



Institutionenkodierung als Grundlage für bibliometrische Indikatoren

Matthias Winterhager, Holger Schwechheimer, Christine Rimmert
Universität Bielefeld, Institute for Interdisciplinary Studies of Science (I²SoS),
Universitätsstr. 25, 33615 Bielefeld, Germany
bibliometrie@uni-bielefeld.de

Abstract

Die nicht standardisierte, fehlerhafte und unvollständige Erfassung von Autorenadressen („Affiliations“) in den großen interdisziplinären Literaturdatenbanken Web of Science (WoS) und Scopus stellt eine erhebliche Herausforderung für bibliometrische Analysen überall dort dar, wo es auf eine genaue und verlässliche Zuordnung der Publikationen zu Institutionen ankommt. Mit den Autorenadressen sind eine Reihe von Problemen verbunden, für die es nicht in jedem Fall einfache Lösungen gibt. Am bekanntesten ist das Problem der unterschiedlichen Schreibweisen („spelling variants“) von Adressen, dem die Datenbankhersteller seit einiger Zeit durch eigene Bemühungen zur (Teil-)Standardisierung der Adresseinträge zu begegnen versuchen. Diesen Ansätzen sind jedoch Grenzen gesetzt, auch weil die Adressangaben schon in den Originalartikeln der wissenschaftlichen Zeitschriften nicht selten ungenau und unvollständig sind. Die Zeitschriften arbeiten mit unterschiedlichen (bzw. gar keinen) Vorgaben für die Autoren hinsichtlich der Nennung der relevanten Adressen und auch die Forschungsinstitutionen selbst machen in Deutschland bisher überwiegend keine entsprechenden Vorgaben für ihre eigenen Mitarbeiter.

Bei der Zuordnung der Autorenadressen zu Forschungseinrichtungen geht es aber um mehr als eine bloße Vereinheitlichung unterschiedlicher Schreibweisen. Neben der grundlegenden Aufgabe einer Zusammenfassung bzw. Standardisierung der auftretenden Varianten von Institutionsnamen stellt vor allem die Identifikation der jeweils „richtigen“ real (in Gegenwart oder Vergangenheit) existierenden Institution zu den Adressen eine Herausforderung dar. Nicht nur fehlende Standards zur Angabe von Adressen in wissenschaftlichen Publikationen, sondern auch komplexe institutionelle Strukturen (z.B. Hierarchien oder Beziehungen zwischen Institutionen) und ihr Wandel über die Zeit können eine eindeutige Zuordnung erheblich erschweren.

In diesem Beitrag werden Ergebnisse eines Projekts vorgestellt, das im Kontext des im Aufbau befindlichen Kompetenzzentrums Bibliometrie für die deutsche Wissenschaft durchgeführt wird. Das Projekt hat die Gewinnung einer möglichst vollständigen, verlässlichen und nachhaltig nutzbaren Zuordnung der in WoS bzw. Scopus erfassten Publikationen mit deutschen Adressen zu real existierenden deutschen Institutionen zum Ziel. Die Ergebnisse werden mit einem teilautomatischen Verfahren erzielt, in dem die Erkennung von Textmustern in den Adressen eine wesentliche Rolle spielt. Zur Abbildung der Institutionendynamik werden auch Daten mit Zeitangaben zur Entwicklung der Institutionen und Sektoren der deutschen Forschungslandschaft erfasst. Ein Informationsaustausch mit interessierten Hochschulen und außeruniversitären Forschungseinrichtungen in Deutschland ist in Bezug auf die sie jeweils betreffenden Datensätze möglich, soweit die lizenzrechtlichen Voraussetzungen erfüllt sind.

1. Problemstellung: Bereinigung und Kodierung institutioneller Adressen

Den Ausgangspunkt des Projekts¹ bildete folgender Sachverhalt: In den beiden für bibliometrische Analysen am häufigsten genutzten interdisziplinären Literaturdatenbanken Web of Science (WoS) und Scopus sind pro Publikation zwar die institutionellen Adressen der Autoren erfasst, soweit diese im Original der jeweiligen Arbeit aufgeführt sind. Die Erfassung dieser Adressen lässt aber hinsichtlich Genauigkeit, Vollständigkeit und Standardisierung sehr zu wünschen übrig.

Die Varianz der Adressen in den Datenbanken hängt mit verschiedenen Problemen zusammen, für die es nicht in jedem Fall eine einfache Lösung gibt. Am bekanntesten ist das Problem der unterschiedlichen Schreibweisen („spelling variants“) von Adressen, dem die Datenbankhersteller seit geraumer Zeit durch eigene Bemühungen zur (Teil-) Standardisierung der Adresseinträge entgegenwirken. Diesen Ansätzen sind jedoch enge Grenzen gesetzt, weil die Adressangaben schon in den Originalartikeln der wissenschaftlichen Zeitschriften nicht selten ungenau und unvollständig sind. Die folgende Liste zeigt beispielhaft Adressvarianten, die in Scopus für das Fraunhofer-Institut für Kurzzeitdynamik zu finden sind:

- Ernst Mach Inst. für Kurzzeitdynamik, Freiburg, DEU
- Ernst Mach-Institute, Fraunhofer Institute for High-Speed Dynamics, Freiburg, DEU
- Ernst Mach Institute, Eckerstr. 4, 79108 Freiburg, DEU
- Fraunhofer-Inst. für Kurzzeitdynamik, Eckerstr. 4, D-79104, Freiburg, DEU
- Fraunhofer EMI, Eckerstr. 4, 79104 Freiburg, DEU
- FhG - Ernst-Mach-Institute, Eckerstrasse 4, 79104 Freiburg, DEU
- ERNST-MACH-INSTITUT, Fraunhofer-Inst. F. Kurzzeitdynamik, Freiburg/Br., DEU
- Fraunhofer Ernst-Mach-Institut, DEU

Die Zeitschriften arbeiten mit unterschiedlichen (bzw. gar keinen) Vorgaben für die Autoren hinsichtlich der Nennung der relevanten Adressen und auch die Forschungsinstitutionen selbst machen in Deutschland bisher überwiegend keine entsprechenden Vorgaben für ihre

eigenen Mitarbeiter (Bornmann et al. 2012). Entsprechend heterogen sieht die Praxis aus. Es gibt sogar Disziplinen, in denen die Nennung von abgekürzten Privatadressen statt der Institution nicht unüblich ist. Ein weiteres Problem sind die Adressangaben von Autoren, deren Arbeiten im Kontext von Gastaufenthalten an „fremden“ wissenschaftlichen Einrichtungen entstehen: in solchen Fällen werden zum Teil zwei Adressen für einen Autor angegeben, manchmal (aber nicht immer) mit entsprechender Qualifizierung.

Problematisch sind auch diejenigen Adressen, die in sich doppeldeutig (verschiedene Institutionen in einem Adressdatensatz) oder zu unspezifisch (Beispiel: „Univ Berlin, Berlin, Germany“) sind. Die Datenlage lässt in diesen Fällen oft keine valide Zuordnung zu. Die an bestimmten Stellen zunehmende, zum Teil politisch ausdrücklich gewünschte Überlappung von Sektoren im deutschen Wissenschaftssystem (z.B. Max-Planck-Gruppen in Hochschulen, gemeinsame Berufungen bei HGF und Hochschulen, Hybrideinrichtungen wie KIT, neue Organisationsformen für Universitätsklinik wie Charité, UKSH etc.) macht eine eindeutige Zuordnung von Publikationen zu Institutionen in vielen Fällen ohne Hinzuziehung weiterer Informationen bzw. Einbeziehung des Analysekontextes unmöglich. Hinzu kommt, dass im Zeitverlauf die Institutionenlandschaft nicht stabil, sondern im Wandel ist.

Vor diesem Hintergrund ist es praktisch unmöglich, valide bibliometrische Indikatoren institutionsbezogen auf der Basis der unbereinigten, in WoS oder Scopus vorgefundenen Adressdatensätze zu generieren. Dass dies in manchen Kontexten dennoch immer wieder versucht wird, trägt zur Gefahr des Missbrauchs von Bibliometrie im politischen Raum bei. Der unreflektierte Einsatz von kommerziellen, „einfach bedienbaren“ Standardprodukten oder anderweitig publizierter bibliometrischer Rankings in Instanzen der Wissenschaftsverwaltung und -politik birgt Risiken der Fehleinschätzung.

Ziel des Projekts Institutionenkodierung ist demgegenüber die Gewinnung einer möglichst genauen, verlässlichen und nachhaltig nutzbaren Zuordnung der in Web of Science bzw. Scopus erfassten Publikationen mit deutschen Adressen zu real existierenden deutschen Institutionen. Das Vorhaben steht im Kontext des Aufbaus eines Kompetenzzentrums Bibliometrie für die deutsche Wissenschaft. Im Rahmen der gemeinsamen Lizenzierung durch das Kompetenzzentrum werden für das Projekt

¹ Gefördert durch: Bundesministerium für Bildung und Forschung (Förderkennzeichen 01PQ08004C).

die Rohdaten von WoS und Scopus mit folgenden Datenbeständen genutzt:

- Web of Science
 - Science Citation Index Expanded (SCIE) ab 1980
 - Social Sciences Citation Index (SCCI) ab 1980
 - Arts & Humanities Citation Index (AHCI) ab 1980
 - Conference Proceedings Citation Index - Science (CPCI-S) ab 1990
 - Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH) ab 1990
- Scopus ab 1996

Damit ist als wesentliche Voraussetzung des Projekts der unlimitierte Zugriff auf die Adressdatensätze von WoS und Scopus sichergestellt.

Die Probleme der Unifizierung von Autorenadressen und ihrer Zuordnung zu realen Institutionen für bibliometrische Analysen sind seit langem bekannt. De Bruin & Moed (1990) haben bereits vor fast 25 Jahren darüber berichtet. Dennoch gibt es bis heute relativ wenig Literatur, die sich dezidiert darauf bezieht. Bekannt ist, dass die Hersteller von WoS und Scopus im eigenen Interesse zwar Anstrengungen zur Unifizierung und Standardisierung unternahmen, dass diese aber an Grenzen stoßen, die ohne detaillierte Kenntnisse der regionalen und lokalen institutionellen Strukturen nicht zu überwinden sind. Eine Bereinigung der Adressen erforderte bisher letztlich immer auch manuelle Eingriffe und Korrekturen. Moed et al. (1995) haben einen Einblick in ihre Erfahrungen mit der Kodierung der niederländischen Adressen im SCI gegeben:

“The unification and classification of Dutch corporate source addresses in our database is almost complete now. Over 99% of the addresses has been unified and classified according to the system mentioned above. However, several of the addresses that have been unified and classified, are still 'problematic'. These institutes belong to overlapping categories, in which their status could not clearly be established or their identification (e.g., as a university department) was not 100% certain. Since these 'problematic' addresses in our database generally have a small output, we believe that we have unambiguously unified and

classified the institutes responsible for well over 95% of the publications. This is sufficient to produce reliable results per institutional sector. Publications that could not be classified yet were not taken into account in the analyses.” (Moed et al. 1995, S. 393)

In der Zwischenzeit sind Zuordnungen von Autorenadressen zu realen Institutionen in zahlreichen bibliometrischen Einzelprojekten meist ad hoc und nur für den jeweiligen Einsatzzweck vorgenommen worden. Entsprechende Erfahrungen liegen vor allem in den Niederlanden und Belgien, Skandinavien, Großbritannien, Australien und Kanada vor. Die Ergebnisse sind jedoch im Allgemeinen nicht öffentlich zugänglich.

Galvez & Moya-Anegón haben 2006 und 2007 über einen neuen Ansatz mit der Anwendung parametrisierter endlicher Graphen (FSG) bzw. Transduktoren (FST) berichtet, der an WoS-Daten entwickelt und später auch mit Daten aus Inspec, Medline, CAB Abstracts getestet wurde. Der Einsatz dieser *Natural Language Processing*-basierten Methoden zur Standardisierung von Autorenadressen erscheint erfolgversprechend. Dennoch bleiben auch in diesem Fall absehbare Grenzen der automatisierten Behandlung, wie die Autoren selbst feststellen:

“It is important to point out that beyond the scope of the present work remain those problems originating in any errors or inconsistencies produced by abbreviations, transliteration differences, differences in spelling, or name changes. Nor do we tackle problems deriving from the absence in the address of the first institutional level, or difficulties in the assignment of each document to a center that may result from ambiguity or inconsistency in the use of different names to refer to a single institution, cases where a single same name may designate two or more separate institutions, or assigneeship reflecting different nationalities. The validation and correct institutional assignment of addresses is a task corresponding to experts.” (Galvez & Moya-Anegón 2007, S. 6)

Morillo et al. (2013) haben einen Ansatz zur Grobkodierung von WoS-Adressen auf der Ebene von Sektoren für Spanien getestet. Doch auch dieser auf der Erkennung von sektorspezifischen Schlüsselworten basierende

Ansatz kommt nicht ohne manuelle Eingriffe in das Verfahren und detaillierte Kenntnisse der Forschungslandschaft des betreffenden Landes aus. Das gilt auch für die von Cuxav et al. (2013) am Beispiel französischer CNRS-Adressen getestete Anwendung von supervised und semi-supervised learning Methoden. Zuletzt haben Huang et al. (2014) einen regelbasierten Algorithmus zur Disambiguierung von Institutionsnamen unter Rückgriff auf Autorenblöcke vorgestellt. Damit konnten zwar hohe Precision-Werte demonstriert werden, die erzielten Recall-Werte sind aber zu niedrig für einen praktischen Einsatz.

Der Vollständigkeit halber sei hier auch noch auf die Probleme verwiesen, die entstehen, wenn die Zuordnung von Publikationen zu Institutionen in bibliometrischen Analysen vorrangig personenbezogen und nicht starr, sondern flexibel gehandhabt werden soll. Dies wird im Kontext forschungspolitischer Anwendungen zunehmend nachgefragt, um z.B. auch virtuelle Einrichtungen wie institutionenübergreifende Projektverbünde, Exzellenzcluster u.ä. angemessen analysieren zu können. Auch die Möglichkeit einer alternativen Betrachtung der Daten unter den Prinzipien von „work-done-at“ einerseits und „current potential“ andererseits erhält zunehmend Bedeutung (vgl. Wissenschaftsrat 2008). Da es hierbei aber um einen grundsätzlich anderen Weg der Zuordnung von Publikationen zu Institutionen geht, nämlich über Personennamen statt Autorenadressen, müssen Lösungen dafür auf einer ganz anderen Ebene gesucht werden. Die Datenbankhersteller haben darauf inzwischen mit der Einführung von personenbezogenen Kennungen reagiert, wie etwa ResearchID² für WoS von Thomson Reuters bzw. herstellerübergreifend: Open Researcher and Contributor ID (ORCID)³. Angesichts des dadurch entstehenden hohen Kontrollpotentials regen sich in der scientific community allerdings Bedenken gegen derartige Systeme (vgl. Enserink 2009); von einer flächendeckenden Umsetzung und Akzeptanz kann bisher keine Rede sein.

² <http://www.researcherid.com/>

³ <http://orcid.org/>

2. Grundzüge der Adress- und Institutionenkodierung

Für eine „richtige“ und verlässliche Zuordnung der Autorenadressen ist sowohl eine Adress- wie auch eine Institutionenkodierung notwendig. Dabei bezeichnet *Adresskodierung* den Vorgang der Zuordnung von konkreten Autorenadressen zum einschlägigen Institutionenschlüssel, während *Institutionenkodierung* die systematische Erfassung der institutionellen Einheiten mit ihren Hierarchiebeziehungen unter Berücksichtigung der Veränderungen in der Zeit (Historisierung) umfasst.

Um eine möglichst genaue, verlässliche und nachhaltig nutzbare Zuordnung der in Web of Science bzw. Scopus erfassten Publikationen mit deutschen Adressen zu real existierenden Institutionen zu gewinnen, wurde ein semiautomatisches Verfahren entwickelt, das wesentlich auf der Erkennung von Textmustern in den Adressen beruht. Zusätzlich wurden Daten zu den Institutionen und Sektoren der deutschen Forschungslandschaft erfasst, soweit sie Relevanz für die in den Datenbanken verzeichneten Publikationen haben. Die Daten sind in verschiedenen Tabellensätzen erfasst und in normalisierter Form in ein relationales Datenbanksystem integriert, so dass der Zugriff zur Bearbeitung sowie zur Nutzung/Anwendung über entsprechende SQL-Schnittstellen erfolgen kann. Das Datenmodell ist so ausgelegt, dass darin die relevanten institutionellen Einheiten auch mit ihren historischen Zuständen und Sektorzugehörigkeiten abzubilden sind. Eine erschöpfende Vollständigkeit ist an dieser Stelle nicht ohne Weiteres möglich, für die Praxis aber auch nicht unbedingt erforderlich. Wichtiger ist, dass für den verwendeten Institutionenschlüssel Konkordanzen zu anderen relevanten Datenbeständen hergestellt werden. Das Hauptaugenmerk liegt dabei auf der Institutionendatenbank der Deutschen Forschungsgemeinschaft.⁴ Auf der Ebene der Hauptinstitutionen (vgl. Abschnitt „Definition der Hauptinstitution“) ist die Konkordanz zu diesem System realisiert und wird regelmäßig aktualisiert.

⁴ Vgl. „Das deutsche Forschungsverzeichnis - [Research Explorer](#)“: Forschungsverzeichnis der Deutschen Forschungsgemeinschaft (DFG) und des Deutschen Akademischen Austauschdienstes (DAAD) in Zusammenarbeit mit der Hochschulrektorenkonferenz (HRK).

Die in den Adressen bezeichneten Einheiten können Teileinheiten unterschiedlicher Hierarchieebenen sein (Beispiel: *Lehrstuhl für Biologische Kybernetik, Fakultät für Biologie, Universität Bielefeld*); nicht immer wird auch die übergeordnete Forschungseinrichtung genannt. Die adressierte Einheit kann bis zum Analysezeitpunkt ihren Namen geändert haben, aus der übergeordneten Einheit ausgegliedert oder in eine andere Forschungseinrichtung eingegliedert worden sein. Für die Institutionenkodierung werden daher die Forschungseinrichtungen (teilweise auch die Teileinheiten) mit verschiedenen Attributen wie Bezeichnungen, Adressen und Sektorzugehörigkeit erfasst, welche jeweils in aktuellen aber auch historischen Varianten relevant sein können. Daher werden alle Attribute mit ihren jeweiligen Gültigkeitszeiträumen versehen (soweit diese recherchiert werden konnten). Damit wird es möglich, den fortlaufenden Wandel der Forschungslandschaft mit Fusionen, Aus- und Eingliederungen, sowie Schließungen abzubilden (Historisierung) und für die Publikationszuordnung zu berücksichtigen. Die in der Adresskodierung verwendeten Einheiten unterschiedlicher Hierarchieebenen werden in den bereitgestellten Tabellen immer auf die Ebene der Hauptinstitution aggregiert (vgl. Abschnitt „Definition der Hauptinstitution“).

Zuordnungsvarianten: Modus A vs. Modus S

Die Veränderungen in der Forschungslandschaft können Einfluss auf die Zuordnung der Publikationen haben und der adäquate Umgang mit diesen Strukturänderungen hängt jeweils von Projekt, Kontext und Fragestellung der jeweiligen bibliometrischen Analyse ab. Angesichts dieser unterschiedlichen Anforderungen an die Zuordnungsregeln kann es für die von Umstrukturierungen betroffenen Forschungseinrichtungen nicht immer ein eindeutiges und allgemein verbindliches Ergebnis der Adress- und Institutionenkodierung geben. So kann es verschiedene Antworten etwa auf die Frage geben, ob ein institutioneller Nachfolger Publikationen seines Vorgängers „erbt“ oder ob Publikationen auch Einrichtungen zugeordnet werden, die nicht mehr existieren bzw. in anderen aufgegangen sind. Es werden daher zwei Tabellensätze zur Zuordnung erstellt: ein Tabellensatz, welcher die Strukturveränderungen nach bestimmten Regeln berücksichtigt und die Publikationen gemäß dieser Regeln der aktuellen bzw. zuletzt gültigen Forschungseinrichtung zuordnet (im Folgenden 'Modus A' genannt) und ein Tabellensatz, der immer die zum Publikationszeitpunkt beste-

hende Situation berücksichtigt ('synchrone' Zuordnung, im Folgenden 'Modus S' genannt).

Während der Zuordnungsmodus A durch die bereits vorgenommene Einarbeitung der Strukturveränderungen dem Anwender viele Entscheidungen abnimmt (somit aber auch eine individuelle Anpassung auf den jeweiligen Kontext durch den Anwender selbst nicht unmittelbar zulässt) und so relativ einfache Abfragen ermöglicht, lässt der Modus S dem Anwender mehr Möglichkeiten offen. So kann beispielsweise eine aktuell nicht mehr als Hauptinstitution existierende, weil in eine andere Einheit eingliederte Forschungseinrichtung in der Analyse mit ihren zugeordneten Publikationen berücksichtigt werden (direkte Verwendung der gelieferten Zuordnung) oder aber deren Publikationen können der aktuell gültigen übergeordneten Einheit zugerechnet werden. Um in diesen Fällen eine individuelle, projektbezogene Übertragung der Publikationen zu ermöglichen, werden die Umstrukturierungen für die betroffenen Hauptinstitutionen mit ihrem jeweiligen Datum sowie dem Typ der Strukturänderung in einer zusätzlichen Tabelle bereitgestellt.

Die konkreten Auswirkungen der beiden Zuordnungsvarianten für Publikationen auf der Ebene der Hauptinstitutionen werden weiter unten noch exemplarisch dargestellt. Welche die 'bessere' ist, muss im jeweiligen Anwendungskontext entschieden werden. Die beiden hier vorgestellten Modi sind in Reaktion auf die bisher artikulierten Bedürfnisse der am Aufbau des Kompetenzzentrums Bibliometrie beteiligten Partner entwickelt worden. Sie stellen nicht die einzig möglichen Lösungen für die Zuordnung dar. Weitere Varianten wären realisierbar, wenn dafür Bedarf entstehen sollte.

Zuordnung zur adressierten Einheit

Ausgangspunkt der Zuordnung sind die Originaldatensätze der Datenbankhersteller, mit ihrer Varianz in den Schreibweisen⁵, den Ungenauigkeiten, Unvollständigkeiten und Fehlern. Die Verwendung der Adressinformationen in bibliometrischen Studien setzt jedoch in den meisten Fällen eine verlässliche Zuordnung der Adressen zu den Basiseinheiten der Analyse, in der Regel Forschungseinrichtungen, Sektoren oder Länder, voraus.

⁵ Die Varianz wird verursacht durch die Verwendung von Abkürzungen, alternativen oder veralteten Namen, richtige bzw. hilfreiche und auch falsche Vorstandardisierungen durch den Datenbankanbieter usw.

Für die Zuordnung der Adressen werden in einem ersten Schritt jeweils die in den Adressen enthaltenen Bezeichnungen (Namensvarianten und Abkürzungen) verwendet. Für jede identifizierte Einheit (real existierende Forschungseinrichtung/Institution oder Teileinheit) wird ein numerischer Kenner vergeben, mit dem verschiedene Attribute wie Namen, Standortadressen, Sektorzugehörigkeit, aber auch Informationen über hierarchische Relationen verknüpft werden können (jeweils mit ihren Gültigkeitszeiträumen). Solange eine erfasste Einheit in ihrer Identität erhalten bleibt, werden Änderungen ihrer hierarchischen Beziehungen oder anderer Attribute zu ihrem Kenner erfasst. Nur wenn die Einheit geschlossen wird, mit anderen Einheiten verschmolzen oder geteilt, endet ihre Existenz und sie erhält ein entsprechendes Enddatum. Von Verschmelzung (Fusion oder Inkorporation) und Aufteilung unterschieden werden jedoch Eingliederungen und Ausgliederungen. Hier verändert die betroffene Einheit nur ihre Hierarchiebeziehung, d.h. eine vorher selbstständige Einheit wird mit der Eingliederung Teileinheit oder für eine ausgegliederte Einheit endet ihre Relation als Teileinheit und sie wird selbstständig. Vom Prinzip her ist das System so angelegt, dass alle adressierten Einheiten unabhängig von ihrem jeweiligen Typ und ihrer Stellung in der Organisationshierarchie erfasst werden können, so dass im theoretischen Fall einer vollständigen Erfassung ein komplettes Organigramm einer Organisation erstellt werden könnte. In der Praxis wurden jedoch überwiegend nur diejenigen Einheiten aufgenommen, die für die Zuordnung der Adressen zur Hauptinstitution relevant sind.⁶

Es gibt Adressdatensätze, in denen mehr als eine Hauptinstitutionen identifiziert werden kann. Beispiel: „Univ Heidelberg, Inst Immunol, Max Planck Inst Biochem, Martinsried, Germany“. Solche Doppeladressen werden grundsätzlich jeweils beiden Hauptinstitutionen zugeordnet.

Historisierung mit Datumsangaben

Angesichts des Wandels der Institutionenlandschaft ist eine statische Institutionenkodierung für bibliometrische Analysen nur bedingt geeignet. Eine solche Kodierung ohne Berücksichtigung von Strukturänderungen ließe sich

⁶ Für einige Hauptinstitutionen wurden jedoch aufgrund spezifischer Projektkontexte in der Vergangenheit auch alle zugehörigen Teileinheiten weiterer Hierarchieebenen erfasst.

schwerlich projektübergreifend und für einen längeren Zeitraum einsetzen. Die Berücksichtigung der Veränderungen erfolgt durch die Erfassung von Start- und Enddatum für alle diejenigen institutionenbezogenen Datensätze, welche für die Veränderungen relevant sind, sofern die Informationen mit vertretbarem Aufwand recherchiert werden konnten. Bei allen anderen institutionenbezogenen Datensätzen wurden die Start- und Enddaten dann aufgenommen, wenn diese im Zuge ihrer Erfassung oder nachträglich bekannt wurden.

Die Zuordnung der Adressen unter Berücksichtigung der Veränderungen der Forschungslandschaft erfordert im Falle der betroffenen Einrichtungen eine Zuordnung in Abhängigkeit vom Publikationszeitpunkt auf der einen Seite und Zeitpunkt der Strukturänderung auf der anderen Seite. Beide Ereignisse, sowohl die Publikation als auch die Änderung einer Einheit, sind im Hinblick auf die Zurechnung einer wissenschaftlichen Arbeit nicht immer unzweifelhaft als fester Zeitpunkt in logischen Abhängigkeiten bzw. Operationen verwendbar. So erfolgt die Publikation je nach Zeitschrift, Reviewprozess und Arbeitsweise des Verlages mit einem unterschiedlich langen zeitlichen Abstand zum Verfassen der Arbeit. Wenn die Zuordnung von Publikationen über Adressen dem in der Bibliometrie meist angewandten Prinzip „work done at“ folgen soll, ist das Publikationsjahr nicht immer zwingend maßgeblich. Hinzu kommt, dass sich eine erfolgte (bzw. bevorstehende) Strukturänderung in der Praxis der Adressangabe unterschiedlich schnell niederschlägt. Damit ist das optimale oder richtige Publikationsjahr, ab dem beispielsweise Publikationen einer ausgegliederten und jetzt eigenständigen Einheit und nicht mehr der ehemaligen übergeordneten Forschungseinrichtung zugeordnet werden sollten, nicht immer pauschal zu bestimmen, sondern erfordert u.U. die Berücksichtigung des Entstehungshintergrundes der einzelnen Publikation. Um diesem Problem Rechnung zu tragen, wurde ursprünglich im Zuordnungsmodus A ein Übergangszeitraum von zwei Jahren festgelegt, innerhalb dessen Publikationen im Falle von relevanten Strukturänderungen sowohl der alten als auch der neuen Hauptinstitution zugeordnet wurden. Dieses Vorgehen wurde inzwischen aber verworfen, weil es u.a. zu Problemen im Zusammenhang mit der Bildung von Kooperationsindikatoren führte (Zuordnung zu alter und neuer Einheit erzeugt „künstliche“ Kooperationen). Stattdessen wurde für das jetzt bereit gestellte Ergebnis in beiden Zuordnungsmodi wie folgt vorgegangen:

- Publikationen, die im Jahr der Veränderung veröffentlicht wurden, werden der alten Hauptinstitution zugeordnet (begrenzter Nachlauf),
- Publikationen, die im darauf folgenden Jahr veröffentlicht wurden, werden der neuen Hauptinstitution zugeordnet.

Eine Ausnahme bilden Fälle, in denen einer Strukturänderung bereits im gleichen Jahr eine weitere folgt. In diesen Fällen wurde das oben beschriebene Verfahren des begrenzten Nachlaufs nicht angewendet, sondern die im Jahr der Veränderung veröffentlichten Publikationen wurden vollständig der nach der ersten Strukturveränderung gültigen Hauptinstitution zugeordnet.

Häufig sind Datumsangaben nur jahresgenau, aber nicht taggenau zu gewinnen. Das gilt für beide Seiten: die Strukturänderungen der Institutionen und die Publikationsdaten. Die Zuordnung erfolgt deshalb in beiden Modi bislang generell nur jahrgenau. Die genauen Datumsangaben der Strukturänderungen wurden jedoch - soweit bekannt - vorsorglich in den Tabellen erfasst (ungenauere Angaben wurden dabei immer mit dem ersten Tag des angegebenen Zeitraums aufgenommen).

Definition der Hauptinstitution

Als Hauptinstitution wird die Organisation bezeichnet, welche in den erfassten Hierarchiebeziehungen an oberster Stelle steht. Die oberste Hierarchieebene ist aber im Einzelfall nicht immer objektiv zu bestimmen (vgl. dazu Hardeman 2013 sowie Praal et al. 2013). Da alle Zuordnungen zu Forschungseinheiten auf die (Hierarchie-) Ebene der Hauptinstitution aggregiert werden, bestimmt diese auch die Analyseebene in den Auswertungen auf Basis der Adresskodierung, da die untergeordneten (oder übergeordneten) Einheiten nicht sichtbar werden. Eine Festlegung der Zuordnungsebene Hauptinstitution kann naturgemäß nicht allen Anforderungen der Anwender gerecht werden, ist aber für die meisten Sektoren des Forschungssystems im Hinblick auf die Standardanforderungen sinnvoll zu bestimmen. Die für die Institutionenkodierung getroffene Festlegung der obersten Hierarchieebene orientiert sich grundsätzlich an den Kriterien weitgehende Eigenständigkeit oder rechtliche Selbstständigkeit (Rechtsfähigkeit), wobei im Zweifelsfall die Anforderungen an die Sichtbarkeit auf der Zuordnungsebene im Hinblick auf den bibliometrischen Verwendungskontext ausschlaggebend waren. Die folgende Tabelle 1 gibt für die wichtigsten Sektoren einen Überblick, welche Einheit jeweils als Hauptinstitution erfasst ist. Ausgewiesen ist je ein Beispiel und in der Spalte Bemerkungen finden sich Hinweise auf Besonderheiten. Die Spezialfälle der 'An-Institute' und 'Lehrkrankenhäuser' sowie der Berliner Universitätsmedizin (Charité) werden weiter unten noch genauer erläutert.



Tabelle 1: Hauptinstitutionen (wichtige Sektoren)

Sektor	Hauptinstitution	Beispiel	Bemerkungen
Universitäten	Universität	Universität Bielefeld	Universitätskliniken werden als Untereinheiten der jeweiligen Universitäten geführt (außer der Berliner Charité). Sog. „An-Institute“ und Lehrkrankenhäuser werden als unabhängige Hauptinstitutionen geführt.
Fachhochschulen	Fachhochschule	Fachhochschule Aachen	
Max-Planck-Gesellschaft (MPG)	Die einzelnen Institute der Gesellschaft	Max-Planck-Institut für Astronomie	
Leibniz-Gemeinschaft (WGL)	Die einzelnen Institute der Gemeinschaft	FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur	
Helmholtz-Gemeinschaft (HGF)	Die einzelnen Mitgliedseinrichtungen (Zentren) der Gemeinschaft	Deutsches Krebsforschungszentrum (DKFZ)	
Fraunhofer-Gesellschaft (FhG)	Die einzelnen Institute der Gesellschaft	Fraunhofer-Institut für System- und Innovationsforschung	
Ressortforschung (Bund/Länder)	Ämter und Anstalten als nachgeordnete Behörden von Bundes- bzw. Landesministerien	Bundesamt für Naturschutz	
Wirtschaft	Größte inländische Einheit	Accurion GmbH	Komplexe, teils intransparente und häufigem Wandel unterworfenen Konzernstrukturen können nicht vollständig abgebildet werden
Kliniken	Das einzelne Krankenhaus	Bezirkskrankenhaus Landshut	Ausnahmen bilden die großen Klinikkonzerne wie Helios, Vivantes oder die Asklepios Kliniken

Um diejenigen Adressen zu erfassen, die nicht eindeutig einer Hauptinstitution, wohl aber einem Sektor bzw. Bereich zugeordnet werden können, wurden für die großen außeruniversitären Forschungsorganisationen, die Universitäten und die Fachhochschulen jeweils "Restkategorien" gebildet, die als "künstliche" Hauptinstitution geführt werden. Damit werden auch diese, auf Ebene der Hauptinstitution nicht eindeutigen Adressen bei sektorbezogenen Analysen berücksichtigt. Ebenso wurden Adressen mit Verweisen auf MPG-Forschungsgruppen an Universitäten und International Research Schools (IMPRS) als "künstliche" Hauptinstitution zusammenge-

fasst, da diese in der Institutionenkodierung bisher nicht alle differenziert als Einheiten erfasst wurden. Die Adressen der MPG-Forschungsgruppen an Universitäten wurden sowohl dem Sektor Max-Planck-Gesellschaft als auch dem Sektor Hochschulen, Bereich Universitäten zugeordnet. Die Sonderstellung der "Sammel"- und Restkategorien ist in den Auswertungen auf Ebene der Institutionen zu berücksichtigen, da sie keine Einheiten im eigentlichen Sinne darstellen. Die Tabelle 2 zeigt die entsprechenden Kategorien je Sektor, jeweils mit einem Adressbeispiel.

Tabelle 2: Rest- bzw. Sammelkategorien für Adressen

Sektor	Rest- bzw. Sammelkategorie	Adressbeispiel
Universitäten	Rest-Kategorie Hochschulen	Univ Berlin, Berlin, Germany
Fachhochschulen	Rest-Kategorie Fachhochschulen	Univ Appl Sci, D-10623 Berlin, Germany
Max-Planck-Gesellschaft (MPG)	MPG-Restkategorie, ohne eindeutige Instituts-Adresse	Max Planck Inst, Leipzig, Germany
	International Max Planck Research Schools	Int Max Planck Res Sch Computat Biol & Sci Comp, Berlin, Germany
MPG und Universitäten	Max-Planck-Forschungsgruppen an Universitäten	Max Planck Res Grp Marine Geochem, Oldenburg, Germany
Leibniz-Gemeinschaft (WGL)	Leibniz-Restkategorie, ohne eindeutige Instituts-Adresse	Leibniz Res Ctr, Dortmund, Germany
Helmholtz-Gemeinschaft (HGF)	HGF-Restkategorie, ohne eindeutige Instituts-Adresse	Helmholtz Gemeinschaft, D-53175 Bonn, Germany
Fraunhofer-Gesellschaft (FhG)	FhG-Restkategorie, ohne eindeutige Instituts-Adresse	Fraunhofer Inst, Stuttgart, Germany

Strukturveränderungen von Hauptinstitutionen: Zuordnungsmodus A vs. S

Strukturveränderungen von Hauptinstitutionen können je nach Modus der Zuordnung von Publikationen zu unterschiedlichen Ergebnissen führen. Im Folgenden werden diese Unterschiede für einige Grundformen exemplarisch erläutert.

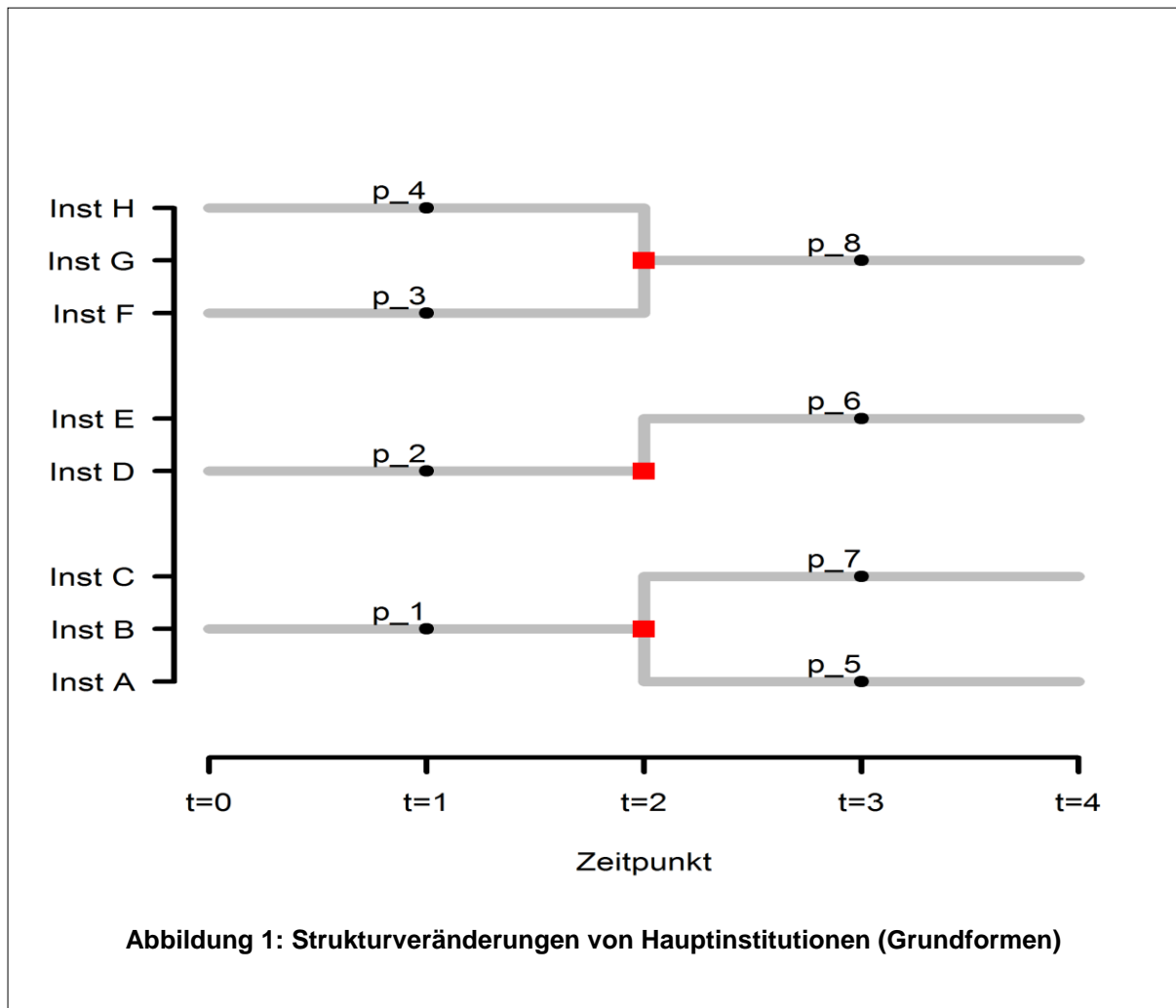
Abbildung 1 zeigt drei Grundformen von Strukturveränderungen bei Hauptinstitutionen: Zusammenführung/Fusion, Teilung/Aufspaltung und Nachfolge. Inst A bis Inst G sind dabei Hauptinstitutionen, deren Existenz über die Zeit jeweils durch eine horizontale Linie dargestellt wird. Die Punkte p_1 bis p_8 stehen für Publikationen, die zu den Zeitpunkten t=1 bzw. t=3 in den jeweiligen Hauptinstitutionen entstanden sind. Die roten Quadrate zum Zeitpunkt t=2 signalisieren Strukturveränderungen. So wird Inst B zum Zeitpunkt t=2 in Inst C und Inst A aufgeteilt, Inst D geht zum Zeitpunkt t=2 in Inst E über und Inst F und Inst H fusionieren zum Zeitpunkt t=2 zu Inst H. Aus Vereinfachungsgründen wird hier angenommen, die Publikationen p_1 bis p_8 seien jeweils nur einer Hauptinstitution zuzuordnen.

Auf diese Grundformen lassen sich auf der Ebene der Hauptinstitutionen alle anderen Typen von Strukturänderungen zurückführen.⁷

Die Publikationen p_1 bis p_8 werden je nach Modus unterschiedlich zugeordnet. Im Modus S (synchrone Zuordnung) wird auf die zum Publikationszeitpunkt gültigen Strukturen fokussiert. Eine Publikation wird grundsätzlich der Hauptinstitution zugeordnet, die zum Zeitpunkt der Publikation Hauptinstitution für die adressierte (Teil-) Einheit ist. So wird beispielsweise p_4 der Hauptinstitution Inst H zugeordnet, auch wenn diese aktuell (zum Zeitpunkt t=4) nicht mehr existiert. Publikationen mit einer Inst H zugeordneten Adresse aus den Jahren nach t=2 werden Inst G zugeordnet, da zum Zeitpunkt der Publikation H in G aufgegangen ist.

erfassten Einheiten kann die Zusammenführung von Einheiten, hier dargestellt als Fusion (H und F verschmelzen in G), auch als Eingliederung (H und F werden Teileinheiten einer neuen Einheit G, oder H wird Teileinheit von F, F bleibt dann Hauptinstitution) oder Inkorporation (H geht in F auf, F bleibt Hauptinstitution) dargestellt werden.

⁷ Im Hinblick auf die Adresszuordnung und die



Im Modus A (aktuelle Sicht) werden dagegen auch die vor $t=2$ entstandenen Publikationen den nach den Strukturveränderungen gültigen Hauptinstitutionen zugeordnet. Jede Hauptinstitution 'nimmt ihre Publikationen mit', wenn sie sich strukturell verändert. Die Publikation p_4 ist zu einem Zeitpunkt veröffentlicht, als Inst H noch Hauptinstitution war. Da diese aber (mit der Inst F zusammen) zur Inst G geworden ist, bringt sie auch ihre Publikationen mit zur Inst G. Die Publikation p_3 wird damit der Inst G zugerechnet.

Publikationen werden in diesem Modus also nur solchen Hauptinstitutionen zugeordnet, die entweder aktuell (im Bild $t=4$) noch existieren oder aber ohne eindeutigen Nachfolger strukturell verändert wurden. 'Ohne eindeutigen Nachfolger' können Institutionen beispielsweise im Fall einer Teilung bleiben. In solchen Fällen ist die eindeutige Zuordnung zu einem der Nachfolger im Allgemeinen nicht möglich. Sofern es sich um eine Ausgliederung handelt (wenn etwa die Einheiten C und A gemäß Abbildung 1 vor der Umstrukturierung Teilein-

heiten von B waren), kann eine Zuordnung theoretisch möglich sein, die Umsetzung scheidet aber in der Praxis häufig wegen mangelnder Vollständigkeit und Genauigkeit der Adressen (keine Nennung der Untereinheiten 'C' und 'A' in der Adresse). Daher werden bei einer Teilung die Publikationen bis zur Teilung dem Vorgänger zugeordnet (hier Inst B), nach der Teilung den Nachfolgern (Inst A, Inst C). Publikationen mit einer Adresse der Inst B aus Publikationsjahren nach $t=2$ werden weiterhin der Inst B zugeordnet, auch wenn diese nicht mehr existiert (Nachlauferfassung). Gleiches gilt auch in einem anderen Fall von Veränderung ohne eindeutigen Nachfolger: der einfachen Schließung einer Hauptinstitution. Durch diese Nachlauferfassung wird sichergestellt, dass für nicht mehr existierende Einrichtungen keine Publikationszuordnungen verloren gehen. Tabelle 3 verdeutlicht noch einmal die Unterschiede, die sich durch die beiden Zuordnungsmodi in Bezug auf die in Abbildung 1 gezeigte Beispielsituation ergeben.

Tabelle 3: Publikationszuordnung Modus A vs. S gemäß Abb. 1

Publikation	Modus S	Modus A	Bemerkungen zu Modus A
p_1	Inst B	Inst B	Kein eindeutiger Nachfolger
p_2	Inst D	Inst E	Der Nachfolger 'erbt' die Publikation vom Vorgänger
p_3	Inst F	Inst G	Der Nachfolger 'erbt' die Publikation vom Vorgänger
p_4	Inst H	Inst G	Der Nachfolger 'erbt' die Publikation vom Vorgänger
p_5	Inst A	Inst A	Aktuell existierende Hauptinstitution
p_6	Inst E	Inst E	Aktuell existierende Hauptinstitution
p_7	Inst C	Inst C	Aktuell existierende Hauptinstitution
p_8	Inst G	Inst G	Aktuell existierende Hauptinstitution

Welche der beiden Zuordnungsvarianten für Publikationen auf der Ebene der Hauptinstitutionen die 'bessere' ist, muss im jeweiligen Anwendungskontext entschieden werden. Jeder Modus hat für unterschiedliche Fragestellungen spezifische Vor- und Nachteile, die insbesondere bei der Bildung von Zeitreihen abzuwägen sind. Eine allgemein gültige Empfehlung ist hier nicht möglich. Der Modus S ermöglicht ggf. eine größere Flexibilität im Einsatz auf der Seite der Anwender, allerdings um den Preis steigender Komplexität in den entsprechenden Abfragen.

Die beiden hier vorgestellten Modi sind in Reaktion auf die bisher artikulierten Bedürfnisse der am Aufbau des Kompetenzzentrums beteiligten Partner entwickelt worden. Weitere Varianten der Zuordnung wären realisierbar, wenn dafür Bedarf entstehen sollte.

Beispiele für Zuordnungen zu Hauptinstitutionen im Modus A vs. S

Im Folgenden wird die Umsetzung der oben dargestellten Zuordnungsprinzipien anhand konkreter Adressbeispiele gezeigt. Damit werden die unterschiedlichen Auswirkungen von

Modus A und Modus S im Einzelfall exemplarisch deutlich.

Die beiden ersten Beispiele dienen der Veranschaulichung von Fusionen: Zwei oder mehrere Hauptinstitutionen werden zu einer neuen Einrichtung zusammengeschlossen. Dabei können die vorher eigenständigen Hauptinstitutionen in der neuen Einheit aufgehen, oder aber ihre Identität als Einheit bewahren und als Teileinheit in der neuen Hauptinstitution weiterexistieren (Eingliederung in eine neu geschaffene Einheit). Im Hinblick auf die Zuordnung zur Hauptinstitution führen beide Fälle zum gleichen Resultat.

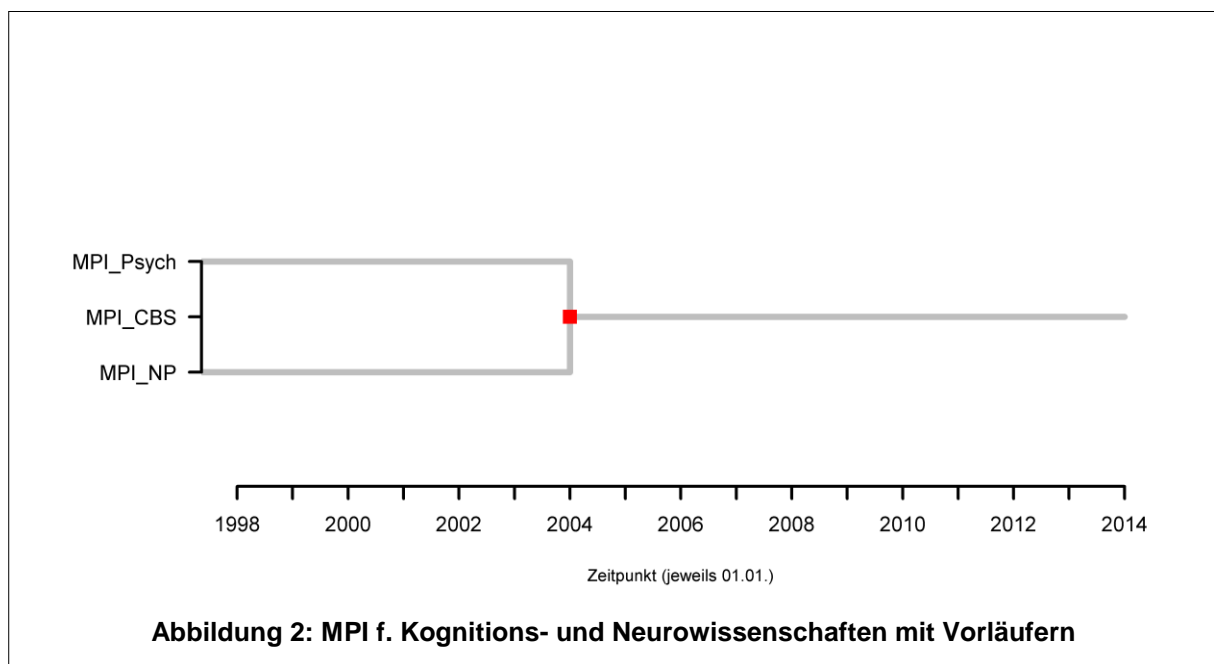
Tabelle 4 zeigt anhand von Beispieladressen die Vorgehensweise im Fall der Fusion der ehemaligen Max-Planck-Institute für psychologische Forschung (MPI_Psych) und für neuropsychologische Forschung (MPI_NP) zum neuen Max-Planck-Institut für Kognitions- und Neurowissenschaften (MPI_CBS) zum 1.1.2004.

Abbildung 2 zeigt die Zusammenführung (Fusion) der beiden Vorläufer-MPI zum MPI-CBS in der schematischen Darstellung.



Tabelle 4: Adressbeispiele für MPI f. Kognitions- und Neurowissenschaften und Vorläufer

Publikationsadresse	PY ⁸	Zuordnung	
		Modus S	Modus A
Max Planck Inst Psychol Res, D-80804 Munich, Germany	2003	MPI_Psych	MPI_CBS
	2004	MPI_CBS	
	2005		
Max Planck Inst Neuropsychol Forsch, D-04103 Leipzig, Germany	2003	MPI_NP	
	2004	MPI_CBS	
	2012		
Max Planck Inst Cognit Neurosci, Leipzig, Germany	1998	MPI_NP	
	2004	MPI_CBS	
	2008		
Max Planck Inst Human Cognit & Brain Sci, Leipzig, Germany	2004	MPI_CBS	



⁸ PY bezeichnet in dieser und den folgenden Tabellen den Publikationsjahrgang (Publication Year).

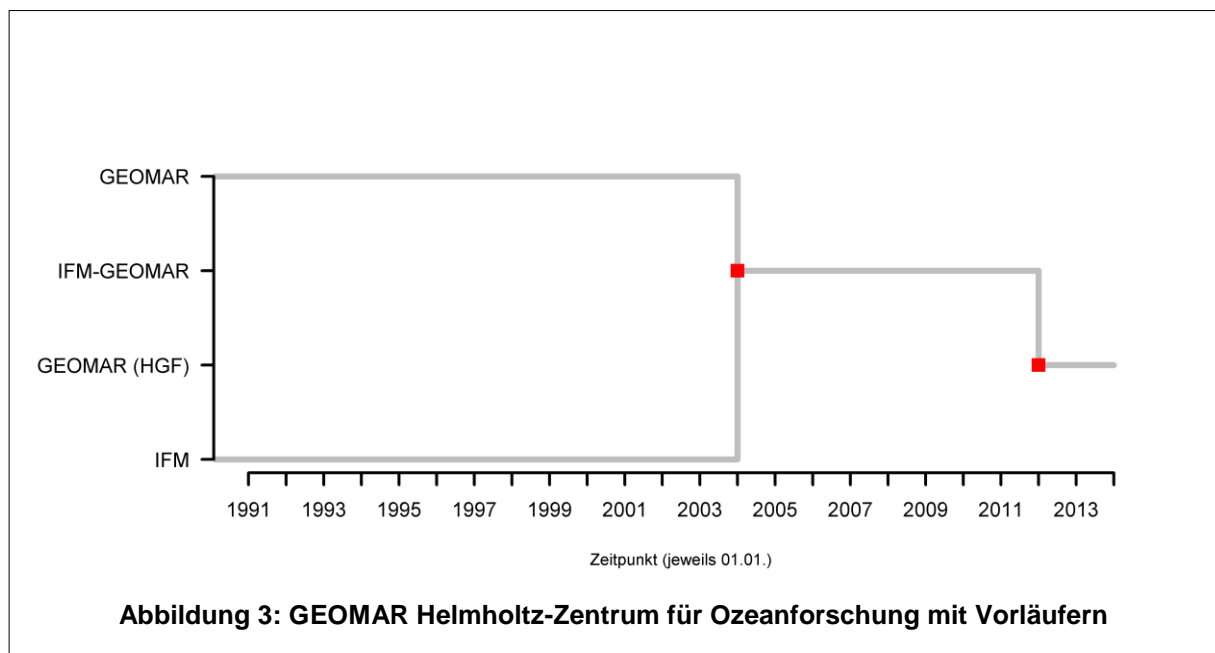


Ein zweites Beispiel (Tabelle 5, Abbildung 3) stellt das am 1.1.2012 etablierte GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel (HGF_GEOMAR) dar, dessen Vorgänger als Leibniz-Institut für Meereswissenschaften an der Universität Kiel (IFM-GEOMAR) am 1.1.2004 durch Fusion des Leibniz-Instituts für Meereswissenschaften an der Universität Kiel (IFM) mit dem Forschungszentrum für Marine Geowissenschaften GEOMAR (GEOMAR) entstanden war.

Das Beispiel zeigt, dass neben den mit Veränderungen der Organisationsform meist verbundenen Namensänderungen auch Wechsel in der Zugehörigkeit zu Sektoren von Bedeutung sein können (im Fall des Geomar sind HGF, WGL und der Hochschulsektor im Spiel). Ebenfalls deutlich wird hier, dass eine Adresse abhängig vom Publikationsjahr zu unterschiedlichen Zuordnungen der Publikation führen kann.

Tabelle 5: Adressbeispiele für GEOMAR (HGF) und Vorläufer

Publikationsadresse	PY	Zuordnung	
		Modus S	Modus A
GEOMAR Res Ctr Marine Geosci, D-24148 Kiel, Germane	2003	GEOMAR	GEOMAR (HGF)
	2004	IFM-GEOMAR	
	2005		
Inst Meereskunde, D-24105 Kiel, Germany	2003	IFM	
	2004	IFM-GEOMAR	
	2008		
Leibniz Inst Marine Sci, D-24105 Kiel, Germany	2004	IFM-GEOMAR	
	2013	GEOMAR (HGF)	
IFM GEOMAR, D-24148 Kiel, Germany	2004	IFM-GEOMAR	
	2013	GEOMAR (HGF)	
Geomar, Kiel, Germany	1991	GEOMAR	
	2004	IFM-GEOMAR	
	2013	GEOMAR (HGF)	
GEOMAR, Helmholtz Ctr Ocean Res, Kiel, Germany	2011	IFM-GEOMAR	
	2013	GEOMAR (HGF)	





Das nächste Beispiel beleuchtet die Verhältnisse im Fall einer Ausgliederung (Tabelle 6, Abbildung 4). Im Zusammenhang mit der Schließung der Gesellschaft für Mathematik und Datenverarbeitung (GMD) Mitte des Jahres 2001 wurden deren Institute ausgegliedert und in selbstständige Fraunhofer-Institute umgewandelt. Die ehemaligen Institute der

GMD (Fokus und FIT) wechseln damit ihre Sektorzugehörigkeit sowie ihre Namen und werden nun als Hauptinstitutionen geführt. Damit liegt nach den Regeln der Institutionenkodierung ein Fall von Ausgliederung (und Verselbstständigung) von Teileinheiten vor, hier mit gleichzeitigem Sektorwechsel.

Tabelle 6: Adressbeispiele für GMD und FhG-Nachfolger

Publikationsadresse	PY	Zuordnung	
		Modus S	Modus A
INST OPEN COMMUN SYST,WIRELESS MOBILE ATM GRP,D-10623 BERLIN,GERMANY	1996	GMD	GMD
GMD Fokus, D-10589 Berlin, Germany	2002	GMD; Fokus (FhG)	GMD; Fokus (FhG)
GMD FOKUS, Berlin, Germany	2008	GMD; Fokus (FhG)	GMD; Fokus (FhG)
Fraunhofer FOKUS, Berlin, Germany	2002	Fokus (FhG)	Fokus (FhG)
	2012	Fokus (FhG)	Fokus (FhG)
FIT, AI Res Div, D-53754 Sankt Augustin, Germany	1996	GMD	GMD
FIT, GMD, D-53754 St Augustin, Germany	2000	GMD	GMD
	2002	GMD; FIT (FhG)	GMD; FIT (FhG)
Fraunhofer Inst Appl Informat Technol, FIT, Schloss Birlinghoven, D-53754 St Augustin, Germany	2001	GMD; FIT (FhG)	GMD; FIT (FhG)
Fraunhofer Inst Appl Informat Technol FIT, St Augustin, Germany	2002	FIT (FhG)	FIT (FhG)
	2012	FIT (FhG)	FIT (FhG)
FIT, D-53757 St Augustin, Germany	2004	FIT (FhG)	FIT (FhG)

Im Modus S werden die den (Teil-) Einheiten (ehemals GMD-Institute, ab Mitte 2001 Fraunhofer-Institute) zugeordneten Adressen nach den Regeln der synchronen Zuordnung ab dem auf die Umstrukturierung folgenden Jahr der neuen Hauptinstitution zugeordnet - in diesem Fall also auch dem neuen Sektor. Bis zum Jahr der Ausgliederung der Institute werden deren Adressen der GMD als Hauptinstitution zugeordnet. Danach, hier also ab dem Publikationsjahr 2002, erfolgt die Zuordnung gemäß den im Publikationsjahr gültigen Hierarchiebeziehungen (vgl. Tabelle 6). Dement-

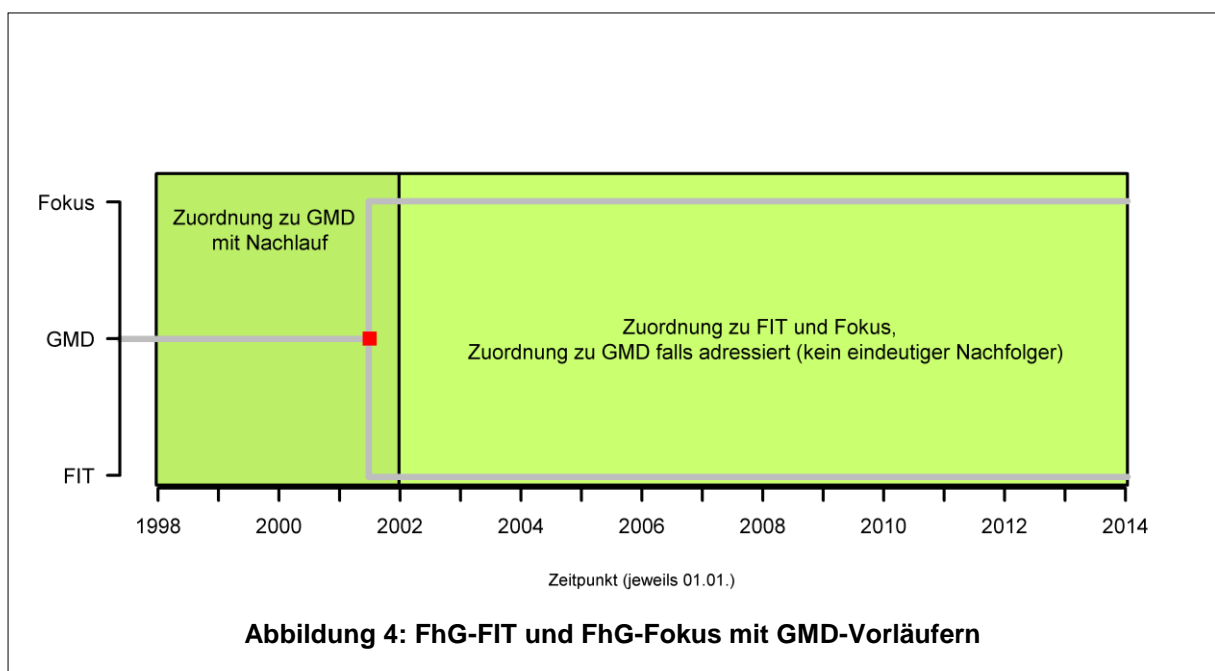
sprechend werden die Adressen der beiden Institute (FIT bzw. Fokus) bis einschließlich 2001 der GMD, danach den verselbstständigten und nun als Hauptinstitutionen geführten Fraunhofer-Instituten zugeordnet. Taucht die GMD auch nach dem Jahr ihrer Schließung noch explizit in Autorendressen auf, so werden diese Adressen auch der GMD und damit der Helmholtz-Gemeinschaft zugeordnet. Publikationen nach 2001 mit diesen Adressen werden also doppelt zugeordnet: sowohl dem Fraunhofer-Institut („Fokus“ bzw. „FIT“) als auch der (bereits geschlossenen) GMD. Damit



ergibt sich für sie auch eine doppelte Sektorzuordnung (FhG und HGF).

Im Modus A, der die aktuelle Sicht priorisiert, führen im Falle der Ausgliederung die Regeln zum gleichen Ergebnis wie in der synchronen Zuordnung. Die Adressen der Teileinheiten werden ab dem auf die Ausgliederung folgenden Jahr den dann selbstständigen Einrichtungen (hier Fraunhofer-Institute) zugeordnet. Das Grundprinzip der Projektion auf die aktuelle Struktur der Forschungseinrichtungen wird im Falle der Ausgliederung nicht umgesetzt, da eine zweifelsfreie Zuordnung der Publikationen

der ehemaligen Teileinheiten zu den nach der Umstrukturierung aus ihnen hervorgegangenen selbstständigen Forschungseinrichtungen häufig (anders als im hier behandelten Beispiel) aufgrund der Adressinformationen nicht möglich ist. Da auch in diesem Modus bei expliziter Adressierung die Zuordnung zur ehemaligen Hauptinstitution auch für die Publikationsjahre nach deren Schließung erfolgt, entstehen hier keine Differenzen zur synchronen Zuordnung.





Mehrfachzuordnung (ausgewählte Fälle)

In einigen Fällen stößt die Anwendung genereller Zuordnungsprinzipien auf Schwierigkeiten, die mit Besonderheiten der jeweiligen Organisationsform zusammenhängen. In der komplexen und sich wandelnden Struktur der Forschungslandschaft entstehen in zunehmendem Maß Einrichtungen, die sich in der Realität einer eindeutigen Zuordnung entziehen, indem sie etwa sektorenübergreifend in Hierarchien eingebunden sind. Für das Prinzip der Adress- und Institutionenkodierung entsteht damit grundsätzlich kein Problem, da Mehrfachzuordnungen vorgesehen sind. Im Einzelfall kann es sich jedoch ergeben, dass die getroffene Entscheidung für die Ebene der Hauptinstitution nicht für alle Anwendungskontexte in gleicher Weise befriedigend ist (ein besonders augenfälliges Beispiel hierfür ist das KIT, s.u.). In den folgenden Abschnitten wird für ausgewählte Fälle das Vorgehen erläutert.

IPP

Ein ganz einfaches Beispiel stellt das Max-Planck-Institut für Plasmaphysik (IPP) dar, das sowohl der Max-Planck-Gesellschaft als auch der Helmholtz-Gemeinschaft angehört. Im Kodierungsergebnis wird das durch einen entsprechenden Doppeleintrag in der Sektorzuordnungstabelle abgebildet.

KIT

Ein anderer Fall einer „sektorhybriden“ Institution ist das Karlsruher Institut für Technologie (KIT). Dabei handelt es sich um die Eingliederung der Universität Karlsruhe (U Karlsruhe) und des Forschungszentrums Karlsruhe (FZK) in eine neu geschaffene Einheit.⁹ Die ehemaligen Hauptinstitutionen bleiben als Teileinheiten des KIT (Campus Süd und Campus Nord) erhalten, verlieren jedoch ihren Status als Hauptinstitution und ihre Sektorzuordnung zu Universitäten bzw. HGF. Die Zuordnungsmodi führen hier zu erheblichen Unterschieden: im Modus A „erbt“ das KIT alle Publikationen seiner beiden Vorläufer; im Modus S dagegen bleiben wegen der synchronen Zuordnung die alten Publikationen den ehemaligen Hauptinstitutionen U Karlsruhe bzw. FZK zugeordnet. Der Anteil von Publikationen, die sich nicht

⁹ Sektormäßig wird das KIT als Spezialfall (mit eigenem Sektor: KIT) geführt. Das kann bei Bedarf geändert werden, ohne in die eigentliche Adresskodierung eingreifen zu müssen.

ohne Weiteres auf der Basis der Adressdatensätze einer dieser beiden Teileinheiten des KIT zuordnen lassen, wird in Zukunft steigen. Damit wird das KIT mittelfristig zu einer Problemzone für sektorbezogene Analysen und erfordert ggf. projektspezifische Nachjustierungen der Kodierung.

Charité

Ein weiterer Problemfall betrifft die Charité - Universitätsmedizin Berlin, die als gemeinsame Einrichtung der Freien Universität Berlin (FU) und der Humboldt-Universität zu Berlin (HU) an vier Standorten inzwischen über 100 Kliniken umfasst. Die beiden Vorläufer *Universitätsklinikum Benjamin Franklin der Freien Universität* und *Universitätsklinikum Charité, Medizinische Fakultät der Humboldt-Universität zu Berlin* wurden 2003 zusammengeführt. Ab diesem Zeitpunkt ist eine trennscharfe Aufteilung der medizinischen Publikationen zwischen FU und HU kaum noch möglich. Stattdessen erfolgt die Zuordnung unmittelbar zur Charité, die jetzt als selbstständige Hauptinstitution im Sektor Universitäten geführt wird. Bei Analysen, die eine getrennte Ausweisung von Indikatoren für die FU und HU erfordern, muss dies ggf. berücksichtigt werden. Die Datenlage ermöglicht hier keine einfache Lösung.

UKSH

Ein ähnlicher, für die Institutionenkodierung allerdings weniger problematischer Fall ist beim Universitätsklinikum Schleswig-Holstein (UKSH) gegeben. Im UKSH sind zum 1.1.2003 die Universitätsklinikum Kiel und Lübeck zusammengeführt worden - unter Beibehaltung der bisherigen Standorte. Da in der Praxis als Publikationsadresse weiterhin jeweils der konkrete Klinikstandort genannt wird, kann die Zuordnung zur entsprechenden Universität als Hauptinstitution unverändert auf dem üblichen Wege erfolgen.

Lehrkrankenhäuser

Als akademische Lehrkrankenhäuser werden Krankenhäuser bezeichnet, an denen Teile der klinischen Ausbildung von Medizinstudenten stattfinden. Zwischen der jeweiligen medizinischen Fakultät und dem Lehrkrankenhaus besteht ein Vertrag, der den Rahmen für die Studentenausbildung und deren Vergütung regelt. Die Krankenhäuser sind damit jedoch nicht Teil der Universität.¹⁰ Durch die Tatsa-

¹⁰ Zur Problematik der institutionellen Abgrenzung zwischen Universitäten und deren Klini-



che, dass in Autorenadressen der Lehrkrankenhäuser häufig explizit auf diesen Status einschließlich der Nennung des Namens der Universität verwiesen wird, besteht das Risiko der Falschzuordnung solcher Publikationen zu den Universitäten. Um dem entgegenzuwirken, wird im Zuge der Institutionenkodierung die Zurechnung zur Universität aufgrund der Adresse eines Lehrkrankenhauses eliminiert, wenn ein entsprechender Bezeichner ("Akad Lehrkrankenhaus", "Acad Teaching Hosp", "Acad Training Hosp", o.a.) enthalten ist. Taucht dagegen der Name der Universität in einer Publikationsadresse zusammen mit dem Krankenhaus *ohne* Verweis auf den Status als Lehrkrankenhaus auf, ist eine Doppeladresse nicht auszuschließen. Deshalb wird in diesen Fällen auch doppelt zugeordnet. In Tabelle 7 sind einige Beispiele aufgeführt, die die Zuordnungen von Autorenadressen in Fällen von Lehrkrankenhäusern verdeutlichen. Am Ende der Tabelle finden sich zwei Beispiele, die auch Doppeladressen darstellen.

Eine generelle Erfassung aller Adressen von Lehrkrankenhäusern, also auch ohne einen expliziten Bezeichner dieses Status in der Adresse, ist nicht ohne Weiteres möglich. Wegen der großen Zahl und der unsicheren Quellenlage¹¹ wäre der Aufwand für eine systematische Erfassung dieses Institutionsmerkmals und der darauf zu beziehenden Adressvarianten zu hoch. Die Anzahl der Adressen von Lehrkrankenhäusern ohne Hinweis auf ihren Status, aber mit Nennung der Universität ist deshalb nicht genau bestimmbar; die Untersuchung von Einzelfällen gibt aber Anlass zur Annahme, dass ihre Zahl gering ist. Falls in Zukunft eine differenzierte Analyse speziell für diesen Bereich gebraucht würde, könnten dafür bei Bedarf die bisher erfassten publikationsrelevanten Lehrkrankenhausrelationen bereitgestellt werden.

ken und Krankenhäusern vgl. auch Praal et al. 2013, 285ff.

¹¹ Wikipedia listet für Deutschland aktuell [über 600 Lehrkrankenhäuser](#) auf. Die Liste kann jedoch nur als Hinweis, nicht als zuverlässiger Nachweis verwendet werden. Eine Rekonstruktion der Vertragsverhältnisse für länger zurückliegende Zeiträume ist häufig nicht möglich, insbesondere im Falle von Fusionen und Eingliederungen von Krankenhäusern.



Tabelle 7: Adressbeispiele für Lehrkrankenhäuser

Publikationsadresse ¹²	Zugeordnete Einheit	Zugeordnete Hauptinstitution
Univ Tübingen, Akad Lehrkrankenhaus , Olgahosp, D-70176 Stuttgart, Germany	Olgahospital, Klinikum Stuttgart	Klinikum Stuttgart
Univ Tübingen, Med Klin, Diakonie Klinikum Stuttgart, Akad Lehrkrankenhaus , D-70176 Stuttgart, Germany	Diakonie Klinikum Stuttgart	Klinikum Stuttgart
Univ Tübingen, Robert Bosch Krankenhaus, Klin Geriatr Rehabil, Akad Lehrkrankenhaus , D-70376 Stuttgart, Germany	Robert-Bosch-Krankenhaus	Robert-Bosch-Krankenhaus
Univ Tübingen, Marienhosp Stuttgart, Unfallchirurg Klin, Akad Lehrkrankenhaus , Stuttgart, Germany	Marienhospital Stuttgart	Marienhospital Stuttgart
Univ Göttingen, Akad Lehrkrankenhaus , Asklepios Kliniken Schildautal, Seesen, Germany	Asklepios Klinik Schildautal	Asklepios Kliniken GmbH
Univ Hamburg, Akad Lehrkrankenhaus , Asklepios Klin Wandsbek, Abt Geriatrie, Hamburg, Germany	Asklepios Klinik Wandsbek	
Univ Frankfurt, Klinikum Darmstadt, Dept Obstet & Gynaecol, Acad Teaching Hosp , Darmstadt, Germany	Klinikum Darmstadt	Klinikum Darmstadt
Univ Heidelberg Mannheim, Acad Teaching Hosp , Dept Diagnost Radiol, Klinikum Darmstadt, Darmstadt, Germany		
Univ Heidelberg, Acad Teaching Hosp , Klinikum Darmstadt, Dept Pathol, D-64283 Darmstadt, Germany		
Univ Frankfurt & Heidelberg Mannheim, Akad Lehrkrankenhaus , Klinikum Darmstadt, D-64276 Darmstadt, Germany		
Bogenhausen Univ Hosp, Clin Neuropsychol Res Grp EKN, Munich, Germany	Klinikum Bogenhausen	Klinikum Bogenhausen
Tech Univ Munich, Dept Plast Surg, Burn Ctr, Klinikum Bogenhausen, Acad Teaching Hosp , D-8000 Munich, Germany		
Charite, Dept Plast & Aesthet Surg, HELIOS Klinikum Emil von Behring, Acad Teaching Hosp , D-14109 Berlin, Germany	HELIOS Klinikum Emil von Behring	HELIOS Kliniken GmbH
Free Univ Berlin, Acad Teaching Hosp , Behring Municipal Hosp, Dept Hand & Plast Surg, Berlin, Germany		
Charite Univ Med Berlin, Inst Pathol, HELIOS Klinikum Emil von Behring, Berlin, Germany	<ul style="list-style-type: none"> • HELIOS Klinikum Emil von Behring • Charité – Universitätsmedizin Berlin 	<ul style="list-style-type: none"> • HELIOS Kliniken GmbH • Charité – Universitätsmedizin Berlin
Free Univ Berlin, City Hosp Berlin Zehlendorf, D-1000 Berlin, Germany	<ul style="list-style-type: none"> • Bezirkskrankenhaus Zehlendorf • Freie Universität Berlin 	<ul style="list-style-type: none"> • HELIOS Kliniken GmbH • Freie Universität Berlin

¹² Für die Erkennung als Lehrkrankenhaus relevante Textmuster sind **fett** dargestellt.



An-Institute

Als angegliederte, aber rechtlich eigenständige Einrichtungen sind die sog. An-Institute an deutschen Hochschulen ein ähnlicher Fall wie die Lehrkrankenhäuser. Auch im Fall dieses Einrichtungstyps findet man häufig die Bezeichnung der Hochschule in der Adresse, obwohl die Einrichtungen nicht Teil der Hochschule sind. Im Unterschied zu den Lehrkrankenhäusern ist hier jedoch die Verwendung eines Bezeichners für den Status nicht üblich. Wenn in einer Publikationsadresse eines An-Instituts auch der Name der Hochschule explizit enthalten ist, geht daraus nicht zweifelsfrei hervor, ob die Universität tatsächlich an der Publikation beteiligt war. Es kann sich schlicht um eine Ergänzung bzw. einen Teil der Bezeichnung des An-Instituts handeln, da diese Institute häufig den Zusatz "an der Universität XY" in ihrem Namen führen. Es kann sich aber auch um eine zusätzliche Adressierung (Doppeladresse) handeln, hinter der in dem konkreten Fall eine Kooperation zwischen der Hochschule und dem An-Institut steckt. Tabelle 8 zeigt einige Beispiele für Autorenadressen von An-Instituten.

Bei vielen An-Instituten ist in ihren Autorenadressen durchgängig kein expliziter Bezug auf den Namen der Hochschule enthalten (was dem selbstständigen Charakter der Institution entspricht). Gibt es in Einzelfällen in den Adressen dieser An-Institute auch einen Adressbestandteil, der auf die Hochschule verweist, so handelt es sich hier eher um eine echte Doppeladresse. In der Adresskodierung wird derzeit entsprechend der tatsächlich in den Adressen vorhandenen Verweise zugeordnet. Alternativ könnten die Autorenadressen mit einer Zuordnung zum An-Institut grundsätzlich nur diesem zugeordnet werden, unabhängig davon, ob in der Adresse auch die Hochschule enthalten ist. Das umgekehrte Vorgehen einer generellen Zuordnung auch zur Hochschule (unabhängig von deren Nennung in der Adresse, nur aufgrund des Status An-Institut der Hochschule) wäre ebenfalls möglich, ist aber angesichts der insgesamt sehr geringen Zahl von echten Doppeladressierungen nicht zu rechtfertigen.

Ähnlich wie bei den Lehrkrankenhäusern ist eine wirklich trennscharfe Abgrenzung der An-Institute von Hochschulen in der Praxis schwierig. Bei insgesamt unsicherer Quellenlage erscheint es kaum möglich, die komplexen und sich im Wandel befindenden Strukturen fehlerfrei abzubilden. Nicht selten finden sich An-Institute, welche in ihrer Entwicklung in Richtung auf andere Sektoren (WGL, HGF, Ausgründungen in die Wirtschaft) im Lauf der Zeit mehrere Organisationsstadien durchlaufen. Die bisher im Rahmen der Institutionenkodierung erfassten publikationsrelevanten Relationen von An-Instituten zu Hochschulen könnten aber herangezogen werden, wenn in Zukunft differenziertere Analysen speziell für diesen Bereich gebraucht werden sollten. Das gilt auch für den Fall, dass Bedarf für eine generelle Zuordnung der An-Instituts-Publikationen zu den jeweiligen Hochschulen entstehen würde.



Tabelle 8: Adressbeispiele für An-Institute (mit und ohne Nennung der Hochschule)

Publikationsadresse ¹³	Adressierte und zugeordnete Einheit	Hochschule
Laser Zentrum Hannover eV, Hollerithallee 8, D-30419 Hannover, Germany	Laser Zentrum Hannover e.V. (LZH)	Universität Hannover
Univ Hannover, Laser Zentrum Hannover LHZ, Hannover, Germany	Laser Zentrum Hannover e.V. (LZH); Universität Hannover	
OFFIS Inst Informat Technol, Oldenburg , Germany	OFFIS e.V. - Institut für Informatik	Universität Oldenburg
Carl von Ossietzky Univ Oldenburg, OFFIS , D-2900 Oldenburg , Germany	OFFIS e.V. - Institut für Informatik; Universität Oldenburg	
Leibniz Inst Arteriosclerosis Res , Munster , Germany	Leibniz-Institut für Arterioskleroseforschung (LIFA)	Universität Münster
Inst Arteriosclerosis Res , Munster , Germany		
Univ Munster, Leibniz Inst Arterioskleroseforsch , Munster , Germany	Leibniz-Institut für Arterioskleroseforschung (LIFA); Universität Münster	
Univ Munster, Inst Arterioskleroseforsch , D-48149 Munster, Germany		
Leibniz Inst Expt Virol, Heinrich Pette Inst , Hamburg , Germany	Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie an der Universität Hamburg	Universität Hamburg
Heinrich Pette Inst Expt Virol & Immunol, D-20251 Hamburg , Germany		
Univ Hamburg, Heinrich Pette Inst Expt Virol & Immunol, D-20251 Hamburg , Germany	Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie an der Universität Hamburg; Universität Hamburg	
DWI eV, D-52056 Aachen , Germany	DWI an der RWTH Aachen e.V.	Rheinisch-Westfälische Technische Hochschule (RWTH)
German Wool Res Inst , D-52062 Aachen , Germany		
Rhein Westfal TH Aachen, DWI RWTH Aachen eV, Lehrstuhl Makromol Mat & Oberflächen, D-52056 Aachen , Germany	DWI an der RWTH Aachen e.V.; RWTH	
RWTH Aachen eV, DWI , D-52056 Aachen , Germany		
Rhein Westfal TH Aachen, DWI , Aachen , Germany		
German Diabet Ctr , Inst Clin Biochem & Pathobiochem, Dusseldorf, Germany	Deutsches Diabetes-Zentrum (DDZ)	Universität Düsseldorf
German Diabet Ctr , Dusseldorf, Germany		
Diabet Forschungsinst , D-40225 Dusseldorf , Germany		
Leibniz Ctr Diabet Res, German Diabet Ctr , Inst Clin Diabetol, Dusseldorf, Germany		

¹³ Für die Erkennung als An-Institut relevante Textmuster sind **fett** dargestellt.



3. Fazit

Die Ergebnisse der Institutionenkodierung liegen für beide Zuordnungsmodi in Form eines Satzes von Tabellen in einem relationalen Datenbanksystem vor. Über entsprechende Identifier ist die Verknüpfung mit den einschlägigen Datensätzen aus WoS bzw. Scopus möglich. Eine laufende Aktualisierung, Vervollständigung und Qualitätsverbesserung des Datenbestandes ist vorgesehen. Die Ergebnisse werden auch in den Datenbestand des Kompetenzzentrums Bibliometrie für die deutsche Wissenschaft eingespist.

In Bezug auf das erzielte Ergebnis des Projekts gilt sowohl für WoS als auch Scopus: Der Erfassungsgrad der Adressen ist für die Kernbereiche des deutschen Wissenschaftssystems mit dem Hochschulsektor, der Max-Planck-Gesellschaft (MPG), der Helmholtz-Gemeinschaft (HGF), der Leibniz-Gemeinschaft (WGL) und der Fraunhofer-Gesellschaft (FhG) deutlich höher als für die übrigen, zum Teil sehr kleinteiligen Bereiche wie Wirtschaft, (nicht-universitäre) Kliniken sowie nicht kommerzielle Einrichtungen (Vereine, Verbände, gemeinnützige GmbH etc.), auf die sich ein Großteil der noch in der Restmenge verbliebenen nicht zugeordneten Adressen verteilt. Nach der weitgehend vollständigen Erfassung der Adressen für die Einrichtungen der großen Bereiche und der weiteren publikationsstarken Einheiten entfallen die restlichen Adressen auf sehr viele kleine Institutionen mit häufig nur einzelnen Publikationen. Für eine weitere Vervollständigung der Kodierung steht deshalb dem steigenden Aufwand ein immer kleiner werdender Ertrag gegenüber.

Ein weiteres Hindernis zur Erreichung von Vollständigkeit stellen diejenigen Adressen dar, die zu unspezifisch (Beispiel: „Univ Berlin, Berlin, Germany“) oder in sich widersprüchlich sind (Beispiel: „Univ Med Ctr Schleswig Holstein, Dept Radiol & Nucl Med, Essen, Germany“). Die Datenlage lässt in diesen Fällen oft keine valide Zuordnung zu.¹⁴

¹⁴ Soweit möglich, wurde in solchen Fällen aber zumindest eine entsprechende Sektorzuordnung (z.B.: Hochschulen) vorgenommen, wenn der fragliche Adressdatensatz in dieser Hinsicht widerspruchsfrei war (vgl. Abschnitt Definition der Hauptinstitution). Damit stehen diese Zuordnungen wenigstens für die häufig auf aggregierten Niveaus gebildeten bibliometrischen Indikatoren zur Verfügung.

Insofern wird mit dem Projektergebnis nicht der Anspruch einer 100%-Kodierung erhoben, die lückenlos und fehlerfrei wäre, denn das ist schon aufgrund der Datenlage nicht möglich. Der erreichte Kodierungsstand bietet aber in jedem Fall eine gute und flexible Basis für zukünftige Anwendungen und die Methode ist vom Prinzip her auf die nutzerorientierte Weiterentwicklung hin angelegt. Ein Informationsaustausch mit interessierten Hochschulen und außeruniversitären Forschungseinrichtungen in Deutschland ist in Bezug auf die sie jeweils betreffenden Datensätze möglich, soweit die lizenzrechtlichen Voraussetzungen erfüllt sind. Das bietet den Vorteil der Transparenz für die betroffenen Einrichtungen und ermöglicht eine Verifizierung der Daten.

Die bisher erzielten Ergebnisse wurden bereits in diversen Projekten in Deutschland genutzt – von am Kompetenzzentrum Bibliometrie beteiligten Partnern ebenso wie darüber hinaus. So konnte z.B. mit der Generalverwaltung der Max-Planck-Gesellschaft (MPG) ein Austausch von Daten zur Qualitätssicherung initiiert werden und mit der Deutschen Forschungsgemeinschaft (DFG) eine gemeinsame Analyse der Kooperationsstrukturen deutscher Hochschulen in der Chemie durchgeführt werden.¹⁵

¹⁵ Die Ergebnisse wurden im Förderatlas Deutschland 2012 der DFG publiziert (Kap. 6: Nationale und internationale Kooperation im Spiegel bibliometrischer Daten – das Beispiel „Chemie“).
<http://dx.doi.org/10.1002/9783527664856.ch6>

**Literaturverzeichnis**

- Bornmann, L., Marx, W., & Winterhager, M. (2012). Wahrheit und Klarheit. Uneinheitliche Namen von Universitäten und ihre Folgen für die Forschungsevaluation. *Forschung & Lehre*, 19 (8), 650–651. http://www.forschung-und-lehre.de/wordpress/Archiv/2012/ful_08-2012.pdf (Stand: 2014-11-10)
- Cuxac, P., Lamirel, J.-C., & Bonvallot, V. (2013). Efficient supervised and semi-supervised approaches for affiliations disambiguation. *Scientometrics*, 97 (1), 47–58. doi:[10.1007/s11192-013-1025-5](https://doi.org/10.1007/s11192-013-1025-5) (Stand: 2014-11-10)
- De Bruin, R. E. & Moed, H. F. (1990), The unification of addresses in scientific publications. In: L. Egghe, R. Rousseau (Eds), *Informetrics 1989/90*. Elsevier Science Publishers, Amsterdam, 65–78. <http://doclib.uhasselt.be/dspace/bitstream/1942/850/1/debruin65.pdf> (Stand: 2014-11-10)
- Enserink, M. (2009), Scientific Publishing: Are You Ready to Become a Number? *Science* 323, Nr. 5922 (27 March): 1662-1664. doi:[10.1126/science.323.5922.1662](https://doi.org/10.1126/science.323.5922.1662) (Stand: 2014-11-10)
- Galvez, C. & Moya-Anegón, F. (2006), The unification of institutional addresses applying parametrized finite-state graphs (P-FSG). *Scientometrics* 69 (2), 323-345. doi:[10.1007/s11192-006-0156-3](https://doi.org/10.1007/s11192-006-0156-3) (Stand: 2014-11-10)
- Galvez, C. & Moya-Anegón, F. (2007), Standardizing formats of corporate source data. *Scientometrics*, 70 (1), 3-26. doi:[10.1007/s11192-007-0101-0](https://doi.org/10.1007/s11192-007-0101-0) (Stand: 2014-11-10)
- Hardeman, S. (2013). Organization level research in scientometrics: a plea for an explicit pragmatic approach. *Scientometrics*, 94 (3), 1175–1194. doi: [10.1007/s11192-012-0806-6](https://doi.org/10.1007/s11192-012-0806-6) (Stand: 2014-11-10)
- Huang, S., Yang, B., Yan, S., & Rousseau, R. (2014). Institution name disambiguation for research assessment. *Scientometrics*, 99 (3), 823–838. doi:[10.1007/s11192-013-1214-2](https://doi.org/10.1007/s11192-013-1214-2) (Stand: 2014-11-10)
- Moed, H., De Bruin, R. & Van Leeuwen, T. (1995), New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics*, 33 (3), 381-422. doi: [10.1007/BF02017338](https://doi.org/10.1007/BF02017338) (Stand: 2014-11-10)
- Morillo, F., Aparicio, J., González-Albo, B. und Moreno, L. (2013) Towards the Automation of Address Identification. *Scientometrics* 94 (1), 207–24. doi: [10.1007/s11192-012-0733-6](https://doi.org/10.1007/s11192-012-0733-6) (Stand: 2014-11-10)
- Praal F.E.W., Kosten M.J.F., Calero Medina C. & Visser M.S. (2013), Ranking universities: The challenge of affiliated institutes. In: Hinze, S., Lottmann A. (Eds.) *Proceedings of the 18th International conference on Science and Technology Indicators: Translational Twists and Turns: Science as a Socio-Economic Endeavour*. Berlin: IFQ. 284-289. http://www.forschungsinform.de/STI2013/download/STI_2013_Proceedings.pdf (Stand: 2014-11-10)
- Wissenschaftsrat (2008), Bericht der Steuerungsgruppe zur Pilotstudie Forschungsrating Chemie und Soziologie. Köln. <http://www.wissenschaftsrat.de/download/archiv/8453-08.pdf> (Stand: 2014-11-10)