

8 Untersuchungsdesign

Bei der Darstellung der Fragestellung im vorangegangenen Kapitel wurde bereits deutlich, dass die Untersuchung explorativ vorgeht und darauf gerichtet ist, die Rolle von Open Access für die Inklusion von Autoren und Rezipienten in das formale Kommunikationssystem umfassend zu analysieren. Dabei ist es Ziel, eine gegenstandsorientierte Theorie der Nutzung von Open-Access-Publikationsmedien zu entwickeln, die als „material“ bezeichnet werden kann. Darunter sind Theorien zu verstehen, die Aussagen über ein bestimmtes Feld oder einen Gegenstandsbereich – hier Open Access in einem Teil der Wissenschaft – treffen (Glaser und Strauss 1998: 42) und sich durch eine begrenzte Reichweite auszeichnen.¹ Diese Zielsetzung spiegelt sich in einem komplexen Untersuchungsdesign wider, das sich aus drei Komponenten zusammensetzt: einer quantitativen Untersuchung des Publikationsoutputs von Wissenschaftlern, einem Mapping der Publikationsinfrastruktur sowie qualitativen Interviews. Zusammengehalten werden die Bestandteile einerseits durch eine geschichtete Zufallsstichprobe von Wissenschaftlern, die den Gegenstand der Untersuchung bestimmt. Andererseits findet eine Integration aber auch auf der Ebene der empirischen Ergebnisse statt, indem die Befunde aus den drei Komponenten der Untersuchung bei ihrer Darstellung zusammengefügt werden. Einen methodologischen Rahmen dafür stellt die Grounded-Theory-Methodik bereit.

Das Untersuchungsdesign wird in vier Abschnitten vorgestellt. Zunächst werden die Struktur und das Zustandekommen der gerade erwähnten geschichteten Stichprobe erläutert (Kap. 8.1). Danach werden die einzelnen Komponenten des Untersuchungsdesigns beschrieben, die sich komplementär zueinander ver-

1 Materiale Theorien besitzen somit eine Referenz im empirischen Feld der Sozialforschung. Davon zu unterscheiden sind formale Theorien, die den Anspruch haben, für einen bestimmten Bereich der Sozialforschung zu gelten und damit auf die interne Differenzierung der Sozialforschung rekurren.

halten und unterschiedliche Aspekte des Gegenstands ausleuchten.² Den Anfang markiert die *quantitativ-bibliometrische Analyse des Publikationsoutputs der Mitglieder der Stichprobe*. Im Mittelpunkt steht hier die Frage, in welchem Umfang und nach welchen Mustern Wissenschaftler in der Rolle des Autors auf die Publikationsinfrastruktur zugreifen und welche Rolle Open-Access-Publikationsmedien dabei spielen (Kap. 8.2). Daran anschließend wird das Vorgehen beim *Mapping der Publikationsinfrastruktur* dargestellt. In diesem Schritt werden die medientechnischen Einrichtungen und ihre Besonderheiten untersucht, die von den Wissenschaftlern zur Mitteilung und Rezeption von Forschungsergebnissen genutzt werden. Kombiniert wird ein quantifizierendes Vorgehen mit einer qualitativen Untersuchung zentraler Bestandteile der Infrastruktur (Kap. 8.3). Den Abschluss bilden die *leitfadengestützten Experteninterviews* mit Wissenschaftlern aus der Stichprobe. Mit ihnen wird nicht nur ein Zugang zu den Handlungsroutrinen gewonnen, in denen die Publikationsinfrastruktur in der Rolle des Autors und des Rezipienten genutzt wird, sondern auch zu den Zuschreibungen und Deutungen, die mit der Nutzung einhergehen (Kap. 8.4).

8.1 KOMPOSITION DER GESCHICHTETEN STICHPROBE

Der Gegenstand der Untersuchung – die Gruppe von Wissenschaftlern, deren Inklusion in das formale Kommunikationssystem der Wissenschaft im Folgenden interessiert – wird durch eine geschichtete Zufallsstichprobe bestimmt. Eine Schichtung wird vorgenommen, um systematische Vergleiche entlang der Dimensionen „Fach“, „Land“ und „Kohorte“ anstellen zu können. Zunächst soll das Ziel der Stichprobenkonstruktion beschrieben, die einzelnen Schichtungsdimensionen vorgestellt, das praktische Vorgehen bei der Realisierung der Stichprobe erläutert und die tatsächlich realisierte Stichprobe dargestellt werden.

Schichtungsdimensionen

Technisch gesprochen besteht das Ziel darin, eine nach drei Dimensionen geschichtete Zufallsstichprobe zu erzeugen. Die Schichtungskriterien sind dabei

-
- 2 Die Verfahrensweise der Untersuchung kann auch als Triangulation bezeichnet werden – und zwar nicht „als Strategie der Validierung [in] der qualitativen Forschung“, sondern als Vorgehen, das „Erkenntnisse durch die Gewinnung weiterer Erkenntnisse zu begründen und abzusichern“ sucht (vgl. Flick 1992a; 1992b; 2007: 311; Denzin und Lincoln 1994: 2).

„Fach“ (mit den Ausprägungen Astronomie und Mathematik), „Herkunftsland“ (mit den Ausprägungen Deutschland und Südafrika) sowie „Kohorte“ (mit den Ausprägungen Etablierung im Fach *vor* beziehungsweise *nach* dem Entstehen der ersten Open-Access-Publikationsmedien). Die Gesamtstichprobe setzt sich demnach aus $2 \times 2 \times 2$ (oder acht) Teilstichproben zusammen (siehe Tabelle 8.1). Um Verfahren der schließenden Statistik im Fall von quantitativen Daten anwenden zu können, wird eine Größe von 30 Personen je Teilstichprobe angestrebt, so dass der Umfang der Gesamtstichprobe bei 240 Personen liegen soll.

Tabelle 8.1: Bezeichnung und Merkmalskombination der Teilstichproben

<i>Schicht</i>	<i>Fach</i>	<i>Kohorte</i>	<i>Herkunftsland</i>
<i>Math Pre GER</i>	Mathematik	vor Open Access	Deutschland
<i>Math Post GER</i>	Mathematik	während Open Access	Deutschland
<i>Math Pre ZA</i>	Mathematik	vor Open Access	Südafrika
<i>Math Post ZA</i>	Mathematik	während Open Access	Südafrika
<i>Astro Pre GER</i>	Astronomie	vor Open Access	Deutschland
<i>Astro Post GER</i>	Astronomie	während Open Access	Deutschland
<i>Astro Pre ZA</i>	Astronomie	vor Open Access	Südafrika
<i>Astro Post ZA</i>	Astronomie	während Open Access	Südafrika

Fach (Astronomie und Mathematik): Als schwierigste der drei Schichtungsdimensionen stellt sich die disziplinäre Herkunft der Wissenschaftler, die bei den theoretischen Überlegungen zur Binnendifferenzierung der Wissenschaft nicht weiter problematisiert wurde³, heraus: Im Grundsatz besteht das Problem darin, *eindeutige* Grenzen zwischen häufig *uneindeutigen* Fällen ziehen zu müssen, deren Forschungsgebiet in den Zuständigkeitsbereich von zwei (oder gar mehreren) Fächern fallen: Wie ist beispielsweise mit einem Statistiker umzugehen, der Methodenentwicklung betreibt und sowohl in statistischen Journalen als auch in soziologischen Fachzeitschriften publiziert und Mitglied von Fachvereinigungen beider Gebiete ist? Ist er der hier interessierenden Gruppe der Mathematiker zuzuordnen oder nicht? Welchem Fach ist ein Professor zuzuschlagen, der einen Lehrstuhl für Hochenergiephysik innehat, dessen Arbeiten aber vor allem von Astrophysikern rezipiert werden? Sollten diese beiden Fälle in eine der Teilstichproben aufgenommen werden – und wenn ja, aus welchem Grund? Da Disziplinen mehrdimensionale Gebilde sind, kommen zur Abgrenzung unterschiedliche Kriterien in Frage, die keineswegs zu deckungsgleichen Fächergrenzen führen: Sie sind *selbstreferentielle Kommunikationsgemeinschaften, reproduzie-*

3 Siehe hierzu Kap. 5.

ren sich durch *Ausbildungsorganisationen*, regulieren ihre Außenbeziehungen über *Fachgesellschaften* und unterliegen der *Außenbeobachtung durch Förderorganisationen* auf der Grundlage von Fächerschemata und Fachsystematiken (Weingart und Schwechheimer 2007: 41ff.).⁴ Vor dem Hintergrund des Interesse an der Frage des Austauschs von Forschungsergebnissen bietet es sich an, bei der Bestimmung von Fächergrenzen auf das formale wissenschaftliche Kommunikationssystem zu rekurrieren und jene Personen als Mitglieder eines Fachs zu verstehen, die sich mit einer bestimmten Häufigkeit in der Rolle des Autors am fachlichen Austausch beteiligt. Praktisch wird zur Bestimmung von Fächergrenzen auf die Fachsystematik einer wissenschaftsübergreifenden Datenbank – dem *Science Citations Index Expanded* (SCIE) – zurückgegriffen. Die Inklusion in der Rolle des Autors sollte dabei in anderen Gebieten nicht größer sein als im interessierenden Fach.⁵

Herkunftsland (Deutschland/Südafrika): Weniger schwierig ist dagegen die Bestimmung der zweiten Schichtungsdimension. Grenzfälle gibt es auch hier durch die Migration von Wissenschaftlern, die in der Regel im Laufe ihrer Karrieren an unterschiedlichen Forschungseinrichtungen beschäftigt sind und zum Teil auch Ländergrenzen überschreiten, um zeitweilig oder dauerhaft im Ausland zu arbeiten. Mit Gastprofessuren und *visiting fellowships* existieren daneben institutionalisierte Formen zeitlich begrenzter Aufenthalte an wissenschaftlichen Einrichtungen, die eine Zuordnung zu einem Herkunftsland erschweren können. Zudem verfügen manche Wissenschaftler zu einem Zeitpunkt über mehrere Affiliationen in unterschiedlichen Ländern. Um hier eine eindeutige Zuordnung vornehmen zu können, findet eine Beschränkung auf eindeutige Fälle statt. In die Stichprobe aufgenommen werden nur Wissenschaftler, deren Publikationen ausschließlich die Adresse von Einrichtungen aus einem Land – entweder Deutschland oder Südafrika – tragen. Ausgeschlossen werden dagegen Personen, die grenzüberschreitend migrierten, zumindest insoweit sich dies auch in ihren Publikationen abbildet.

4 Die genannten Kriterien korrespondieren mit Datenquellen, die zur Ziehung einer fachspezifischen Stichprobe genutzt werden könnten. So gibt es Mitgliederverzeichnisse von Fachgesellschaften (wie das World Directory of Mathematicians, IMU 2002), Stellenverzeichnisse von Forschungseinrichtungen (z.B. Longman Ed. Team 1985), übergreifende Verzeichnisse von Wissenschaftlern (z.B. Gelehrtenkalender 2016) oder Datenbanken geförderter Wissenschaftler, denen eine Fachsystematik zugrunde liegt (DFG 2016a). Allerdings unterscheiden sich die Datenquellen deutlich hinsichtlich ihrer Eignung zur Ziehung der anvisierten Stichprobe.

5 Die Schichtungskriterien werden weiter unten operationalisiert.

Kohorte (vor/nach Entstehen der ersten frei zugänglichen Publikationsmedien): Völlig unproblematisch ist das Schichtungskriterium Kohorte, mit dem Wissenschaftler, die sich vor beziehungsweise nach der Entstehung der ersten Open-Access-Publikationsmedien etabliert haben, unterschieden werden sollen. Weiter oben wurde bereits darauf hingewiesen, dass das Entstehen der ersten frei zugänglichen elektronischen Publikationsmedien gut dokumentiert ist⁶ und in einen relativ kurzen Zeitraum zwischen 1991–1995 fällt⁷. Daher kann dieser Zeitraum zur Abgrenzung der beiden Wissenschaftlerkohorten verwendet werden.

Datenquelle

In Vorbereitung auf die Stichprobenziehung wurden die Vor- und Nachteile verschiedener Datenquellen abgewogen, darunter auch die Publikationsnachweissysteme der Astronomie (*NASA Astrophysics Data System*) und der Mathematik (*zbMATH* und *MathSciNet*).⁸ Die Entscheidung fiel zugunsten des *Science Citation Index Expanded* (SCIE), da sich die Zitationsdatenbank durch die folgende Kombination von Merkmalen auszeichnet:

- *Dokumentierte Einschlusskriterien:* Der SCIE entscheidet nach einheitlichen und vergleichsweise gut dokumentierten Kriterien über die Aufnahme eines Journals in die Zitationsdatenbank.⁹
- *Einheitliche Zuordnungskriterien zu Fächern:* Die Zuordnung von Journalen (und damit auch von Publikationen) zu „Subject Categories“ (Fächern) findet nach einheitlichen, wenn auch nicht sonderlich gut dokumentierten Kriterien statt. Zu diesen zählen neben dem Titel des Journals auch die Zitationsmuster (Leydesdorff und Rafols 2009: 3).
- *Fachübergreifende Datenbank:* Da es sich um eine fachübergreifende Datenbank handelt, können sämtliche Teilstichproben auf der Grundlage derselben Datenbank und mithilfe einer einheitlichen Prozedur gezogen werden.¹⁰

6 Siehe für eine Übersicht Suber (2009).

7 Siehe Kap. 7.2.

8 Siehe zur Darstellung der beiden Systeme ausführlicher die Kap. 10.2 und 10.3 in dieser Arbeit.

9 Die aktuelle Auswahlprozedur und die dabei zum Einsatz kommenden Kriterien sind in Testa (2016) beschrieben, die Gründe für die Selektivität des SCIE in Garfield (1990) dargelegt.

Vorgehen bei der Ziehung der Stichprobe

Das praktische Verfahren der Stichprobenziehung setzt sich aus vier Schritten zusammen:

- *Beschaffung der Rohdaten:* In einem ersten Schritt wurden sämtliche Publikationen, die den folgenden drei Kriterien entsprechen, aus dem SCIE heruntergeladen und in eine Datenbank verbracht¹¹: Erstens müssen die Publikationen den *Subject Categories* „Astronomy/Astrophysics“, „Mathematics pure“ oder „Mathematics applied“ zugeordnet sein. Zweitens muss der Veröffentlichungszeitpunkt der Publikation zwischen 1972–2005 liegen. Drittens muss sich im Adressfeld des SCIE mindestens einer der folgenden Einträge befinden: „south africa“, „germany“, „fed rep ger“, „bundesrepublik“ oder „ger dem rep“.
- *Erstellung von Teilauswahlgesamtheiten / Vorbereitung der Stichprobenziehung:* Aus den Publikationen wurden sämtliche Autorennamen extrahiert und diese gemäß der Kombination der Ausprägung der Merkmale „Fach“, „Herkunftsland“ und „Kohorte“ einer von acht Listen zugeordnet. Zur Zuordnung von Autoren zu einer Kohorte wurden zwei Kriterien angewandt: Die Autoren dürfen keine Publikation in einem definierten Ausschlusszeitraum (1971–1980 für die ältere und 1985–1994 für die jüngere Kohorte) veröffentlicht haben. Ihnen müssen mindestens drei oder mehr Publikationen in einem der beiden Einschlusszeiträume (1981–1991 oder 1995–2005) zugeordnet werden können.
- *Ziehung der Elemente:* Für jede der acht Listen wurden Zufallszahlen aus einem Zahlenbereich generiert, der der Anzahl der Rangplätze in den Teilauswahlgesamtheiten entsprach.¹² Die gezogenen Elemente wurden in die jeweilige Teilstichprobe aufgenommen.
- *Überprüfung der Autoren:* Jeder Autor wurde auf die Einhaltung der drei Schichtungsmerkmale hin überprüft. Diese Überprüfung war aufgrund von zwei Problemen notwendig: Erstens gelangte durch ko-autorierte Publikationen gerade im Fall der Astronomie, die große Autorenkollektive kennt, eine

10 Dies wäre beispielsweise im Fall der Nutzung von Fachdatenbanken nicht möglich. Ein weiterer Nachteil der Fachdatenbanken besteht darin, dass die Einschlusskriterien für die Aufnahme von Publikationsmedien variieren.

11 Dazu wurde das Tool ISI.exe verwendet (Leydesdorff 1989).

12 Zu diesem Zweck wurde der Online-Dienst <http://www.random.org> (Zugriff am 18. April 2017) genutzt, der echte Zufallszahlen produziert.

große Zahl von Wissenschaftlern in die Auswahllisten, die nicht aus den gesuchten Herkunftsländern stammen.¹³ Daher wurde geprüft, ob der gezogenen Person eine Affiliation aus dem interessierenden Herkunftsländ zugeordnet werden kann und ob dies für sämtliche ihrer Publikationen gilt. Wurde der Autor im Zuge der Überprüfung auch einem anderen Herkunftsländ zugeordnet, wurde er aus der Teilstichprobe ausgeschlossen. Zweitens ist es möglich, dass Wissenschaftler weitere Publikationen verantworten, die anderen „Subject Categories“ zugeordnet sind. Daher wurde überprüft, ob der SCIE für den Autor weitere Publikationen ausweist. Ein Ausschluss des Wissenschaftlers aus der Teilstichprobe fand statt, sofern vorgängige Veröffentlichungen im SCIE verzeichnet waren, die in den Ausschlusszeitraum fielen oder wenn Publikationsaktivitäten des Wissenschaftlers in einer anderen als der gesuchten „Subject Category“ umfangreicher waren.

Definition der Grundgesamtheit und der Schichtungskriterien

Gemäß der Operationalisierung der Schichtungskriterien und dem hier praktizierten Vorgehen der Stichprobenziehung kann die Grundgesamtheit wie folgt definiert werden: *Die Grundgesamtheit setzt sich zusammen aus allen Astronomen und Mathematiker aus Deutschland und Südafrika, die entweder einer Kohorte von Wissenschaftlern angehören, die sich vor dem Entstehen von Open-Access-Publikationsmedien haben etablieren können oder deren Etablierung in einen Zeitraum fällt, in dem bereits Open-Access-Publikationsmedien existierten.* Die drei für die Grundgesamtheit konstitutiven Kriterien sind dabei wie folgt bestimmt:

- *Fach*: Der Schwerpunkt der Publikationsaktivität des Wissenschaftlers liegt in den *Subject Categories* „Astronomy/Astrophysics“ oder „Mathematics (pure and applied)“ des SCIE. In einer dieser *Subject Categories* weist der SCIE für den Autor mindestens drei Publikationen nach.
- *Herkunftsland*: Sämtliche Publikationen nennen für den Autor eine Institution oder mehrere Institutionen mit entweder ausnahmslos deutschen oder ausnahmslos südafrikanischen Adressen.

13 Eine Bereinigung der Listen, aus denen die Ziehung der Teilstichproben vorgenommen wurde, war vor Ziehung der Stichprobe nicht praktikabel, da der SCIE erst seit dem Jahr 2008 Autorennamen mit Adressen verknüpft. Eine manuelle Überprüfung der Einhaltung der Kriterien hätte den Arbeitsaufwand gegenüber dem hier praktizierten Vorgehen erheblich vergrößert.

- *Kohortenzugehörigkeit*: Der SCIE weist in den Zeiträumen von 1971–1980 oder von 1985–1994 für den Autor keine und in den Zeiträumen von 1981–1991 beziehungsweise 1995–2005 drei oder mehr Publikationen nach.

Aufgrund der Kombination der Merkmale ist jeder Wissenschaftler genau einer Teilstichprobe zugeordnet.

Ergebnis der Stichprobenziehung

Als Größe der Gesamtstichprobe wurden 240 Wissenschaftler angestrebt, die sich nach Möglichkeit gleichmäßig auf 8 Teilstichproben mit jeweils 30 Personen verteilen sollten. Dieses Ziel konnte nicht vollständig realisiert werden, wie die nachstehende Tabelle 8.2 zeigt.

Tabelle 8.2: Umfang der Schichten in der Gesamtstichprobe

<i>Schicht</i>	<i>Anzahl Wissenschaftler</i>	<i>Anteil an der Stich- probe</i>
<i>Math Pre GER</i>	35	15,63
<i>Math Post GER</i>	39	17,41
<i>Math Pre ZA</i>	25	11,16
<i>Math Post ZA</i>	23	10,27
<i>Astro Pre GER</i>	40	17,86
<i>Astro Post GER</i>	30	13,39
<i>Astro Pre ZA</i>	18	8,04
<i>Astro Post ZA</i>	14	6,25

Für die deutschen Teilstichproben gelang es, eine ausreichend große Anzahl von Personen zu finden, die den Kriterien entspricht. Bei sämtlichen südafrikanischen Gruppen wurden die Auswahllisten voll ausgeschöpft, der angestrebte Umfang der Teilstichprobe aber verfehlt.¹⁴ Die Ursache dafür liegt zum einen in der vergleichsweise geringen Größe der Kommunikationsgemeinschaft der Astronomen und Mathematiker in Südafrika und zum anderen an einem beachtlichen Anteil an Wissenschaftlern, die eine gewisse Zeit im Ausland geforscht haben. Bei den südafrikanischen Gruppen handelt es sich dementsprechend nicht

14 In der weiteren Untersuchung führt dies aber nicht zu grundsätzlichen Problemen, da der verschieden große Umfang der Teilstichproben durch Gewichtungsfaktoren korrigiert werden kann. Siehe dazu Kap. 9.2.

um Stichproben, sondern um sämtliche Mitglieder der Grundgesamtheit mit den oben genannten Merkmalskombinationen.

8.2 BIBLIOMETRISCHE UNTERSUCHUNG DES PUBLIKATIONSOUTPUTS

Nachdem im ersten Schritt die geschichtete Stichprobe vorgestellt wurde, sollen nun die verschiedenen Bestandteile des Untersuchungsdesigns erläutert werden. Die erste Komponente bildet die Analyse des Publikationsoutputs der Mitglieder der Stichprobe. Die bibliometrische Analyse gewährt einen quantifizierenden Einblick in den Umfang, die Frequenz und die Publikationsorte der Inklusion in das Kommunikationssystem in der Rolle des Autors und erlaubt es, einen Vergleich entlang der drei Schichtungsdimensionen vorzunehmen.

Um die Inklusion von Autoren möglichst vollständig zu erfassen, wurde nicht auf die Zitationsdatenbank, sondern auf die Publikationsnachweissysteme der beiden Fächer zurückgegriffen. Im Unterschied zum SCIE sind sie nicht selektiv, sondern decken – zumindest ihrem Anspruch nach – alle wesentlichen Publikationen des Fachs ab. Zur Erfassung des Publikationsoutputs wurde für die Astronomen auf das bereits erwähnte *NASA Astrophysics Data System*¹⁵, für die Mathematiker auf die beiden Review-Organe *zbMATH*¹⁶ und *MathSciNet*¹⁷ zurückgegriffen. Für jedes Mitglied der Stichprobe wurden die Metadaten der Publikationen heruntergeladen und in einer SQL-Datenbank organisiert. Danach fanden Plausibilitätsprüfungen statt, ob die entsprechende Publikation auch von dem gesuchten Autor stammt, und die Datenbanken wurden um Doubletten bereinigt, die – in der Mathematik – aus der Verwendung von zwei Datenquellen resultierten. Der so gewonnene Datensatz enthält u.a. Informationen zum Publikationsjahr, zum Publikationstyp, zur Anzahl der Autoren und zum Publikationsort, jedoch keine zum hier interessierenden Merkmal „Zugang“. Um diese Informationen zu ergänzen, wurde für jede der 12.567 Publikationen eine *empirische Zugänglichkeitsprüfung* durchgeführt.¹⁸ Dazu wurden teilautomatisiert Anfragen an die Suchmaschinen Google Scholar¹⁹, BASE²⁰ und OIASTER²¹ abge-

15 Siehe: <http://www.adsabs.harvard.edu/> (Zugriff am 18. April 2017).

16 Siehe: <http://www.ams.org/mathscinet/> (Zugriff am 18. April 2017).

17 Siehe: <https://zbmath.org/> (Zugriff am 18. April 2017).

18 Die Zugänglichkeitsprüfung fand zwischen Januar und Juli 2012 statt.

19 Siehe: <http://scholar.google.de/> (Zugriff am 18. April 2017).

setzt.²² Wurde eine Publikation aufgefunden, wurde manuell geprüft, ob sie frei zugänglich ist. Dazu wurde das Dokument von einem Computer aus aufgerufen, der mit einer einfachen Verbindung zum Internet ausgestattet und nicht Teil eines universitären Computernetzes war. Ausgeschlossen wurde damit, dass ein Zugang zu einer Publikation durch von der Universitätsbibliothek erworbenen Lizenzen bestand. Konnte das Dokument geöffnet werden, wurde geprüft, welcher Ort einen Zugang herstellt. Zum Ende jedes Prüfungsvorgangs wurde das Ergebnis anhand der folgenden Ausprägungen des Merkmals „Zugang“ klassifiziert.

- Als *nur gedruckt* wurden Publikationen klassifiziert, bei denen kein elektronischer Volltext aufgefunden wurde.
- Als *zugangsbeschränkt* wurden all jene Publikationen bezeichnet, bei denen zwar ein digitaler Volltext gefunden wurde, auf den aber nicht zugegriffen werden konnte.
- Bei *Gold Open Access* bestand ein Zugang zu digitalen Volltexten am originären Publikationsort.
- Lag dagegen ein Zugang zu elektronischen Volltexten an einem anderen als dem originären Publikationsort vor, wurde die Publikation als *Green Open Access* klassifiziert.
- In Fällen, in denen ein digitaler Volltext sowohl am originären Publikationsort als auch an einem anderen Ort zugänglich war, wurde die Publikation als *Green & Gold Open Access* bezeichnet.
- Im Zuge der Zugänglichkeitsprüfung stellte sich heraus, dass bei einem kleinen Teil der Publikationen nicht zu entscheiden war, ob es sich bei dem Ort, an dem ein freier Zugang bestand, um den originären Publikationsort handelt

-
- 20 Bielefeld Academic Search Engine (<http://www.base-search.net/>, Zugriff am 18. April 2017).
- 21 Siehe: <http://www.oclc.org/en/oaister.html> (Zugriff am 18. April 2017).
- 22 Die Nutzung von drei Suchmaschinen ist der Überlegung geschuldet, dass Suchmaschinen nicht den gesamten im Internet verfügbaren Content indexieren, sondern jeweils nur einen Teil. Mit der Verwendung mehrerer Suchmaschinen sollte die Auffindungswahrscheinlichkeit erhöht werden. Ausgeschlossen werden kann aber nicht, dass Publikationen an Orten abgelegt sind, die von allen drei Suchmaschinen nicht indiziert werden. Die im Folgenden berichteten Anteile digital vorliegender Publikationen sind daher Mindestanteile; die tatsächlichen Werte können etwas höher sein.

oder nicht.²³ Für diese Fälle wurde die Residualkategorie *Open Access Undecided* geschaffen.

Auf der Grundlage dieses Datensatzes wurden der Umfang und die Nutzungsmuster von Open-Access-Publikationen untersucht. Eine Analyse fand dabei auf der Ebene einzelner Wissenschaftler, der unterschiedlichen Teilstichproben und der Gesamtstichprobe statt. Für die statistischen Auswertungen wurde das Programmpaket *STATA 11* (Kohler und Kreuter 2008) genutzt.

8.3 MAPPING DER PUBLIKATIONSINFRASTRUKTUR

Der zweite Teil des Untersuchungsdesigns richtet sich auf die weiter oben als Publikationsinfrastruktur bezeichneten technischen Einrichtungen. Ziel des Mappings ist es nicht etwa, die Publikationsmedien aus dem interessierenden Bereich der Wissenschaft vollständig abzubilden, sondern besteht – vom Anspruch her bescheidener – darin, die wesentlichen Charakteristika der Publikationsinfrastruktur herauszuarbeiten und Aussagen über die technischen und sozialen Zugangsregeln sowie die von ihr bereitgestellten Ressourcen zu treffen. Im Rahmen des Mappings der Publikationsinfrastruktur wurden sowohl quantitative als auch qualitative Daten erhoben und ausgewertet. Um die wesentlichen Publikationsmedien zu identifizieren, wurde der gerade vorgestellte bibliometrische Datensatz daraufhin untersucht, in welchen Publikationsmedien fünf oder mehr Publikationen erschienen sind. Dies ist bei 239 Journalen und 16 Conference Proceedings der Fall. Um diese Publikationsmedien herum wurde ein weiterer Datensatz aufgebaut, der die wesentlichen Merkmale zusammenführt. Als Datenquellen wurden neben dem SCIE und *Ulrich's Web*²⁴ auch das *Directory of Open Access Journals* genutzt.²⁵ Zudem wurde eine Recherche auf Websites der Zeitschriften durchgeführt und der Datensatz wurde um weitere Merkmale ergänzt. Er versammelt Informationen über die Art und den Umfang eines freien Zugangs, etwaig angewandte hybride Open-Access-Modelle, die Höhe der Publikationsgebühren, den Verlag bzw. die verlegerisch tätige Organisation, die An-

23 Dies galt zum Beispiel für Konferenzbeiträge, die auf der Website des Veranstalters einer Konferenz hinterlegt waren.

24 *Ulrich's Web* (<https://ulrichsweb.serialssolutions.com/>, Zugriff am 18. April 2017) ist eine umfassende Datenbank für Zeitschriften und Serials aus allen Bereichen der Wissenschaft.

25 Siehe: <http://www.doaj.org/> (Zugriff am 18. April 2017).

zahl der jährlich publizierten Beiträge und den *Journal Impact Factor* (JIF). Auch dieser Datensatz wurde mithilfe des Statistik-Programmpakets *STATA 11* ausgewertet.

Daneben wurden Dokumente zu bedeutenden Komponenten der Publikationsinfrastruktur, wie etwa große fachspezifische Publikations-Repositorien, wichtige Journale, Daten-Repositorien und Fachdatenbanken, gesammelt und qualitativ analysiert. Zu diesen Dokumenten zählen Websites von Zeitschriften und Fachdatenbanken, technische Dokumentationen sowie Fachpublikationen, in denen die Infrastrukturkomponenten beschrieben werden. Mit diesem Material wurde ein vertieftes und stärker kontextualisiertes Verständnis über deren Bedeutung gewonnen. Umfangreich fiel die Erhebung und Auswertung von qualitativem Material in der Astronomie aus, da diese über eine stark zentralisierte Publikationsinfrastruktur mit einem kleinen und abgegrenzten Kern von Journalen verfügt und eine quantitative Untersuchung hier wenig aussagekräftig ist.

8.4 LEITFADENGESTÜTZTE EXPERTENINTERVIEWS

Die Untersuchung des Publikationsoutputs von Wissenschaftlern und die Analyse der Eigenschaften der Publikationsinfrastruktur offerieren gewissermaßen eine objektivierte Perspektive auf Publikationsaktivitäten, da hiermit Publikationen gezählt, deren Merkmale verglichen sowie Charakteristika und Eigentümlichkeiten der Publikationsinfrastruktur ermittelt werden können. In den Blick kommen dabei vor allem die Ergebnisse der Inklusion von Autoren in das Kommunikationssystem, die verwendeten Publikationsmedien und die dabei hinterlassenen Spuren. Verborgен bleiben diesem Zugang dagegen alle vorgelagerten Prozesse, die zu einer Veröffentlichung von Forschungsergebnissen führen, sowie alle Anschlusshandlungen wie etwa der Zugang zu und die Lektüre von Publikationen im Rahmen von Rezeptionsvorgängen. Ebenfalls außen vor bleiben die Orientierung der Wissenschaftler, die bei der Veröffentlichung und Rezeption leitend ist, sowie die Deutungen und Sinnzuschreibungen, die sie gegenüber einzelnen Bestandteilen der Publikationsinfrastruktur im Zuge der Verwendung vornehmen.

Um einen Zugang zu dieser Ebene des Phänomens zu gewinnen, wurden als dritte Komponente leitfadengestützte Experteninterviews durchgeführt. Das Untersuchungsdesign folgt hier der Grounded-Theory-Methodik (GTM), die gleichzeitig auch den Rahmen für die Integration der Ergebnisse der bibliometrischen Untersuchung des Publikationsoutputs und des Mappings der Publikationsinfrastruktur bereitstellt. Forschungsstrategisch dienen die Interviews einem

doppelten Zweck. Zum einen liefern sie Beschreibungen davon, wie Wissenschaftler in der Rolle des Autors und des Rezipienten mit Publikationsmedien umgehen, und lassen die Klärung der Frage zu, welchen eingespielten Routinen der Umgang folgt. Daneben bilden Erfahrungen mit Zugangsproblemen zu Publikationen einen Schwerpunkt der Interviews. In dieser Hinsicht werden die Experteninterviews eingesetzt, um in komprimierter Form das „Betriebswissen“ (Meuser und Nagel 2005: 75) von Wissenschaftlern im Umgang mit der Publikationsinfrastruktur zu erheben. Zum anderen werden die Experteninterviews aber auch genutzt, um einen Zugang zum Deutungswissen der Experten zu gewinnen. Hier interessieren insbesondere die Zuschreibungen gegenüber Publikationsvorgängen, den beiden Spielarten von Open Access sowie gegenüber E-Prints. Diesem zweiten Zweck liegt die Überlegung zugrunde, soziale Deutungen würden unmittelbar Einfluss auf Publikations- und Rezeptionshandlungen nehmen, etwa dergestalt, dass sie Quellen des Renommees von Publikationsmedien sind und Entscheidungen über den Ort der Veröffentlichung von Forschungsergebnissen mitprägen. In dieser zweiten Hinsicht gelten Wissenschaftler als Experten²⁶ für die von ihnen verwendete Publikationsinfrastruktur, sobald sie sich erfolgreich im formalen Kommunikationssystem etabliert haben.

Eine Auswahl der Wissenschaftler fand kriteriengeleitet unter Berücksichtigung der im Verlauf der Untersuchung gewonnenen Erkenntnisse sowie der theoretischen Überlegungen zum wissenschaftlichen Kommunikationssystem (Kapitel 5) und zu Open Access (Kapitel 6) statt. Die Sampling-Strategie verfolgte dabei nicht etwa das Ziel, die Grundgesamtheit der Wissenschaftler statistisch zu repräsentieren, sondern die ‚richtigen‘ Personen – hier verstanden als Träger des interessierenden Wissens – auszuwählen (Morse 2010: 234). Um diese Wissensbestände möglichst breit zu erfassen, variieren die Merkmale der Interviewpartner des Samples systematisch entlang der Dimensionen „Fach“, „Herkunftsland“ und „Kohorte“. Ausgewählt und interviewt wurden dementsprechend jeweils zwei Wissenschaftler aus jeder der in Tabelle 8.2 genannten acht Teilstichproben. Dabei wurde darauf geachtet, dass die Interviewpartner an möglichst unterschiedlichen Forschungsorganisationen beschäftigt sind, um verschiedene durch Organisationen gegebene Rahmenbedingungen einer Beteiligung am wissenschaftlichen Kommunikationssystem abzubilden.²⁷ Die Inter-

26 Siehe hierzu ausführlicher Bogner und Menz (2005: 46). Konstitutiv für den Experten ist dem dort vertretenen Verständnis nach die Chance zur Durchsetzung der Expertenorientierung gegenüber anderen, im Feld vertretenen Orientierungen.

27 Durchgeführt wurden Interviews an folgenden Einrichtungen: *Universität Hamburg, Universität Bielefeld, RWTH Aachen, European Southern Observatory, Zuse Insti-*

views erbrachten Hinweise auf weitere interessante Gesprächspartner, so dass das Sample erweitert wurde und Gespräche mit insgesamt 20 Wissenschaftlern geführt wurden. Die Interviews mit der deutschen Gruppe von Mathematikern und Astronomen wurden im Januar 2012, die Interviews mit den südafrikanischen Gesprächspartnern im Rahmen einer Interviewreise im Februar 2012 geführt. Die Länge der Interviews schwankt zwischen 39 und 122 Minuten.

Die Auswertung der Interviews nutzt mit den verschiedenen Formen der Kodierung den prozeduralen Kern der Grounded-Theory-Methodik (Breuer 2009: 52). Die GTM bietet sich dabei aus den folgenden Gründen an: Erstens ist sie nicht nur zur Analyse manifester Informationen, sondern auch zur Untersuchung latenter Sinnstrukturen geeignet (Strauss und Corbin 1996). Zweitens ist das Vorgehen hinreichend offen, um auch nicht antizipierte Zusammenhänge zwischen dem Kommunikationssystem, der Publikationsinfrastruktur, seiner Nutzung und der Deutung der Komponenten aufzudecken und zu untersuchen. Dies gilt insbesondere für die erste Phase der Auswertung, dem offenen Kodieren. Zweitens legt es die komparative Anlage der Untersuchung mit ihren drei Vergleichsdimensionen nahe, auch bei der Analyse der qualitativen Daten auf eine Methodik zurückzugreifen, die den Vergleich an prominenter Stelle berücksichtigt. Bei dem Grounded-Theory-Verfahren ist dies der Fall. Die Kontrastierung bildet die „Leitidee des Kodierprozesses“ (Strübing 2008: 18) und sieht bei der Ausarbeitung der Kategorien vor, die einzelnen kodierten Abschnitte mit ihrem Sinngehalt systematisch zu vergleichen. Trotz seiner Offenheit sorgt das Vorgehen drittens für eine Fokussierung der Analyse, wenn in der Phase des axialen Kodierens eine Ausarbeitung vorrangig derjenigen Konzepte und Kategorien gefordert wird, die nach dem jeweiligen Erkenntnisstand für die Beantwortung der Forschungsfrage von Bedeutung sind (ebd.: 21).

Die Interviewdaten wurden nach einem einfachen Transkriptionsschema verschriftlicht, die Auswertung und Interpretation der Daten wurde durch das Analyseprogramm *Atlas.ti* 6.2 unterstützt. Das Programm diente dabei zur Organisation des Materials, zur offenen Kodierung, zur Entwicklung von Code-Familien im Rahmen des axialen Kodierens, zum Schreiben von Memos und zum Vergleich der verschiedenen Textpassagen. Die Zerlegung der Interviews durch die Kodierung von kleineren Interviewpassagen und deren Zusammenstellung durch Abfragen in *Atlas.ti* bergen die Gefahr, kleinere Abschnitte des Materials isoliert zu betrachten und außerhalb des Kontexts ihrer Äußerung zu interpretieren. Dem

tute Berlin, Max-Planck-Institut für Astronomie, South African Astronomical Observatory, University of Cape Town, University of Stellenbosch, North-West University (Potchefstroom), University of Pretoria (Tshwane) und University of South Africa.

Risiko, durch ein solches Vorgehen bei der Analyse zu ungesicherten Aussagen zu gelangen, wurde begegnet, indem die Passagen im Verlauf der Interpretationsarbeit immer wieder auch in den Kontext ihres Auftretens gerückt wurden.

Abschließend soll auf einen wesentlichen Unterschied zwischen dem Lehrbuchverständnis der Grounded Theory und dem hier beschrittenen Vorgehen aufmerksam gemacht werden. Die GTM sieht eine Verschränkung der Phasen der Erhebung von Daten, ihrer Auswertung, der Generierung neuer Theorien und der weiteren Erhebung von Daten vor (z.B. Krotz 2005: 167) – also ein mehrfaches Durchlaufen eines kreisförmigen Erkenntnisprozesses, der dem hermeneutischen Zirkel ähnelt (Breuer 2009: 48). Zwar fand im Rahmen dieser Untersuchung die Planung der einzelnen Datenerhebungsphasen und die Festlegung der mit dem Material zu klärenden Fragen auf Basis der bis dahin abgeschlossenen Schritte und dem erzielten Erkenntnisstand statt, die Erhebung der Interviews selbst folgte aber nicht der kreisförmigen Bewegung zwischen Datenerhebung, Auswertung und erneuter Datenerhebung. Dies kam aus forschungspragmatischen Gründen nicht in Frage, da die Erhebung der Interviews sich auf mehrere Forschungsaufenthalte in Deutschland und Südafrika hätte verteilen müssen. Die Datenerhebung folgte stattdessen dem oben dargestellten und vorab entwickelten Auswahlplan und nicht den analytischen Fragen, die sich im Zuge der Interpretation vorangegangener Interviews auf der Hintergrundfolie des dabei erreichten Stands der Theoriebildung ergaben (Strübing 2008: 30). Entsprechend war die Vorbereitung der Interviews deutlich stärker durch eine planende Antizipation der sich möglicherweise als relevant erweisenden Kategorien geprägt, als es dem Verständnis der GTM nach der Fall sein sollte. Neben der Identifikation relevanter Vergleichsdimensionen zählte dazu auch die Ausarbeitung eines Interviewleitfadens, der in den Gesprächen allerdings nicht Schritt für Schritt abgearbeitet wurde, sondern von dem je nach Verlauf auch abgewichen wurde. Dies geschah etwa in Situationen, in denen ein Gesprächspartner eigene Relevanzgesichtspunkte einbrachte, oder um Deutungen, Einschätzungen, Beobachtungen und der Veranschaulichung durch Beispiele entsprechenden Raum zu geben.

Durch eine Vorab-Festlegung der Größe und Merkmale des Interviewsamples konnte auch das Abbruchkriterium der GTM nicht angewandt werden. Dem zufolge soll der Kreis aus Datenerhebung, Auswertung und Veränderung des (theoretischen) Vorverständnisses so lange durchlaufen werden, bis eine „Sättigung“ der Theorieentwicklung erreicht wird, bei der also eine weitere Datenerhebung und -interpretation keine Ausarbeitung der Theorie zulässt und sich die neu erhobenen Daten in die bereits entwickelte Theorie einordnen lassen (Krotz 2005: 178). Zwar scheint das Abbruchkriterium auf den ersten Blick dazu geeignet, die praktische Interpretationsarbeit begründet zu beenden, es überzeugt aber

aus forschungslogischen Gründen nicht vollständig. Prinzipiell ist es zu jedem Zeitpunkt möglich, neue Fragen im Rahmen einer Theorie zu entwickeln, mit dem Ziel, ihre Komplexität, thematische Reichweite und den Grad ihrer Allgemeinheit zu erhöhen oder die Durchdringung eines empirischen Gegenstands zu vertiefen. Daher ist das Abbruchkriterium allein kaum geeignet, um den kreisenden Erkenntnisprozess der GTM zu einem Abschluss zu bringen. Mindestens hinzukommen muss eine – implizite oder explizite – Entscheidung, auf eine weitere Ausarbeitung oder Ausweitung der Theorie zu verzichten und es beim erreichten Stand zu belassen. In der vorliegenden Untersuchung wurde dem Umfang der Datenerhebung und Interpretation durch den Auswertungsplan und die darin festgelegte Anzahl an Interviews eine Grenze gesetzt. Durch die Antizipation möglicherweise relevanter Unterscheidungen wurden aber Vorkehrungen getroffen, dass die Rahmenbedingungen variiert und Fälle kontrastierend miteinander verglichen werden können.

Was die Darstellung der Theorie angeht, finden sich wesentliche Kategorien bereits weiter oben im theoretischen Teil und werden bei der Darstellung der empirischen Ergebnisse wieder aufgegriffen und weiter ausgearbeitet. Diese getrennte Darstellung soll allerdings nicht darüber hinwegtäuschen, dass im Zuge des Forschungsprozesses eine fortlaufende Rückkopplung zwischen empirischem Material, Theorie und dem Stand der Forschung in der Literatur stattgefunden hat.²⁸

28 Siehe in ähnlicher Weise die Darstellungsstrategie von Reisenauer (2017: 102).