziert und angestoßen, in dem die Datenerfassung und Datenhaltung von Forschungsdaten unter Nutzung des CERIF-Standards (Common European Research Information Format) vereinheitlicht werden soll. Bei dem Projekt geht es nicht um den Aufbau neuer zentraler Datenbanken, die vor allem in skandinavischen Ländern als sogenannte Forscherregister geführt werden (Kapitel 4.2.4), sondern um die Verknüpfung von Daten die bereits an verschiedenen Stellen innerhalb des Forschungssystems vorliegen, z. B. Hochschulen, Forschungseinrichtungen, Publikationsdatenbanken, Forschungsförderer. Diese sollen durch Standardisierung miteinander verknüpft werden. Vom Wissenschaftsrat genannte Inhalte eines solchen Kerndatensatzes sind Informationen über Wissenschaftler, Nachwuchsförderung, Drittmittel und Projekte, Publikationen, Patente, Forschungspreise und Spin-Offs. Das bislang dominierende Ziel des Kerndatensatz Forschung ist die Schaffung einer verbesserten Grundlage für nach Einrichtungen, Fächern und Lehrstühlen differenzierende Evaluationen.

Untersuchungen der Mobilität von Forschern und ihrer Karriereverläufe sind bislang nicht explizit als Anwendungsfeld des Kerndatensatz Forschung genannt. Ein solcher Datensatz bietet jedoch potenziell sehr gute Möglichkeiten, die im Rahmen der Machbarkeitsstudie aufge-

worfenen Fragestellungen für Forscher an Hochschulen und öffentlichen Forschungseinrichtungen zu untersuchen. Dafür müssen allerdings einige Voraussetzungen erfüllt sein:

- Die Informationen müssen als echte Individualdaten abrufbar sein. Die Erfassung auf der Ebene von Lehrstühlen, die vom Wissenschaftsrat als Minimalforderung gestellt wird, ist für diesen Zweck unzureichend.
- Es muss eine Möglichkeit gefunden werden, eine eindeutige Zuordnung von Personen auch beim Wechsel der Einrichtung vorzunehmen. Dies kann beispielsweise über eine eindeutige Identifikationsnummer geschehen, die dem Forscher für seine gesamte Karriere zugeordnet wird. Ein solches System existiert in den Niederlanden mit dem Digital Author Identifier (DAI).
- Die gespeicherten Individualdaten sollten Informationen zum Karriereverlauf in Form von Lebenslaufdaten enthalten, die Aussagen über Orte und Einrichtungen der Ausbildung sowie früherer Tätigkeiten beinhalten. Anhand dieser Daten lassen sich Aussagen über Mobilitätsereignisse in der Vergangenheit treffen. Die Forschungsinformationssysteme einiger Universitäten, z. B. der WWU Münster (Herwig und Becker 2012), beinhalten bereits Lebenslaufinformationen.
- Abwanderung aus Deutschland selbst ist auch in einem Kerndatensatz Forschung schwer erfassbar. Hier wäre die Erfassung des Grundes für das Ausscheiden einer Person aus dem Bereich der erfassten Einrichtungen in Deutschland sinnvoll. Diese Information könnte durch die abgebende Einrichtung erfasst werden, sofern diese darüber Kenntnis hat. Alternativ könnte in diesem Fall eine automatisierte Versendung einer Kurzbefragung ausgelöst werden, um sicherzustellen, dass Mobilität ins Ausland eindeutig von anderen Gründen für das Verschwinden einer Person aus dem Kerndatensatz abgegrenzt werden kann.

Es wird empfohlen, sich am Prozess der Spezifikation des Kerndatensatzes zu beteiligen. Der vom Wissenschaftsrat vorgeschlagene Projektablauf sieht vor, dass in den Projektmonaten drei bis sechs Expertenwissen zur Festlegung der Inhalte im Rahmen von Anhörungen eingeholt wird. Hier sollte die Relevanz des Themas Forschermobilität für den Kerndatensatz artikuliert und auf die dazu notwendigen Inhalte hingewiesen werden.

Von einer schnellen bundesweiten Einführung des Kerndatensatzes ist jedoch vorerst nicht auszugehen. Die Ergebnisse der ersten Projektphase sollen zunächst an Piloteinrichtungen getestet werden. Selbst wenn diese Tests erfolgreich verlaufen, ist aufgrund der Zuständigkeit der Länder für die Hochschulen und der großen Autonomie der Einrichtungen nicht von einer schnellen flächendeckenden Einführung auszugehen. Gerade wegen der noch nicht absehbaren Barrieren für eine Implementierung sollte das Vorhaben in seiner Frühphase von Seiten der Hochschul- und Innovationsforschung intensiv unterstützt werden, da es sich um einen äußerst vielversprechenden Ansatz zur Behebung des Mangels an zentral vorliegenden Forschungsinformationen in Deutschland handelt.

#### 3.7 Zwischenfazit

Die Diskussion deutscher Datensätze und Datenquellen zur Forschermobilität hat gezeigt, dass an vielen Stellen Informationen zu Teilaspekten des Phänomens vorliegen, es aber keinen Datensatz gibt, der das Phänomen vollständig für alle in der Einleitung genannten Fragen erfasst. Ein zentrales Problem bei der Beschränkung auf deutsche Daten besteht darin, dass ein internationales Phänomen darin nur teilweise abgebildet werden kann. Deutsche, die ins Ausland abwandern, sind in keinem der Datensätze in zufriedenstellender Weise erfasst. Daher werden im folgenden Kapitel internationale Datensätze und Datenquellen dahingehend geprüft, inwiefern dort Informationen zu deutschen Forschern im Ausland enthalten sind. Ein besonderes Augenmerk wird dabei auf solchen Datenquellen liegen, die entweder sowohl Forscher in Deutschland als auch im Ausland enthalten (Kapitel 4.1) oder die Hauptzielländer der deutschen Forschermobilität abdecken und eine Vergleichbarkeit mit deutschen Datensätzen besitzen (Kapitel 4.2).

Von den deutschen Datensätzen ist der Mikrozensus am besten zur Erfassung der Forschermobilität geeignet. Er erfüllt als repräsentative Umfrage der deutschen Wohnbevölkerung am ehesten das Kriterium einer verzerrungsfreien Erfassung des Phänomens für alle Tätigkeitsfelder und Berufsgruppen. In Kapitel 4.2.1 wird daher diskutiert, ob und wie sich ein internationaler Vergleich methodisch ähnlich gelagerter Befragungen durchführen lässt, der z. B. auf Arbeitskräfteerhebungen in Hauptzielländern der deutschen Forschermobilität zurückgreift.

Nicht weiter verfolgt werden die folgenden Datenquellen vor allem aus den hier kurz zusammengefassten Gründen:

- SOEP aufgrund der zu geringen Zahl an erfassten Forschern.
- IEB aufgrund der fehlenden Identifikation von Mobilität und der Nicht-Erfassung wichtiger Berufsgruppen (Beamte, Selbständige).
- AZR aufgrund der geringen Fallzahl von Forschern, der Begrenzung auf Nicht-EU-Ausländer und fehlender personenbezogener Informationen.
- Absolventenpanels aufgrund der starken Eingrenzung der erfassten Personengruppe, die Zuwanderer nach dem Studium nicht berücksichtigt, und der Probleme bei der Nachverfolgung von Abwanderern. Es wird allerdings empfohlen, die enthaltenen Daten zum Thema Mobilität in Sonderauswertungen in regelmäßigen Abständen zur Verfügung zu stellen und ihnen in den folgenden Wellen einen hinreichend großen Stellenwert einzuräumen.
- Kerndatensatz Forschung aufgrund der noch nicht absehbaren Implementierung. Es wird allerdings empfohlen, sich aktiv während der Phase der Projektdefinition einzubringen, damit das Thema Forschermobilität mit dem Datensatz abgebildet wird und in Zukunft auch tatsächlich untersucht werden kann.

### 4 Internationale Datensätzen und Datenquellen

#### 4.1 Internationale Daten

##### 4.1.1 Publikationsdaten

Die für die Forschermobilität relevante Information in Publikationsdatenbanken ist die Angabe der Mitgliedschaft der individuellen Autoren zu bestimmten Organisationen. Über den Sitz der angegebenen Organisation können für die Untersuchung räumlicher Mobilität Rückschlüsse auf den Arbeitsort der Autoren getroffen werden. Mobilität zwischen verschiedenen Organisationen kann über die in einer späteren Publikation angegebene veränderte Mitgliedschaft direkt abgelesen werden.

Für die Untersuchung von Mobilität in Publikationsdatenbanken wird zunächst eine Ausgangspopulation an Forschern definiert, z. B. alle Forscher, die in einem bestimmten Jahr bzw. Zeitraum in der Vergangenheit mit einer Publikation in der genutzten Datenbank vermerkt sind. Zu den so identifizierten Forschern werden alle Publikationen zugeordnet, die sie in der Folgezeit bis zum aktuellen Rand veröffentlicht haben. Aus den darin enthaltenen Informationen zur organisationalen Mitgliedschaft und zum Datum der Veröffentlichung wird der Karriereverlauf rekonstruiert. Auf Basis dieser Informationen können Aussagen zur Forschermobilität nach folgendem Muster getroffen werden:

- Abwanderung: Wie viele der Forscher, die im Ausgangsjahr in Land x publiziert haben, sind in der Folgezeit in ein anderes Land gewandert? In welche Länder sind sie gewandert?
- Zuwanderung: In welchen Ländern waren die Forscher in der Vergangenheit aktiv, die heute in Land x aktiv sind? Wie groß ist ihr Anteil an den gesamten Forschern von Land x?
- Rückwanderung: Wie viele Forscher die im Ausgangsjahr und am aktuellen Rand in Land x publiziert haben, waren zwischenzeitlich in einem anderen Land aktiv?

Analysen zu Fragestellungen dieser Art sind von Moed et al. (2013) nach dem Informationsstand des Autors erstmals in einem wissenschaftlichen Journal publiziert worden und werden für Deutschland aktuell durch das Fraunhofer-Institut für System und Innovationsforschung (ISI) im Rahmen des Kompetenzzentrums Bibliometrie durchgeführt.

Als Datenbasis für Untersuchungen dieser Art bietet sich die Publikationsdatenbank SciVerse Scopus an, die von der Elsevier-Gruppe betrieben wird. Theoretisch wäre auch die Web of Science-Datenbank von Thomson Reuters geeignet, allerdings besitzt Scopus Vorteile bei der eindeutigen Zuordnung von Publikationen zu Autoren. Für alle Publikationsanalysen ist die korrekte Unterscheidung von namensgleichen Personen (Homonyme) und die korrekte Zuordnung von Publikationen zu Personen die größte Herausforderung. In Scopus werden Publikationen durch einen Algorithmus automatisch einem Autorenprofil zugeordnet. Dieses wird mit einer eindeutigen Nummer versehen und kann für weitere Analysen genutzt werden.

Entscheidend für die Qualität der auf Basis von Publikationsdaten durchgeführten Analysen ist die Informationsdichte, die aus diesen Daten gewonnen werden kann. Scopus deckt dabei mehr Zeitschriften ab, als die Web of Science-Datenbank und bleibt trotz der Zugehörigkeit zur Elsevier-Gruppe, nicht auf Zeitschriften dieses Verlags beschränkt. Auch nicht-englischsprachige Zeitschriften und solche, zu denen online nicht einmal die Inhaltsverzeich-

nisse verfügbar sind, werden erfasst, so dass davon auszugehen ist, dass zu Autoren, die regelmäßig in wissenschaftlichen Fachzeitschriften publizieren, in Scopus eine hinreichende Informationsdichte für die genannten Analysen vorliegt. Nicht erfasst werden jedoch Disziplinen, in denen Forschungsergebnisse nicht in wissenschaftlichen Fachzeitschriften, sondern in Buchform publiziert werden. Hier handelt es sich häufig um geisteswissenschaftliche Disziplinen. Außerdem sind es vor allem Forscher an Universitäten und öffentlichen Forschungseinrichtungen, die ihre Forschungsergebnisse in Fachzeitschriften publizieren, so dass die Erfassbarkeit von Forschern in der Privatwirtschaft mit dieser Methode nicht in umfassendem Maße möglich scheint.

Die Methode bietet jedoch gegenüber den bislang genannten nationalen Quellen einige wesentliche Vorteile. Es handelt sich um Daten, die unabhängig vom Aufenthaltsort des Forschers seinem Autorenprofil zugeordnet werden können. Damit können Karriereverläufe abgebildet werden und Zuwanderung, Abwanderung und Rückwanderung untersucht werden. Auswirkungen von Mobilität können ebenfalls direkt mit Hilfe der Daten untersucht werden, indem internationale Kooperationen (Ko-Autorenschaften) oder Produktivitätsmaße (bibliometrische Kennziffern) ausgewertet werden (vgl. Kapitel 6.1 und 6.2).

Dennoch unterliegt auch diese Methode einigen wichtigen Einschränkungen bezüglich der möglichen Aussagen. Ein wesentlicher Nachteil ist, dass Personen erst durch ihre erste Publikation erfasst werden. Informationen zum Geburtsort, zur Nationalität oder zu Orten der Ausbildung liegen daher nicht vor. Damit können keine Aussagen über die Mobilität in der Frühphase der Forscherkarriere getroffen werden, da die erste Publikation wissenschaftlicher Ergebnisse in der Regel erst nach oder während der Promotion erfolgt. Eine Aussage darüber, wie viele deutsche Forscher in den USA tätig sind, ist auf Basis dieser Daten also nicht möglich. Die Aussage müsste sich korrekterweise auf die Anzahl der in den USA tätigen Forscher, die in der Vergangenheit in Deutschland publiziert haben, beziehen. Ob es sich dabei um abgewanderte Deutsche, rückgewanderte Amerikaner oder weitergewanderte Forscher anderer Nationen handelt, ist anhand der Daten nicht feststellbar.

Weitere Probleme können sich aus der Nutzung der organisationalen Zugehörigkeit, die in der Publikation angegeben ist, ergeben. Diese kann durch die lange Dauer von Einreichung bis Publikation von Artikeln in vielen Fächern nicht mehr aktuell sein bzw. Artikel, die zu einem späteren Zeitpunkt unter einer neuen Zugehörigkeit geschrieben worden sind, können früher veröffentlicht werden als solche die noch unter einer alten Zugehörigkeit geschrieben worden sind. Die Auswertungslogik würde dies fälschlicherweise als mehrfache Mobilität erkennen. Eine weitere Schwierigkeit besteht im Falle multipler Zugehörigkeiten einzelner Autoren (z. B. im Rahmen von Gastprofessuren), die dann zum Problem werden, wenn beide Zugehörigkeiten im Artikel genannt werden oder wenn nur die Gastprofessur genannt wird. In diesen Fällen könnte Mobilität angenommen werden, obwohl die betreffende Personen dauerhaft an ein und demselben Ort tätig war. Da die Informationsdichte zusätzlich von der Anzahl der Publikationen abhängt, werden kürzere Aufenthalte an Standorten, an denen keine Publikationen entstanden sind, nicht erfasst.

Trotz der genannten Nachteile handelt es sich bei der Nutzung von Publikationsdatenbanken zur Untersuchung der Forschermobilität prinzipiell um eine sehr vielversprechende Methode, da die Daten personenscharf über einen langen Zeitraum und für alle Arten der Zuwanderung vorliegen. Es sollte daher in der Praxis getestet werden, welche methodischen Probleme in

welchem Umfang auftreten und ob diese gelöst werden können. Außerdem sind die aus Publikationsdatenbanken gewonnenen Informationen in sehr guter Weise geeignet, um mit anderen Verfahren der automatischen Generierung von Autorenprofilen unter Nutzung von im Internet frei zugänglichen Informationen kombiniert zu werden (vgl. Kapitel 4.1.6).

In diesem Zusammenhang ist noch anzumerken, dass GoogleScholar im Vergleich zu Scopus und der Web of Science-Datenbank keine Publikationsdatenbank ist, die man in gleicher Weise auswerten kann. Es werden nicht alle Artikel systematisch gespeichert und Personen zugeordnet, sondern Personen können sich Artikel zu ihrem Profil zuordnen lassen. Diese Zuordnung erfolgt durch Google automatisch durch Suchalgorithmen, sofern der betreffende Autor aktiv ein Profil anlegt. In Scopus erfolgt die Erstellung eines Autorenprofils automatisch, sobald eine erste Publikation eines neuen Autors erkannt wird. Die Informationen stehen bei GoogleScholar außerdem nicht für externe Analysen zur Verfügung. Die begrenzten Möglichkeiten, GoogleScholar für Analysen der Forschermobilität zu nutzen, werden in Kapitel 4.1.6 unter dem Thema Natural Language Processing noch einmal ausführlich aufgegriffen.

### 4.1.2 Patentdaten

Mit Hilfe von Patentdaten sind grundsätzlich ähnliche Analysen möglich wie mit Publikationsdaten. Über den Wohnort des Erfinders können ebenfalls Angaben zur räumlichen Mobilität von Forschern gemacht werden. Die organisationale Zugehörigkeit des Erfinders kann auch den Wechsel von Arbeitgebern abbilden. Patentdaten ermöglichen es außerdem besser, Forscher in der Privatwirtschaft zu erfassen, da diese kommerziell verwertbares Wissen häufiger in Form von Patenten schützen lassen.

In stärkerem Maße als bei den Publikationsdaten schlägt hier allerdings die Selektivität negativ zu Buche. Patente sind fast ausschließlich für naturwissenschaftlich-technische Disziplinen von Bedeutung und verteilen sich auch dort sehr unterschiedlich zwischen den einzelnen Teilgebieten. Außerdem finden Patentierungen im Vergleich zu Publikationen seltener statt, so dass die Informationsdichte dieser Daten deutlich geringer ist.

Da alle übrigen positiven und negativen Bewertungen der Methode im Wesentlichen aus der Diskussion zu Publikationsdaten übertragbar sind, erscheinen Patentdaten weniger gut als Publikationsdaten geeignet, um als alleinige Datenquelle für Mobilitätsanalysen herangezogen zu werden. Aufgrund der besseren Abdeckung von Forschern in der Privatwirtschaft lassen sie sich jedoch mit anderen Informationen zu diesem Personenkreis kombinieren. In Kapitel 3.3 wurde bereits darauf hingewiesen, dass in einem aktuellen Projekt des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) und dem Max-Planck-Institut für Immaterialgüter- und Wettbewerbsrecht Patentdaten mit den IEB-Daten verknüpft werden.

### 4.1.3 Forscherdatenbanken

In den letzten Jahren wurden zwei große Projekte gestartet, die versuchen über die auf Basis der Publikationsdaten erstellten Autorenprofile hinaus Forscherdatenbanken mit weitergehenden Informationen zu Lebensläufen, Forschungsprojekten, Publikationen, Patenten und anderen wissenschaftlichen Leistungskriterien zu erstellen. Seit 2008 existiert die Plattform ResearcherID von ThomsonReuters und seit 2012 ist die Plattform ORCID am Markt, die als Non-Profit-Organisation angelegt und auf Spenden der großen Verlagshäuser und wissenschaftlicher Vereinigungen angewiesen ist. ResearcherID umfasst dabei bereits mehrere hun-



dertausend Personen (davon ca. 9.000 aus Deutschland), während ORCID für November 2012 erst eine Zahl von 12.000 Profilen berichtete.

Diese Datenbanken bieten den Vorteil, dass neben den automatisch aus dem Web of Science bzw. der Scopus-Datenbank eingelesenen Publikationen weitere Informationen vom Forscher selbst eingetragen werden können. Von besonderem Interesse für die Forschermobilität sind dabei Lebenslaufdaten. Für die Untersuchung von Interdependenzen zwischen Mobilität und wissenschaftlicher Leistungsfähigkeit bieten sich Angaben zu Forschungsprojekten an. Über die Angaben zu Ko-Autoren oder Partnern in Forschungsprojekten sind auch Informationen zur Vernetzung enthalten.

Die Nutzung von Daten aus diesen Forschungsdatenbanken kann zukünftig eine interessante Ergänzung zu den bereits diskutierten Publikations- und Patendatenbanken darstellen. Allerdings müssten dafür zwei Voraussetzungen erfüllt sein, die zum jetzigen Zeitpunkt noch nicht gegeben sind. Erstens muss die Teilnehmerzahl an diesen Datenbanken in quantitativer Hinsicht noch stark anwachsen, um wirklich belastbare Aussagen machen zu können. ResearcherID ist aufgrund der längeren Existenz hier deutlich weiter. Durch die breitere Abdeckung, den Non-Profit-Ansatz und die Kooperation mit ResearcherID sind die Wachstumschancen für ORCID allerdings künftig durchaus vorhanden. Da die Registrierung jedoch freiwillig stattfindet und von der Initiative der einzelnen Forscher abhängig ist, wird die Abdeckung immer selektiv bleiben.

Zweitens ist eine Erhöhung des Informationsgehalts in qualitativer Hinsicht notwendig. Bislang werden lediglich Publikationen automatisch und damit systematisch einem Profil zugeordnet. Um einen Mehrwert gegenüber Publikationsdatenbanken zu erzielen, werden jedoch insbesondere Lebenslaufdaten benötigt. Da diese Angaben freiwillig sind und keinem festen Format folgen, ist die Herausbildung einer für systematische Auswertungen hinreichend belastbaren Informationsbasis momentan nicht absehbar. Aus Sicht des Autors ist es aber durchaus lohnenswert, die Entwicklungen bei ResearcherID und ORCID zu verfolgen und in regelmäßigen Abständen die Qualität der enthaltenen Daten zu prüfen. Auch diese Daten eignen sich für eine Integration in automatisierte Verfahren der Informationsgewinnung (vgl. Kapitel 4.1.6).

Auf Basis eines Vorläuferdatensatzes von ResearcherID, der HighlyCited-Datenbank von Thomson Reuters, haben der Autor und andere Forscher bereits erfolgreich Auswertungen zum Mobilitätsverhalten von Spitzenforschern durchgeführt (Schiller und Revilla Diez 2008, 2010, 2012; Trippel 2013). In dieser Datenbank wurden Informationen zu den meistzitierten Forschern nach Fachbereichen abgelegt. Insgesamt umfasste die Datenbank ca. 5.500 Forscher, von denen ca. 250 zum Zeitpunkt der Erfassung in Deutschland tätig waren. Der Vorteil dieser Datenbank war, dass die Profile auf Basis der in der Web of Science-Datenbank gespeicherten Publikationen von Thomson Reuters selbst angelegt wurden. Die erfassten Forscher wurden dann informiert, dass sie zu den meistzitierten ihres Fachbereichs zählen und weitere Informationen zum Lebenslauf wurden angefragt. Auf diesem Weg konnten immerhin zu 2.278 Personen Informationen zu ihrem Geburtsort und weiteren Lebenslaufdaten erfasst werden, die für die Analyse des Migrationsverhaltens genutzt werden können. Eine Weiterführung dieser Datenbank hat nach Einführung von ResearcherID allerdings nicht mehr stattgefunden.

### 4.1.4 Professionelle soziale Netzwerke

Professionelle soziale Netzwerke gewinnen zunehmend für die Darstellung der eigenen Qualifikationen an Bedeutung. Im internationalen Raum besitzt momentan LinkedIn die größte Verbreitung im Bereich der allgemeinen professionellen sozialen Netzwerke. Für Forscher hat sich ResearchGate zur am weitesten verbreiteten Plattform entwickelt. Da bei ResearchGate vor der Aufnahme in das Netzwerk eine Prüfung erfolgt, ob die betreffende Person tatsächlich ein Forscher ist, kann davon ausgegangen werden, dass der dort registrierte Personenkreis unmittelbar zur Untersuchung der Forschermobilität herangezogen werden kann. Bei LinkedIn sind dagegen Beschäftigte aus verschiedenen Berufsfeldern aktiv, so dass zunächst die Teilgruppe der Forscher abgegrenzt werden müsste, bevor die Daten analysiert werden können (z. B. über die Beschreibung der aktuellen Tätigkeit durch den Nutzer). Bei ResearchGate sind bereits 2,5 Millionen Forscher erfasst. Damit ist das Netzwerk bei weitem größer als die zuvor beschriebenen Forscherdatenbanken. Die Zahl der Nutzer bei LinkedIn wird auf 200 Millionen beziffert, so dass die Zahl der dort registrierten Forscher unter der Annahme eines Anteils von Forschern an den Gesamtbeschäftigten von 1 % und einem Bias unter den LinkedIn-Nutzern zugunsten von Tätigkeiten mit höherer Qualifikation eher größer sein dürfte als bei ResearchGate.

Beide Datenbanken bieten die Möglichkeit, neben der aktuellen Tätigkeit auch weitere Lebenslaufdaten in strukturierter Form zu hinterlegen. Hierbei kann es sich um Angaben zur Ausbildung oder zu früheren Tätigkeiten handeln. An den dabei genannten Orten könnte eine Untersuchung der Mobilität ansetzen. Bei ResearchGate werden außerdem Publikationen und weitere Angaben zu wissenschaftlichen Leistungen teilweise automatisch, teilweise durch den Nutzer selbst gespeichert. Die Erfassung erscheint allerdings weniger treffsicher und automatisierbar als bei den Forscherdatenbanken ResearcherID und ORCID. Grundsätzlich ist jedoch auf diese Weise eine Verknüpfung mit Indikatoren der wissenschaftlichen Leistungsfähigkeit möglich. Das zentrale Element sozialer Netzwerke sind jedoch die Verbindungen zwischen Personen, die sowohl bei LinkedIn als auch bei ResearchGate vorhanden sind. Hier könnten Untersuchungen über den Zusammenhang von Forschermobilität und Forschernetzwerken ansetzen. Bislang sind dem Autor jedoch keine Untersuchungen bekannt, die Daten aus sozialen Netzwerken zur Mobilitätsanalyse nutzen.

Ein zentrales Problem bei der Nutzung von Daten aus professionellen sozialen Netzwerken ist die Möglichkeit des Zugangs zu den Daten. Für viele Profile sind nur wenige oder gar keine Informationen durch die Nutzer als öffentlich zugänglich definiert worden, so dass mit den Betreibern der Netzwerke selbst ein Zugang zu den geschützten Daten vereinbart werden müsste. Dies schließen jedoch die Datenschutzvereinbarungen oder das Geschäftsgebaren der Betreiber aus. Darüber hinaus sind alle enthaltenen Angaben freiwillig und durch den Nutzer einzugeben. Daher kann nicht davon ausgegangen werden, dass vollständige Informationen zu den Karriereverläufen der Personen vorliegen. Die Stärken dieser Daten dürften daher eher nicht in ihrer alleinigen Nutzung als Datenquelle liegen, sondern in der Zusätzlichkeit als Quelle für standardisierte Lebenslaufinformationen. Die Nutzung solcher Daten aus öffentlichen Profilen von LinkedIn und ResearchGate wird im folgenden Kapitel diskutiert.

### 4.1.5 Natural Language Processing (NLP)

In den vorangehenden Kapiteln zu internationalen Datenquellen ist deutlich geworden, dass der Vorteil dieser Ansätze in der internationalen Abdeckung der Forschermobilität liegt. Auch Personen, die aus dem Erfassungskreis nationaler Daten abwandern, können weiter verfolgt werden. Häufig auftretende Probleme sind allerdings die Selektivität der erfassten Informationen (z. B. nur organisationale Zugehörigkeiten anhand von Publikationen) und die Freiwilligkeit der Angabe von weiteren mobilitätsrelevanten Informationen (z. B. Lebenslaufdaten bei ORCID bzw. ResearcherID oder LinkedIn bzw. ResearchGate). Da allerdings eine personenscharfe, namentliche Erfassung von Forschern erfolgt, bietet sich bei diesen Daten anders als bei anonymisierten Daten aus dem Mikrozensus die Möglichkeit, Informationen aus verschiedenen Quellen zu verknüpfen.

Ein besonders großes Potenzial scheint aufgrund der immer weiter wachsenden Zahl frei zugänglicher persönlicher Informationen im Internet im automatisierten Auslesen (Data Mining) aus mehr oder weniger strukturierten Datenbestände mit Hilfe der Technik des Natural Language Processing (NLP) zu liegen. Durch diese Verfahren werden zu einer vorher definierten Ausgangspopulation von Forschern Informationen zu Mobilitätsereignissen aus verschiedenen Quellen im Internet gesucht. Auf Basis der gefundenen Informationen wird dann versucht, den Karriereweg zu rekonstruieren. Ein erster Machbarkeitsnachweis für dieses Verfahren wurde von García Flores et al. (2012) erbracht.

Es ist davon auszugehen, dass vor allem für Wissenschaftler an Universitäten und Forschungseinrichtungen eine Vielzahl von Informationen frei zugänglich ist, da in diesen Bereich die Notwendigkeit zur individuellen Selbstvermarktung und Erhöhung der Sichtbarkeit besonders groß ist. Es besteht ein persönlicher Anreiz, Informationen wie Lebensläufe auf persönlichen Homepages anzugeben. Darüber hinaus beschränken Universitäten und öffentliche Forschungseinrichtungen die öffentliche Darstellung von Informationen über Forschungsaktivitäten ihrer Mitarbeiter nicht, denn die Reputation einer Einrichtung ergibt sich letztlich aus der Summe der Leistungen der Mitarbeiter. In der Privatwirtschaft sieht dies tendenziell anders aus. Hier sind in der Regel keine Angaben über einzelne Mitarbeiter in FuE-Abteilungen verfügbar, da die Arbeiten aufgrund der kommerziellen Verwertung der Forschungsergebnisse der Geheimhaltung unterliegen. In professionellen sozialen Netzwerken sind jedoch Forscher aus der Privatwirtschaft registriert, so dass auch hier einige Informationen vorliegen dürften. Viele der vorhandenen Daten liegen außerdem in vergleichsweise strukturierter Form vor, z. B. tabellarische Lebensläufe. Diese können durch die in Kapitel 5.2 näher beschriebenen technischen Verfahren ausgelesen werden. Damit würde die Lebenslaufanalyse, die in der Wissenschaftsevaluation in jüngster Zeit an Bedeutung gewonnen hat (Cañibano und Bozeman 2009), auf eine weitaus größere Basis gestellt werden.

Insgesamt besitzen NLP-Verfahren aufgrund der Vielzahl von persönlichen Angaben im Internet ein großes Potenzial, die Selektivität von einzelnen Datenbanken und Netzwerken durch Kombination mobilitätsrelevanter Informationen zu einzelnen Personen zu überwinden. Daher wird in Kapitel 5.2 erläutert, wie ein Projekt zur Nutzung von NLP-Verfahren für die Forschermobilität aufgebaut sein könnte, welche Chancen und Risiken sich dabei ergeben und welche Ressourcen dafür nötig sind.

### 4.2 Nationale Daten der Hauptzielländer deutscher Forschermobilität

#### 4.2.1 Arbeitskräfteerhebungen anderer Länder

Neben der Nutzung internationaler Datensätze bieten sich auch nationale Daten der Hauptzielländer deutscher Forschermobilität für eine Untersuchung der Abwanderung von Forschern aus Deutschland an. Im Zwischenfazit zu Kapitel 3 wurde herausgearbeitet, dass die Daten des deutschen Mikrozensus potenziell die beste nationale Datenquelle für Forschermobilität in Deutschland sind. Da diese Erhebung einem internationalen Standard unterliegt, sind die Ergebnisse mit ähnlich gelagerten Erhebungen anderer Länder vergleichbar. Besonders gut geeignet sind hierfür die nationalen Arbeitskräfteerhebungen der Hauptzielländer deutscher Forschermobilität bzw. die konsolidierten Daten der europäischen Arbeitskräfteerhebung. Letztere haben jedoch den Nachteil, dass aufgrund von Anonymisierungsvorschriften statt einzelner Länder nur Ländergruppen hinterlegt sind, die eine eindeutige Zuordnung von Personen zu einem bestimmten Geburtsland oder einer bestimmten Nationalität nicht ermöglichen. In der europäischen Arbeitskräfteerhebung sind daher deutsche Forscher oder Forscher aus Deutschland nicht erkennbar. Es können lediglich die Zahlen der zugewanderten Forscher und die Herkunftsstruktur nach Ländergruppen im Vergleich ausgewertet werden. In den nationalen Erhebungen sind jedoch detailliertere Länderdaten enthalten. Allerdings eignen sich hier vor allem die Daten größerer Länder, da nur dort für kleine Subpopulationen wie ausländische Forscher, eine hinreichend große Fallzahl erreicht wird. Folgende drei mit dem Mikrozensus vergleichbare Datensätze sind dafür beispielhaft zu nennen:

- USA: Public Use Microdata Sample (PUMS) des American Community Survey (ACS), ca. 3.000.000 Personen pro Jahr, ca. 90.000 Forscher
- Großbritannien: Labour Force Survey, ca. 90.000 Personen pro Jahr, ca. 2.700 Forscher
- Frankreich: Enquête sur l'Emploi, ca. 85.000 Personen pro Jahr, ca. 2.500 Forscher

Eine Abgrenzung von Forschern ist auch in diesen Datensätzen wie im deutschen Mikrozensus aus der Kombination von Beruf, Branche und Qualifikation möglich. Auch für diese Länder kann die Fallzahl durch das Poolen verschiedener Erhebungen erhöht werden.

Untersuchungen für kleinere Länder scheitern an den zu geringen Fallzahlen, die eine Differenzierung der Forscherpopulation nach weiteren Merkmalen (z. B. Herkunftsland, Nationalität) unmöglich machen. Dies gilt insbesondere für Deutschlands kleinere Nachbarländer, in denen deutsche Forscher einen hohen Anteil an der Gesamtheit der Forscher ausmachen (Schweiz, Österreich, Niederlande).

In Kapitel 5.1 wird ein international vergleichendes Forschungsdesign auf Basis von Daten des Mikrozensus und verschiedener Arbeitskräfteerhebungen detailliert erläutert. Zu den wenigen Publikationen, die bisher diesen Ansatz genutzt haben, zählen Diehl und Dixon (2005) sowie im Rahmen des EU-Projekts MORE (Mobility and career paths of researchers in Europe) die Studie zur Mobilität von Forschern in der Industrie (IDEA Consult 2010a), in denen jedoch noch kein systematischer Vergleich des Mobilitätsverhaltens deutscher Forscher in verschiedenen Ländern durchgeführt wurde.

### 4.2.2 Daten von Einwanderungsbehörden

Analog zu Daten über Nicht-EU-Ausländer im Ausländerzentralregister des BAMF werden auch in anderen Ländern Einreiseinformationen über Deutsche erhoben, die für Analysen der Forschermobilität von Interesse sein können. Da innerhalb der EU für alle Deutschen die freie Wahl des Wohnsitzes und des Arbeitsortes gilt, sind diese Analysen nur für Nicht-EU-Staaten geeignet. Unter den Hauptzielländern der deutschen Forschermobilität beschränkt sich der Kreis der relevanten Länder damit alleine auf die USA.

Für die Forschermobilität sind hier insbesondere die temporäre Arbeitserlaubnis der Visaklasse H-1B (Specialty Occupations) und die permanente Aufenthaltserlaubnis der Klasse EB-1 (Employment-Based Immigration: First Preference) von Bedeutung, in denen auch Forscher erfasst werden. Für die EB-1 Visa existiert z. B. eine Unterklasse ‚Outstanding professors and researchers‘ (E12, E17). Hier sind detaillierte Daten zu Herkunftsländern, Berufen und sogar Arbeitgebern abrufbar, so dass eine Auswertung dieser Daten hinsichtlich deutscher Forscher in den USA möglich scheint. In der Vergangenheit wurde diese Methode von Calmfors et al. (2003) genutzt, um den Umfang der Einwanderung von Hochqualifizierten in die USA mit dem H-1B-Visum zu quantifizieren. Insgesamt wurden im Jahr 2011 knapp 500.000 H-1B Visa bewilligt (Homeland Security 2012), davon die meisten an Inder (ca. 150.000) und ca. 8.000 an Deutsche, von denen natürlich nur eine Teilgruppe den Forschern zuzuordnen ist und in denen auch Folgebewilligungen erfasst sind. Bei den Erstbewilligungen nennen Calmfors et al. (2003) einen Wert von ca. 2.000 bis 2.500 Deutschen jährlich.

Die Daten der amerikanischen Einwanderungsbehörden bieten darüber hinaus genauso wie die Daten des deutschen Ausländerzentralregisters keine umfassenden sozio-demographischen Informationen. Die Individualdaten unterliegen darüber hinaus besonderen Geheimhaltungsvorschriften. Dennoch könnten die amerikanischen Einwanderungsdaten eine gute komplementäre Quelle bieten, um die Abwanderung Deutscher zu quantifizieren, die bislang in der Forschung nur unregelmäßig genutzt wird.

### 4.2.3 Studien zu Absolventen und Promovierten

In den Hauptzielländern deutscher Forschermobilität existieren genauso wie in Deutschland (Panel)Studien zu Hochschulabsolventen im Allgemeinen und zu Promovierenden im Speziellen. An Ihnen kann abgelesen werden, wie viele Deutsche sich bereits während der universitären Ausbildung in einem andern Land aufhalten. Sofern die Erhebungen als Panel organisiert sind bzw. mit ausreichend zeitlichem Abstand zum Abschluss des Studiums oder der Promotion erfolgen, kann auch abgelesen werden, wie viele von ihnen nach dem Abschluss nach Deutschland zurückkehren. Hierbei gelten jedoch die gleichen Einschränkungen wie bei deutschen Absolventenstudien bezüglich der Erreichbarkeit von Panelteilnehmern, die zwischenzeitlich in ein anderes Land abgewandert sind (vgl. Kapitel 3.5). Beispielhaft werden hier aufgrund der quantitativen Bedeutung vor allem Erhebungen aus den USA genannt. Entsprechende Studien liegen jedoch auch für andere europäische Länder vor.

Der National Survey of College Graduates (NSCG) wendet sich an Hochschulabsolventen und ist mit den Erhebungen des HIS-Absolventenpanel in Deutschland vergleichbar. Auf Basis des NSCG von 2003 wurden im Rahmen des MORE-Projekt umfangreiche Auswertungen zu europäischen Hochqualifizierten in den USA durchgeführt und methodische Aspekte diskutiert

(IDEA Consult 2010b). Momentan ist die im Jahr 2010 gebildete Kohorte aktuell, die wiederum über eine Dekade lang befragt wird. Die Personen werden anhand des Zensus‘ identifiziert und gelangen in das Panel, sofern sie mindestens einen Bachelorabschluss besitzen und zum Zeitpunkt des Zensus‘ in den USA leben. Die Rücklaufquoten liegen mit knapp 80 % bei der Erstbefragung und ca. 90 % bei den Folgebefragungen sehr hoch. Die Fragen sind im Zeitverlauf vergleichbar geblieben.

Der Survey of Earned Doctorates (SED) und der Survey of Doctorate Recipients (SDR) wenden sich explizit an Promovierte. Eine vergleichbare Untersuchung existiert für Deutschland nicht, da sich das Promovierendenpanel des iFQ vor allem an noch Promovierende richtet und daraus erst in der Zukunft umfassendere Informationen über Promovierte gewonnen werden können. Die Grundgesamtheit der Befragung speist sich aus allen promotionsberechtigten Institutionen der USA, die gebeten werden, teilzunehmen und die Fragen an die Personen weiterzugeben, die erfolgreich ihre Promotion abgeschlossen haben. Trotz freiwilliger Teilnahme wurde 2010 und 2011 ein Rücklauf von 93 % erreicht. Es werden ca. 50.000 Personen aus 412 promotionsberechtigten Forschungseinrichtungen befragt. Die Befragung enthält u. a. Angaben über Geburtsland, Staatsangehörigkeit, schulische Laufbahn, Orte vorangegangener Schul- und Hochschulabschlüsse, Studienfach, Tätigkeitsfelder und Hauptaufgaben bei derzeitiger Tätigkeit, Typ und Sitz des Arbeitgebers, jährliches Grundgehalt und die Einrichtung an der der Doktorgrad erreicht wurde. Der SED wird jährlich nur an die neuen Absolventen verschickt, während der SDR als Panel angelegt ist.

Die Daten wurden bereits in verschiedenen Studien genutzt, um Aussagen über die Anzahl der ausländischen Promovierten in den USA zu treffen (Moguéro und Di Pietrogioacomo 2008), den Verbleib von ausländischen Doktoranden in den USA zu untersuchen (Kim et al. 2011) oder die Auswanderung Hochqualifizierter aus Deutschland in die USA zu quantifizieren (Ette und Sauer 2010). Auf Grundlage von Daten des SED lässt die Zahl der deutschen Absolventen abschätzen, die in den USA einen Dokortitel erwerben. Ihre Zahl ist zwischen 1989 und 2006 von ca. 100 auf etwas über 200 pro Jahr gestiegen (Schütte 2006).

Das Statistische Bundesamt (2011) hat anhand von amtlichen Datensätzen aus 21 Zielländern deutscher Promovierender eine Zahl von 10.622 im Ausland registrierter Promotionsstudenten ausgemacht. Dabei ist die Schweiz das wichtigste Zielland mit 3.436 Promovierenden. Auf den Plätzen zwei und drei liegen Österreich und Großbritannien mit jeweils über 2.000 deutschen Promovierenden. Mit großem Abstand folgt Frankreich mit ca. 500 deutschen Promovierenden auf dem vierten Rang.

Der amerikanische OpenDoors Survey des Institute of International Education Research ist für weitergehende Untersuchungen der Forschermobilität leider nicht brauchbar, da es sich hier vor allem um eine Befragung zur Internationalität während des Studiums handelt, die in ihrer Methodik der Datensammlung „Wissenschaft Weltoffen“ von DAAD und HIS (2012) ähnelt. Unter den 764.000 ausländischen Studierenden in den USA befinden sich 9.347 Deutsche, die damit die zwölftgrößte Nation sind. Mit großem Abstand führt China mit 194.000 Studierenden (opendoors 2012, Fast Facts). Die Anzahl der deutschen Dozenten in den USA ist hier ebenfalls erfasst. Von den 117.000 ausländischen Wissenschaftlern stammen ca. 5.500 aus Deutschland, die damit die viertgrößte Gruppe stellen. Auch hier führt China mit großem Abstand und 32.000 Personen. Weitere Informationen liegen zur Forschermobilität in diesem

Datensatz allerdings nicht vor, da die Daten nicht von den Personen selbst, sondern von den Universitäten gemeldet werden.

Ergänzend ist an dieser Stelle noch zu erwähnen, dass es gemeinsam von UNESCO, OECD und Eurostat auch Bemühungen gibt, international vergleichbare Studien zu Promovierten im Rahmen des Careers of Doctorate Holders-Projekt (CDH) durchzuführen. Eines der Ziele dieses Projekts ist es, eine international vergleichbare Basis an Indikatoren bezüglich der Mobilität von Promovierten zu schaffen. Dabei wird zum Teil auf nationalen Erhebungen aufgebaut bzw. versucht, diese vergleichbar zu machen. Zum Teil werden auch neue Daten erhoben. Das Projekt greift vor allem auf Absolventenbefragungen, Zensusdaten, Arbeitskräfteerhebungen, Migrationsstudien und Forschungs- und Entwicklungsstatistiken in den OECD-Ländern zurück. Für Deutschland wurde der Labour Force Survey genutzt. Die ersten Ergebnisse für Promovierte der Jahrgänge 1990 bis 2006 wurden 2007 erhoben und publiziert (Auriol 2010). Da die Gruppe der Promovierten jedoch nicht mit Forschern gleichzusetzen ist, sind diese Informationen nur als Vergleichsgröße für die Auswertungen zur Forschermobilität nutzbar. Ein Großteil der Promovierten arbeitet nach der Promotion nicht in der Forschung, während viele Forscher nicht promoviert sind.

### 4.2.4 Nationale Forscherregister, Forscher- und Lebenslaufdatenbanken

In diesem Abschnitt soll kurz ein Überblick über nationale Register und Datenbanken gegeben werden, in denen mobilitätsrelevante Informationen über Forscher abgelegt sind. Da diese Ansätze jedoch in absehbarer Zeit nicht auf Deutschland übertragbar sind und daher für ein Berichtssystem Forschermobilität nicht nutzbar sind, werden diese Ansätze nicht vertiefend diskutiert. Sie sollen vor allem als Anregung dienen, in welche Richtung man bei der Erfassung von Daten über Forscher und Karriere noch denken kann. Ein an die deutschen Verhältnisse angepasster Prozess der Verbesserung der Datenlage wurde kürzlich mit dem Kerndatensatz Forschung (vgl. Kapitel 3.6) initiiert.

In vielen Ländern gibt es verschiedene Formen der verpflichtenden Erfassung aller Forscher. Im skandinavischen Raum existieren Register, die Informationen zu allen Forschern des Landes enthalten. Ein bekanntes Beispiel ist das Forscherpersonalregister in Norwegen. Darüber hinaus sind in nationalen Forscherdatenbanken und Informationssystemen weitere Angaben wie Lebenslaufdaten, Publikationen oder Forschungsprojekte hinterlegt. Folgende Systeme können beispielhaft genannt werden, werden aber aufgrund ihrer fehlenden unmittelbaren Übertragbarkeit auf Deutschland nicht weiter kommentiert:

- Lebenslaufdatenbanken: ANECA-CV (Spanien), DeGOIS (Portugal), ARBIS (Türkei)
- Forscherdatenbanken: Star Metrics (USA, im Aufbau), CRISStin (Norwegen), LATTES (Brasilien)

In den Niederlanden gibt es eine eindeutige Identifikationsnummer, die Digital Author Identification (DAI), unter der alle Leistungen einem Forscher verwechslungsfrei zugeordnet werden können. Das Prinzip ähnelt den in internationalen Forscherdatenbanken genutzten Identifikationsnummern, die z. B. bei ORCID oder ResearcherID zum Einsatz kommen.

### 4.3 Zwischenfazit

Die Betrachtung internationaler Datensätze hat gezeigt, dass sich hier verschiedene interessante Quellen finden lassen, um die Mobilität deutscher Forscher auch dann abbilden zu können, wenn diese Deutschland verlassen. Besonders vielversprechend sind Analysen mit Hilfe von Publikationsdatenbanken, internationale Vergleiche nationaler Zensusdaten und Arbeitskräfteerhebungen sowie automatisierte Verfahren zur Generierung von Karriereverläufen zu einzelnen Forschern auf Basis frei im Internet zugänglicher Informationen und mit Hilfe von Techniken des Natural Language Processing. Die anderen internationalen Datensätze bieten für sich allein genommen entweder keine genügenden Informationen zum Thema Forschermobilität, umfassen (noch) eine zu geringe Fallzahl (internationale Forscherdatenbanken) oder entsprechen nicht mit hinreichender Genauigkeit der Definition von Forschern (internationale Absolventen- und Promoviertenbefragungen).

Da Publikationsdatenbanken aktuell bereits durch das Fraunhofer ISI zur Untersuchung der Forschermobilität genutzt werden, erfolgt im folgenden fünften Kapitel nur die detaillierte Erläuterung von zwei Vorgehensweisen, die zukünftig für eine Untersuchung der Forschermobilität besonders gut geeignet erscheinen. Kurzfristig ist es mittels bestehender Datensätze möglich, eine international vergleichende Auswertung von Mikrozensus und Arbeitskräfteerhebungen vorzunehmen. Eine dafür geeignete Vorgehensweise wird in Kapitel 5.1 erläutert. Mittelfristig besteht erhebliches Innovationspotenzial in der Nutzung von Natural Language Processing für die Generierung eines neuen Datensatzes zur Forschermobilität. Das dafür notwendige umfassendere Projektdesign wird in seinen Grundzügen in Kapitel 5.2 erläutert.



## **5 Ausführliche Darstellung von zwei vielversprechenden Forschungsszenarien**

In diesem Kapitel werden zwei Forschungsszenarien detaillierter erläutert, die in unterschiedlicher Weise dazu beitragen können, die bestehende Daten- und Forschungslücke im Bereich der Forschermobilität zu schließen.

Die Auswertung von Daten des Mikrozensus und aus Arbeitskräfteerhebungen ist kurzfristig umsetzbar, da die Datensätze bereits vorliegen und vergleichsweise leicht zugänglich sind. Diese Daten ermöglichen es, einen Überblick über das Ausmaß und die Struktur der Forschermobilität zu geben. Im deutschen Mikrozensus können Zu- und Rückwanderung nach Deutschland abgebildet werden, in vergleichbaren ausländischen Erhebungen kann die Abwanderung deutscher Forscher quantifiziert werden. Allerdings beinhalten diese Daten keine Informationen, mit denen einzelne Personen identifiziert werden können. Eine solche Identifizierung wäre notwendig, wenn Karriereverläufe abgebildet werden sollen oder eine Verknüpfung mit anderen Datenbeständen vorgenommen werden soll.

Informationen zum Karriereverlauf und der damit verbundenen Mobilität liegen für Forscher vielfach in öffentlich zugänglichen Quellen im Internet vor. Natural Language Processing (NLP) ist daher eine vielversprechende Methode, die erhebliches innovatives Potenzial für die Erstellung von Datensätzen zur Forschermobilität birgt. Es wird erläutert, wie ein Projekt, das NLP-Verfahren nutzt, strukturiert sein sollte. Außerdem werden Aussagen zu den benötigten Ressourcen und möglichen Risiken getroffen. Das Ergebnis eines solchen Projekts ist die weitgehend automatisierte Erstellung von Datensätzen zum Karriereverlauf einzelner Forscher, die mit zusätzlichen Informationen zu Auswirkungen und Determinanten der Forschermobilität (Kapitel 6) verknüpft werden können.

### **5.1 International vergleichende Auswertung von Zensusdaten**

In diesem Forschungsansatz werden die in den Kapiteln 3.1 (deutscher Mikrozensus) und 4.2.1 (Arbeitskräfteerhebungen anderer Länder) beschriebenen Datensätze so kombiniert, dass ein Überblick über Ausmaß und Struktur der Forschermobilität in Deutschland und von Deutschen gegeben werden kann. Die Vor- und Nachteile bezüglich der Aussagekraft über internationale Forschermobilität wurden bereits in den entsprechenden Teilkapiteln diskutiert und werden an dieser Stelle nicht wiederholt.

In den Daten des deutschen Mikrozensus werden die Aspekte Zu- und Rückwanderung von Forschern nach Deutschland abgebildet. Die Daten stehen momentan bis zum Jahr 2009 für Auswertungen zur Verfügung, frühestens ab Mitte 2013 auch für 2010. In den internationalen Datensätzen kann das Phänomen der Abwanderung unter Nutzung vergleichbarer Datenquellen anderer Länder erfasst werden. Die USA sind dabei das wichtigste Zielland der deutschen Forschermobilität. Die empirische Analyse der Forschermobilität in den USA erfolgt mit den Daten des seit 2006 verfügbaren American Community Survey (ACS), der als Ergänzung zum 10-jährigen US-Zensus eine gute Vergleichbarkeit der Ergebnisse mit dem deutschen Mikrozensus ermöglicht. Für die Personengruppen der Forscher werden Ausmaß und Struktur der Zuwanderung in die USA insgesamt untersucht. Das besondere Augenmerk liegt dabei auf den deutschen Zuwanderern innerhalb dieser Gruppe.

Zur Analyse des American Community Survey (ACS) wird auf das Public Use Microdata Sample (PUMS) zurückgegriffen, das bis einschließlich 2011 zur Verfügung steht. Die Fragebögen des ACS werden monatlich an ca. 250.000 Adressen versandt (unterjährige Erhebung). Das 1-year ACS PUMS File umfasst etwa 1 % der US-Bevölkerung (rund 3 Mio. Beobachtungen), einschließlich entsprechender Hochrechnungsfaktoren zur Schätzung der interessierenden Bevölkerungsgrößen und -anteile. Sollte sich im Bearbeitungsprozess die Zahl der Beobachtungen in der betrachteten Subpopulation der jahresweisen Daten als zu klein erweisen, kann auf die extra für diese Fälle bereitgestellten 3- und 5-Jahres-Files (rollende Stichproben mit angepassten Hochrechnungsfaktoren) zurückgegriffen werden.

Zusätzlich können Eckzahlen zu deutschen Forschern in Großbritannien und Frankreich dargestellt werden. Der Zugang zu den Datensätzen des UK Labour Force Survey und des französischen Enquête sur l'Emploi ist kostenfrei möglich. Zu diesen Ländern erscheint eine nach weiteren Merkmalen als der absoluten Anzahl differenzierende Analyse der zugewanderten Forscher aus Deutschland aufgrund der zu geringen absoluten Fallzahlen allerdings nicht möglich. Die niederländische Variante (Dutch Labour Force Survey) ist aufgrund der geringen Stichprobengröße genauso wie die Daten aus Österreich und der Schweiz nicht nutzbar. Anhand der zusammengeführten nationalen Daten der europäischen Arbeitskräfteerhebung ist eine Untersuchung deutscher Forscher nicht möglich, da sämtliche Angaben einzelner Länder (für Nationalität und Herkunftsland) aus Geheimhaltungsgründen durch Ländergruppen (z. B. EU-15) ersetzt worden sind.

Da die genutzten Erhebungen Mikrozensus und ACS jährlich durchgeführt werden, ist es mit diesen Daten möglich, ein standardisiertes Vorgehen zu entwickeln, das in ein laufendes Berichtssystem überführt werden kann. Die dafür notwendigen Schritte werden im Folgenden erläutert.

### **1. Schritt: Datenzugang und Datenpooling**

Sowohl der Mikrozensus als auch der ACS können für wissenschaftliche Untersuchungen genutzt werden. Die aktuellsten nutzbaren Befragungen liegen beim Mikrozensus im Moment nur für das Jahr 2009 vor; für den ACS ist bereits das Jahr 2011 verfügbar. Das Scientific Use File für 2010 wird erst Mitte des Jahres 2013 veröffentlicht werden. Auch der Zugang zum ACS ist deutlich leichter als zum Mikrozensus. Der Zugang zum ACS erfolgt über den kostenlosen Download der Daten beim amerikanischen Census Bureau und unterliegt keinen weiteren Beschränkungen. Der Mikrozensus ist auf zwei verschiedene Arten nutzbar: zum einen als Scientific Use File (70 % Unterstichprobe der gesamten Daten) zum Preis von 250 Euro je Erhebung, zum anderen auf Basis der gesamten Stichprobe mittels Fernrechnen bei den Forschungszentren der statistischen Ämter des Bundes und der Länder für ebenfalls 250 Euro je projektspezifischer Nutzung ([www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de)).

Trotz der insgesamt hohen Fallzahlen für jedes Jahr in beiden Befragungen (Mikrozensus: ca. 830.000, ACS: ca. 3.000.000) können die relevanten Untergruppen (in Deutschland: zugewanderte Forscher, in den USA: deutsche Forscher) für nach einer Vielzahl weiterer Merkmale differenzierende Fragestellungen zu klein sein. Eine weitere Erhöhung der Fallzahl ist durch das Poolen einzelner Jahre möglich.

Beim Mikrozensus ist ein Poolen nur bedingt möglich, da es sich bei der Auswahl der Befragten um eine rotierende Stichprobe handelt, bei der jeder Haushalt vier Jahre in Folge befragt

wird. Jedes Jahr findet dabei ein Austausch von einem Viertel der Stichprobe statt. Ein vollständiger Austausch der Stichprobe ist damit erst nach vier Jahren gegeben. Eine Identifizierung der neu aufgenommenen Personen ist nicht möglich. Ein Poolen ist daher nur von Datensätzen im Abstand von vier Jahren möglich, d.h. die Daten für 2005 und 2009 bzw. für 2006 und 2010 können gepoolt werden. Andernfalls würde es zur Doppelzählung eines Teils der Befragten kommen. Beim Poolen ist zu beachten, dass die Hochrechnungsfaktoren für gewichtete bzw. auf die Grundgesamtheit hochgerechnete Ergebnisse neu berechnet werden müssen. Eine andere Möglichkeit, mehrere Jahrgänge zur Verbesserung der Qualität der Ergebnisse gleichzeitig zu nutzen, besteht darin, die einzelnen Jahre nebeneinander zu betrachten. Durch den Vergleich der Ergebnisse zwischen den Jahren und der Berechnung des Stichprobenfehlers bzw. seiner Spannweite ist es möglich, einen einigermaßen sicheren Eindruck über die Robustheit der interessierenden Anteilswerte bei geringen zugrundeliegenden Fallzahlen zu erhalten.

Der ACS hat auch beim Poolen von Daten deutliche Vorteile, da bereits gepoolte Datensätze für Zeiträume von drei und fünf Jahren für kleinere Teilpopulationen zur Verfügung stehen. In diesen Datensätzen sind auch die Hochrechnungsfaktoren bereits entsprechend angepasst.

### **2. Schritt: Abgrenzung von mobilen Forschern**

Mobilität kann in beiden Datensätzen auf sehr ähnliche Weise abgebildet werden. Die zentrale Variable ist die Angabe eines Zuzugsjahres, seitdem sich eine Person in Deutschland befindet. Die Variable erlaubt differenzierte Informationen bis in die 1950er Jahre. Im Mikrozensus wird der Zuzug nur erfasst, sofern der Aufenthalt im Ausland mindestens sechs Monate betragen hat. Da dies jedoch auch in dieser Machbarkeitsstudie als Untergrenze für die Erfassung von Mobilität definiert wurde, ergeben sich daraus keine Einschränkungen. Weitere wichtige Informationen sind der Geburtsort und die Nationalität, durch die sich annäherungsweise zwischen Zu- und Rückwanderern unterscheiden lässt und die Herkunft der Zuwanderer bestimmt werden kann. Weitere mobilitätsrelevante Indikatoren sind der Wohn- und Arbeitsort der Personen sowie die Herkunft der Vorfahren. Letztere wird in den USA immer erfasst, so dass auch der Anteil von Forschern mit deutschen Wurzeln ermittelt werden kann. In Deutschland ist diese Frage einmal im Ad-hoc-Modul des Jahres 2008 gestellt worden.

Die Abgrenzung von Forschern ist deutlich anspruchsvoller. In beiden Datensätzen ist zwar der Beruf der Personen erfasst, die in beiden Ländern als offene Frage gestellt wird und bei der Dateneingabe einer Berufsklassifikation zugeordnet wird. Die zugrunde liegenden Klassifikationen (Deutschland: International Standard Classification of Occupations (ISCO), USA: Standard Occupational Classification (SOC)) lassen aber nur eine Annäherung an forschende Tätigkeiten zu, da der Beruf des Forschers in den gängigen Berufsklassifizierungen nicht enthalten ist. Für die Identifikation von Forschern wird eine weitestgehende Annäherung an die Definition des Frascati-Manuals auf Basis der 3-steller der ISCO im Mikrozensus bzw. der 4-steller der SOC vorgenommen. Im Rahmen des MORE-Projekts (IDEA Consult 2010a: 22) wurde vorgeschlagen, auf Basis der ISCO 3-steller die Berufsklassen 214, 221, 222, 231, 241 und 244 für die Abgrenzung von Forschern zu nutzen. Diese Abgrenzung kann als Ausgangspunkt für die hier skizzierte Untersuchung genutzt werden, bietet aber noch Potenzial für Verfeinerungen. Aus der Verwendung leicht abweichender Klassifikationen ergeben sich keine Probleme, da eine Umsteigetabelle zwischen ISCO und SOC existiert, deren Unschärfen sich nicht auf potenziell forschende Berufe auswirken. Der britische Labour Force Survey nutzt für die Klas-

sifizierung der Tätigkeiten ebenfalls die SOC, während die französischen Daten auf der ISCO beruhen.

Um die Ungenauigkeiten bei der Abgrenzung von Forschern anhand der Berufsklassifikation weiter zu verringern, kann die Variable des höchsten Bildungsabschlusses genutzt werden. Personen mit Hochschulabschluss bzw. darunter Personen mit Promotion sind in beiden Datensätzen identifizierbar. Ein Hochschulabschluss kann als Minimalvoraussetzung für forschende Tätigkeiten genutzt werden. Eine weitere Möglichkeit ist die Nutzung der Angabe des Wirtschaftszweigs zum weiteren Ausschluss nicht-forschender Wirtschaftszweige. So können z. B. praktizierende Ärzte oder Lehrer an Schulen ausgeschlossen werden, die aufgrund ihres Berufs und ihres Bildungsabschlusses fälschlicherweise als Forscher klassifiziert werden könnten. Auch die Art des Arbeitgebers ist in den USA und Deutschland in den Daten enthalten. So können auf einfache Weise Forscher in Privatunternehmen, im öffentlichen Dienst und in selbständigen Tätigkeiten voneinander unterschieden werden. Auch Angaben zur Höhe des Einkommens sind in beiden Datensätzen vorhanden.

Zu prüfen ist, ob durch die Angabe der konkreten Tätigkeit, eine genauere Identifizierung von Forschern möglich ist. In Deutschland wurden im Mikrozensus des Jahres 2007 die Tätigkeit (Forscher in der Gruppe „Forschen, Entwerfen, Konstruieren, Gestalten von Produkten, Plänen, Programmieren“), die Bezeichnung der Abteilung im Betrieb (Forscher in der Abteilung „Entwicklung, Konstruktion, Forschung, Design, Musterbau“) und die Stellung im Betrieb bislang letztmalig abgefragt (vorher jeweils im Abstand von vier Jahren). In den USA wird die tatsächliche Tätigkeit jedes Jahr mit einer offenen Frage neben dem Beruf abgefragt.

### **3. Schritt: Differenzierte Untersuchung der Forschermobilität**

Die aufbereiteten Datensätze mit vergleichbarer Abgrenzung von mobilen Forschern werden dann für die differenzierte Untersuchung der Forschermobilität genutzt. Mit den Daten sind unter anderem Antworten auf folgende Arten von Fragen möglich:

- Wie groß ist der Anteil zu- und rückgewanderter Forscher an den gesamten Forschern eines Landes?
- Wie viele deutsche Forscher sind in den USA tätig? Welchen Anteil machen sie an den Forschern der betreffenden Länder aus?
- Welchen Stellenwert hat Zu- und Rückwanderung von Forschern in Deutschland im Vergleich mit anderen Ländern?
- Aus welchen Ländern stammen die zu- und rückgewanderten Forscher in Deutschland?
- Wie ist diese Verteilung der Herkunftsländer im Vergleich mit anderen großen Industrieländern zu bewerten?
- Welche Qualifikationen weisen die zugewanderten Forscher in Deutschland auf?
- Wie hat sich die Zuwanderung differenziert nach Qualifikationen und Berufsfeldern im Zeitablauf entwickelt?
- In welchen Berufen und Wirtschaftszweigen (Wirtschaft, Hochschulen, Staat) werden sie eingesetzt und in welcher Form (insb. Vollzeit/Teilzeit, befristet)?

- Sind die zugewanderten Forscher in Deutschland bzw. die abgewanderten deutschen Forscher in den USA familiär verwurzelt; wie groß ist die Wahrscheinlichkeit der Rückwanderung?
- Welche sozio-demographischen Merkmale besitzen deutsche Forscher und Hochqualifizierte in den USA im Vergleich zu den entsprechenden Personen in Deutschland?
- Seit wann sind die nach Deutschland zugewanderten Forscher bereits im Land? Während welcher Karrierephasen erfolgte die Zuwanderung?
- Wie alt sind die ausländischen Forscher in Deutschland im Vergleich zu den deutschen Forschern in Deutschland und im Ausland?
- Welche Hinweise über den Ort der Ausbildung können aus der Kombination von Informationen zum Alter und zum Zuzugsjahr getroffen werden?

Mit Hilfe der Daten aus dem Ad-hoc Modul des Mikrozensus 2008 zur "Arbeitsmarktsituation von Zuwanderern und ihren Nachkommen" können weitere Informationen zu Motiven der Zuwanderung, zu Familienmitgliedern im Herkunftsland sowie zu bestimmten Maßnahmen der Arbeitsmarktintegration in Deutschland (Anerkennung von Abschlüssen, Sprachkurse, Suchverhalten) ausgewertet werden.

Der Aufwand und die Risiken der hier skizzierten Untersuchung sind aufgrund des Rückgriffs auf vorhandene Datensätze, zu denen ein relativ einfacher Zugang besteht, überschaubar. Das größte Risiko liegt in der Größe der Teilstichproben für ausländische Forscher in Deutschland bzw. deutsche Forscher in den USA. Hier kann es zu Einschränkungen bei der mehrdimensionalen Differenzierung der Ergebnisse kommen. Außerdem ergibt sich durch das Poolen von Datensätzen des Mikrozensus die Herausforderung, neue Hochrechnungsfaktoren zu entwickeln. Auch dies ist jedoch mit Hilfe statistischer Verfahren lösbar. Da die genannten Datensätze für Deutschland, die USA sowie (zur Berechnung von Eckzahlen) für Großbritannien und Frankreich bislang noch nicht vergleichend analysiert worden sind, ist für die erstmalige Analyse ein erhöhter Aufwand bei der Entwicklung einer vergleichbaren Abgrenzung von Forschern in den Klassifikationen SOC und ISCO zu veranschlagen. Erschwerend kommt hinzu, dass die Klassifikationen selbst im Zeitverlauf an veränderte Berufsbilder angepasst werden und daher bei Längsschnittanalysen ebenfalls Umsteigetabellen für einzelne Berufe genutzt werden müssen. Aufbauend auf diesen einmal erstellten Abgrenzungen haben die Datensätze allerdings ein großes Potenzial, Indikatoren für ein laufendes Berichtssystem zur Forschermobilität zu liefern, die mit geringerem Aufwand fortgeschrieben werden können. Für die erstmalige Aufbereitung der Datensätze und die Entwicklung von Auswertungsroutinen und Indikatoren sollte ein Zeitraum von ca. vier Personenmonaten veranschlagt werden.

Die wichtigsten Einschränkungen bei den Daten aus Mikrozensus und Arbeitskräfteerhebungen sind das Fehlen von Informationen über Erwerbsbiografien bzw. Karriereverläufe und die Anonymität der befragten Personen, die eine Verknüpfung mit weiteren Informationen zur individuellen Forschungsleistung oder zur Zuordnung zu einzelnen Organisationen unmöglich macht. Diese Schwächen können mit dem zweiten vielversprechenden Verfahren, der Generierung eines Datensatzes zur Forschermobilität durch Methoden des Natural Language Processing (NLP), der im folgenden beschrieben wird, ausgeglichen werden.

### 5.2 Generierung eines Datensatzes durch Natural Language Processing (NLP)<sup>1</sup>

Natural Language Processing (NLP) ist der Versuch, aus strukturierten und unstrukturierten Fließtexten relevante Informationen auszulesen, zu speichern und mit Verfahren der semantischen Analyse weiterzuverarbeiten. Grundlage für dieses Verfahren sind elektronisch zu verarbeitende Dokumente wie z. B. Internetseiten, Online-Datenbanken und PDF-Dokumente. Das Verfahren erscheint für die Analyse der Forschermobilität potenziell geeignet zu sein, da über die betreffenden Personen vergleichsweise viele Informationen frei zugänglich im Internet verfügbar sind. Der Grund dafür ist die hohe Sichtbarkeit der individuellen Forscherleistung insbesondere in der öffentlich geförderten Forschung und der damit verbundene Anreiz zur individuellen Selbstvermarktung. Personenbezogene Informationen zu Forschern und ihrem Aufenthaltsort finden sich unter anderem auf Institutshomepages, persönlichen Homepages, frei zugänglichen Lebensläufen, öffentlichen Profilen in sozialen Netzwerken, biographischen Angaben in Publikationen (vor allem frei zugängliche Arbeitspapiere) und Forschungsprojekten, Medienberichten und vielen weiteren elektronischen Dokumenten im Internet. Etwas eingeschränkter ist der Umfang frei zugänglicher Informationen für Forscher in der Privatwirtschaft, da ihre Arbeit häufig einem hohen Maß an Geheimhaltung unterliegt. Aber auch hier ist davon auszugehen, dass einige Informationen verfügbar gemacht werden können.

Für Fragestellungen zur Forschermobilität können Methoden des NLP dazu genutzt werden, personenbezogene Informationen zu Arbeitsorten im zeitlichen Verlauf einer Karriere zu sammeln. Die so erhobenen Daten können dann für die Analyse des Mobilitätsverhaltens genutzt werden. Damit kann das Verfahren dazu genutzt werden, die in den letzten Jahren an Bedeutung gewinnende Lebenslaufanalyse (Cañibano und Bozeman 2009) auf eine weitaus umfassendere Datenbasis anzuwenden.

Die entscheidenden Herausforderungen einer solchen Untersuchung liegen in der Umsetzung einer weitgehend automatisierten Generierung entsprechender Datensätze. Dazu bietet sich folgendes Vorgehen in fünf Schritten an:

#### 1. Schritt: Erstellung einer Referenzdatenbank

Zu Beginn der Analyse muss festgelegt werden, zu welchen Personen Mobilitätsspuren gesammelt werden sollen. Dieser Schritt ist entscheidend, da hier bereits die Grundgesamtheit für die Untersuchung definiert wird. Hier bieten sich sowohl deutsche Quellen an (z. B. Hochschullehrerverzeichnis), um das Mobilitätsverhalten von Forschern, die sich aktuell in Deutschland aufhalten, zu untersuchen. Um einen Einblick in das internationale Geschehen zu erhalten, ist aber auch hier die Nutzung internationaler Datenquellen ratsam (z. B. Forscherverzeichnisse anderer Länder oder fachspezifische internationale Vereinigungen).

Die Eingrenzung auf einzelne Wissenschaftsfelder kann an dieser Stelle sinnvoll sein, da dann auch Publikationsdatenbanken wie das Web of Science oder Scopus genutzt werden können, um Forscher eines bestimmten Feldes zu identifizieren. Für eine fachliche Eingrenzung spricht außerdem, dass man dadurch eher einen kompletten Überblick über eine Teilpopulation der Forscher im In- und Ausland erhalten kann und das Problem der Unterscheidbarkeit von Per-

---

<sup>1</sup> Die Ausführungen zu NLP-gestützten Verfahren wurden mit fachlicher Beratung durch Dr. Stefan Hennemann, Institut für Geographie an der Justus-Liebig-Universität Gießen erstellt.

sonen mit gleichem Namen (Homonyme) geringer wird. Für eine Eingrenzung sind besonders gut die Fächer geeignet, deren Publikationen im Web of Science oder der Scopus-Datenbank aufgeführt werden, d. h. eher keine Geisteswissenschaften, die auf einer bestimmten Technologie beruhen, aus der ein gemeinsames Suchwort abgeleitet werden kann und die wenig interdisziplinär sind. Letztlich wird die inhaltliche Fragestellung und die Verfügbarkeit geeigneter Datenbanken mit Namen von Forschern über die Auswahl der Referenzdatenbank entscheiden.

Bei der Auswahl der Referenzdatenbank ist für die weitere Untersuchung zu beachten, dass die Angabe eines Namens alleine nicht ausreichend ist, da dadurch keine eindeutige Unterscheidung von Homonymen möglich ist. Idealerweise sollten zusätzliche Informationen wie der aktuelle Arbeitsort, ausgewählte frühere Arbeitsorte/Arbeitgeber oder eine oder mehrere Publikationen bereits bekannt sein.

### **2. Schritt: Sammeln elektronischer Dokumente zu den Personen der Referenzdatenbank**

Zu den in der Referenzdatenbank aufgeführten Personen werden in diesem Schritt Dokumente gesammelt, die frei zugänglich im Internet vorliegen oder aus anderen vorher definierten Datenbanken ausgelesen werden. Dabei muss im Vorfeld keine Einschränkung auf bestimmte Dokumente erfolgen. Die Suche nach Dokumenten erfolgt mit Hilfe eines speziell programmierten Webcrawlers, der auf der Google-Suchmaschine aufsetzt und eine sogenannte erschöpfende Suche nach der Brute-Force-Methode durchführt, d. h. es werden verschiedene Suchkombinationen durchgespielt und immer ein bestimmter Prozentsatz der relevantesten Dokumente aus der Google-Suche zu der betreffenden Person abgelegt. Durch den hohen Rechenaufwand erhöht sich dadurch zwar die Dauer des Einsammelns von Dokumenten, die Vollständigkeit wird aber deutlich erhöht. Durch diese Methode werden mit relativ großer Sicherheit alle öffentlich zugänglichen Informationen erfasst. Dazu zählen auch die öffentlichen Profile in sozialen Netzwerken (z. B. die in Kapitel 4.1.4 beschriebenen Dienste LinkedIn und ResearchGate).

Zusätzlich kann dann noch in geschützten Datenbanken gesucht werden (z. B. Scopus). GoogleScholar ist hier leider als Grundlage für die Recherche nicht geeignet, da es ein automatisiertes Auslesen von Informationen durch Webcrawler verhindert. Man ist sich hier offenbar über den Wert der dort gesammelten Daten bewusst und nach Erfahrungsberichten anderer, an den Daten interessierter Wissenschaftler auch nicht zur Zusammenarbeit bereit. Auch eine Zusammenarbeit mit Betreibern von sozialen Netzwerken ist für diese Fragestellung, bei der die Namen der Personen benötigt werden, aus Geheimhaltungsgründen nicht realistisch. Hier wird die Recherche auf vom Benutzer als öffentlich eingestufte Informationen beschränkt bleiben müssen. Interessanterweise hat Facebook seine Daten vor kurzer Zeit wissenschaftlich mit Methoden der Netzwerkanalyse ausgewertet, um herauszufinden, über wie viele Kontakte jeder Nutzer von Facebook durchschnittlich jeden anderen erreichen kann (Ugander et al. 2011, Backstrom et al. 2012). Hier stand aber offenbar das Eigeninteresse des Unternehmens an der Nutzung des Ergebnisses von *4,74 degrees of separation* im Vordergrund.

### 3. Schritt: Auslesen von Mobilitätsspuren aus den Dokumenten

In diesem Schritt, der den Kern der NLP-Methode darstellt, geht es darum, Mobilitätsspuren, d. h. Textstellen mit mobilitätsbezogenen Informationen, von anderen Informationen zu trennen. Für die Identifizierung der relevanten Informationen zur Mobilität (Orte, Länder, Organisationen, Zeitangaben) gibt es verschiedene Verfahren. Es ist bereits möglich, während des Durchlaufs des zweiten Schritts die Textmenge zu reduzieren, in dem man mit sogenannten regulären Ausdrücken nach bestimmten syntaktischen Regeln nur potenziell mobilitätsrelevante Informationen abspeichert. Dadurch ließe sich die zu analysierende Textmenge bereits im Vorfeld verringern und die semantische Klassifizierung beschleunigen.

Die gespeicherte Textmenge wird nun mit Hilfe geeigneter Programme (z. B. dem Python Paket Natural Language Toolkit NLTK) analysiert. Dabei werden die Sätze einzeln nach grammatikalischen und syntaktischen Regeln klassifiziert. Zunächst erfolgt eine Bereinigung und Normalisierung, bei der z. B. Plural durch Singular ersetzt wird (das sogenannte *Stemming*) und Artikel sowie Füllwörter gelöscht werden (anwenden von sogenannten *Stopwords*). Der bereinigte Datensatz wird dann nach grammatikalischen Regeln klassifiziert. Anhand existierender Wörterbücher für die Erkennung von Ort- und Zeitangaben werden hier mobilitätsbezogene Informationen extrahiert und unter dem Namen des entsprechenden Forschers abgelegt, um sie im nächsten Schritt weiterzuverarbeiten.

### 4. Schritt: Bewertung der Mobilitätsspuren

Der technisch anspruchsvollste Schritt ist die nun folgende semantische Analyse auf Basis der geordneten mobilitätsbezogenen Informationen, bei der die Informationen bewertet und weitgehend automatisiert in einen Bedeutungszusammenhang gebracht werden müssen. Hier kann man z. B. nach Zeitangaben in der Nähe von Ortsangaben suchen lassen, um tabellarische Lebensläufe und andere relativ systematisch strukturierte Angaben auszuwerten, oder Erscheinungsjahre von Publikationen mit den zur entsprechenden Person abgelegten Ortsangaben verknüpfen. Diese Klassifizierungen werden vom Programm in der Regel mit einer Treffer-Wahrscheinlichkeit versehen, auf die man automatisierte Entscheidungsregeln anwenden kann.

Wie das Programm für die semantische Analyse genau zu justieren ist, wird sich erst im Laufe der tatsächlichen Durchführung der Untersuchung ergeben. Es ist dafür notwendig, das Programm mit Hilfe von Testdatensätzen zu trainieren. Die Analyse wird dazu zuerst für einen Personenkreis angewandt, zu dem das richtige Ergebnis z. B. durch Vorliegen eines kompletten Lebenslaufes bekannt ist. Es wird dann so lange versucht, die Algorithmen zu verbessern, bis das Programm die benötigten Informationen mit einer möglichst hohen Zuverlässigkeit automatisch ausliest.

Der Machbarkeitsnachweis für dieses Vorgehen wurde jüngst auf Basis eines Testdatensatzes mit Wissenschaftlern aus Uruguay erbracht (García Flores et al. 2012). Hier wurde eine Zuverlässigkeit der Bewertung mobilitätsrelevanter Informationen von 80 % erreicht. Die Zuverlässigkeit hängt selbstverständlich stark von der Beschaffenheit der Datenquelle ab. Tabellarische Lebensläufe oder öffentliche LinkedIn-Profilen dürften mit einer Trefferquote von nahezu 100 % analysierbar sein, bei anderen Dokumenten wie Fließtexten auf Institutshomepages oder Medienberichten über einzelne Forscher dürften die Quoten deutlich geringer sein. Auch hier



kann man Regeln definieren, welche Angaben mit höherer Priorität als korrekte Angabe eines Aufenthaltsorts zu einem bestimmten Zeitpunkt gewertet werden.

### **5. Schritt: Erzeugen von Karriereverläufen für die Personen der Referenzdatenbank**

Im letzten Schritt werden aus den mit Wahrscheinlichkeiten versehenen mobilitätsrelevanten Informationen die Karriereverläufe der Personen in der Referenzdatenbank rekonstruiert. Dabei soll es das Ziel sein, möglichst vollständige Orts- und Zeitangaben zur Karriere vom Studium bis zum aktuellen Stand zu erhalten. Diese Datenbank kann dann mit statistischen Auswertungsmethoden analysiert werden.

### **Bewertung der Chancen und Risiken sowie benötigter Ressourcen für die Umsetzung**

Unter der Bedingung, dass das geschilderte Vorgehen tatsächlich hinreichend zuverlässige Ergebnisse produziert, bietet es die große Chance, mit Hilfe eines innovativen Verfahrens und frei zugänglicher Informationen eine neuartige Datenbank zu erzeugen. Der Machbarkeitsnachweis für das Verfahren in seinen Grundzügen ist bereits durch García Flores et al. (2012) erbracht worden. Eine Anwendung auf eine größere Forscherpopulation wäre hingegen bislang einmalig. Die automatisierte Erstellung einer Datenbank zu Karrieren deutscher Forscher wäre dabei gerade vor dem Hintergrund der bislang schlechten Datenverfügbarkeit in diesem Bereich wünschenswert (vgl. Einschätzung zur Einführung eines Kerndatensatz Forschung in Kapitel 3.6). Eine durch NLP-Methoden erzeugte Datenbasis kann nicht nur für die Analyse der Forschermobilität genutzt werden, sondern auch für weitere Fragestellungen im Zusammenhang mit dem Verlauf wissenschaftlicher Karrieren. Durch die weitestgehend eindeutige Identifizierung von Personen und die personenscharfe Zuordnung der Karriereverläufe ist eine Verknüpfung mit Informationen aus anderen Quellen möglich. Im folgenden Kapitel 6 wird erläutert, welche Möglichkeiten hinsichtlich der Analyse von Determinanten und Auswirkungen der Forschermobilität bestehen.

Aufgrund des innovativen Charakters der Methode der Datengewinnung ist das Projekt jedoch auch mit gewissen Risiken verbunden. Bei der Datenerhebung ist es von entscheidender Bedeutung, dass das Crawlen, d. h. das automatisierte Auslesen von Informationen aus dem Internet, für wichtige Quellen nicht im Laufe des Datenabrufs blockiert wird. GoogleScholar verhindert einen solchen Zugriff grundsätzlich. Andere Datenbanken erfordern eine Zugangsberechtigung (z. B. Web of Science) oder es muss ein Offline-Abbild erstellt werden (z. B. Scopus). Die Programmierung der Algorithmen für die Analyse der Dokumente ist eine weitere große Herausforderung. Es muss trotz heterogener Quellen und teilweise unstrukturierter Syntax gelingen, die notwendigen Informationen mit einer hinreichenden Treffsicherheit auszulesen. Hier können programmiertechnische Probleme, aber auch unzureichende Rechnerkapazitäten auftreten, da die Verfahren gerade bei großen Textmengen sehr rechenintensiv sind. Diese Probleme sind allerdings nicht im Vorfeld zu antizipieren, sondern treten erst im Forschungsprozess auf, z. B. beim Trainieren der Programme anhand eines Testdatensatzes. Im Bereich der Informatik sind ähnliche Anwendungen jedoch seit einigen Jahren für andere Fragestellungen erfolgreich implementiert worden, so dass davon auszugehen ist, dass sich auch für die bei diesem Forschungsansatz auftretenden Herausforderungen kreative, fallbezogene Lösungen entwickeln lassen werden.

Für die Durchführung einer NLP-basierten Analyse der Forschermobilität wird ein interdisziplinäres Forscherteam benötigt, das sowohl inhaltliche Kenntnisse der Materie, als auch technische Fertigkeiten für die Programmierung der Webcrawler und semantischen Analysetools mitbringt. Das Team sollte aus in Fragen der Forschermobilität erfahrenen Wirtschafts- und Sozialwissenschaftlern bzw. Wirtschaftsgeographen und an der Schnittstelle von Informatik und Sprachwissenschaften arbeitenden Computerlinguisten bestehen.

Da in dem skizzierten Projekt in vielen Aspekten zunächst Grundlagenarbeit betrieben werden muss, ist eine hinreichend lange Laufzeit einzuplanen. Für die vollständige Bearbeitung der fünf genannten Schritte sowie die inhaltliche Auswertung der dadurch erstellten Datenbank sind zwei Jahre nötig. Um ein stabil laufendes System wie im Beispiel für Uruguay (García Flores et al. 2012) zu trainieren und zu testen, ist ungefähr ein halbes Jahr einzuplanen. In einem weiteren halben Jahr kann dann die Implementierung für einen großen Personenkreis erfolgen. Daraufhin muss die erstellte Datenbank umfassenden Tests bezüglich der Validität der Ergebnisse unterzogen werden. Ggf. können weitere Daten aus anderen Quellen zur Forschungsleistung zugespielt werden (vgl. Kapitel 6). In der letzten Projektphase erfolgt dann die statistische Analyse und Erstellung des Berichts. Die inhaltliche Begleitung durch Wirtschafts- und Sozialwissenschaftler bzw. Wirtschaftsgeographen erfolgt über die gesamte Laufzeit des Projekts, mit einer zunehmenden Intensität in der zweiten Projekthälfte. Die technische Umsetzung durch Computerlinguisten ist stärker auf die erste Projekthälfte konzentriert. In der zweiten Projekthälfte könnte eine Standardisierung des Verfahrens zur Übertragbarkeit auf andere Teilpopulationen angestrebt werden.

Zusammenfassend kann festgehalten werden, dass die Generierung eines Datensatzes zur Forschermobilität durch Methoden des Natural Language Processing (NLP) im Vergleich zu allen anderen im Rahmen der Machbarkeitsstudie vorgestellten Verfahren die größten Chancen bietet, einen neuartigen Datensatz zu diesem Themenfeld hervorzubringen. Ein solcher Datensatz ermöglicht die Abbildung von Zu-, Rück- und Abwanderung, da durch die Nutzung des Internets keine Einschränkung auf Quellen einzelner Länder besteht. Die gesammelten Informationen sind personenscharf und erlauben die Verknüpfung mit weiteren Informationen. Da auf öffentlich zugängliche Informationen zugegriffen wird, gibt es keine Probleme bei der Umsetzung, wie sie bei der Verknüpfung von Daten im Zuge der Erstellung des Kerndatensatzes zu erwarten sind. Die NLP-Methode ist dadurch deutlich schneller realisierbar und flexibler einsetzbar als der Kerndatensatz Forschung. Diese Vorteile wiegen die Risiken und die nicht zu 100 % genaue Zuordnung der mobilitätsrelevanten Informationen aus Sicht des Autors der Machbarkeitsstudie auf. Daher wird der Expertenkommission Forschung und Innovation empfohlen, die Ausschreibung eines solchen Projektes gemeinsam mit dem BMBF zu forcieren.

### 6 Auswirkungen und Determinanten der Forschermobilität im Innovationssystem

In diesem abschließenden Kapitel werden Möglichkeiten diskutiert, die bisher beschriebenen Datensätze und -quellen zur Forschermobilität mit weiteren Informationen anzureichern, um Auswirkungen und Determinanten der Forschermobilität im Innovationssystem zu untersuchen. Da dazu in der Regel eine Verknüpfung verschiedener Datensätze nötig ist, können hier nur Datensätze genutzt werden, in denen personenscharfe Informationen mit Angabe des Namens und weitere Merkmale zur eindeutigen Unterscheidung von Homonymen vorliegen. Von den beschriebenen Verfahren sind nur diese geeignet: Integrierte Erwerbsbiografien der BA, Kerndatensatz Forschung, Publikationsdaten, Patentdaten, internationale Forscherdatenbanken, NLP-basierte Datenbanken, nationale Forschungsregister und Lebenslaufdatenbanken. Insbesondere Daten aus Mikrozensus und Arbeitskräfteerhebungen ermöglichen es nicht, auf einzelne Personen zurückzuschließen und sind daher für Untersuchungen, die über Ausmaß und Struktur der Forschermobilität hinausgehen, ungeeignet.

Bei allen Untersuchungen zu den Auswirkungen der Forschermobilität ist zu beachten, dass Mobilität nur eine Determinante der Forschungsleistung ist, die methodisch von anderen Determinanten getrennt werden müsste, um die tatsächliche Wirkung von Mobilität identifizieren zu können.

#### 6.1 Individuelle Karriereentwicklung

Mobilität ist eng mit der individuellen Karriereentwicklung verbunden. Leistungsfähigere Forscher erhalten beispielsweise eher Angebote von anderen Einrichtungen und haben damit potenziell mehr Möglichkeiten zur Mobilität. Mobilität kann gleichzeitig ein wichtiges Element der Karriereentwicklung sein, da der Karrierefortschritt durch die Arbeit in anderen Umfeldern, das Knüpfen neuer Kontakte und die Reputation der Arbeitsorte beeinflusst werden kann. Außerdem können Aussagen zur qualitativen Dimension der Forschermobilität durch die Kenntnis der individuellen Forschungsleistungen getroffen werden, da die Mobilität von Nachwuchsforschern am Anfang ihrer Karriere anders einzuordnen ist als die Mobilität von Spitzenforschern im Höhepunkt ihrer Schaffenskraft oder die Mobilität von älteren Forschern zum Ende ihrer Karriere. Um diese und ähnliche Zusammenhänge empirisch herausarbeiten zu können, ist es eine Verknüpfung der Informationen zur Forschermobilität mit Daten zur individuellen Forschungsleistung notwendig.

Die Karriereentwicklung von Wissenschaftlern ist in jüngster Zeit durch Verfahren der Lebenslaufanalyse (Cañibano und Bozeman 2009) untersucht worden. Diese Daten wurden in der Vergangenheit häufig Lebenslaufdatenbanken entnommen, die an einzelnen Einrichtungen vorlagen und sich daher auf kleine Fallzahlen bezogen. In der Machbarkeitsstudie sind dazu zwei Verfahren genannt worden, die eine automatisierte Erzeugung von Lebenslaufdaten in großer Zahl aus anderen Quellen ermöglichen: NLP-basierte Verfahren der Datenextraktion und Auswertung der organisationalen Zugehörigkeiten in Publikationsdatenbanken. Die so gewonnenen Informationen können ebenfalls für die Verknüpfung mit anderen Datenbeständen zur Forschungsleistung genutzt werden. Die automatisierte Zuordnung von Informationen zur Forschungsleistung könnte technisch ähnlich ablaufen wie in den Schritten zwei und drei der NLP-gestützten Verfahren geschildert (Kapitel 5.2).

Neben der reinen Analyse des Karriereverlaufs können Datenbanken zu Publikationen, Patenten und Forschungsprojekten genutzt werden, um personenscharfe Informationen zu gängigen wissenschaftlichen Leistungskriterien zu ergänzen. Auf diese Datensätze können dann z. B. Verfahren der bibliometrischen Analyse angewandt werden. Diese werden hier nicht näher beschrieben, die damit verbundenen Herausforderungen wurden von Ferrara und Salini (2012) in einem Überblicksartikel dargestellt.

Sofern internationale Forschungsdatenbanken (ORCID, ResearcherID) eine weitere Verbreitung finden und von den Nutzern mit Lebenslaufdaten angereichert würden, wäre unmittelbar eine Datenquelle gegeben, die eine Verknüpfung von Karriereverlauf und Forschungsleistung ermöglicht. Aufgrund der Freiwilligkeit von Registrierung und Angabe von Lebenslaufdaten ist damit aber nicht in umfassendem Maße zu rechnen. Der Kerndatensatz Forschung würde diese Verknüpfungen auch zulassen – er hat explizit die Schaffung einer besseren Datenbasis für die Forschungsevaluation zum Ziel. Hier ist allerdings nicht mit einer flächendeckenden Umsetzung in näherer Zukunft zu rechnen. Außerdem ist hier noch nicht klar, ob die Daten tatsächlich auf der individuellen Ebene vorgehalten werden oder auf anderen organisationalen Einheiten (z. B. Lehrstühle, Institute, Fachbereiche) aggregiert werden.

### 6.2 Organisationale Innovationsleistung und Wissensspillover

Die forschungs- und innovationspolitische Relevanz der Untersuchung von Umfang und Struktur der Forschermobilität ergibt sich vor allem aus der plausiblen Vermutung, dass die Zu- und Abwanderung von Forschern Auswirkungen auf die Innovationsleistung der betroffenen Organisationen (Universitäten, Forschungseinrichtungen und Unternehmen) und Innovationssysteme hat. Die konzeptionellen Bezüge zu den Stichworten implizites Wissen, Wissensspillover, *brain drain*, *brain gain* und *brain circulation* sind bereits im einleitenden Kapitel dieser Machbarkeitsstudie genannt worden. Es ist daher wünschenswert, die Datensätze zur Forschermobilität mit Informationen zur Innovationsleistung von Organisationen oder Innovationssystemen zu verknüpfen.

Prinzipiell können dafür die gleichen Datensätze genutzt werden, wie für die Analyse der individuellen Auswirkungen. Die Daten müssen hier allerdings auf der Ebene einer je nach Fragestellung abzugrenzenden Organisation erhoben werden. Dazu werden neben Informationen zum Arbeitsort des Forschers auch Informationen zur organisationalen Zugehörigkeit benötigt, die nicht in allen genannten Datenquellen vorliegen. Außerdem ist die Zuordnung zu einer Organisation (z. B. einem Unternehmen oder einer Universität) möglicherweise zu grob, um tatsächliche Effekte der Mobilität einer Person messen zu können. Eine detailliertere Zuordnung der Forscher zu einzelnen Abteilungen, Fachbereichen oder gar Lehrstühlen ist aber auf Basis keiner der genannten Datenquellen realistisch. Folgende Verfahren könnten jedoch geeignet sein, dieses methodische Problem teilweise zu lösen und Aussagen über die Wirkungen der Forschermobilität auf organisationaler oder regionaler Ebene ermöglichen:

- Da die Abteilungsstruktur von Organisationen in keinem der Datensätze nachgebildet werden kann, ist die Abgrenzung über Wissenschaftsdisziplinen oder Forschungsthemen eine Alternative. So könnte untersucht werden, wie sich die Forschungsleistung einer Organisation in Feldern, die von starker Zu- oder Abwanderung von Forschern geprägt waren, entwickelt hat. Gleiches kann für die regionale Ebene durch eine Verknüpfung der For-

schemobilität für ein bestimmtes Wissenschaftsfeld in einer Region und den korrespondierenden Leistungsindikatoren erreicht werden.

- Zusätzlich zu Leistungsindikatoren, die auf die reine Anzahl an Publikationen, Patenten und Forschungsmitteln bzw. deren Qualität abzielen, können Vernetzungsindikatoren herangezogen werden, um den Wissenstransfer durch mobile Forscher zu messen. Ko-Publikationen, Ko-Patente oder kooperative Forschungsprojekte können Hinweise auf das Ausmaß der Vernetzung an einem Standort geben und – auch im Sinne des Ansatzes der *brain circulation* – Belege dafür liefern, ob durch die Präsenz eines international mobilen Forschers neue interregionale Vernetzungen (z. B. in die Herkunftsregion) entstanden sind.

### 6.3 Determinanten der Forschermobilität

Zusätzlich zu den Auswirkungen der Forschermobilität sind auch die zugrunde liegenden Determinanten von Bedeutung, da durch diese Information unmittelbar politische Handlungsempfehlungen abgeleitet werden können, wie die Attraktivität einzelner Einrichtungen, Regionen oder des Wissenschaftssystems für Forscher gesteigert werden kann. Unmittelbare Informationen zu den Beweggründen für die Mobilität einzelner Individuen sind jedoch in keinem der diskutierten Datensätze enthalten und könnten nur durch zusätzliche direkte Befragungen der mobilen Forscher erhoben werden. Mittelbar kann jedoch eine Annäherung an Informationen zu den Determinanten der Mobilität aus einigen der beschriebenen Datensätze gelingen.

Durch den Saldo der Forschermobilität für einzelne Einrichtungen, Regionen oder Wissenschaftssysteme kann quantitativ oder (im Falle einer Verknüpfung mit individuellen Leistungsindikatoren) qualitativ eine Aussage darüber getroffen werden, welche Einrichtungen, Regionen oder Wissenschaftssysteme besonders attraktiv bzw. unattraktiv für Forscher sind. Durch einen Vergleich der Eigenschaften dieser Einrichtungen, Regionen oder Wissenschaftssysteme kann zumindest indirekt auf Merkmale geschlossen werden, die tendenziell für das Attrahieren von Forschern förderlich sind, auch wenn dadurch die tatsächlichen individuellen Beweggründe verborgen bleiben.

Weitere Hinweise auf Motive der Mobilität lassen sich erschließen, wenn Karriereverläufe differenziert ausgewertet werden. Voraussetzung dazu sind Daten, die mit den Informationen eines Lebenslaufes vergleichbar sind und z. B. durch die beschriebenen NLP-gestützten Verfahren rekonstruiert werden können. Einige Hinweise zu den Motiven der Forschermobilität lassen sich sogar aus Daten des Mikrozensus oder aus Arbeitskräfteerhebungen ableiten, sofern aus Kombination des Alters und des Zuzugsjahres die Karrierephase identifiziert werden kann, in der ein Forscher zu- oder rückgewandert ist. Unter anderem können folgende Aspekte untersucht werden:

- Welche in der jeweiligen Datenquelle identifizierbaren Eigenschaften der Forscher führen zu welchen Mobilitätsentscheidungen (Ziel der Mobilität, Dauer der Mobilität, Anzahl der Mobilitätsereignisse)? Durch die Ergebnisse können individuelle Determinanten für Mobilität abgeleitet werden. Eine Übertragung der Ergebnisse auf bestimmte Regionen ist jedoch nicht möglich. Allerdings können Teilpopulationen der gesamten Forscher abge-

grenzt werden, an die sich die Forschungs- und Innovationspolitik mit zielgruppenspezifischen Mobilitätsangeboten wenden kann.

- Welche Wissenschaftssysteme, Regionen oder Einrichtungen werden in welchen Karrierephasen (z. B. Nachwuchsphase, Etablierung, Spätphase) besonders häufig als Ziel mobiler Forscher gewählt? Anhand der Kenntnis der Kriterien, nach denen im allgemeinen Forscher bestimmter Karrierephasen ihre Standortentscheidung treffen, lässt sich ableiten, wo diese Anforderungen überdurchschnittlich gut erfüllt werden.
- Wie haben sich Veränderungen bestimmter Aspekte von Wissenschaftssystemen und des Regulierungsrahmens für Mobilität auf die Forschermobilität ausgewirkt? Sind Änderungen in den Mobilitätsmustern im Vergleich der Situationen vor und nach einer Veränderung der Rahmenbedingungen erkennbar? Sollte dies der Fall sein, ist zumindest implizit davon auszugehen, dass die Änderung eine gewisse positive oder negative Auswirkung auf die Forschermobilität hatte.

Insgesamt ist eine Untersuchung der Mobilitätsmotive auf Basis der vorliegenden Datensätze nur in eingeschränktem Maße. Aufgrund des stark individuellen Charakters sind detaillierte Mobilitätsmotive nur durch direkte Befragungen der Forscher selbst zu identifizieren, die wiederum nur mit erheblichem Aufwand und unter Inkaufnahme der in Kapitel 2 geschilderten methodischen Einschränkungen möglich sind.

### 7 Zusammenfassung

Das Ziel der Machbarkeitsstudie „Forschermobilität und Innovation“ war es, Datenquellen und methodische Ansätze zu identifizieren, durch welche die internationale Mobilität von Forschern mit Deutschland als Ziel- und Herkunftsland besser als bislang erfasst und analysiert werden kann. Es wurden solche Ansätze gesucht, mit denen Antworten auf gesellschaftlich relevante Fragen bezüglich Wissensbilanz, ökonomischer Bilanz, Auswirkungen und Determinanten der Forschermobilität gegeben werden können und die das Potenzial besitzen, in ein regelmäßiges Berichtssystem überführt zu werden. Die Machbarkeitsstudie ist in fünf Teilkapitel untergliedert, deren Kernergebnisse in diesem Kapitel zusammengefasst werden:

- Definition von inhaltlichen und technischen Anforderungen an ein Berichtssystem zur Erfassung der Forschermobilität (Kapitel 2),
- Bewertung deutscher Datensätze/-quellen und empirischer Forschungsansätze hinsichtlich ihrer Einsatzmöglichkeiten als Basis quantitativer Untersuchungen zur Forschermobilität (Kapitel 3),
- Bewertung internationaler Datensätze/-quellen und empirischer Forschungsansätze hinsichtlich ihrer Einsatzmöglichkeiten als Basis quantitativer Untersuchungen zur Forschermobilität (Kapitel 4),
- Ausführliche Bewertung von zwei besonders vielversprechenden Ansätzen hinsichtlich Kosten, Ressourceneinsatz und Risiken, die mit dem jeweiligen Ansatz verbunden sind (Kapitel 5) und
- Möglichkeiten der Verknüpfung des Berichtssystems mit anderen Informationen zu Auswirkungen und Determinanten der Forschermobilität (Kapitel 6).

#### Definition von Anforderungen

Zum Einstieg in die Untersuchung der Forschermobilität müssen die Begriffe Forscher und Mobilität abgegrenzt werden. Das Frascati-Manual bietet eine international anerkannte Definition für Forscher, die z. B. anhand der International Standard Classification of Occupations (ISCO) operationalisiert werden kann. Im Rahmen der Machbarkeitsstudie wurde ausschließlich internationale Mobilität untersucht. Es wird außerdem empfohlen, Mobilität von einer Dauer von unter sechs Monaten nicht zu berücksichtigen und die Untersuchung auf kurzfristige (sechs Monate bis ein Jahr), mittelfristige (ein bis drei Jahre) und langfristige (mehr als drei Jahre) Mobilität zu beschränken.

Die entscheidenden Anforderungen an geeignete Datensätze und Datenquellen setzen sich aus inhaltlichen und methodischen Aspekten zusammen. Inhaltlich sollten im Idealfall alle Informationen in der Datenquelle enthalten sein, die für die Beantwortung der gesellschaftlich relevanten Fragen im Zusammenhang mit Forschermobilität notwendig sind. Es konnten jedoch keine Quellen identifiziert werden, für die das vollständig zutrifft. Von zentraler Bedeutung für Analysen, die über die Abbildung von Ausmaß und Struktur der Mobilität hinausgehen sollen, ist die personenscharfe Identifizierbarkeit von Forschern. Nur sie ermöglicht es, eine Verknüpfung mit Daten aus anderen Quellen vorzunehmen. Für Analysen des Mobilitätsverhaltens ist außerdem die Abbildung individueller Karriereverläufe im Zeitverlauf notwendig. Methodisch

wird ein Verfahren gesucht, das für ein regelmäßiges Berichtssystem geeignet ist. Fallspezifische Befragungen fallen als Methode aus, da diese häufig keine verzerrungsfreien Informationen liefern können und einen hohen Ressourceneinsatz benötigen.

### **Bewertung deutscher Datensätze**

Die Diskussion deutscher Datensätze und Datenquellen zur Forschermobilität hat gezeigt, dass an vielen Stellen Informationen zu Teilaspekten des Phänomens vorliegen, es aber keinen Datensatz gibt, der das Phänomen vollständig für alle in der Einleitung genannten Fragen erfasst. Ein zentrales Problem bei der Beschränkung auf deutsche Daten besteht darin, dass ein internationales Phänomen darin nur teilweise abgebildet werden kann. Deutsche, die ins Ausland abwandern, sind in keinem der Datensätze in zufriedenstellender Weise erfasst. Von den deutschen Datensätzen ist der Mikrozensus am besten zur Erfassung der Forschermobilität geeignet. Er erfüllt als repräsentative Umfrage der deutschen Wohnbevölkerung am ehesten das Kriterium einer verzerrungsfreien Erfassung des Phänomens für alle Tätigkeitsfelder und Berufsgruppen. Außerdem ist bei Nutzung des Mikrozensus ein internationaler Vergleich mit methodisch ähnlich gelagerten Befragungen, z. B. Arbeitskräfteerhebungen in Hauptzielländern der deutschen Forschermobilität, möglich.

Zukünftig könnte der geplante Kerndatensatz Forschung großes Potenzial besitzen, eine umfassende Quelle für Forschungsinformationen zu werden. Es wird daher empfohlen, sich aktiv während der Phase der Projektdefinition einzubringen, damit Forschermobilität mit dem Datensatz abgebildet und in Zukunft auch tatsächlich untersucht werden kann.

### **Bewertung internationaler Datensätze**

Internationale Datensätze, in denen auch die Mobilität deutscher Forscher abgebildet ist, sind besonders geeignet, um Abwanderung aus Deutschland zu untersuchen und internationale Vergleiche durchzuführen. Besonders vielversprechend sind (i) die Nutzung von Publikationsdatenbanken, (ii) internationale Vergleiche nationaler Zensusdaten und Arbeitskräfteerhebungen sowie (iii) automatisierte Verfahren zur Generierung von Karriereverläufen zu einzelnen Forschern auf Basis frei im Internet zugänglicher Informationen und mit Hilfe von Techniken des Natural Language Processing. Die anderen untersuchten internationalen Datensätze bieten für sich allein genommen entweder keine genügenden Informationen zum Thema Forschermobilität, umfassen (noch) eine zu geringe Fallzahl (internationale Forscherdatenbanken) oder entsprechen nicht mit hinreichender Genauigkeit der Definition von Forschern (internationale Absolventen- und Promoviertenbefragungen).

### **Ausführliche Darstellung von zwei vielversprechenden Ansätzen**

Kurzfristig ist es mittels bestehender Datensätze möglich, eine international vergleichende Auswertung von Zensusdaten und Arbeitskräfteerhebungen vorzunehmen. Mittelfristig besteht erhebliches Innovationspotenzial in der Nutzung von Natural Language Processing für die Generierung eines neuen Datensatzes zur Forschermobilität.

Der Rückgriff auf Zensusdaten und Arbeitskräfteerhebungen ermöglicht es, Ausmaß und Struktur der Forschermobilität international vergleichend darzustellen und tiefergehend zu analysieren. Dabei können deutsche Mikrozensusdaten mit den entsprechenden Erhebungen anderer Länder kombiniert werden. Der Schwerpunkt der Untersuchung sollte auf den Haupt-



zielländern der deutschen Forschermobilität liegen, deren Stichprobe hinreichend groß für die Analyse von kleinen Teilpopulationen ist. Daher sind insbesondere die Daten des American Community Survey (ACS) und mit Abstrichen die Arbeitskräfteerhebungen Großbritanniens und Frankreichs geeignet. Die notwendigen Analysen können kurzfristig umgesetzt werden und bieten aufgrund der jährlichen Datenerhebung das größte Potenzial, die Grundlage für ein Berichtssystem zu bilden.

Die Generierung eines Datensatzes zur Forschermobilität durch Methoden des Natural Language Processing (NLP) bietet im Vergleich zu allen anderen im Rahmen der Machbarkeitsstudie vorgestellten Verfahren die größten Chancen, einen innovativen und einzigartigen Datensatz zu diesem Themenfeld hervorzubringen. Ein solcher Datensatz ermöglicht die Abbildung von Zu-, Rück- und Abwanderung, da durch die Nutzung des Internets keine Einschränkung auf Quellen einzelner Länder besteht. Aus den gesammelten Informationen können personenscharfe Karriereverläufe rekonstruiert werden. Auch eine Verknüpfung mit weiteren Datensätzen zur Forschungsleistung ist möglich. Da auf öffentlich zugängliche Informationen zugegriffen wird, gibt es keine Probleme bei der Umsetzung, wie sie bei der Verknüpfung von Daten im Zuge der Erstellung des Kerndatensatzes zu erwarten sind. Die NLP-Methode ist dadurch deutlich schneller realisierbar und flexibler einsetzbar als der Kerndatensatz-Forschung. Diese Vorteile wiegen die Risiken und die nicht zu 100 % genaue Zuordnung der mobilitätsrelevanten Informationen aus Sicht des Autors der Machbarkeitsstudie auf. Daher wird der Expertenkommission Forschung und Innovation empfohlen, die Ausschreibung eines solchen Projektes gemeinsam mit dem BMBF zu forcieren.

### **Daten zu Auswirkungen und Determinanten der Forschermobilität**

Informationen zu Auswirkungen und Determinanten der Mobilität liegen in den untersuchten Datenquellen selbst nicht vor. Sie müssen entweder durch Verknüpfung der Mobilitätsdaten mit anderen Datenquellen erzeugt werden oder indirekt aus den Strukturen der Mobilität abgeleitet werden. Eine wichtige Voraussetzung für die Verknüpfung mit anderen Daten ist die namentliche Identifizierung von Forschern in der genutzten Datenquelle. Dies ist vor allem bei Datensätzen, die durch NLP-gestützte Methoden generiert werden, oder bei Publikationsdaten gegeben. Informationen zu den Auswirkungen der Mobilität können durch die Forschungsleistung (Publikationen, Patente, Forschungsprojekte) oder die Vernetzung der Forscher (Ko-Publikationen, Ko-Patente, Forschungskooperationen) abgebildet werden. Sie sind auf individueller Ebene leichter auszuwerten als auf organisationaler Ebene, da sich sowohl Probleme bei der eindeutigen Zuordnung der Daten zu einer bestimmten Einrichtung oder Abteilung ergeben und die Auswirkungen eines mobilen Forschers im Falle großer Organisationen (Universitäten, Großunternehmen) oder ganzer Innovationssysteme von anderen Einflussfaktoren überlagert wird. Determinanten der Forschermobilität können direkt nur durch zusätzliche Befragungen erfasst werden. Eine Analyse der Strukturen der Forschermobilität und von Karriereverläufen kann lediglich indirekt Hinweise zu Determinanten auf Basis der Mobilitätsdaten geben.

Abschließend bleibt die hohe gesellschaftliche und ökonomische Relevanz der internationalen Forschermobilität hervorzuheben. Wissen ist die wichtigste Ressource für den Erhalt der Wettbewerbsfähigkeit Deutschlands und Forscher sind die wichtigsten Wissensträger. Da aufgrund des Fehlens einer belastbaren Datenbasis bislang nur unzureichende empirische Erkenntnisse zur Forschermobilität vorliegen, wird die weitere Förderung der Forschung zu die-

sem Thema empfohlen. Dabei sind insbesondere die international vergleichende Auswertungen von Mikrozensus und Arbeitskräfteerhebungen, Auswertung von Publikationsdaten und die Generierung von Mobilitätsdaten durch Methoden des Natural Language Processing vielversprechend. Zusätzlich ist eine Beteiligung am Projekt „Kerndatensatz Forschung“ sinnvoll, da hier in der Frühphase darauf hingewirkt werden kann, dass auch mobilitätsrelevante Informationen erfasst werden.

### 8 Literaturverzeichnis

- Arndt, C.; Christensen, B.; Gurka, N. (2010): Abwanderung von Hochqualifizierten aus Baden-Württemberg. IAW Policy Report 3/2010. Tübingen: Institut für Angewandte Wirtschaftsforschung.
- Auriol, L. (2010): Careers of Doctorate Holders: Employment and Mobility Patterns. OECD Science, Technology and Industry Working Papers 2010/04 Paris: OECD.
- Backstrom, L.; Boldi, P.; Rosa, M.; Ugander, J.; Vigna, S. (2012): Four Degrees of Separation. <http://arxiv.org/abs/1111.4570>.
- BAMF (2012): Wanderungs- und Freizügigkeitsmonitoring: Migration nach Deutschland – Halbjährlicher Bericht mit Junizahlen (Stand 30.06.2012). Nürnberg: Bundesamt für Migration und Flüchtlinge.
- BAMF (2013): Migrationsbericht 2011. Nürnberg: Bundesamt für Migration und Flüchtlinge.
- Calmfors, L.; Corsetti, G.; Flemming, J.; Honkapohja, S.; Kay, J.; Leibfritz, W.; Saint-Paul, G.; Sinn, H.-W.; Vives, X. (2003): Should We Worry about the Brain Drain?. EEAG European Economic Advisory Group at CESifo: Report on the European Economy 2003, 118-130.
- Cañibano, C.; Bozeman, B. (2009): Curriculum vitae method in science policy and research evaluation: the state-of-the-art. *Research Evaluation* 18(2): 86-94.
- DAAD und HIS (2012): Wissenschaft weltoffen 2012. Daten und Fakten zur Internationalität von Studium und Forschung in Deutschland. Bielefeld.
- Diehl, C.; Dixon, D. (2005): Zieht es die Besten fort? Ausmaß und Formen der Abwanderung deutscher Hochqualifizierter in die USA. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 57(4):714-734.
- Docquier, F. und Marfouk A. (2004): Measuring the International Mobility of Skilled Workers (1990-2000) – Release 1.0. Policy Research Working Paper Number WPS3381. The World Bank: Washington, DC.
- Enders, J.; Bornmann, L. (2002): Internationale Mobilität bundesdeutscher Promovierter – Eine Sekundäranalyse der Kasseler Promoviertenstudie. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 35(1): 60-73.
- Engin, T.; Reifenberg, D. (2012): Studie über den Verbleib von Teilnehmerinnen und Teilnehmern der GAIN-Jahrestagungen 2004-2011. Abschlussbericht für den DAAD im Auftrag des Bundesministeriums für Bildung und Forschung. Umfragezentrum Bonn.
- Erlinghagen, M.; Stegmann, T. (2009): Goodbye Germany – und dann? Erste Ergebnisse einer Pilotstudie zur Befragung von Auswanderern aus Deutschland. SOEP paper 193. Berlin: DIW.
- Ette, A.; Sauer, L. (2010): Auswanderung aus Deutschland. Daten und Analysen zur internationalen Migration deutscher Staatsbürger. Wiesbaden: VS Verlag.

- Ferrara, A. und Salini, S. (2012): Ten challenges in modeling bibliographic data for bibliometric analysis. *Scientometrics* 93: 765-785.
- Fritsch, M. und Schwirten, C. (1998): Öffentliche Forschungseinrichtungen im regionalen Innovationssystem. *Raumforschung und Raumordnung* 56(4): 253-263.
- García Flores, J.J.; Zweigenbaum, P.; Zhao, Y.; Turner, W. (2012): Tracking Researcher Mobility on the Web Using Snippet Semantic Analysis. *Advances in Natural Language Processing. Lecture Notes in Computer Science* 7614: 180-191.
- Hauss, K.; Kaulisch, M.; Zinnbauer, M.; Tesch, J.; Fräßdorf, A. Hinze, S.; Hornbostel, S. (2012): Promovierende im Profil: Wege, Strukturen und Rahmenbedingungen von Promotionen in Deutschland. Ergebnisse aus dem ProFile-Promovierendenpanel. iFQ-Working Paper 13. Berlin.
- Herwig, S.; Becker, J. (2012): Einführung eines Forschungsinformationssystems an der Westfälischen Wilhelms-Universität Münster – Von der Konzeption bis zur Implementierung. in: Bittner, S.; Hornbostel, S.; Scholze, Frank (Hrsg.): *Forschungsinformation in Deutschland: Anforderungen, Stand und Nutzen existierender Forschungsinformationssysteme. Workshop Forschungsinformationssysteme 2011. iFQ-Working Paper 10.* Berlin. S. 41-53.
- Homeland Security (2012): *Yearbook of Immigration Statistics 2011.* Office of Immigration Statistics.
- Hornbostel, S. (Hrsg.) (2012): *Wer promoviert in Deutschland? Machbarkeitsstudie zur Doktorandenerfassung und Qualitätssicherung von Promotionen an deutschen Hochschulen.* iFQ-Working Paper 14. Berlin.
- Horowitz, I. (1966): Some aspects of the effects of the regional distribution of scientific talent on regional economic activity. *Management Science* 13: 217-232.
- IDEA Consult (Hrsg.) (2010a): *Study on mobility patterns and career paths of EU researchers. Technical Report 2 – Part II: Industrial researcher mobility study.*
- IDEA Consult (Hrsg.) (2010b): *Study on mobility patterns and career paths of EU researchers. Final Technical Report 3: Extra-EU mobility pilot study.*
- Ioannidis, J.P.A. (2004): Global estimates of high-level brain drain and deficit. *Journal of the Federation of American Societies for Experimental Biology* 18: 936-939.
- Jaksztat, S.; Preßler, N.; Briedis, K. (2011): *Die internationale Ausrichtung des wissenschaftlichen Nachwuchses. HIS:Forum Hochschule F10/2011.* Hannover: Hochschul-Informations-System.
- Jöns, H. (2007): Transnational mobility and the spaces of knowledge production: A comparison of global patterns, motivations and collaborations in different academic fields. *Social Geography* 2: 97-114.
- Kim, D.; Bankart, C.; Isdell, L. (2011): International doctorates: trends analysis on their decision to stay in US. *Higher Education* 62(2): 141-161.

- Klingert, I.; Block, A.H. (2013): *Ausländische Wissenschaftler in Deutschland. Analyse des deutschen Arbeitsmarktes für Forscherinnen und Forscher*. BAMF Working Paper 50. Bundesamt für Migration und Flüchtlinge.
- Moed, H.F.; Aisati, M.; Plume, A. (2013): *Studying scientific migration in Scopus*. *Scientometrics* 94(3): 929-942.
- Moguérou, P.; Di Pietrogiacomo, M.P. (2008): *Stock, Career and Mobility of Researchers in the EU*. JRC Scientific and Technical Report.
- OECD (1995): *Canberra Manual. Manual on the Measurement of Human Resources Devoted to S&T*. Paris: OECD.
- OECD (2002): *Frascati Manual. Proposed Standard Practice for Surveys on Research and Experimental Development*. 6. Aufl. Paris: OECD.
- OECD (2008): *The Global Competition for Talent. Mobility of the Highly Skilled*. Paris.
- Reiner, C. (2010): *Brain competition policy as a new paradigm of regional policy. A European perspective*. *Papers in Regional Science* 89(2): 449-461.
- Saxenian, A. (2005): *From brain drain to brain circulation: transnational communities and regional upgrading in India and China*. *Comparative International Development* 40(2): 35-61.
- Schiller, D. (2011): *Institutions and practice in cross-sector research collaboration: conceptual considerations with empirical illustrations from the German science sector*. *Science and Public Policy* 38(2): 109-121.
- Schiller, D. und Kiese, M. (2010): *Hochschulen und öffentliche Forschungseinrichtungen als Motoren regionaler Cluster- und Wirtschaftsentwicklung*. *Berichte zur deutschen Landeskunde* 84(2): 105-113.
- Schiller, D. und Revilla Diez, J. (2008): *Mobile star scientists as regional knowledge spillover agents*. IAREG Working Paper WP2/7.
- Schiller, D. und Revilla Diez, J. (2010): *Local embeddedness of knowledge spillover agents: Empirical evidence from German star scientists*. *Papers in Regional Science* 89(2): 275-294.
- Schiller, D. und Revilla Diez, J. (2012): *The Impact of Academic Mobility on the Creation of Localized Intangible Assets*. *Regional Studies* 46(10): 1319-1332.
- Schütte, G. (2006): *Diplomatie der Forscher. Wenn Deutschland international mithalten will, braucht es eine Außenwissenschaftspolitik. Ein Plädoyer*. *Die Zeit*, 12.04.2006.
- Statistisches Bundesamt (2011): *Deutsche Studierende im Ausland. Statistischer Überblick 1999-2009*. Statistisches Bundesamt: Wiesbaden.
- Trippl, M. (2013): *Islands of Innovation as Magnetic Centres of Star Scientists? Empirical Evidence on Spatial Concentration and Mobility Patterns*. *Regional Studies* 47(2): 229-244.
- Ugander, J.; Karrer, B.; Backstrom, L.; Marlow, C. (2011): *The Anatomy of the Facebook Social Graph*. <http://arxiv.org/abs/1111.4503>.

Wissenschaftsrat (2013): Empfehlungen zu einem Kerndatensatz Forschung. Drucksache 2855-13. Berlin.

Zucker, L. und Darby, M. (2006): Movement of star scientists and engineers and high-tech firm entry. NBER Working Paper No. 12172