

Präsenz und Sichtbarkeit von Meeresforschungsinstituten im World Wide Web

Eine Fallstudie zum Potential der Webometrie zur Untersuchung der internationalen Einbettung wissenschaftlicher Einrichtungen

Dissertation
zur Erlangung des Grades einer Doktorin der Philosophie (Dr. phil.)

vorgelegt an der Fakultät für Soziologie
der Universität Bielefeld

von
Tina Ruschenburg

Erstgutachter: Prof. em. Dr. Peter Weingart
Universität Bielefeld, Fakultät für Soziologie /
Institut für Wissenschafts- und Technikforschung (IWT)

Zweitgutachter: Prof. Dr. Stefan Hornbostel
Institut für Forschungsinformation und Qualitätssicherung (iFQ), Bonn
Humboldt-Universität zu Berlin, Institut für Sozialwissenschaften

Bielefeld, November 2010

Danksagung

Viele Personen und Einrichtungen haben einen wesentlichen Beitrag zu dieser Dissertation geleistet. Ihnen möchte ich herzlich danken.

Im Rahmen eines Projektes zur Globalisierung der Forschung, gefördert von der Deutschen Forschungsgemeinschaft, konnte ich die Webometrie kennenlernen und ausprobieren. Das Evangelische Studienwerk e.V. Villigst förderte meine Dissertation mit einem Stipendium und gab mir die Möglichkeit, Promotion und Familiengründung zu vereinbaren.

Meinen beiden Gutachtern Peter Weingart und Stefan Hornbostel danke ich für die gute Betreuung: Sie ließen mir Freiraum bei Auswahl und Entfaltung des Themas und gaben in kritischen Situationen wertvollen Rat.

Auf den richtigen Weg halfen diesem Dissertationsprojekt die ExpertInnen des Doctoral Forum der Konferenzen der „International Society for Scientometrics and Informetrics“ 2005 in Stockholm und 2007 in Madrid, die das Konzept und die geplante Datenerhebung kommentierten. Weitere hilfreiche Hinweise verdanke ich dem Kolloquium des Graduiertenkollegs „Auf dem Weg in die Wissensgesellschaft“ des Instituts für Wissenschafts- und Technikforschung der Universität Bielefeld sowie der Statistical Cybermetrics Research Group der University of Wolverhampton in England. Besonders hervorzuheben ist an dieser Stelle Mike Thelwall, dessen technische Beratung für mich so essentiell war wie seine persönliche Ermunterung und seine Bemühungen, die Doktoranden im Bereich Webometrie zu vernetzen.

Darüber hinaus haben Anita Engels, Philipp Mayr, Frank Meier, Andre Müller, Andreas Pöge, Holger Schwechheimer und Matthias Winterhager das Dissertationsprojekt in verschiedenen Phasen durch ihre Kommentare um wesentliche Schritte vorangebracht.

Für die Kodierung von Webseiten mit Inhalten in mir unverständlichen Sprachen danke ich Łukasz Derdziuk, Anja Grebe, Kim Holmberg, Kwang-Jin Jung, Kayvan Kousha, Tao Liu, Ismael Rafols, Hürrem Tezcan-Güntekin und Tatjana Zimenkova.

Stephanie Zuber hat mein Promotionsprojekt vom Anfang bis zum Abschluss begleitet und die Zweitkodierung der Webseiten im inhaltsanalytischen Teil übernommen. Ohne ihre sorgfältige, zuspitzende und ermutigende Kritik wäre diese Dissertation sicher ungeschrieben.

Meinen Eltern und Geschwistern danke ich für ihre unbegrenzte Unterstützung. Meinem Mann und meinen Kindern bin ich für ihre Geduld ebenso dankbar wie für die Abwechslung und Unterhaltung.

Inhaltsverzeichnis

Abbildungsverzeichnis	V
Tabellenverzeichnis	VI
Abkürzungsverzeichnis	IX
Abkürzungen und Lage der untersuchten Meeresforschungseinrichtungen	XI
Glossar	XII
1. Einleitung	1
1.1 Hintergrund.....	2
1.2 Ziel und Vorgehensweise	3
1.3 Gliederung	4
1.4 Grundlegende Begriffe und Abkürzungen	5
2. Grundlagen der webometrischen Forschung	6
2.1 Eine neue Datenquelle für die Wissenschaftsforschung: das World Wide Web	6
2.1.1 Die Entwicklung von Internet und World Wide Web.....	6
2.1.2 Grundlegende Elemente und Eigenschaften des World Wide Webs	10
2.1.3 Die Nutzung des World Wide Webs durch die Wissenschaft.....	13
2.2 Grundzüge der Webometrie	16
2.2.1 Die Entstehung der Webometrie.....	17
2.2.2 Die Webometrie im Kontrast zur Bibliometrie	21
2.3 Die webometrische Analyse wissenschaftlicher Einrichtungen.....	29
2.3.1 Funktionen der Websites von Forschungseinrichtungen	29
2.3.2 Funktionen und Aussagekraft von Links im wissenschaftlichen Kontext	31
2.3.3 Disziplinäre Unterschiede in der Gestaltung von Internetauftritten	36
2.3.4 Die Identifikation wissenschaftlicher Websites im World Wide Web	38
2.4 Erhebungsinstrumente für Linkanalysen.....	41
2.4.1 Die Verwendung von Webcrawlern zur Erhebung von Links	41
2.4.2 Die Nutzung kommerzieller Suchmaschinen zur Linksuche	43
2.4.3 Eigenheiten von Linkerhebungen mit Webcrawlern	46
2.4.4 Vor- und Nachteile von Linkerhebungen mit Suchmaschinen.....	50
2.5 Die Nutzung der Webometrie zur Evaluation.....	58
2.6 Zwischenfazit.....	62
2.6.1 Grundlegende Eigenschaften der Webometrie	62
2.6.2 Relevante Aspekte im Hinblick auf internationale Fragestellungen	63

3. Die webometrische Untersuchung internationaler Zusammenhänge in der Wissenschaft	65
3.1 Der Stellenwert von Internationalität in der Wissenschaft	67
3.1.1 Die internationale Ausrichtung der Wissenschaft.....	67
3.1.2 Die nationale Ausrichtung der Wissenschaft.....	69
3.1.3 Weiterer Forschungsbedarf	72
3.2 Die Rolle des Internets in der internationalen wissenschaftlichen Kommunikation	73
3.2.1 Der Einfluss des Internets auf die internationale Ungleichheit – zwei konträre Einschätzungen	74
3.2.2 Globale Ungleichheit im Internetzugang	76
3.2.3 Internationale Unterschiede in der Webpräsenz	80
3.2.4 Global research village oder global digital divide?	85
3.3 Die webometrische Untersuchung internationaler Vernetzung	88
3.3.1 Inlinkbasierte Internationalitätsmaße	90
3.3.2 Outlinkbasierte Internationalitätsmaße.....	95
3.3.3 Netzwerkmaße.....	97
3.3.4 Offene Fragen bei der Untersuchung internationaler Zusammenhänge im World Wide Web.....	100
3.4 Zwischenfazit.....	103
4. Leitfragen, Fallauswahl und Datenerhebung der Fallstudie zu internationalen Linkstrukturen in der Meeresforschung	105
4.1 Leitfragen für die empirische Fallstudie zur webometrischen Untersuchung internationaler Zusammenhänge.....	105
4.2 Die Auswahl der Fälle	107
4.2.1 Ein Fachgebiet mit starken internationalen Bezügen: die Meeresforschung.....	107
4.2.2 Kriterien für die Auswahl der Fälle	111
4.2.3 Auswahlverfahren	113
4.3 Datengrundlage und -erhebung.....	117
4.3.1 Struktur der Websites	117
4.3.2 Outlinkdaten.....	119
4.3.3 Inlinkdaten	121
4.3.4 Publikationsdaten	127
4.3.5 Weitere Daten.....	129
4.3.6 Exkurs: Daten zur Rezeption von Websites.....	130
4.4 Zwischenfazit.....	132
4.4.1 Schlussfolgerungen aus der Fallauswahl.....	132
4.4.2 Schlussfolgerungen aus der Linkdatenerhebung.....	132
4.4.3 Schlussfolgerungen zu Rezeptionsdaten.....	134

5. Die Ergebnisse der empirischen Fallstudie	136
5.1 Wie aussagekräftig sind Top-Level-Domains als Indikatoren der nationalen Anbindung?	136
5.1.1 Verwendete Daten	137
5.1.2 Ergebnisse.....	140
5.1.3 Zusammenfassung und Diskussion	154
5.2 Inwieweit ähneln sich die Bilder, die webometrische und bibliometrische Indikatoren von der internationalen Einbettung von Forschungseinrichtungen zeigen?	158
5.2.1 Verwendete Daten	158
5.2.2 Ergebnisse.....	161
5.2.3 Zusammenfassung und Diskussion	174
5.3 Hängen Präsenz und Vernetzung wissenschaftlicher Einrichtungen im World Wide Web bzw. im Web of Science mit dem Entwicklungsstand ihrer Sitzländer zusammen?	180
5.3.1 Verwendete Daten	180
5.3.2 Ergebnisse.....	181
5.3.3 Zusammenfassung und Diskussion	195
6. Fazit	201
6.1 Zum Potential der Webometrie zur Untersuchung der internationalen Einbettung von Forschungseinrichtungen	203
6.2 Zur Praktikabilität webometrischer Analysen	210
6.3 Ausblick	211
7. Literaturverzeichnis	213
8. Anhang	228
8.1 Abfrage der Inlinks bei Yahoo nach Top-Level-Domains	228
8.2 Datentabellen	231
8.3 Codebuch zur Codierung der nationalen Anbindung verlinkter Seiten	235

Abbildungsverzeichnis

Abbildung 1: Oberfläche des Webcrawlers SocSciBot	42
Abbildung 2: Ansicht der Ergebnisausgabe der kommerziellen Suchmaschine Live Search	45
Abbildung 3: Webometrische Standardisierungsgrößen	90
Abbildung 4: Leitfragen für die empirische Fallstudie	106
Abbildung 5: Kriterien, nach denen Outlink-URLs und -domains zu einem Ziel zusammengefasst wurden	120
Abbildung 6: Abfragen nach dem durchschnittlichen Anteil der gezeigten distinkten URLs an der von Yahoo angegebenen Gesamttrefferzahl	123
Abbildung 7: Unter der Inlinkquellen auftretende Typen von Spam	143
Abbildung 8: Fiktives Beispiel zur Unterscheidung von Inlinks, Inlinkdomains, Outlinks und Outlinkdomains	159

Tabellenverzeichnis

Tabelle 1:	Inlinkbasierte Internationalitätsmaße und ihre bisherige Anwendung in webometrischen Studien	92
Tabelle 2:	Outlinkbasierte Internationalitätsmaße und ihre bisherige Anwendung in webometrischen Studien	95
Tabelle 3:	Netzwerkmaße und ihre bisherige Anwendung in webometrischen Analysen	98
Tabelle 4:	ausgeschlossene außeruniversitäre Meeresforschungsinstitute in Deutschland, den USA und englischsprachigen Entwicklungsländern.....	114
Tabelle 5:	geeignete Meeresforschungsinstitute in Deutschland, den USA und englischsprachigen Entwicklungsländern	115
Tabelle 6:	Startseiten und Zeiten des Crawlens der Websites	117
Tabelle 7:	Von der Erhebung ausgeschlossene Webbereiche	118
Tabelle 8:	Umfang der erhobenen Websites.....	119
Tabelle 9:	Outlinkwerte nach Instituten	121
Tabelle 10:	Inlinkabfragen bei Yahoo nach Zahl der Treffer.....	122
Tabelle 11:	Inlinkabfragen bei Yahoo mit mehr als 1.000 Treffern	123
Tabelle 12:	Inlinkwerte nach Instituten.....	124
Tabelle 13:	Durchschnittlicher Anteil der Inlinkwerte an allen Yahoo-Treffern	125
Tabelle 14:	Durchschnittlicher Anteil der Inlinkwerte an der jeweils vorausgehenden Bereinigungsstufe.....	126
Tabelle 15:	Abfrage der Publikationen aus dem Web of Science 2005-2007	127
Tabelle 16:	Rohe und standardisierte Zitationen im Web of Science bis zum 1.10. 2008	128
Tabelle 17:	Anzahl der wissenschaftlichen MitarbeiterInnen an den ausgewählten Meeresforschungseinrichtungen	129
Tabelle 18:	Human Development Index 2007/08 für die Sitzländer der ausgewählten Meeresforschungseinrichtungen	130
Tabelle 19:	Auskünfte der Institute zur Website-Nutzung.....	131
Tabelle 20:	Inlink- und Outlinkstichprobe für die Inhaltsanalyse.....	137
Tabelle 21:	Übereinstimmungsquotienten nach Holsti für die Outlinkziele	138
Tabelle 22:	Übereinstimmungsquotienten nach Holsti für die Inlinkquellen	139
Tabelle 23:	Outlinkziele nach inhaltlichem und äußerlichem Länderbezug.....	141
Tabelle 24:	Inlinkquellen nach inhaltlichem und äußerlichem Länderbezug	141
Tabelle 25:	Outlinkziele und Inlinkquellen mit Spaminhalten.....	144
Tabelle 26:	Inlinkquellen mit und ohne Spaminhalt nach TLD-Typ.....	144
Tabelle 27:	Inlinkquellen mit und ohne Spaminhalt nach inhaltlichem Länderbezug	145

Tabelle 28:	Outlinkziele mit widersprüchlichen nationalen Bezügen in Seiteninhalt und Top-Level-Domain	145
Tabelle 29:	Inlinkquellen mit widersprüchlichen nationalen Bezügen in Seiteninhalt und Top-Level-Domain	146
Tabelle 30:	Outlinkziele nach inhaltlichem Länderbezug und Top-Level-Domain mit und ohne Länderhinweis	148
Tabelle 31:	Inlinkquellen nach inhaltlichem Länderbezug und Top-Level-Domain mit und ohne Länderhinweis	148
Tabelle 32:	Outlinkziele mit übernationalem inhaltlichem Bezug	150
Tabelle 33:	Inlinkquellen mit übernationalem inhaltlichem Bezug	150
Tabelle 34:	Outlinkziele mit supranationalen Top-Level-Domains	151
Tabelle 35:	Outlinkziele mit inhaltlichem Bezug zur Europäischen Union und anderen zwischenstaatlichen Organisationen und Programmen	152
Tabelle 36:	Inlinkquellen mit supranationalen Top-Level-Domains	153
Tabelle 37:	Inlinkquellen mit inhaltlichem Bezug zur Europäischen Union und zu zwischenstaatlichen Einrichtungen und Programmen	154
Tabelle 38:	Spearmans p für Maße der Präsenz im World Wide Web und im Web of Science	162
Tabelle 39:	Spearmans p für Maße der Sichtbarkeit im World Wide Web und im Web of Science	163
Tabelle 40:	Spearmans p für grundlegende Maße der internationalen Orientierung im World Wide Web und im Web of Science	164
Tabelle 41:	Spearmans p für Maße der Orientierung an Ländergruppen nach Entwicklungsstand im World Wide Web und im Web of Science	165
Tabelle 42:	Spearmans p für Maße der Orientierung an einzelnen Kontinenten im World Wide Web und im Web of Science	166
Tabelle 43:	Spearmans p für grundlegende Maße der internationalen Sichtbarkeit im World Wide Web und der internationalen Orientierung im Web of Science	167
Tabelle 44:	Spearmans p für Maße der Sichtbarkeit im World Wide Web und der Orientierung im Web of Science, bezogen auf Ländergruppen nach Entwicklungsstand	168
Tabelle 45:	Spearmans p für Maße der Sichtbarkeit im World Wide Web und der Orientierung im Web of Science, bezogenen auf einzelne Kontinente	169
Tabelle 46:	Spearmans p für grundlegende Maße der internationalen Orientierung und Sichtbarkeit im World Wide Web	170
Tabelle 47:	Spearmans p für Maße der Orientierung und Sichtbarkeit im World Wide Web, bezogen auf Ländergruppen nach Entwicklungsstand	171
Tabelle 48:	Spearmans p für Maße der Orientierung und Sichtbarkeit im World Wide Web, bezogen auf einzelne Kontinente	171
Tabelle 49:	Am häufigsten auftretende Kontinente unter den Publikationen, Outlink- und Inlinkdomains der Institute	172
Tabelle 50:	Spearmans p für den Stellenwert der Weltregionen in den Publikations-, Inlink- und Outlinkprofilen	173
Tabelle 51:	Spearmans p für HDI und standardisierte Zitationen resp. Inlinkdomains	181

Tabelle 52:	Spearman's ρ für HDI und Maße der allgemeinen Außenorientierung im World Wide Web.....	182
Tabelle 53:	Spearman's ρ für HDI und Maße der internationalen Orientierung im World Wide Web.....	183
Tabelle 54:	Spearman's ρ für HDI und Maße der Orientierung im World Wide Web an Ländergruppen nach Entwicklungsstand	183
Tabelle 55:	Spearman's ρ für HDI und Maße der Orientierung im World Wide Web nach einzelnen Kontinenten.....	184
Tabelle 57:	Spearman's ρ für HDI und Maße der allgemeinen Sichtbarkeit im World Wide Web.....	185
Tabelle 58:	Spearman's ρ für HDI und Maße der internationalen Sichtbarkeit im World Wide Web.....	186
Tabelle 59:	Spearman's ρ für HDI und Maße der Sichtbarkeit im World Wide Web bei Ländergruppen nach Entwicklungsstand	187
Tabelle 60:	Spearman's ρ für HDI und Maße der Sichtbarkeit im World Wide Web nach einzelnen Kontinenten	188
Tabelle 62:	Publikationen der Institute aus dem Jahr 2006 und ihre Abdeckung durch das Web of Science	190
Tabelle 63:	Spearman's ρ für HDI und die Publikationen resp. Zitationen der Institute im Web of Science	191
Tabelle 64:	Spearman's ρ für HDI und verschiedene Maße der internationalen Orientierung der Institute im Web of Science.....	192
Tabelle 65:	Spearman's ρ für HDI und Maße der Orientierung der Institute im Web of Science an Ländergruppen nach Entwicklungsstand	193
Tabelle 66:	Spearman's ρ für HDI und Maße der Orientierung der Institute im Web of Science an einzelnen Kontinenten.....	193
Tabelle 67:	Outlinkdomains nach TLD-Typ sowie nach nationaler und internationaler TLD.....	231
Tabelle 68:	Outlinkdomains mit internationaler TLD nach eigenem und anderen Kontinenten sowie nach Entwicklungsstatus	231
Tabelle 69:	Outlinkdomains mit internationaler TLD nach Kontinenten.....	231
Tabelle 70:	Inlinkdomains nach TLD-Typ sowie nach nationaler und internationaler TLD.....	232
Tabelle 71:	Inlinkdomains mit internationaler TLD nach eigenem und anderen Kontinenten sowie nach Entwicklungsstatus	232
Tabelle 72:	Inlinkdomains mit internationaler TLD nach Kontinenten.....	232
Tabelle 73:	Publikationen nach Koautorenschaftsstatus	233
Tabelle 74:	Internationale Kopublikationen mit dem eigenen und anderen Kontinenten, sowie mit Übergangsländern, Entwicklungs- und Industrieländern	233
Tabelle 75:	Internationale Kopublikationen nach Kontinenten.....	233
Tabelle 76:	Kopublikationen, Outlinks und Inlinks von BLOS, IFM, IOW, MBARI und NIO nach Weltregionen (die eigene Region ist jeweils grau unterlegt)	234

Abkürzungsverzeichnis

Dieses Verzeichnis enthält alle verwendeten Abkürzungen mit Ausnahme der Akronyme für die untersuchten Forschungseinrichtungen, die in einem separaten Verzeichnis zusammengestellt sind (S. XI).

API	Application Programming Interface
ARPA	Advanced Research Projects Agency
BBS	Bulletin Board System
ccTLD	country-code Top-Level-Domain
CERN	Organisation Européenne pour la Recherche Nucléaire (ursprünglich: Conseil Européen pour la Recherche Nucléaire)
CSS	Cascading Style Sheets
DNS	Domain Name System
EC	European Commission
EDV	Elektronische Datenverarbeitung
EU	Europäische Union
FAO	Food and Agriculture Organization of the United Nations
FTP	File Transfer Protocol
FuE	Forschung und Entwicklung
GCRMN	Global Coral Reef Monitoring Network
GDP	Gross Domestic Product
GLOSS	Global Sea-Level Observing System
GOOS	Global Ocean Observing System
gTLD	generic Top-Level-Domain
GUS	Gemeinschaft Unabhängiger Staaten
HDI	Human Development Index
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IANA	Internet Assigned Numbers Authority
ICANN	Internet Corporation for Assigned Names and Numbers
ICSU	International Council of Scientific Unions
IF	Journal Impact Factor
IGY	International Geophysical Year
ILQ	Inlinkquellen
IOC	Intergovernmental Oceanographic Commission
IODE	International Oceanographic Data and Information Exchange

IP	Internet Protocol
ISC	Internet Systems Consortium
ISI	Institute for Scientific Information
ISO	International Organization for Standardization
IuK	Information und Kommunikation
JOI	Joint Oceanographic Institutions
NGO	Non-governmental organisation
NSB	National Science Board
NSF	National Science Foundation
NTIA	National Telecommunications and Information Administration
ODP	Ocean Drilling Program
OECD	Organisation for Economic Co-operation and Development
OLZ	Outlinkziele
PPP	Purchasing Power Parity
RAE	Research Assessment Exercise
SCI	Science Citation Index
SCOR	Scientific Committee on Oceanic Research
SNA	Soziale Netzwerkanalyse
SPSS	Statistical Package for the Social Sciences
SSCI	Social Science Citation Index
TCP	Transmission Control Protocol
TLD	Top-Level-Domain
UN	United Nations
UNDP	United Nations Development Programme
UNEP	United Nations Environmental Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
URL	Uniform Resource Locator
UT	Unique Article Identifier
VSAT	Very Small Aperture Terminals
WIF	Web Impact Factor
WMO	World Meteorological Organization
WoS	Web of Science

Abkürzungen und Lage der untersuchten Meeresforschungseinrichtungen

Akronym	Institut	Sitzland	Kontinent	Weltregion	Ländergruppe
BLOS	Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor	USA	Nordamerika	Nördliches Amerika	Industrieland
HBOI	Harbor Branch Oceanographic Institution, Harbor Branch	USA	Nordamerika	Nördliches Amerika	Industrieland
IFM	Leibniz-Institut für Meereswissenschaften IfM-Geomar, Kiel	Deutschland	Europa	Westliches Europa	Industrieland
IOW	Institut für Ostseeforschung, Warnemünde	Deutschland	Europa	Westliches Europa	Industrieland
KMFRI	Kenya Marine and Fisheries Research Institute, Mombasa	Kenia	Afrika	Östliches Afrika	Entwicklungsland
MBARI	Monterey Bay Aquarium Research Institute, Moss Landing	USA	Nordamerika	Nördliches Amerika	Industrieland
MOI	Mauritius Oceanography Institute, Quatre Bornes	Mauritius	Afrika	Östliches Afrika	Entwicklungsland
NIO	National Institute of Oceanography, Goa	Indien	Asien	Südliches Asien	Entwicklungsland
NIOPK	National Institute of Oceanography, Karachi	Pakistan	Asien	Südliches Asien	Entwicklungsland
ORI	Oceanographic Research Institute, Durban	Südafrika	Afrika	Südliches Afrika	Entwicklungsland

Glossar

country-code vs. generic Top-Level-Domains

Country-code Top-Level Domains (ccTLDs) sind TLDs, die für eine bestimmte Nation stehen, wie z.B. „de“ für Deutschland. Dagegen stehen generische Top-Level-Domains (gTLDs) nicht für Länder, sondern für einzelne Sektoren, wie z.B. „com“ für die Wirtschaft oder „org“ für den nicht-kommerziellen Bereich.

Directory

Directories sind Unterverzeichnisse von Websites, die durch einen Schrägstrich vom Domainnamen getrennt sind (z.B. „www.uni-bielefeld.de/iwt“).

Domain

Die Domain bildet den Kern jeder URL (z.B.: „www.uni-bielefeld.de“). Sie dient der Identifikation des Servers, auf dem eine Seite abgelegt ist. Jede von ihnen besteht aus mindestens zwei Bestandteilen, die durch Punkte getrennt sind und hierarchisch angeordnet sind.

Domain Name System

Das Domain Name System (DNS) ist ein weltweiter Verzeichnisdienst, der jede Domain dem entsprechenden Server zuordnet. Dessen IP-Adresse wird gefunden, indem die Domain von der Top-Level-Domain ausgehend rückwärts aufgeschlüsselt wird.

Homepage

Wenngleich dieser Begriff in der deutschen Umgangssprache als Synonym für „Website“ eingesetzt wird, bezieht er sich ausschließlich auf die Startseiten von Internetauftritten.

Inlinks vs. Outlinks

Links, die auf einen definierten Webbereich verweisen, werden als Inlinks bezeichnet, während die davon ausgehenden Verweise Outlinks genannt werden. Jeder Link ist gleichzeitig In- und Outlink, je nachdem, ob man ihn aus der Perspektive seines Ziels oder seiner Quelle betrachtet. Diese Unterscheidung entspricht der von Zitationen und Referenzen in der Bibliometrie.

interne vs. externe Links

In der Webometrie werden interne und externe Links unterschieden: Erstere sind Verweise innerhalb eines definierten Webbereichs, bei letzteren liegen Quelle oder Ziel außerhalb des definierten Bereichs. Webometrische Untersuchungen sind in der Regel auf externe Links beschränkt, da interne Links primär Navigationszwecken dienen und keine Informationen über die äußere Vernetzung bieten.

Klient

Ein Computer, der über das Internet oder ein anderes Netzwerk die Dienste eines Servers in Anspruch nimmt, wird als Klient bezeichnet.

Link

Ein Link ist virtueller Querverweis. Er ist in eine Webseite eingebettet (die Linkquelle) und verweist auf ein anderes Webdokument (das Linkziel), das durch Anklicken des Links aufgerufen werden kann. Die Begriffe Link, Hyperlink, Verweis und Verknüpfung werden im Folgenden synonym benutzt.

Linkquelle / Linkquelle

Die Quelle eines Links ist die Webseite, in die er eingebettet ist. Das Ziel eines Links ist die Webseite, auf die er verweist.

Outlinks

> siehe Inlinks

Second-Level-Domain

In vielen Ländern gibt es zusätzlich zur landesspezifischen TLD eine Second-Level-Domain, die Aufschluss über die sektorale Zugehörigkeit gibt, z.B. „.gov.uk“ für Einrichtungen der Regierung in Großbritannien.

Server

Ein Server ist ein System aus Computer und Software, das einen Dienst anbietet (z.B. das Abrufen einer Website) und in der Regel dauerhaft zum Datenaustausch mit anderen Computern (Klienten) bereit steht.

Subdomain

Subdomains sind Unterverzeichnisse von Websites, die dem Domain-Name-System gemäß hierarchisch in die Domain eingefügt werden. So besitzt die Bielefelder Universitätsbibliothek mit www.ub.uni-bielefeld.de eine eigene Subdomain.

Suchmaschine

Suchmaschinen indexieren Teile des World Wide Webs und bieten die Möglichkeit, die erfassten Dokumente nach Stichworten zu durchsuchen. Da einige Suchmaschinen auch die Suche nach Inlinks erlauben, werden sie als Erhebungsinstrument für Linkanalysen genutzt. Die Ergebnisse werden in Form einer nach Relevanz geordneten Liste von URLs ausgegeben. Die am häufigsten genutzten Suchmaschinen wie Google, Yahoo oder Windows Live werden von Unternehmen betrieben.

Top-Level-Domain

Die Top-Level-Domain (TLD) ist der letzte Teil einer Domain (z.B.: „.de“), der in der Hierarchie des Domain-Name-Systems an oberster Stelle steht.

URL

Eine URL, kurz für Uniform Resource Locator, ist die Adresse, mit deren Hilfe der Speicherort eines einzelnen Dokuments im World Wide Web lokalisiert werden kann. So kann das Impressum des IWT unter der URL <http://www.uni-bielefeld.de/iwt/Impressum.html> aufgerufen werden.

Webbereich

Ein Webbereich ist ein beliebiger Teil des World Wide Webs, der üblicherweise anhand von URLs definiert wird. Hierbei kann es sich beispielsweise um eine einzelne Seite, eine Website oder alle Websites innerhalb einer Top-Level-Domain handeln.

Webcrawler

Webcrawler sind Programme, die einen definierten Bereich des World Wide Webs absuchen und die darin enthaltenen URLs, Links, Emailadressen, Text- oder sonstigen Inhalte dokumentieren. Dazu gehen sie von vorgegebenen Startseiten aus, identifizieren die darin enthaltenen Links, rufen deren Zielseiten auf und wiederholen diese Prozedur, bis jedes Linkziel für den angegebenen Webbereich überprüft wurde. Webcrawler werden u.a. als Erhebungsinstrument in der Webometrie eingesetzt, aber von Suchmaschinen zum Aufbau von Datenbanken verwendet.

Webdokument

Ein Webdokument ist ein Dokument, das im World Wide Web unter einer eigenen URL abgelegt ist und das mit einem beliebigen Programm geöffnet werden kann. Dieser Begriff umfasst neben Webseiten beispielsweise pdf-, Excel- oder Word-Dateien.

Webseite

Eine Webseite ist ein Dokument im World Wide Web, das unter einer eigenen URL abgelegt ist und sich mit Hilfe eines Webbrowsers öffnen lässt.

Website

Dieser etwas diffuse Begriff definiert eine zusammenhängende Sammlung einzelner Webseiten, die über eine gemeinsame Domain definiert wird. Häufig repräsentiert eine Website eine bestimmte Person oder Organisation. Als Synonyme werden „Internetauftritt“ und „Webpräsenz“ verwendet.

Webometrie

Als Webometrie wird die quantitative Analyse kommunikativer Strukturen anhand des World Wide Webs bezeichnet. Sie ist international unter dem englischen Namen *webometrics* bekannt, der die Nähe zu anderen quantitativen Methoden der Wissenschaftsforschung wie *scientometrics* und *bibliometrics* ausdrückt. Eine genaue Definition enthält Kapitel .

World Wide Web

Das World Wide Web ist ein Informationssystem, das auf dem Prinzip des Hypertext basiert, also der direkten Verknüpfung verwandter Informationseinheiten durch Links. Es wird über das Internet genutzt, mit dem es – anders als es die deutsche Umgangssprache suggeriert – nicht identisch ist. Eine detailliertere Beschreibung enthält Kapitel

1. Einleitung

Universalismus, Kommunalismus, Uneigennützigkeit und organisierter Skeptizismus sind die vier zentralen Normen des wissenschaftlichen Ethos', die Robert K. Merton (1972) nennt. Das World Wide Web wurde zu dem Zweck entwickelt, Information sinnvoll zu organisieren, und baut auf Prinzipien auf, die sehr gut mit diesen Grundwerten der Wissenschaft harmonieren: Als offenes, nicht-kommerzielles und globales Medium steht es grundsätzlich allen Menschen zur Speicherung, Vernetzung und Suche beliebiger Inhalte zur Verfügung. Eine institutionalisierte Kontrolle der Inhalte im Sinne des organisierten Skeptizismus' gibt es hier zwar nicht, sie lässt sich jedoch bei Bedarf über die Gemeinschaft der WebnutzerInnen herstellen. So bietet das Web eine einzigartige technische Grundlage für kollektive Projekte wie die Online-Enzyklopädie Wikipedia oder die Entwicklung verschiedenster Open-Source-Software. Sie folgen den eingangs genannten Prinzipien, die für die Wissenschaft gelten.

Angesichts dieser günstigen Grundvoraussetzungen ist es nicht verwunderlich, dass sich die Nutzung des Webs durch die Wissenschaft seit seiner öffentlichen Einführung im Jahr 1993 rasant ausgeweitet hat. Seine Einsatzmöglichkeiten sind äußerst vielfältig, so dass es als Mittel zur Selbstdarstellung, Publikation, Recherche, Datenablage und Vernetzung in Zeiten von elektronischen Zeitschriften, Open Access, digitalen *repositories* und Preprintarchiven nicht mehr aus dem Wissenschaftsbetrieb wegzudenken ist. Indirekt wird diese Entwicklung auch durch die inzwischen allgegenwärtige, wenngleich oft diffuse Forderung der Forschungspolitik nach mehr Internationalität weiter vorangetrieben.

Indem das World Wide Web ein beliebtes Werkzeug der Wissenschaft wurde, ist es auch zu einer neuen Datengrundlage für die Wissenschaftsforschung geworden. Eines ihrer zentralen Themen ist von jeher die Messung, Abbildung und Bewertung wissenschaftlicher Produktivität, die klassischerweise mit Hilfe bibliometrischer Verfahren anhand von Publikationen und deren Zitationen erhoben wird. Mit der Webometrie existiert nun eine weitere, eigenständige Methode zur quantitativen Analyse kommunikativer Strukturen, deren Datenquelle das Web ist. Sie ist nicht auf die Wissenschaft begrenzt, wurde jedoch bisher überwiegend in diesem Bereich empirisch eingesetzt. Wie bei anderen Verfahren auch, ist jedoch sorgfältig zu prüfen, welche Fragestellungen mit ihrer Hilfe überhaupt erforscht werden können und welche Anforderungen an die webometrischen Daten und Erhebungsinstrumente zu stellen sind.

1.1 Hintergrund

Das Internet wurde erst spät als Forschungsgegenstand von den Sozialwissenschaften angenommen (DiMaggio et al. 2001). Die Frage, inwiefern das World Wide Web Aufschluss über wissenschaftliche Strukturen geben kann, wird seit Mitte der 1990er Jahre diskutiert. Um die Jahrtausendwende entwickelte sich die Webometrie zur eigenständigen Methode, die zunächst stark mit der Bibliometrie verglichen wurde und insbesondere in der quantitativen Wissenschaftsforschung sowie in der Bibliotheks- und Informationswissenschaft rasant an Popularität gewann. Zum vorläufigen Schwerpunkt wurde die Linkanalyse, die virtuelle Verknüpfungen zwischen verschiedenen Homepages, Websites oder nationalen Domains untersucht (Thelwall 2004a). Sie steht auch im Mittelpunkt dieser Dissertation.

Die Webometrie wurde bisher primär für Studien innerhalb einzelner Nationen eingesetzt. Aus verschiedenen Gründen liegt es jedoch nahe, sie gerade für internationale Fragestellungen einzusetzen: Im Unterschied zu früheren Datenbanken erlaubt das Web die Verknüpfung von Inhalten unabhängig von der räumlichen Nähe. Ob eine gemeinsame Website von KollegInnen aus dem gleichen Büro oder auf verschiedenen Kontinenten eingerichtet wird, spielt keine Rolle. Anders als bei den „alten Medien“ wie Post, Telefon oder Fax hat die Entfernung bei den neuen Medien auch keine Auswirkung auf die Dauer oder Kosten des Informationstransfers. Diese Orientierung an der globalen Dimension haben bereits seine Urheber im Namen World Wide Web zum Ausdruck gebracht. Aufgrund seiner räumlichen Durchlässigkeit bildet das Web internationale Kontakte möglicherweise besonders gut ab.

In diesem Zusammenhang wurde die Vermutung geäußert, dass dieser Effekt speziell Entwicklungsländern zugute kommt. Dafür spricht, dass viele von ihnen weit von den wissenschaftlichen und wirtschaftlichen Zentren in Nordamerika und Europa entfernt sind. Das World Wide Web könnte dortigen ForscherInnen die Chance bieten, weltweiten Zugang zu Informationen zu bekommen, ihre Arbeit einer globalen Öffentlichkeit zu präsentieren und sich international besser zu vernetzen.

Aguillo et al. (2006) postulieren daher die Überlegenheit webometrischer Verfahren über bibliometrische für Studien, die Entwicklungsländer einschließen:

„Some of the shortcomings of the traditional bibliometric measurements noted in different studies can be targeted by the cybermetric indicators. Webometrics could be useful to evaluate Third World universities with less financial resources and less access to mainstream scientific publications“ (Aguillo et al. 2006: 1299).

Diese These beinhaltet zwei Aspekte, die es zu trennen gilt: Zum einen nehmen die Autoren an, dass ForscherInnen aus peripheren Regionen bei begrenzten Forschungsmitteln leichter

Zugang zum World Wide Web bekommen als zu den Fachzeitschriften des wissenschaftlichen Mainstreams. Zum anderen vermuten sie, dass die strukturelle Einbindung dieser WissenschaftlerInnen sich mit Hilfe der Webometrie besser abbilden lässt als mit der Bibliometrie. Hintergrund ist, dass bibliometrische Studien häufig auf den Publikationsdaten des Science Citation Index' basieren, der in den letzten Jahrzehnten dafür kritisiert wurde, einen Bias zugunsten der USA und anderer hochentwickelter Forschungsnationen zu haben.

Belege für beide Teilaspekte der These bleiben Aguillo et al. (2006) jedoch schuldig, so dass hier Klärungsbedarf besteht. Zudem gibt es einige Unwägbarkeiten, die vorwiegend technischer Art sind: Die Qualität der kommerziellen Suchmaschinen, die für webometrische Zwecke genutzt werden, und ihre internationale Abdeckung des World Wide Webs sind nach wie vor Unsicherheitsfaktoren. Es stellt sich zudem die Frage nach geeigneten Indikatoren für nationale Zugehörigkeit.

1.2 Ziel und Vorgehensweise

Die vorliegende Dissertation beschäftigt sich mit dem Beitrag, den die Webometrie speziell zur Untersuchung internationaler Fragestellungen in der Wissenschaftsforschung leisten kann. Konkret geht es darum, die besonderen Stärken, aber auch aktuellen Schwachpunkte der Webometrie im Hinblick auf internationale Erhebungsdesigns aufzuzeigen.

Universitäten und außeruniversitäre Forschungseinrichtungen bilden den organisationalen Kern der Wissenschaft. Daher stehen sie in dieser Arbeit wie auch in zahlreichen anderen webometrischen Studien im empirischen Fokus. Die Forschungsfrage lautet entsprechend:

Welches Potential besitzt die Webometrie zur Untersuchung der internationalen Einbettung wissenschaftlicher Einrichtungen?

Zur empirischen Klärung dieser Forschungsfrage werden auf der Basis von Literatur zur Webometrie sowie zur internationalen Kooperation und Kollaboration in der Wissenschaft drei besonders relevante Leitfragen herausgearbeitet. In einer Fallstudie werden sie an einem weltweit verteilten Sample von Meeresforschungseinrichtungen geprüft.

Die erste Leitfrage bezieht sich auf mögliche Nationalitätsindikatoren für Linkanalysen. Will man beispielsweise die internationale Einbettung einer Website untersuchen, ist es essentiell zu wissen, in welche Länder sich die darin enthaltenen Links richten oder aus welchen Ländern Links auf diese Website verweisen. Das einzige formale Merkmal von Links bzw. von Webadressen im Allgemeinen, das einen Hinweis auf ihre nationale Zuordnung enthält, sind die sogenannten länderspezifischen Top-Level-Domains. Ihre Aussagekraft ist jedoch ungewiss. Um sie zu prüfen, werden inhaltsanalytische Verfahren eingesetzt. Da es abgesehen von den Top-Level-Domains keine Indikatoren der Nationalität im World Wide Web gibt,

die mit geringem Aufwand zu erschließen sind, hängt der Möglichkeitsraum der Webometrie in Bezug auf internationale Untersuchungen wesentlich von dieser Frage ab.

Zweitens werden die internationalen webometrischen und bibliometrischen Profile der Forschungseinrichtungen verglichen. Dies geschieht vor dem Hintergrund, dass sich die Entwicklung der Webometrie ursprünglich stark an der Bibliometrie orientierte. Zudem wurden in mehreren Studien starke Korrelationen zwischen der Präsenz oder Sichtbarkeit von Forschungseinrichtungen im World Wide Web und ihrer Forschungsleistung – meist gemessen an der Präsenz im Web of Science, Universitätsrankings oder dem britischen Research Assessment Exercise – gefunden. Hier ist kritisch zu hinterfragen, inwiefern die Gestaltung und Einbettung der Institutswebsites tatsächlich in einem direkten Zusammenhang zum wissenschaftlichen Output stehen. Daher erscheint es sinnvoll, im Detail zu prüfen, ob sich die internationale Orientierung von Forschungseinrichtungen in den Publikationsprofilen (d.h. primär in Koautorenschaftsmustern) der internationalen Einbettung ihrer Website (d.h. vorwiegend im Hinblick auf Linkmuster) entspricht. Für die ausgewählten Meeresforschungsinstitute wird dazu der Zusammenhang zwischen verschiedenen, komplementären web- und publikationsbasierten Indikatoren der internationalen Einbettung untersucht.

Die dritte Leitfrage fokussiert die globale Ungleichheit. Hier geht es insbesondere darum, ob der Umfang und die Vernetzung der Internetauftritte wissenschaftlicher Einrichtungen mit dem Entwicklungsstand ihrer Sitzländer zusammenhängen. In diesem Rahmen wird auch die in Abschnitt 1.1 erwähnte These geprüft, dass webometrische Maße möglicherweise ein angemesseneres Bild von Forschung in Entwicklungsländern geben als bibliometrische. Diese Leitfrage wird operationalisiert, indem verschiedene web- und publikationsbasierte Maße mit dem Human Development Index der Vereinten Nationen – einem Indikator der allgemeinen sozio-ökonomischen Entwicklung einzelner Staaten – korreliert werden.

In Kombination geben diese Leitfragen somit Aufschluss über drei grundlegende, bisher jedoch ungeklärte Aspekte, die besondere Relevanz für die Einsatzmöglichkeiten webometrischer Verfahren und die Interpretation von Webdaten besitzen.

1.3 Gliederung

Im Folgenden wird zunächst der historische Hintergrund der Webometrie und aktuelle Stand der Forschung zusammengetragen (Kapitel 2). Neben der Entwicklung des World Wide Webs (2.1) werden die Grundlagen der Webometrie vorgestellt (2.2). Ferner enthält dieser Abschnitt eine Übersicht über andere webometrische Studien zu Universitäten und außer-

universitären Forschungseinrichtungen (2.3), eine kritische Betrachtung der Erhebungsinstrumente (2.4) sowie einen Exkurs zum Thema Webometrie und Evaluation (2.5).

Ein weiteres Kapitel basiert auf Literatur zur Untersuchung internationaler Zusammenhänge in der Wissenschaft (Kapitel 3). Hier wird sowohl die Bedeutung von Internationalität in der Wissenschaft herausgearbeitet (3.1) als auch die spezielle Rolle des Internet in der internationalen wissenschaftlichen Kommunikation beleuchtet (3.2). Außerdem enthält dieses Kapitel eine Übersicht über die bisherige empirische webometrische Untersuchung der grenzüberschreitenden Vernetzung in der Wissenschaft (3.3).

Auf der Grundlage der beiden vorangegangenen Kapitel werden in Kapitel 4 zunächst drei Leitfragen für die empirische Fallstudie formuliert (4.1), um anschließend ein geeignetes Fachgebiet zu bestimmen und eine Auswahl von Forschungseinrichtungen zu treffen (4.2). In diesem Teil wird außerdem die Erhebung sämtlicher Daten dokumentiert (4.3).

Kapitel 5 beinhaltet die empirischen Ergebnisse der Fallstudie für die drei Leitfragen. Zunächst geht es um die Aussagekraft von Top-Level-Domains als Indikatoren der nationalen Anbindung (5.1). Im nächsten Abschnitt werden die Bilder verglichen, die webometrische und bibliometrische Indikatoren von der internationalen Einbettung der untersuchten Forschungseinrichtungen zeichnen (5.2). Im letzten Abschnitt steht die Frage im Mittelpunkt, inwiefern Umfang und Vernetzung der Internetauftritte mit dem Entwicklungsstand des Landes zusammenhängen, in dem die Forschungseinrichtungen angesiedelt sind (5.3).

Das abschließende Fazit bündelt und reflektiert die zentralen Ergebnisse dieser Dissertation (Kapitel 6). Der Fragestellung entsprechend wird ein Überblick über die Möglichkeiten und Grenzen der Webometrie bei der Untersuchung der internationalen Einbettung von Forschungseinrichtungen gegeben (6.1). Als weiterer wesentlicher Aspekt zur Bewertung dieser Methode wird der Aufwand bewertet, der erforderlich ist, um hochwertige Daten zu gewinnen (6.2). Ein Ausblick zeigt zukünftigen Forschungsbedarf auf (6.3).

1.4 Grundlegende Begriffe und Abkürzungen

Die Kenntnis einiger technischer Begriffe ist für das Verständnis der vorliegenden Dissertation unerlässlich. Sie werden jeweils im Verwendungskontext erläutert. Zusätzlich bietet das Glossar (S. XII) eine Übersicht über die Fachbegriffe, das Abkürzungsverzeichnis (S. IX) über die verwendeten Akronyme. Daneben gibt es ein spezielles Verzeichnis der Akronyme der untersuchten Forschungseinrichtungen, das auch Angaben zu ihrem Sitzland und Kontinent enthält (S. XI).

2. Grundlagen der webometrischen Forschung

In diesem Kapitel wird die webometrische Methode vorgestellt. Die Nachvollziehbarkeit von Anlage und Interpretation der später vorgestellten empirischen Fallstudie setzt neben der Kenntnis elementarer Prinzipien der Webometrie auch etwas technisches Grundwissen zu Internet und World Wide Web voraus. In diesem Kapitel wird in beides auf Grundlage eines kurzen historischen Abrisses eingeführt, um abstrakte Definitionen soweit wie möglich zu vermeiden.

Zunächst werden relevante Aspekte des World Wide Webs (2.1) und der Webometrie (2.2) dargestellt, wobei die webometrische Linkanalyse mit ihrem ursprünglichen Vorbild – der bereits weithin etablierten Bibliometrie – verglichen wird. Es folgt eine Auswertung des aktuellen Forschungsstandes zu den Funktionen von Internetauftritten wissenschaftlicher Einrichtungen und den darin enthaltenen Links unter Berücksichtigung internationaler und disziplinärer Unterschiede (2.3). Anschließend werden die beiden gängigen Instrumente zur Erhebung von Linkdaten – kommerzielle Suchmaschinen und Webcrawlerprogramme – vorgestellt und auf ihre spezifischen Vorzüge und Limitierungen hin untersucht (2.4), bevor ein erstes Zwischenfazit gezogen wird (2.5), das neben grundlegenden Eigenschaften der Linkanalyse auch ihr Potential für die evaluative Nutzung und für die Untersuchung internationaler Fragestellungen beleuchtet.

2.1 Eine neue Datenquelle für die Wissenschaftsforschung: das World Wide Web

Das Internet und die verschiedenen Dienste, die es umfasst (wie E-Mail, das World Wide Web, Newsgroups, Mailinglisten etc.), erscheinen heute in vielen gesellschaftlichen Bereichen als unentbehrlich. In der Wissenschaft hatte die Verwendung des Internets bereits eine lange Tradition, insbesondere in den USA, bevor sich seine kommerzielle, politische, administrative und private Nutzung allgemein verbreitete. Das World Wide Web wurde eigens für Forschungszwecke konzipiert, hat inzwischen jedoch maßgeblich zur allgemeinen Popularität des Internets beigetragen.

2.1.1 Die Entwicklung von Internet und World Wide Web

Ursprünglich ging die Entwicklung des Internets auf eine Initiative der US-Luftwaffe in den 1960er Jahren zurück. Vor dem Hintergrund des Kalten Krieges und möglicher Atomangriffe wünschte sie sich ein stabiles und sicheres Kommunikationssystem. Ein dezentrales Rechnernetzwerk sollte auch bei Ausfall einzelner Knoten die Übertragung von Daten über weite Entfernungen gewährleisten. 1969 begann die Advanced Research Projects Agency (ARPA) des US-Verteidigungsministeriums die Vernetzung von zunächst vier akademischen

Forschungseinrichtungen in Kalifornien und Utah zu fördern. In dem so entstandenen Internetvorläufer ARPANET wurden Daten über Telefonleitungen ausgetauscht. Zu den früh verfügbaren Diensten gehörten Telnet¹, FTP² sowie E-Mail, die in den Jahren 1971 bzw. 1972 entwickelt wurden. Die Zahl der Knoten im ARPANET stieg exponentiell: 1971 waren es 15, 1977 gab es 111 und 1984 bereits 1000 vernetzte Computer. Die National Science Foundation (NSF) gründete 1984 das NSFNET, um den US-amerikanischen Forschungseinrichtungen die Vernetzung durch ein Backbone – also eine besonders leistungsstarke Hauptleitung zur Anbindung großer Rechenzentren – zu ermöglichen. Nachdem der militärische Teil des ARPANET in ein separates, geschlossenes Netzwerk namens MILNET ausgegliedert worden war und immer mehr Datenverkehr über das NSFNET lief, wurde der Betrieb des technisch veralteten ARPANET 1990 eingestellt. Etwa zur selben Zeit erhielten mit den Universitäten in Dortmund und Karlsruhe die ersten deutschen Forschungseinrichtungen Zugang zum NSF-Backbone (vgl. Koll 2004: 6ff; Castells 2005: 20ff).

Die Basis für die elektronische Übermittlung von Daten im Internet ist seit Anfang der 1980er Jahre das Transmission Control Protocol (TCP), das die Kommunikation zwischen Servern regelt. Um vernetzte Rechner identifizieren und erreichen zu können, werden sogenannte IP-Adressen vergeben, die zumeist aus von drei Punkten unterbrochenen Zahlenkombinationen bestehen.³ Seit 1984 erleichtert das Domain Name System (DNS) die Adressierung, indem es ermöglicht, den numerischen IP-Adressen einprägsamere Domainnamen zuzuordnen (z.B. www.uni-bielefeld.de an Stelle der IP-Adresse 129.70.240.4). Es ist hierarchisch organisiert, wobei die Top-Level-Domains (TLD) die oberste Ebene darstellen. Unter ihnen sind landesspezifische TLDs (ccTLDs für country-code Top-Level-Domain) wie „.de“ und allgemeine TLDs (gTLDs für generic Top-Level-Domain) wie „.com“ zu unterscheiden. Letztere werden im weiteren Verlauf dieser Dissertation auch als sektorale Top-Level-Domains bezeichnet, um die ihnen zugedachte Funktion zur Adressierung einzelner gesellschaftlicher Sektoren zu unterstreichen. Das DNS ist nach Bedarf erweiterbar und wird dezentral verwaltet. Um die Eindeutigkeit von Domainnamen zu gewährleisten, ist ihre Registrierung notwendig. Die Verwaltung von Internetadressen wurde von Beginn an weltweit von der Internet Assigned Numbers Authority (IANA) koordiniert, die inzwischen in die Non-Profit-Organisation Internet Corporation for Assigned Names and Numbers (ICANN) integriert wurde. ICANN ist insofern nicht unumstritten, als sie dem US-Wirtschaftsministerium unter-

¹ Telnet ist ein Protokoll, mit dessen Hilfe man sich über das Internet auf anderen Rechnern einloggen und Programme ausführen kann.

² Das File Transfer Protocol (FTP) ermöglicht die Übertragung von Dateien, also z.B. das Herauf- und Herunterladen von einem Server.

³ IP steht für das „Internet Protocol“, das die Vergabe, Interpretation und Auffindung von Rechneradressen regelt. Heute wird zusätzlich zu dem beschriebenen (dem sog. IPv4) ein weiteres, komplexer strukturiertes Internetprotokoll (das sog. IPv6) verwendet, um der enormen Nachfrage nach Internetadressen nachkommen zu können.

stellt ist (vgl. Wass 2003b; Castells 2005: 43; Hofmann 2005). Die Vergabe von Adressen innerhalb der einzelnen landesspezifischen TLDs erfolgt jedoch durch Einrichtungen in den jeweiligen Nationen, die die Angelegenheit unterschiedlich restriktiv handhaben (vgl. Wass 2003b).

Im Vergleich zu anderen Diensten des Internets blickt das World Wide Web nur auf eine recht kurze Geschichte zurück. Tim Berners-Lee beschrieb erstmals 1989 seine Idee eines internetbasierten Hypertextsystems (vgl. Berners-Lee 1989). Aufhänger für seine Skizze war das Informationsmanagement am europäischen Teilchenbeschleuniger CERN, an dem er zu dieser Zeit beschäftigt war. Dort wurden große Mengen von Daten und Informationen generiert und gespeichert, wobei auch verschiedene Internetanwendungen genutzt wurden. Die Informationssuche gestaltete sich – insbesondere für nur vorübergehend am CERN tätige GastforscherInnen – voraussetzungsvoll, aufwendig und nicht selten erfolglos. Als Gründe dafür identifiziert Berners-Lee die Vielfalt an Speichermedien und Datenzugängen sowie die Ablagesystematik (vgl. Berners-Lee 1989).

Die damaligen elektronischen Informationssysteme wiesen typischerweise eine hierarchische Baumstruktur mit Kapiteln und Unterkapiteln auf, die zwar Querverweise, jedoch keine direkten Verbindungen enthielten. Verwandte Dokumente, die einem anderen Zweig zugeordnet waren, ließen sich nicht direkt, sondern nur durch erneutes Abwandern des ganzen Hierarchiepfades erreichen:

„Many systems are organised hierarchically. [...] A tree has the practical advantage of giving every node a unique name. However, it does not allow the system to model the real world. For example, in a hierachical HELP system [...], one often gets to a leaf on a tree [...] only to find a reference to another leaf [...] and it is necessary to leave the system and re-enter it. What was needed was a link from one node to another, because in this case the information was not naturally organised into a tree“ (Berners-Lee 1989: o.S.).

Auch Schlagworte können Berners-Lee zufolge nur denen weiterhelfen, die bereits gute Kenntnisse über das jeweilige Fachgebiet besitzen. Seiner Ansicht nach erlaubt die Lagerung von Informationen in einem Netzwerk eine einfachere und zugleich authentischere Abbildung komplexer Zusammenhänge (vgl. Berners-Lee 1989).

Ein weiterer Kritikpunkt lag in der Vielfalt der genutzten internetbasierten Dienste, die sich jeweils eigener Kommandosprachen bedienten. Hinzu kam, dass es am CERN eine Pluralität von Betriebssystemen gab und UNIX-, Macintosh-, Dumb-Rechner und PCs sich auch im Hinblick auf die Zugangsoberfläche der Informationssysteme unterschieden. Folglich erforderte die Informationssuche ein hohes Maß an Systemkenntnissen:

„When data is available on the net, the average person is not privy to it, but must consult a ‚guru‘ who understands the ins and outs of anonymous FTP, telnet, stty, and the command systems of the various information servers (Berners-Lee/Cailliau/Groff 1992: 454).

Deshalb schlug Berners-Lee die Entwicklung eines einheitlichen, dezentralen und hypertext-basierten Informationssystems für das CERN vor – zunächst unter dem Arbeitstitel *mesh* (Berners-Lee 1989; zu deutsch: Netz), später unter dem allseits bekannten Namen World Wide Web (Berners-Lee/Cailliau 1990).

1990 stellte er gemeinsam mit seinem Kollegen Robert Cailliau einen Projektantrag, um einen internen Prototyp des Webs samt der nötigen Software entwickeln zu können (Berners-Lee/Cailliau 1990). Ihr Ziel war es, Zusammenarbeit und Informationsaustausch im CERN durch die Integration der nebeneinander bestehenden Datensysteme und durch die Vernetzung von Dokumenten zu erleichtern:

„The current incompatibilities of the platforms and tools make it impossible to access existing information through a common interface, leading to waste of time, frustration and obsolete answers to simple data lookup. There is a potential large benefit from the integration of a variety of systems in a way which allows a user to follow links pointing from one piece of information to another one. This forming of a web of information nodes rather than a hierarchical tree or an ordered list is the basic concept behind HyperText“ (Berners-Lee/Cailliau 1990: o.S.).

Das 1994 gegründete W3 Consortium verfolgte ähnliche Ziele mit einer offenen, über das CERN hinausgehenden Perspektive: Das World Wide Web sollte generell die Abfrage sowie die Ablage von Informationen dadurch erleichtern, dass bestehende Benutzeroberflächen, Datenformate und Protokolle vereinheitlicht und die Einbindung von Informationen vereinfacht würden:

„The aims of the W3 initiative are twofold: firstly to make a single, easy, user-interface to all types of information so that all may access it, and secondly to make it so easy to add new information that the quantity and quality of online information will both increase. [...] if we can solve the problems of heterogeneity of platform, data format and access protocol the resulting universe of knowledge will considerably enhance our working together“ (Berners-Lee/Cailliau/Groff 1992: 454).

Heute gestaltet sich der Zugang zum Web tatsächlich äußerst einfach. Zwar nimmt die Gestaltung von Websites aufgrund der Weiter- und Neuentwicklung von Software immer komplexere Formen an, und die erfolgreiche Suche nach Informationen erfordert wegen der immensen Menge an Webdokumenten einschließlich Spamseiten etwas Erfahrung. Dennoch können selbst AnfängerInnen grundlegende Anwendungen gut bewältigen. Zu dieser Vereinfachung haben verschiedene konzeptionelle und technologische Elemente des Webs beigetragen.

2.1.2 Grundlegende Elemente und Eigenschaften des World Wide Webs

Als Berners-Lee (1989) erstmals seine Grundidee des World Wide Webs beschrieb, gab es bereits einige Hypertextsysteme. Sie beruhten jedoch meist auf einer einzigen geschlossenen Datenbank. Berners-Lee strebte ein verteiltes System mit beliebigen Verknüpfungs-, Zusammenführungs- und Einbettungsmöglichkeiten an, das die Entwicklung neuer Technologien sowie ein gewisses Maß an Standardisierung erforderte. Dieses „Wissensuniversum“ basiert auf drei Elementen: dem Prinzip des Hypertexts, der Programmiersprache HTML und der Kombination mit dem Internet.

Hypertext bietet die Möglichkeit, beliebige thematische Verbindungen durch Verweise zu markieren und so die hierarchische Organisation von Inhalten zu umgehen. Dieses Prinzip wurde bereits vor über sechzig Jahren von Vannevar Bush (1945) beschrieben, der versuchte, technologische Entwicklungen vorherzusehen, und dabei die fiktive Maschine *memex* skizzierte: Äußerlich einer Mischung aus Schreibtisch und PC ähnelnd, sollte das Gerät der individuellen Speicherung und Verknüpfung von Informationen jeglicher Art dienen. Der Begriff Hypertext wurde von Ted Nelson (1967) geprägt, dessen Konzept die Verknüpfung von elektronischen Dokumenten auf einem einzelnen Rechner vorsah. So sollte ein großer Korpus von Informationen entstehen, der nicht den üblichen Beschränkungen des Information Retrieval anhand von Schlagworten unterliegt. Berners-Lee adaptierte diese grundlegende Idee für sein geplantes Informationssystem. Von Beginn an war allerdings auch die Verwendung von Suchmaschinen, die die Suche nach Stichworten erlauben und die Ergebnisse ebenfalls in Form von Links ausgeben, als Zugangsalternative eingeplant (vgl. Berners-Lee 1989).

Um elektronische Dokumente als Hypertext vernetzen zu können, entwickelten Berners-Lee und Kollegen am CERN die Auszeichnungssprache Hypertext Markup Language (HTML). Sie ist vergleichsweise einfach zu erlernen und vielfältig einsetzbar. Grundprinzip ist die Trennung von Text und Formatierung: Anstatt Text wie in Textverarbeitungsprogrammen zu formatieren, werden hier unformatierte Texte u.a. mit kodifizierten Schrift- oder Positionsangaben ausgestattet. Seit der zweiten Hälfte der 1990er Jahre ist es üblich, diese Formatangaben in separate Dateien, sogenannte Cascading Style Sheets (CSS), auszulagern. Durch diese Aufspaltung ließen sich Daten auf Rechnern mit verschiedenen Betriebssystemen (Unix, PC, Mac, Dumb etc.) interpretieren und auf unterschiedlichen Browsern vergleichbar darstellen. Die Einbettung von Links, die auf andere Text- oder sonstige Dateien verweisen, ist leicht möglich. Heute werden neben weiterentwickelten Versionen wie XHTML auch Skriptsprachen (wie JavaScript, ASP, PHP oder Perl) sowie grafikbasierte Sprachen (wie Flash) zur Erstellung von Webseiten verwendet.

Die vermutlich größte Innovation besteht in der Verbindung des Hypertext-Prinzips mit dem Internet. Zwar war Ende der 1980er bereits hypertextbasierte Datenbanksoftware verfügbar (z.B. die Hypercard von Macintosh), sie war jedoch ausschließlich für die Nutzung auf einzelnen Rechnern oder internen Netzwerken konzipiert. Berners-Lee schuf die Voraussetzungen für ein verteiltes und über das Internet verbundenes, offenes Hypertext-Netzwerk. Grundlegend ist dabei die Trennung von Speicher- und Abrufrechnern: Während die Daten auf sogenannten Servern abgelegt werden – also auf permanent mit dem Internet verbundenen Rechnern – erfolgt der Abruf von Daten durch Programme, die temporär mit ihnen Kontakt aufnehmen und als Klienten bezeichnet werden. Die Eingabe von Anfragen geschieht über Browser-Programme, die eine schlichte, einheitliche Oberfläche bieten und ohne große Vorkenntnisse zu bedienen sind. Das eigens entwickelte Hypertext Transfer Protocol (HTTP) regelt bei jeder Anfrage die Kommunikation zwischen Klienten zum Server. Zur Identifikation der Server sowie zur Übertragung der Daten wird das bereits vorher etablierte TCP/IP in Kombination mit dem Domain Name System genutzt. Um den exakten Speicherort von Dateien identifizieren und mit Link adressieren zu können, werden Uniform Resource Locators (URLs) verwendet, die neben der Domain des Servers ggf. Unterverzeichnisse, Dateinamen und Dokumententypen enthalten.

Dem Internet Domain Survey zufolge stieg die Anzahl der mit dem Internet verbundenen Server seit 1994 nahezu exponentiell an. Mitte 2009 konnten Daten von knapp 700 Millionen Rechnern abgerufen werden (vgl. ISC 2009). Der enorme Anstieg ist vor allem der Ausbreitung des World Wide Webs geschuldet. Mehrere Eigenschaften des Webs, die z.T. aus der Kombination der beschriebenen Elemente, z.T. aus Entscheidungen seiner Urheber resultieren, tragen maßgeblich zu seinem Erfolg bei. Dies sind vor allem seine Flexibilität, Variabilität und Offenheit.

Flexibilität

Das Web lässt sich für ein breites Spektrum von Anwendungen nutzen: HTML bietet eine anpassungsfähige Grundlage, mit der sich beliebige Inhalte aufbereiten lassen und in die andere Dokumente verschiedenster Formate, passwortgeschützte Bereiche etc. integriert werden können. Diese Flexibilität kommt voll zum Tragen, weil es keinerlei verbindliche Vorgaben für die Gestaltung von Websites gibt. Vorgegeben sind lediglich das TCP/IP-System sowie die Nutzung von HTML oder einer alternativen Programmiersprache, die das Einfügen von Links erlaubt (wie xHTML, php, flash, JavaScript etc.). Kehrseite der Abwesenheit jeglicher Standards beim Design von Websites ist die daraus resultierende Unstrukturiertheit: Sowohl der Aufbau von Internetauftritten als auch von Dokumenten variieren immens. Zwar erlaubt HTML über sogenannte *meta tags* die Angabe von Dokumenttiteln, Autoren, Schlagworten etc., diese Option wird jedoch kaum genutzt (Thelwall 2000; Jepsen

et al. 2004). Dies erschwert die automatisierte, quantitative Auswertung und den Vergleich von Webdokumenten.

Variabilität

Die Flexibilität der verwendeten Programmiersprachen und die Abwesenheit formaler Vorgaben führen auch dazu, dass sich Webdokumente leicht aktualisieren und an veränderte Situationen anpassen lassen. Das Einfügen von Links, das Ändern des Inhalts sowie das Hinzufügen oder Entfernen ganzer Dokumente ist jederzeit und ohne großen Aufwand möglich. Diese besondere Wandelbarkeit war ein ausdrückliches Ziel bei Entwicklung des Webs, sie impliziert jedoch auch die Flüchtigkeit von Webstrukturen:

„The Web is a new type of environment with changes occurring constantly. The Web of today is different from the Web of yesterday and that of tomorrow“ (Bar-Ilan 2004b: 256).

Eine Archivierung früherer Versionen aktualisierter Websites gehörte nicht zum ursprünglichen Konzept. Ein Internetarchiv (unter <http://www.archive.org>, zuletzt abgerufen am 19.11.2009) versucht diese Lücke zu füllen, indem es kontinuierlich Teile des Webs erfasst und speichert, was angesichts von dessen Umfang eine unvollendete Aufgabe bleiben muss. Bereits hier zeigt sich, dass Erhebungen im Web bestenfalls Momentaufnahmen liefern können, was dem Begriff *snapshot* in der webometrischen Literatur zu einiger Beliebtheit verhilft (z.B. Haas/Grams 2000; Bar-Ilan 2001; Bharat et al. 2001; Li 2003; Thelwall et al. 2003; Bar-Ilan 2004b).

Offenheit

Das World Wide Web wurde von Berners-Lee und Kollegen als offenes Informationsnetzwerk angelegt. Der freie und kostenlose Zugang (Open Access) war von Beginn an ein essentieller Bestandteil. Die Kommunikation über das Web sollte öffentlich stattfinden. Die passwortgeschützte Lagerung von Daten im Web nahmen die Autoren in ihren ersten Projektanträgen am CERN ausdrücklich aus. Die Offenheit des Webs beinhaltet auch, dass es jeder beliebigen Person frei steht, Beiträge zu veröffentlichen (vgl. Berners-Lee 1989; Berners-Lee/Cailliau 1990). Die notwendige Browser-Software ist bis heute größtenteils gratis erhältlich, so dass die einzige Zugangsvoraussetzung in einem Rechner mit Netzanschluss besteht.

Prinzipiell ist das World Wide Web somit demokratisch angelegt. Dennoch kann faktisch nicht von einem gleichen Zugang ausgegangen werden: Soziale, politische und ökonomische Barrieren schränken die Internetnutzung ein; dies gilt insbesondere in Entwicklungsländern. Außerdem begrenzen mangelnde Sprachkenntnisse die Möglichkeit, Informationen

abzurufen. Auch ist das Veröffentlichen im Web nach wie vor voraussetzungsvoller als die rezeptive Nutzung, so dass auch technische Hürden auftreten.

Aufgrund seiner Offenheit, Flexibilität und Variabilität ist das Web in der Lage, den in der Wissenschaft so essentiellen kommunikativen Austausch umfassend zu unterstützen.

„Because science is 'many-brained', as Derek Price used to like to say, the open and accessible nature of sites and links on the Web is tailor made to suit that character" (Beaver 2001: 375).

Entsprechend rasch und vielfältig wurde dieses Medium in der wissenschaftlichen Kommunikation eingesetzt. Als Datenquelle birgt das Web daher interessante neue Perspektiven für die Wissenschaftsforschung, aber auch spezielle Herausforderungen, wie später noch detaillierter gezeigt wird (vgl. Abschnitt 2.4).

2.1.3 Die Nutzung des World Wide Webs durch die Wissenschaft

WissenschaftlerInnen waren sowohl die erste Nutzergruppe des Internets als auch des World Wide Webs. Zwar wurde der Internetvorläufer ARPANET mit militärischen Intentionen entwickelt, die ersten vernetzten Organisationen waren jedoch US-amerikanische Forschungseinrichtungen. Auch in Deutschland verfügten als erstes Universitäten über einen Zugang zum Backbone (vgl. Koll 2004). Die Idee des World Wide Webs sowie die grundlegende Technologie dazu entstanden komplett im Rahmen der Grundlagenforschung des CERN mit dem Zweck, wissenschaftliche Kooperation und Dokumentation zu erleichtern (vgl. Berners-Lee/Cailliau 1990). Die akademische Forschung hat – mit Unterstützung durch das US-Militär, die NSF und das CERN – eine Vorreiterrolle bei der Entwicklung und Verbreitung internetbasierter IuK-Technologien eingenommen, was zum einen auf das vorhandene technologische Fachwissen, zum anderen auf den enormen Bedarf an Möglichkeiten der Suche nach und des Austauschs von Informationen zurückzuführen sein dürfte.

Inzwischen werden im Web zahlreiche andere Anwendungszwecke – vor allem kommerzielle, aber auch politische, administrative und private – verfolgt, so dass die wissenschaftliche Nutzung nur noch eine Möglichkeit von vielen ist. Einer Schätzung zufolge, die auf einer Stichprobe zufällig generierter IP-Adressen beruht, waren 1999 lediglich 6 Prozent aller Internetauftritte der Wissenschaft zuzurechnen (Lawrence/Giles 1999: 107), seitdem dürfte ihr Anteil angesichts des ungebrochenen Trends zur Einrichtung neuer, vorwiegend kommerzieller und privater Websites eher noch abgenommen haben. Aufgrund der vielfältigen Nutzungsmöglichkeiten ist die Bedeutung des World Wide Webs für die Wissenschaft nach wie vor enorm: In der Forschung dienen Websites zur Präsentation von Personen, Projekten und Einrichtungen, zur Veröffentlichung von Dokumenten sowie zum öffentlichen oder internen Archivieren von Daten. Auch der Wissenschaftsjournalismus hat das Web als

spezifisches Medium entdeckt, sowohl als Kanal zur Verbreitung von Berichten (vgl. Bernhofer 2001) als auch als Informationsquelle (Dumlao/Duke 2003).

ForscherInnen verwenden das Web intensiv und auf vielfältige Weise. Zu nennen ist zunächst seine allgemeine Recherchefunktion, der jedoch aufgrund der ungewissen Verlässlichkeit der im Web enthaltenen Informationen eine gewisse Skepsis entgegengebracht wird, sofern die Informationsquelle nicht bekannt ist und als hochwertig eingeschätzt wird (vgl. Herring 2001; Zhang 2001; Liu/Huang 2005). Außerdem sind Webquellen insofern eine unsichere Zitiergrundlage, als ungewiss ist, wie lange sie noch abrufbar sind. Eine Untersuchung der zitierten Webadressen in den Jahrgängen 1997 und 2001 der renommierten „Annual Review“-Reihe ergab, dass sechs bzw. zwei Jahre nach ihrem Erscheinen zwei Drittel bzw. ein Viertel aller Webreferenzen ungültig waren. Gleichzeitig zeigte sich statistisch der folgende negative Zusammenhang: Beiträge, die Webquellen zitieren, werden selbst seltener zitiert (Brown 2004). Unter Berücksichtigung der ungewissen Güte und Langlebigkeit von Inhalten des World Wide Webs wundert es nicht, dass Referenzen auf Webquellen in Fachartikeln nach wie vor eine Seltenheit sind, sofern es sich nicht um Verweise auf Beiträge in elektronischen Zeitschriften handelt, denen sowohl geprüfte Qualität als auch Stabilität zugeschrieben werden kann (vgl. Wouters/de Vries 2004).

Über die allgemeine Recherchefunktion hinaus gibt es vielmehr eine Reihe weiterer wissenschaftsspezifischer Anwendungen: Große Datenbestände lassen sich in Online-Datenbanken lagern, so dass gezielte individuelle Abfragen des aktuellen Stands jederzeit von mehreren Standorten aus möglich sind. Sofern sie allgemein zugänglich sind, könnten sie insbesondere ForscherInnen in marginalisierten Ländern einen leichteren und umfassenderen Datenzugang bieten (vgl. Hurd 2000). Sie erleichtern das *data sharing* zwischen Forschungsinstituten erheblich, so dass sich neue Formen von Großprojekten realisieren lassen und auch in datenintensiven Disziplinen Kooperationen entstehen können, die früher kaum möglich gewesen wären (vgl. van Raan 2001; Gläser 2003). Für diese neuartige, erst durch das Internet ermöglichte Art der Zusammenarbeit, bei der räumliche Nähe keine Rolle mehr spielt, hat sich ein eigener, aus *collaboration* und *laboratory* zusammengesetzter Begriff herausgebildet: *collaboratory* (Finholt 2002).

Auch die Publikationskultur bekommt neue Impulse durch das World Wide Web: Über Bibliotheken zugängliche Literaturdatenbanken erleichtern die Suche nach Fachartikeln. E-Journals mit Peer-Review – als exklusive Ausgabe oder als Doppelung der Printversion – sind weithin etabliert und werden aufgrund ihres bequemen, zeitlich und räumlich uneingeschränkten Zugangs vom Großteil der ForscherInnen gedruckten Bänden vorgezogen (vgl. Bar-Ilan/Fink 2005). Durch Multi-Media-Technologien lassen sich u.a. Text-, Bild- und Tondateien verknüpfen, so dass selbst komplexere Phänomene anschaulich und dynamisch

darstellt werden können. Stärkere Auswirkungen auf das wissenschaftliche Publikationssystem hat vermutlich die mit der Entstehung des Webs verbundene Idee des Open Access, des freien Zugangs zu Informationen. Die in einigen naturwissenschaftlichen Disziplinen verbreitete Tradition, bereits vor dem Erscheinen von Artikeln in Fachzeitschriften Preprints zu verbreiten, hat durch die im Web mögliche Einrichtung elektronischer Archive eine immense Ausweitung erfahren. Sie bieten einen zeitnahen und kostenlosen Zugriff auf Forschungsergebnisse. Da keine Gebühren für Zeitschriftenabonnements anfallen, können Preprintforen auch ForscherInnen in Entwicklungsländern einen neuen Informationszugang eröffnen. Gleichzeitig werden Wartezeiten für die Begutachtung und den Druck der Manuskripte umgangen (vgl. Hurd 2000). Cronin (2001) befürchtet jedoch, dass die Verbreitung von Preprints den Stellenwert des Peer-Review-Verfahrens als Qualitätssicherungsmechanismus schwächen und zu einer Pluralisierung des wissenschaftlichen Publikationssystems führen könnte. Üblicherweise erfolgt jedoch zu einem späteren Zeitpunkt eine Veröffentlichung in einer referierten Fachzeitschrift. Aufgrund des zentralen Stellenwerts, den Journals für die Zuweisung von Anerkennung und Reputation spielen, ist die Annahme plausibel, dass ihr Status auch in Zukunft erhalten bleibt (vgl. van Raan 2001; Gläser 2003). Offensichtlich stellen Preprints also nur eine neue Zwischenstufe dar, bei der die unmittelbare Aktualisierbarkeit des Webs dazu genutzt wird, um neue Forschungsergebnisse schneller zirkulieren zu lassen. Diese Strategie erscheint durchaus Wirkung zu zeigen: Einer Studie zu Tagungsbänden im Bereich Informatik zufolge wirkt sich der offene Zugang zu Artikeln im Web positiv auf die Sichtbarkeit sowie die Zitationsraten der Beiträge in späteren Publikationen aus (vgl. Lawrence 2001).

Die Anwendungsmöglichkeiten des Webs oder anderer Internettechnologien für die Wissenschaft sind zahlreich, und die Mehrzahl der ForscherInnen nutzt sie in irgendeiner Form: Eine Befragung israelischer WissenschaftlerInnen von 2001 ergab, dass 92,2 Prozent von ihnen über einen Internetzugang im eigenen Büro verfügten. Entsprechend hoch war der Anteil derer, die das Web täglich (69,0%) oder wöchentlich (23,9%) nutzten. 60,2 bzw. 72,9 Prozent der Befragten verwendeten gelegentlich oder häufig E-Journals bzw. elektronische Literaturdatenbanken. Insbesondere das Web of Science hielten viele (52,3%) für unabdingbar (vgl. Bar-Ilan/Peritz/Wolman 2003: 350ff). Auch in Deutschland gehörten Internetanwendungen wie E-Mail oder das World Wide Web längst zum wissenschaftlichen Alltag, wie aus einer Erhebung unter Berliner SozialwissenschaftlerInnen aus dem Jahr 1999 deutlich wird (vgl. Eisend 2002). Für naturwissenschaftliche und technische Disziplinen dürfte dies in noch stärkerem Maße gelten.

So wie die Wissenschaft erheblichen Anteil an der Entwicklung und Verbreitung von Internet und World Wide Web hatte, so haben diese Technologien neue Optionen für die Forschung

eröffnet und einen gewissen Einfluss auf wissenschaftliche Arbeits- und Kommunikationsweisen genommen:

„[...] we are witnessing qualitative and quantitative changes in the ways scholars communicate with each other for informal conversations, for collaborating locally and over distances, for publishing and disseminating their work, and for constructing links between their work and that of others. [...] In technological terms, scholarly communication is being transformed through the use of personal and portable computers, electronic mail, word processing software, electronic publishing, digital libraries, the Internet, the World Wide Web, mobile phones, wireless networks, and other information technologies“ (Borgman/Furner 2002: 3f).

Die inzwischen langjährige und intensive Nutzung des World Wide Webs durch die Wissenschaft macht es zu einem „obvious research field for bibliometrics, scientometrics and informetrics“ (Björneborn/Ingwersen 2004: 1217). Entsprechend hat sich die Vermessung der Forschung inzwischen unter dem Namen Webometrie auf die vielfältige wissenschaftliche Nutzung des Webs ausgedehnt.

2.2 Grundzüge der Webometrie

Als *webometrics* wird die quantitative Analyse kommunikativer Strukturen anhand des World Wide Webs bezeichnet. Innerhalb der Forschungslandschaft befindet sich dieses Verfahren an der Schnittstelle von Wissenschaftsforschung, Informatik sowie Informations- und Bibliothekswissenschaften. Björneborn definiert Webometrie als:

„The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web, drawing on bibliometric and informetric approaches“ (Björneborn 2004: 12).

Sie umfasst vier Bereiche: die Analyse von Webinhalten, von Linkstrukturen, von Webnutzung (z.B. anhand von Logfiles) sowie von Webtechnologie (wie z.B. Suchmaschinen) (vgl. Björneborn 2004). Der eindeutige Schwerpunkt lag dabei bisher auf dem Bereich der Linkanalyse (vgl. Thelwall 2004a: 2).

Andere, anstelle von *webometrics* vorgeschlagene Bezeichnungen wie *netometrics* (Bossy 1995), *internetometrics* (Almind/Ingwersen 1996), *influmetrics* (Elisabeth Davenport nach Cronin 2001: 2), *webometry* (Abraham 1996), *web bibliometry* (Chakrabati et al. 2002) oder *www-metrics* (Persson o.J.) haben sich nicht durchgesetzt.

Der Begriff *cybermetrics*, der Isidro Aguillo zugeschrieben wird (vgl. Thelwall 2004a: 2), wird mitunter als Synonym für *webometrics* gebraucht (z.B. Bar-Ilan 2004b: 259), im Allgemeinen bezeichnet er jedoch die übergeordnete Kategorie, die neben der Webanalyse auch andere internetbasierte Datenquellen wie E-Mail, Newsgroups, Mailinglisten sowie die Auswertung

von Internetbeschaffenheit und Datenverkehr einschließt (vgl. Aguillo 2002; Björneborn 2004). Entsprechend ist für Cybermetrie eine breitere Definition zu wählen, nämlich als:

„The study of the quantitative aspects of construction and use of information resources, structures and technologies on the whole Internet, drawing on bibliometric and informetric approaches“ (Björneborn 2004: 13, Hervorhebung im Original).

Die vorliegende Arbeit beschäftigt sich ausschließlich mit der Analyse von Linkstrukturen und gehört somit zu einem Teilbereich der Webometrie.

2.2.1 Die Entstehung der Webometrie

Seit Mitte der 1990er Jahre befassen sich verschiedene Fachrichtungen mit dem Internet und dem World Wide Web. Ein Beispiel ist die „Geographie des Cyberspace“, die verschiedene Arten der Kartographie des Internets umfasst (vgl. Dodge 1999; Dodge/Kitchin 2002), andere Varianten der Webforschung sind *web dynamics*, *web intelligence*, *web ecology* und *web mining* (Björneborn/Ingwersen 2004). In einigen Disziplinen werden Linkanalysen aus der jeweils spezifischen Perspektive betrieben: In der Informatik dienen sie u.a. der Entwicklung effizienter Suchmaschinen, in der statistischen Physik werden sie zur Modellierung des Webwachstums gebraucht, während man sich in der Mathematik der Analyse von Webstrukturen und Komplexität widmet (vgl. Thelwall 2004a: 2). Der Grund für das Interesse der Wissenschaftsforschung und der Informationswissenschaft liegt in dem dramatischen Popularitätszuwachs, den die neuen Medien von Beginn an gerade in der Wissenschaft erfahren haben. Hiermit wurde die Annahme verbunden, dass auch Publikationen zunehmend in E-Journals statt in Printausgaben erscheinen würden, so dass die Bibliotheken der Zukunft nicht mehr in realen Räumen, sondern in virtuellen Datenbanken zu finden sein würden.

„If we assume that the WWW is a prototype of the distributed digital libraries of the future, it would be helpful to know if the tools and techniques developed for the analysis of intellect structure in paper-based libraries will be able to make the transition to this network-based environment“ (Larson 1996: o.S.).

Almind und Ingwersen begründeten ihre webometrischen Bemühungen ähnlich, indem sie auf William Paisley verwiesen, der bereits 1990 – also noch ohne Kenntnis des World Wide Webs – einen Trend von Printmedien zu elektronischen Datenbanken und eine entsprechende Neuausrichtung der Bibliometrie prognostizierte:

„In the future, a large proportion of all text that now appears in books, journals, magazines, and newspapers will be contained in electronic databases. In fact, the electronic databases will contain more than the published record, because there will be unpublished data collected just for the databases. This vast collection of electronic information is the future domain of bibliometric research“ (Paisley 1990: 286).

Es lag somit nahe, eine webspezifische Version der erprobten bibliometrischen Verfahren zu entwickeln, um dieser medialen Verlagerung Rechnung zu tragen. Einen ersten Versuch dazu unternahm Ray R. Larson (1996) auf der Konferenz der American Society for Information Science 1996 mit seiner Präsentation „Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace“, in der er in Anlehnung an die Kozitationsanalyse eine Kolinkanalyse durchführte. Bereits im Jahr darauf prägten Tomas C. Almind und Peter Ingwersen (1997) für informationswissenschaftliche Linkanalysen den Begriff *webometrics*. Im gleichen Jahr wurde das Online-Journal *Cybermetrics* gegründet. Ein entscheidender Auslöser für diese ersten Schritte in Richtung Linkanalyse bestand darin, dass sich ab Mitte der 1990er Jahre über die Suchmaschinen AltaVista und AllTheWeb erstmals Verweise auf bestimmte Websites abfragen ließen. Diese Option eröffnete den Weg zur empirischen Erhebung von Links und inspirierte somit vermutlich auch die ersten konzeptionellen Ansätze zur Entwicklung der Linkanalyse (vgl. Thelwall/Vaughan/Björneborn 2005).

Zu Beginn stand das Ausprobieren des Möglichen im Vordergrund (Larson 1996; Almind/Ingwersen 1997). Einen wesentlichen Stellenwert nahm hierbei die Übertragung bibliometrischer Maße auf das World Wide Web ein. So prüfte Rousseau (1997) die Gültigkeit von Lotkas Gesetz für Weblinks. Mit Hilfe dieses *power law* wird u.a. die Verteilung von Zitationen auf Dokumente beschrieben. Der von Ingwersen (1998) entwickelte Web Impact Factor (WIF) – ein Indikator, der sich aus der Anzahl der Inlinks pro Webseite errechnete und zur Messung von Sichtbarkeit und Qualität von Webseiten dienen sollte – war direkt dem Journal Impact Factor (IF) entlehnt, den das Institute for Scientific Information (ISI) jährlich im Journal Citation Report für die im Web of Science enthaltenen Fachzeitschriften ermittelt. Er ist definiert als die durchschnittliche Anzahl von Zitationen pro Artikel, die zwei aufeinander folgende Jahrgänge einer Zeitschrift im darauf folgenden Jahr erhalten (vgl. Thomson Reuters o.J.).

Im Anschluss daran beschäftigten sich zahlreiche Publikationen mit technischen Aspekten der Erhebung und Auswertung von Linkdaten, wobei insbesondere die Qualität der Erhebungswerkzeuge stark diskutiert wurde (vgl. z.B. Bar-Ilan 1998/99; Rousseau 1999; Snyder/Rosenbaum 1999; Thelwall 2001a; Bar-Ilan 2002a; Bar-Ilan 2002b). Ein weiterer Gegenstand der Methodenentwicklung war die Prüfung unterschiedlicher Referenzebenen. Durch das sogenannte *Alternative Document Model* wurde die Qualität von Linkanalysen auf

Basis einzelner Seiten, Directories, Domains oder ganzer Universitätswebsites verglichen (Thelwall 2003a; Thelwall/Wilkinson 2003). Zudem erschienen diverse explorative empirische Studien, beispielsweise zur Linkanalyse (Thelwall 2002a; Thelwall 2002b), zur Ko-Wortanalyse im Web (Leydesdorff/Curran 2000) und der Aussagekraft von Linkwerten (Thelwall 2002d).

Später ist ein Trend zur Verfeinerung festzustellen: Erhebungswerkzeuge wurden nun auch im Hinblick auf speziellere Anwendungen geprüft (z.B. ihre Tauglichkeit für Abfragen in verschiedenen Sprachen, vgl. Bar-Ilan/Gutman 2005). Empirische Studien beschäftigten sich beispielsweise mit den Funktionen von Links (vgl. Thelwall 2003b; Wilkinson et al. 2003; Bar-Ilan 2004a) sowie mit disziplinären und internationalen Vergleichen (vgl. Smith/Thelwall 2002; Thelwall/Tang 2003; Thelwall/Tang/Price 2003; Tang/Thelwall 2004; Li et al. 2005a; Li et al. 2005b; Park/Thelwall 2006; Thelwall/Zuccala 2008). Auch die Methodengrundlage erweiterte sich: In den Sozialwissenschaften entstand ein ethnographischer Zugang zum Web, der an den Linkstrukturen ansetzt (Beaulieu 2005). Ferner wurde eine linkbasierte Adaption der sozialen Netzwerkanalyse (SNA) entwickelt (Park/Thelwall 2005). Die SNA wurde auch verwendet, um Koautorenschaftsmuster in Publikationen und Webreferenzen zu vergleichen (Kretschmer/Aguillo 2004). Im Zuge der zunehmenden Diversifizierung erschien eine Reihe von Überblicksartikeln (z.B. Li 2003; Wilkinson/Thelwall/Li 2003; Björneborn/Ingwersen 2004; Bar-Ilan 2004b; Thelwall/Vaughan/Björneborn 2005; Thelwall/Ruschenburg 2006; Thelwall 2008a) sowie zwei Lehrbücher zur Linkanalyse (Thelwall 2004a) und zur Webometrie (Thelwall 2009a).

Mit der Emergenz des Web 2.0 sind neue Formen der Webnutzung zu beobachten: WebnutzerInnen machen zunehmend von der Möglichkeit Gebrauch, selbst Webinhalte zu kreieren, indem sie Blogs schreiben oder twittern, an wiki-basierten Informationssystemen mitwirken, sich an virtuellen sozialen Netzwerken wie Second-Life, Facebook oder MySpace beteiligen oder Beiträge auf Flickr oder Youtube einstellen. Mit den vielfältigen neuen Möglichkeiten für die WebnutzerInnen ging eine Verschiebung des inhaltlichen und methodologischen Schwerpunkts in der Webometrie einher: In den letzten Jahren lag der Fokus weniger auf der akademischen Webnutzung als auf der Analyse von Blogs und RSS feeds (Thelwall 2007; Prabowo/Thelwall/Alexandrov 2007), Wikipedia (Voß 2005) und virtuellen sozialen Netzwerken (vgl. Thelwall 2008c und 2008e). Im informationswissenschaftlichen Bereich wurde ferner die Verschlagwortung von Datenbeständen durch die NutzerInnen (*folksonomy*) untersucht (vgl. Angus/Stuart/Thelwall 2008 und Peters/Stock 2008). Ferner wurden webbasierte Publikationsdatenbanken wie Google Scholar oder Google Books auf ihre szientometrische Qualität geprüft (Mayr/Walter 2007; Kousha/Thelwall 2008; Kousha/Thelwall 2009). Gleichzeitig hat die Linkanalyse an Gewicht verloren. Hier wurden die Methoden nicht mehr so stark weiterentwickelt wie vorher, sondern lediglich auf neue

Felder und Länder angewandt. Entsprechende Studien behandeln die Sichtbarkeit medizinischer Universitäten im Iran (Aminpour et al. 2009), bibliotheks- und informationswissenschaftlicher Departments in Großbritannien (Arakaki/Willett 2009) oder südkoreanischer Top-Universitäten (Chung et al., im Erscheinen). Im nicht-akademischen Bereich wurden u.a. die Vernetzung lateinamerikanischer Wissenschaftsmuseen (Gouveia/Kurtenbach 2009) oder finnischer Kommunen (Holmberg/Thelwall 2009) untersucht.

Auffällig an der Entwicklung der Webometrie ist die zunächst starke, später durchaus umstrittene begriffliche und konzeptionelle Nähe zur Bibliometrie. In der zweiten Hälfte der 1990er Jahre überwog zunächst die optimistische Erwartung, durch die schlichte Übertragung einer erprobten Methode rasch und einfach die neuen Medien erschließen zu können. Bereits in Larsons Initialstudie von 1996 diente die Bibliometrie als Vorbild. Obgleich er sogar explizit darauf hinwies, dass die Bedeutungen von Links und Zitationen aufgrund des unterschiedlichen Kontextes deutlich differieren können, zog er das folgende enthusiastische Fazit:

„Although this is only one preliminary examination of a single field, I believe that it shows that the method does work using the hypertext linkages of the WWW, much in the same way as it works using journal or author cocitation in the scholarly literature of a field. This is an exciting and provocative discovery“ (Larson 1996: o.S.).

Auch Almind und Ingwersen gingen, als sie das Forschungsfeld *webometrics* skizzierten, von der Übertragbarkeit bibliometrischer Verfahren auf das World Wide Web aus und begründeten dies ebenfalls mit der Austauschbarkeit von Links und Zitationen:

„It is obvious that informetric methods using word counts and similar techniques can be applied to the www. What is new is to regard the www as a citation network where the traditional information entities, and citations from them, are replaced by Web pages. These pages are the entities of information on the Web with hyperlinks from them acting as citations“ (Almind/Ingwersen 1997: 404).

Rousseau wies in seiner frühen explorativen Studie von 1997 zwar darauf hin, dass Websites zumeist keine wissenschaftlichen Artikel seien und die enthaltenen Links weniger Zitationen im engeren Sinn als vielmehr Hinweise auf weiterführende Informationen darstellen. Dennoch hob er die Übertragbarkeit der bibliometrischen Verfahren auf das Internet hervor:

„Studying this kind of links is conceptually the same as studying citations between published articles. The meaning, however, is probably slightly different. [...] We will present here another illustration that ‚old bibliometrics‘ is applicable to the ‚new‘ Internet phenomenon“ (Rousseau 1997: o.S.).

Der frühe Optimismus zeigt sich auch in einer Analogie, die Josep Rodríguez i Gairín wählte: Die Suchmaschine AltaVista, die als erste die Möglichkeit zur Linksuche bot, nannte er den

„Citation Index des Internet“⁴ (Rodríguez i Gairín 1997: 178). In eine ähnliche Richtung ging Ronald Rousseau (1997), als er Weblinks in Anlehnung an *citations* als *sitations* bezeichnete und damit einen Begriff von Gerry McKiernan (o.J.) einem größeren Publikum bekannt machte:

„We are, however, not concerned with the number of links that can be found on a given web page, but we are interested in the number of times a given web page is referred to. In terms of classical citation analysis: we do not study references, but citations. We will use the term ‚sitation‘ to designate this relation between sites on the Internet“ (Rousseau 1997: o.S.).

Sinnvoll sind diese Vergleiche insofern, als sowohl das klassische wissenschaftliche Publikationssystem als auch das World Wide Web ihre kohärente Struktur durch Verweise zwischen Dokumenten (Zitate resp. Links) gewinnen:

„The notion of citation is fundamental to both the scholarly enterprise and to hypertext networks where it provides the primary mechanism for connection and traversal of the information space (or ‚cyberspace‘)“ (Larson 1996: o.S.).

Links stellen das grundlegende Merkmal des Webs dar: Ohne die Verlinkung würden die Dokumente schließlich kein Netz bilden (vgl. Bar-Ilan 2005c). Somit erscheint das Vorhaben vielversprechend, die auf Zitationsindizes basierten Analyseverfahren auf das World Wide Web zu übertragen. Der Begriff *sitation* wird zwar häufig zitiert, hat sich aber nie wirklich durchgesetzt, was daran liegen mag, dass er im mündlichen Gebrauch nicht von *citation* zu unterscheiden ist. Ferner vermittelt er zwar äußerst treffend die beschriebene Analogie zwischen Links und Zitationen, dies ändert jedoch nichts an der Tatsache, dass es sich dabei *nur* um eine Analogie handelt (Björneborn/Ingwersen 2004). Beide Datengrundlagen unterscheiden sich deutlich in Bezug auf eine Reihe zentraler Eigenschaften, die die Eignung der Bibliometrie als Vorbild für die Webometrie fraglich erscheinen lassen. Entsprechend mehrten sich etwa ab der Jahrtausendwende die skeptischen Hinweise darauf, dass die Anwendung bibliometrieähnlicher Verfahren der Datenquelle World Wide Web nicht gerecht wird und eine spezifische Methode erforderlich ist.

2.2.2 Die Webometrie im Kontrast zur Bibliometrie

Sowohl Bibliometrie als auch Webometrie dienen in der Wissenschaftsforschung dazu, Publikationen und die Verweise zwischen ihnen zu analysieren und auf dieser Grundlage Schlussfolgerungen über Strukturen der Wissenschaft zu ziehen. Trotz dieser formalen Ähnlichkeit und dem Vorbildcharakter, den die Bibliometrie von Beginn an für die Webometrie innehatte, gibt es entscheidende Differenzen zwischen beiden Methoden. Sie resultie-

⁴ Im Original: „El Citation Index de Internet“ (Rodríguez i Gairín 1997: 178).

ren aus den unterschiedlichen Dokumententypen und den Eigenschaften der zugrundeliegenden Medien Fachzeitschrift und World Wide Web. Im Wesentlichen lassen sich sechs elementare Unterschiede ausmachen, die im Folgenden näher benannt werden:

Funktionale Homogenität der Dokumente

Fachzeitschriften enthalten ihrer klar vorgegebenen und begrenzten Funktion entsprechend vorwiegend Fachartikel sowie Editorials, Buchbesprechungen u.ä. Entsprechend bestehen Zitationsdatenbanken wie z.B. das Web of Science aus wenigen Publikationsformen (wissenschaftlichen Artikeln, Reviews, Conference Proceedings etc.). Im Gegensatz dazu gibt es im Web eine Vielzahl unterschiedlicher Dokumente, von denen nur ein äußerst geringer Teil Form und Funktion einer Fachpublikation aufweist. Die Internetauftritte wissenschaftlicher Einrichtungen wenden sich an diverse soziale Gruppen (vgl. Abschnitt 2.3.1) und enthalten meist ein breites Spektrum an Seiten, das auch administrativen und organisatorischen Aufgaben dient. Nicht in jedem Fall ist somit ein inhaltlicher Wissenschaftsbezug gegeben.

Wenn man sich nicht auf die Websites von Forschungsinstituten und die Links zwischen ihnen beschränkt, ist von einer noch größeren funktionalen Heterogenität auszugehen (vgl. Abschnitt 2.3). Hierin liegt zunächst eine gewisse Unsicherheit, jedoch zugleich auch eine Stärke der Webometrie: Durch die fehlende Beschränkung auf wissenschaftliche Dokumente, lassen sich mit ihrer Hilfe auch Beziehungen zwischen Wissenschaft und anderen gesellschaftlichen Sektoren wie Wirtschaft, Politik, Medien oder Öffentlichkeit untersuchen (vgl. Bossy 1995; Leydesdorff/Curran 2000; Wormell 2001; Aguillo 2002; Thelwall 2002c; Heimeriks/Hörlesberger/van den Besselaar 2003; Thelwall/Harries 2004; Stuart/Thelwall 2005), wobei der Anteil der intersektoralen Links, der eine Kooperation im engeren Sinn repräsentiert, gering zu sein scheint (Stuart/Thelwall/Harries 2007).

Zu beachten ist auch, dass das Web sich deutlich von klassischen Informationsdatenbanken unterscheidet. Offensichtlich werden darin nicht nur wissenschaftlich geprüfte Inhalte abgelegt, sondern auch jede Art individueller Ansichten. Dabei sind viele Beiträge weniger auf kooperative Unterstützung bei der Informationssuche ausgelegt, sondern auf möglichst hohe Sichtbarkeit. Dies führt auch zum Phänomen Spam, wenn z.B. die Gestaltung von Webseiten darauf ausgerichtet wird, so häufig und so weit oben wie möglich in Suchmaschinenergebnissen zu erscheinen. Im Web herrscht nicht nur inhaltliche Diversität, sondern es wird auch eine Vielfalt unterschiedlicher Ziele verfolgt.

Funktionen der Referenzen

Ebenso wie die Vielfalt der Dokumententypen im Web wesentlich stärker gestreut ist als in Publikationsdatenbanken, ist zu erwarten, dass auch die Funktionen von Links im Web wesentlich breiter gestreut sind als die von Zitationen in Fachartikeln. Eine Reihe von empirischen Studien zu Links zwischen den Websites wissenschaftlicher Einrichtungen hat inzwischen ein vielfältiges Spektrum von Funktionen aufgezeigt (vgl. Kim 2000; Thelwall 2003b; Thelwall/Harries/Wilkinson 2003; Bar-Ilan 2004a; für Details vgl. Abschnitt 2.3.2). Zu beachten ist grundsätzlich, dass ein gewisser Anteil von Links zu Navigationszwecken oder aus trivialen Gründen eingerichtet wird (vgl. Thelwall 2003b). Entsprechend sind die Schlussfolgerungen, die aus Linkanalysen gezogen werden, noch stärker zu reflektieren, als bei bibliometrischen Studien. Dies gilt insbesondere für evaluative Zusammenhänge (vgl. Abschnitt 2.5.2).

Institutionalisierte Qualitätskontrolle

Da es eine der zentralen Eigenschaften des Webs ist, jeder Person die Veröffentlichung von Beiträgen und das Abrufen von Informationen zu ermöglichen (vgl. Abschnitt 2.1.2), gibt es im Web nur selten Qualitätssicherungsverfahren, wie sie bei wissenschaftlichen Veröffentlichungen üblich sind (vgl. Bar-Ilan 2004b; Scharnhorst/Thelwall 2005). Das Web kann somit zu einer Stärkung der Informationsfreiheit beitragen, gerade auch in Teilen der Welt mit schwach ausgeprägten demokratischen Rechten und erschwertem Zugang zu anderen Medien. Gleichzeitig bietet es jedoch eine nicht immer klar zu unterscheidende Mischung von Meinung, Spekulation und wissenschaftlich geprüfter Wahrheit:

„Obviously, the breakthrough for everybody to express themselves, practically without control from authorities, to become visible world wide [...] and to obtain access to data, information, values and knowledge in many shapes and degrees of truth, has generated a reality of freedom of information – also in regions and countries otherwise poor of infrastructure. The other side of the coin is that the Web increasingly becomes a web of uncertainty to its users; the thin red line between opaqueness, shading truth, misinformation, beliefs, opinions, visions or speculation and reliability, validity, quality, relevance or truth becomes increasingly thinner. It becomes a matter of trust“ (Björneborn/Ingwersen 2001: 69).

Dieses Problem trifft besonders stark Laien, die nach Informationen zu einem bestimmten Thema suchen und dabei wissenschaftliche Einschätzungen, die einer Peer-Review unterzogen wurden, nicht ohne Weiteres von weniger fundierten Äußerungen unterscheiden können, die über das Web verbreitet werden. Doch selbst bei Online-Artikeln, die gedruckten Fachartikeln gleichen, die in E-Journals oder Preprintforen zu finden sind und sich explizit an ein Fachpublikum wenden, kann nicht davon ausgegangen werden, dass sie ein Peer-

Review-Verfahren durchlaufen haben. Durch die Möglichkeit, im *Web work in progress*-Befunde mit der *scientific community* zu teilen oder durch Preprints die Wartezeit bis zur endgültigen Veröffentlichung eingereicherter Artikel zu umgehen, könnten die traditionellen Publikationsstandards verwischen:

„What is unclear in all of the ongoing experimentation and attendant speculation is the extent to which established peer review practices may be subverted by open peer review or, indeed, the extent to which different scientific communities and sub-fields may wish to adopt a tiered model of peer review, ranging from, say, full-bodied, double-blind refereeing through ‚peer review lite‘ to zero peer review. It does, however, seem likely that the present monolithic publishing and peer review system will become rather more flexible and pluralistic in character, as new publishing, posting, critiquing and archiving behaviours establish themselves“ (Cronin 2001: 3).

Auch van Raan (2001) erwartet eine Diversifizierung des Publikationssystems, bei der unreferierte Webdatenbanken klassische Medien ergänzen, jedoch nicht ersetzen: Er nimmt an, dass Veröffentlichungen in Fachzeitschriften mit Peer-Review weiterhin ein essentielles Kriterium wissenschaftlicher Reputation bleiben und ihr Status deshalb nicht geschwächt wird. Die neuen technologischen Optionen werden seiner Ansicht nach aber dazu führen, dass Journale zunehmend Print- und elektronische Ausgaben parallel herausgeben, um sowohl schnelle Verfügbarkeit als auch dauerhafte Archivierung zu ermöglichen.

Während also viele E-Journals am klassischen Peer-Review-Verfahren festhalten, unterliegen andere Bereiche des wissenschaftsbezogenen Webs, wie z.B. die Internetauftritte von Universitäten, in der Regel keiner Kontrolle durch externe Mitglieder der *scientific community*. Bei ihnen geht es weniger um die Präsentation der Forschungsergebnisse als der Einrichtung im Allgemeinen. Webometrische Untersuchungen haben folglich einen anderen Gegenstand als bibliometrische Studien: Es werden nicht die sozialen Strukturen der Wissenschaft analysiert, wie sie in qualitätsgeprüften Produkten wie Publikationen oder auch Patenten dokumentiert sind. Vielmehr bietet das Medium World Wide Web eine vielfältigere und alltäglichere Perspektive auf die Wissenschaft (vgl. Bossy 1995), in der die Entwicklung und Darstellung von Ideen, Projekten und Netzwerken Platz findet.

Formale Standardisierung der Dokumente

Die starke Homogenität der Dokumente in Publikationsdatenbanken hat Nebeneffekte, die für die Erhebungsverfahren von großer Bedeutung sind: Wissenschaftliche Artikel stellen aufgrund ihrer sowohl funktionalen als auch formalen Homogenität eine naheliegende Bezugsgröße für bibliometrische Berechnungen dar (z.B. kann für eine Zeitschrift oder einen Autoren die durchschnittliche Zitationsrate pro Artikel ermittelt werden). Dokumente im World Wide Web sind dagegen nicht nur funktional heterogen, sie können auch sehr verschieden strukturiert sein (vgl. Lewandowski 2005). Der gleiche Beitrag kann in einem Webdokument

oder verteilt auf mehrere untereinander verlinkte Seiten veröffentlicht werden. Wenn beide Versionen die gleiche Menge an Links von anderen Websites erhielten, würde dies im ersten Fall zu einer deutlich höheren Zahl von Inlinks pro Seite führen als im zweiten. Dieses Problem lässt sich zum Teil umgehen, in dem Inlinkzahlen nicht auf die Zahl der Seiten, sondern auf die der MitarbeiterInnen bezogen wird (vgl. Thelwall 2001b). Folglich sind Webindikatoren weniger aussagekräftig und wesentlich manipulationsanfälliger als wissenschaftliche Publikationsmaße (vgl. Bar-Ilan 2002a).

Gedruckte Fachartikel enthalten außerdem eine Reihe von Strukturdaten, die bei der bibliometrischen Auswertung genutzt werden. Hierzu gehören neben Namen, Jahrgang und Ausgabe des Journals auch Name und institutionelle Anschrift der AutorInnen sowie die gewählte Sprache. Vergleichbare Standards gibt es für Webdokumente nicht. Im HTML-Code können mit Hilfe sogenannter *meta tags* z.B. Angaben zum Autor, zum Titel, zur Sprache, in der das Dokument verfasst ist, oder zu Keywords gemacht werden. Diese Möglichkeit wird jedoch nur bei einer kleinen Minderheit der HTML-Dokumente genutzt, und es ist unwahrscheinlich, dass sie sich in absehbarer Zeit allgemein durchsetzen wird (vgl. Lawrence/Giles 1999: 108; Thelwall 2001c: 190). Zudem werden *meta tags* häufig nicht korrekt verwendet oder sogar in manipulativer Absicht eingesetzt, um häufiger in Suchmaschinen gefunden zu werden (vgl. Brooks 2003). Andere Daten finden sich zum Teil im Inhalt der Webdokumente, wie Informationen zum Autor oder zum Entstehungs- bzw. Aktualisierungsdatum. Sie erfordern jedoch eine manuelle Auswertung, deren Mühe sich allenfalls bei der Untersuchung einzelner Internetauftritte lohnt (vgl. Almind/Ingwersen 1997). Aufgrund der flexiblen Gestaltbarkeit von HTML-Seiten sind alle diese Angaben jedoch häufig gar nicht, in variierender Form oder sogar fehlerhaft angegeben.

„[...] the WWW is a distributed hypertext system. This fact has the effect that all data are unstructured, and there are no controls or requirements for the use of mark-up-codes“ (Almind/Ingwersen 1997: 408).

Die Flexibilität des HTML-Codes, die verteilte Struktur des Netzes und die Wandelbarkeit von Webdokumenten sind zwar wesentliche Gründe für den immensen Erfolg des World Wide Webs, sie stellen jedoch die Informationssuche im Allgemeinen und webometrische Webanalysen im Besonderen vor große Schwierigkeiten, die in klassischen Informations- oder Publikationsdatenbanken nicht auftreten:

„One of the great advantages of the Web is that it has no rules and no rulers. This advantage becomes a shortcoming when trying to retrieve information from the Web. The search engines have to cope with disorder, dynamic changes, malicious users, communication problems, and server failures, to mention just a few obstacles“ (Bar-Ilan 2002b: 308).

Als einziges universelles Strukturmerkmal bleiben die Domainnamen, die Anhaltspunkte zur institutionellen und nationalen Anbindung geben können. Ihre Aussagekraft ist jedoch noch zu prüfen (vgl. Abschnitt 2.3.4. und Kap. 3).

Für die Zukunft sind bereits Konzepte vorgeschlagen worden, um eine höhere Vereinheitlichung der Dokumentbeschreibungen zu erreichen, ihre automatisierte Auswertung zu verbessern und vertrauenswürdige Seiten besser von Spam unterscheiden zu können. Dazu gehören die Dublin Core Metadata Initiative (vgl. <http://dublincore.org>, zuletzt abgerufen am 19.11.2009) sowie die Idee des *semantic Web* (Berners-Lee/Hendler 2001). Beide treten für verstärkte Anwendung dokumentbeschreibender Metadaten ein. Da beide Ansätze auf Freiwilligkeit und die Einsicht in einen erhöhten Nutzen setzen, wird sich zeigen müssen, ob sich auf diese Weise eine stärkere Standardisierung von Teilen des Webs durchsetzen lässt.

Offenheit des Referenzraumes

Das auf Printmedien basierende wissenschaftliche Publikationssystem ist stark in sich geschlossen. Referenz- und Zitationsraum fallen hier zusammen, da in wissenschaftlichen Veröffentlichungen fast ausschließlich auf andere Fachpublikationen verwiesen wird. Obgleich Datenbanken wie das Web of Science nur einen Teil aller wissenschaftlichen Publikationen umfassen, finden sich zahlreiche Verweise innerhalb des dort erfassten Ausschnitts. Die Geschlossenheit von Publikationsdatenbanken und ihre Beschränkung auf Fachzeitschriften sind wesentliche Voraussetzungen für die Analyse wissenschaftsinterner Zitationsstrukturen.

Da es sich bei dem World Wide Web grundsätzlich um ein Medium handelt, das jeder Art von Publikation offensteht, findet sich hier eine entsprechende Vielfalt verschiedener Linkziele, -quellen und -funktionen (vgl. Abschnitt 2.3.2). Dies erschwert insofern die Analyse wissenschaftlicher Strukturen, als sie zunächst von anderen Zusammenhängen isoliert werden müssen. Gleichzeitig erlaubt sie jedoch die Untersuchung von Wissenschaft in ihrem sozialen Kontext, da sich im Web ein- oder wechselseitige Linkbeziehungen zwischen Wissenschaft und Wirtschaft, Politik, Medien, Nichtregierungsorganisationen (NGOs) oder interessierten Laien finden lassen.

Zeitliche Ordnung der Daten

Ein weiterer Unterschied liegt in der Zeitdimension: Während eine Publikationsdatenbank eine stabile und dauerhaft abrufbare Menge an Einträgen enthält, die um Neuerscheinungen ergänzt wird, können Websites jederzeit umstrukturiert werden. Dabei geben Webdokumente weder Aufschluss über ihren Entstehungszeitpunkt noch darüber, ob die enthaltenen Informationen und Verweise aktuell oder veraltet sind. Dokumente können nicht nur hinzukom-

men, sondern auch wegfallen. Entsprechend können Ziele von Verweisen zum Zeitpunkt der Analyse bereits verschwunden sein. Eine Studie zu Webseiten, die 1996 „zufällig“ über eine Suchmaschine ausgewählt und bis 2001 regelmäßig aufgerufen wurden, ergab eine Halbwertszeit von nur zwei Jahren (vgl. Koehler 2002). Im wissenschaftlichen Kontext scheint jedoch eine etwas höhere Stabilität zu herrschen. Beispielsweise ermittelte eine Studie über zwei Informatikzeitschriften eine Halbwertszeit von vier Jahren für die darin enthaltenen Webreferenzen (vgl. Spinelli 2003). Eine weitere Erhebung (Bar-Ilan/Peritz 2004) liefert ebenfalls Hinweise auf eine höhere Langlebigkeit von Dokumenten im Themenbereich Szientometrie sowie auf Unterschiede zwischen verschiedenen Top-Level-Domains.

Die ständige Möglichkeit zur Aktualisierung führt auch dazu, dass im Web auch wechselseitige Verweise zwischen zwei Seiten zu finden sind. Dagegen sind in Zeitschriften, von seltenen Ausnahmen abgesehen, nur Zitationen in eine Richtung zu finden: von neueren Artikeln zu älteren (vgl. Egghe 2000). Problematisch ist dies insofern, als sich die geringere Standardisierung auch auf die Datierung der Dokumente erstreckt, die meist gar nicht vorhanden oder uneindeutig ist: Sie kann den Zeitpunkt der Entstehung ebenso anzeigen wie den einer Aktualisierung oder einer rein technischen Umstrukturierung der Website anzeigen (vgl. Prime/Bassecouard/Zitt 2002). Eine zeitliche Abfolge von Links lässt sich im Web somit nicht ermitteln.

Dies hat eine entscheidende Konsequenz für die Untersuchung von Entwicklungsprozessen: Eine rückwirkende Studie, wie sie mit Hilfe eines Zitationsindexes möglich ist, erlaubt das Web nicht, da es keine immanente archivarische Funktion hat. Prime, Bassecouard und Zipf sowie Egghe weisen deshalb darauf hin, dass der Webometrie mit der Zeitdimension eine wesentliche Voraussetzung für Zitations- und vor allem Kozitationsanalysen fehlt (vgl. Prime/Bassecouard/Zitt 2002: 303f; Egghe 2000: 335). Diese bedeutende Lücke könnte in Zukunft durch das *Internet Archive*⁵ gefüllt werden, in dem bisher jedoch nicht alle Bereiche des Webs gleichmäßig repräsentiert sind (vgl. Thelwall/Vaughan 2004).

Längsschnittstudien im Web sind derzeit also nicht retrospektiv, sondern nur durch zukünftig wiederholte Erhebungen möglich. Die Tatsache, dass nur Momentaufnahmen der aktuellen Struktur verfügbar sind, schließt das Replizieren von Daten durch andere Forscher aus:

„As the contents of sites on the Internet are very volatile, an investigation such as this cannot be repeated and provide exactly the same results. This means that people who do this kind of studies must always be prepared to forward their data to other scientists in case they want to check the data, reinvestigate the statistical analysis or reuse the data set for other purposes“ (Rousseau 1997: o.S.).

⁵ Vgl. <http://www.archive.org>, zuletzt abgerufen am 19.11.2009

Die „real-time nature“ (Almind/Ingwersen 1997: 408) des Webs hat jedoch auch einen Vorteil: Während man bei Publikationsdatenbanken erst mit einer gewissen Verzögerung auf rezente Jahrgänge zugreifen kann, die der Erfassung aktueller Publikationen geschuldet ist, lässt sich die gegenwärtige Struktur des Webs zu jedem Zeitpunkt erheben (vgl. Scharnhorst/Thelwall 2005). Lediglich bei der Nutzung von Suchmaschinen ist eine gewisse Zeitverzögerung zu erwarten, da ihre Datengrundlage zum Teil nicht kontinuierlich, sondern in Intervallen aktualisiert wird. Webometrische Verfahren bieten somit die Möglichkeit, „science in action“ (Bossy 1995: o.S.) zu beobachten.

Qualität des Datenzugangs

In Bibliometrie wie Webometrie ist es üblich, auf Sekundärdaten zurückzugreifen. Während jedoch Publikationsdatenbanken wie das Web of Science speziell auf die akademische Nutzung ausgerichtet sind, orientieren sich die zur Linksuche genutzten kommerziellen Suchmaschinen am Markt. Dies hat weitreichende Konsequenzen für die Datenqualität (vgl. Snyder/Rosenbaum 1999).

Publikationsdatenbanken verfügen, obgleich sie z.B. für ihre begrenzte Abdeckung weltweit erscheinender Fachzeitschriften kritisiert wurden, über eine vergleichsweise hohe Reliabilität. Trotz einer gewissen Variabilität aufgrund des internen Datenmanagements sowie äußerer Faktoren (wie z.B. die Änderung von Zeitschriftentiteln) bieten sie eine im Wesentlichen stabile, nachvollziehbare und umfassende Datengrundlage (vgl. Paisley 1990). Zum Qualitätsmanagement gehört, dass bei der Eingabe eine Prüfung auf Dubletten erfolgt, so dass Dokumente üblicherweise nur einmal in wissenschaftlichen Datenbanken vorhanden sind (vgl. Brooks 2003; Lewandowski 2005). Im Web dagegen sind Dokumente häufig mehrfach zu finden, weil ganze Websites dupliziert werden (sogenannte *mirror sites*), um den Datenverkehr auf einzelnen Servern zu begrenzen, oder weil einzelne Dateien kopiert und an anderen Stellen abgelegt werden. Aus diesem Grund erfordern Linkanalysen eine besonders sorgfältige Dublettenprüfung.

Webstrukturen sind nicht nur, wie bereits beschrieben, grundsätzlich volatil und kaum standardisiert, auch den Instrumenten zur Erhebung von Webdaten – Webcrawlern und kommerziellen Suchmaschinen – mangelt es zum Teil an der nötigen Reichweite, Stabilität und Transparenz. Diese Schwächen führen dazu, dass webometrische Studien derzeit häufig eher Anhaltspunkte als reliable Resultate liefern (Snyder/Rosenbaum 1999; für Details vgl. Abschnitt 2.4). Partiiell sind sie jedoch dem geringen Alter der Webometrie geschuldet, so dass in Zukunft verbesserte, speziell auf akademische Zwecke zugeschnittene Erhebungsinstrumente verlässlichere Ergebnisse bringen könnten.

Abschließend lässt sich das Fazit ziehen, dass Bibliometrie und Webometrie sich sowohl in der Beschaffenheit als auch in der Funktion ihrer jeweiligen Datengrundlagen deutlich unterscheiden:

„Finally, it should be remembered that ‚citation‘ and citation are different matters entirely. Beyond formal analogies, structuring a landscape of communication through web page networks in science, and structuring a landscape of scientific outputs on calibrated databases, are very distinct in nature.“ (Prime/Bassecouard/Zitt 2002: 306).

Die Webometrie ist somit weder Konkurrenz noch zukünftiger Ersatz für die Bibliometrie. Es bleibt zu zeigen, inwiefern sie eine sinnvolle Ergänzung sein und Aufschluss über Aspekte von Wissenschaft geben kann, die durch die reine Auswertung wissenschaftlicher Produkte, wie Publikationen oder Patente, nicht zu erschließen sind.

2.3 Die webometrische Analyse wissenschaftlicher Einrichtungen

Um Informationen über wissenschaftliche Strukturen aus dem Web gewinnen zu können, ist die Entwicklung spezifischer Methoden und Fragestellungen notwendig. Als allererste Voraussetzung ist zu klären, wozu das Web in der Wissenschaft genutzt wird, wozu Links gesetzt werden und worüber man folglich durch ihre Analyse Erkenntnisse gewinnen kann.

2.3.1 Funktionen der Websites von Forschungseinrichtungen

Wie in dem Vergleich mit bibliometrischen Quellen (vgl. Abschnitt 2.2.2) deutlich wurde, weist das Web eine große Vielfalt an Dokumententypen auf. Auch die Websites von Forschungseinrichtungen sind von Heterogenität gekennzeichnet: Meist werden sie von einer nicht klar definierten Gruppe von Personen erstellt und gepflegt, zu der neben diversen Webmastern auch ProfessorInnen, wissenschaftliche MitarbeiterInnen und Studierende gehören. Ihre Inhalte reichen von Instituts- und Projektbeschreibungen sowie organisatorischen Hinweisen (Kontakt Daten, Wegbeschreibungen etc.) über Lehrmittel, Entwürfe und Endversionen von Artikeln bis hin zu Lebensläufen und persönlichen Informationen über Hobbies, Familie oder Weltanschauung einzelner Institutsangehöriger. Die Aktualität der abrufbaren Inhalte kann stark variieren. Unabhängig von der konkreten Gestaltung der Internetauftritte kann ihre Existenz für jede wissenschaftliche Einrichtung oder Hochschule mit Sicherheit vorausgesetzt werden. Die Website dient nicht nur zur Informationsvermittlung, sondern auch als Beleg ihrer Existenz und Vertrauenswürdigkeit (vgl. Engels/Ruschenburg 2006).

Bislang gibt es keine empirische Forschung zum Aufbau und Inhalt akademischer Internetauftritte. Middleton, McConnell und Davidson (1999) haben modellhaft aufgezeigt, welche

Funktionen ein Internetauftritt für Universitäten einnehmen kann, wobei ihre normative Beschreibung auf die zentral verwaltete Rahmenstruktur fokussiert und die wesentlich vielfältigeren und äußerst umfangreichen Möglichkeiten der individuellen Ausgestaltung außer Acht lässt: Die Website einer Forschungseinrichtung lässt sich erstens zur Kommunikation nutzen, wie es dem ursprünglichen Zweck des Webs entspricht, also z.B. zur Verbreitung von aktuellen Nachrichten, Ausschreibungen oder Arbeitsaufgaben für Studierende. Zweitens kann sie einen Zugang zu Informationen bieten, wie z.B. zu Bibliothekskatalogen, Mitarbeiter- und Telefonverzeichnissen, aber auch zu umfangreichen Datenbanken. Eine dritte Option ist die Nutzung der Website zu PR-Zwecken. Dabei halten Middleton, McConnell und Davidson die Präsentation aktueller und nützlicher Informationen für werbewirksamer als Selbstdarstellungen nach Art einer Hochglanzbroschüre. Entsprechend umfasst dieser Punkt nicht nur die Präsentation der Einrichtung sowie ihrer Unterheiten, Projekte und MitarbeiterInnen, sondern auch das Angebot nützlicher Tools, Linklisten etc., die das Image einer Einrichtung erhöhen und zum erneuten Besuch motivieren können. Die identifizierten Zielgruppen sind so unterschiedlich wie zahlreich: Neben den internen studentischen und wissenschaftlichen Nutzergruppen kann die Website zukünftige StudentInnen und MitarbeiterInnen ebenso adressieren wie ForscherInnen an anderen Einrichtungen, Personen aus Wirtschaft, Medien und Politik oder Alumni und Sponsoren. Die Kommunikation mit (potentiellen) SpenderInnen ist gegenwärtig in den USA üblicher als beispielsweise in Deutschland. Bereits die Startseiten von Universitätswebsites enthalten häufig Hinweise auf Informationen und Angebote für diese Gruppen.

Es ist anzunehmen, dass die Internetauftritte von öffentlichen Forschungseinrichtungen denen von Universitäten im Hinblick auf mögliche Zwecke und Zielgruppen ähneln, abgesehen davon, dass gegenwärtigen oder zukünftigen Studierenden weniger bzw. kein Gewicht eingeräumt wird, je nachdem ob das betreffende Institut in Lehrtätigkeiten involviert ist oder nicht. Da in der außeruniversitären Forschung häufig eine stärkere thematische Fokussierung als an Hochschulen herrscht, sind weitere Mitglieder im Adressatenkreis denkbar: Dazu gehören Fachgesellschaften, interessierte Laien oder NGOs.

Sowohl für Universitäten als auch für andere Forschungseinrichtungen sind die beschriebenen Nutzungsmöglichkeiten und Zielgruppen idealtypische Optionen, die nicht in jedem Fall und nicht auf der ganzen Website wahrgenommen werden. Insbesondere auf der Ebene der individuellen Seiten von WissenschaftlerInnen herrscht eine große Vielfalt von Gestaltungsweisen. Bei der quantitativen Analyse von Websites ist daher zu berücksichtigen, inwiefern persönliche Inhalte ohne Wissenschaftsbezug ins Gewicht fallen.

Fest steht, dass die Darstellung von Forschungsergebnissen in Form vollständiger Artikel, die sich an andere WissenschaftlerInnen richtet, nur eine Nutzungsmöglichkeit von vielen ist.

Akademische Websites besitzen ein deutlich anderes, vielfältigeres Profil als die klassischen wissenschaftlichen Printmedien. Auch haben technische und Navigationsaspekte einen hohen Stellenwert, was sich u.a. darin zeigt, dass die entsprechende Terminologie zu den am häufigsten auf britischen, australischen und neuseeländischen Universitätswebsites vertretenen Wörtern gehört (vgl. Thelwall 2005). Ein Vorbehalt gegenüber der Webometrie lautet dementsprechend, dass sich anhand akademischer Websites keine Erkenntnisse über die Wissensproduktion, also den Kern wissenschaftlicher Aktivität, gewinnen ließen. Studien über die Funktionen von Zitationen und Links offenbaren deutliche Unterschiede zwischen den beiden Medien (vgl. Abschnitt 2.3.2). Ein vorläufiger Konsens in der Webometrie scheint die Annahme zu sein, dass Publikationen und Zitationen die formelle Kommunikation im Erkenntnisprozess abbilden, während Webdokumente und die darin enthaltenen Links Resultate der informellen Kommunikation sind (vgl. Li 2003; Tang/Thelwall 2003; Thelwall/Tang 2003; Wilkinson et al. 2003; Li et al. 2005a). Folglich könnten sie Aufschluss über (nicht durch Koautorenschaft formalisierte) Kontakte und Kooperationen innerhalb und außerhalb der Wissenschaft, (noch nicht in Artikeln behandelte) aktuelle Themen sowie (sich nicht in Zitationen niederschlagende) Informationsquellen bieten. Deshalb verbindet Cronin (2001) mit der Webometrie die Hoffnung, den Stellenwert von *silent scientists*, die einen Beitrag zur Forschung leisten, aber wenig veröffentlichen, sichtbar machen zu können:

„[...] more of the world's ‚silent scientists‘ to use Meadows‘ [...] label for those who are not particularly visible in terms of their publication or citation records, may finally receive due recognition for some of their unseen and unsung contributions. In addition to citations, we can track other forms of invocation, which, individually or conjointly may provide more finely textured accounts of an individual's (or research team's) influence on peer group thinking, both locally and globally“ (Cronin 2001: 3).

Als vorteilhaft erweist sich dabei insbesondere, dass Webstrukturen, im Vergleich zu Telefonaten oder persönlichen Gesprächen, dauerhafter erhalten bleiben und somit ausgewertet werden können.

2.3.2 Funktionen und Aussagekraft von Links im wissenschaftlichen Kontext

Wie bereits in Abschnitt 2.2 herausgestellt, lag die Inspiration für die ersten webometrischen Studien in der offensichtlichen Ähnlichkeit von Links und Zitationen. Die unterschiedlichen strukturellen Eigenschaften von Fachpublikationen mit Peer-Review-Verfahren und des sich ohne äußere Regeln ausbreitenden World Wide Webs haben jedoch Zweifel an der Vergleichbarkeit beider Medien hervorgerufen. Zitationen erfüllen eine Reihe verschiedener Zwecke (für eine Übersicht vgl. Liu 1993). Sie alle dienen dabei letztlich der Stützung der eigenen Forschungsbefunde. Da das Web, wie in den vorangegangenen Abschnitten erläutert, diverse Nutzergruppen anspricht und eine wesentlich größere Vielfalt von Funktionen

erfüllt, unter denen die Darstellung von Forschungsergebnissen nur eine von vielen ist, ist von einer großen Streuung der Funktionen von Verweisen auszugehen.

Studien zur Klassifikation von Linkzwecken thematisieren zumeist die Motivationen für das Einsetzen der jeweiligen Verweise. Im Folgenden wird in dieser Arbeit der Begriff „Linkfunktion“ verwendet, da es weniger um individuelle Motive als um die Ergebnisse dieser Handlungen und ihren strukturellen Zweck geht. Es gibt im Wesentlichen zwei Zugänge, um präziseres Wissen zu den Funktionen von Links zu erhalten. Eine Möglichkeit besteht darin, Interviews mit den AutorInnen von Dokumenten, die Links enthalten, zu führen und sie nach ihren Beweggründen zu fragen. Diese Vorgehensweise ist sehr aufwendig und nicht in jedem Fall erfolgversprechend, da bei zahlreichen Webdokumenten nicht zu ermitteln ist, wer sie verfasst hat, und die Ergebnisse vom Erinnerungsvermögen der Befragten abhängen. Folglich ist bis dato nur eine solche Studie bekannt, in der VerfasserInnen von Artikeln in E-Journals interviewt wurden (Kim 2000). Die verbreitetere Alternative ist die Verwendung inhaltsanalytischer Verfahren. Auf diese Weise wurden bereits eine Reihe detaillierter Untersuchungen durchgeführt, in denen Linkfunktionen und z.T. auch andere wesentliche Aspekte – wie die Art der Ursprungsseite, die Art der Zielseite oder der Linkkontext – inhaltsanalytisch klassifiziert wurden. Ihre Ergebnisse werden im Folgenden kurz referiert, wobei sich die ersten vier vorgestellten Studien auf Links zwischen Hochschulen beschränken und Links anderer Herkunft ausschließen.

Eine Klassifikation von 100 zufällig ausgewählten Links auf die Startseiten britischer Universitäten (Thelwall 2003b) ergab vier grobe, als webspezifisch identifizierte Kategorien. Hierzu zählen Navigationslinks (z.B. in Linklisten enthalten), Links zur Kennzeichnung der Urheberschaft (Hinweise auf die AutorInnen oder Ko-AutorInnen eines Dokuments), soziale Links (Verweise auf andere Gruppen oder Personen) sowie unbegründete Links. Letzteres sind Verweise, die nur der Vollständigkeit halber eingefügt werden, wann immer eine Person oder Organisation mit eigener Webpräsenz erwähnt wird, ohne dass damit ein besonderer Zweck oder die Erwartung des Aufrufens verbunden wird. Zwar wurden auch bei der Analyse von Fachpublikationen soziale Einflüsse auf die Auswahl der Referenzen festgestellt; die zur Navigation, aus sozialen Gründen oder quasi automatisch eingefügten Hyperlinks beinhalten jedoch nach Ansicht des Autors keinerlei kognitive Nebenaspekte. Es gibt also eine Reihe von Linkfunktionen, die sozialer oder organisatorischer Art sind und wenig inhaltliche Aussagekraft besitzen, wenngleich die beschriebene Erhebung keine Angaben über ihre Häufigkeit enthält.

Eine etwas detailliertere Klassifikation einer Zufallsstichprobe von gut 400 Links zwischen britischen Hochschulen (Wilkinson et al. 2003: 53ff) führte zu dem Resultat, dass knapp 90 Prozent der Links als forschungsbezogen bezeichnet werden können, wenn dieser Begriff

breit ausgelegt wird und die für Forschung und Lehre nützliche Ressourcen- und Service-seiten einbezogen werden. Als zitationsäquivalent wurden nur zwei Links kategorisiert (0,5%). Links auf nicht-akademische Inhalte nahmen eine untergeordnete Rolle ein („tourist information“: 1,0%; „recreational“: 6,5%). Die AutorInnen zogen die Schlussfolgerung, dass Links als „natural by-product of scholarly activity“ (Wilkinson et al. 2003: 54) anzusehen sind und eine gute Grundlage für die Untersuchung informeller wissenschaftlicher Strukturen darstellen. Allerdings gelte es, den möglichen Einfluss anderer Faktoren, wie z.B. disziplinäre Unterschiede oder Defizite in der Website-Gestaltung, zu berücksichtigen, da er die Vergleichbarkeit der Fälle beeinträchtigen könne.

Die beiden detailliertesten Linkklassifikationen stammen von Judit Bar-Ilan und erstrecken sich jeweils auf Zufallsstichproben aus allen Verweisen zwischen den acht israelischen Hochschulen. Die erste Studie (Bar-Ilan 2004a: 398ff) umfasste eine inhaltsanalytische Kategorisierung der Links selbst sowie ihrer Ursprungs- und Zielseiten. Im Ergebnis zeigte sie, dass fast die Hälfte aller Links von Linklisten stammen (49,6%). Der Großteil von ihnen wurde im Zusammenhang mit der Erwähnung einer Organisationseinheit (39,6%), einer Person (24,4%) oder einer Textressource (17,2%) eingefügt. Entsprechend waren die Ziele in den meisten Fällen Seiten, die eine Organisationseinheit (49,2%) oder eine Person (29,4%) beschrieben oder die einen Text enthielten (17,0%). Den Linkursprungsseiten wurden in den meisten Fällen Forschungs-, Lehr- oder andere professionelle Zwecke zugeschrieben (insgesamt 73,0%), wobei die letztgenannte Kategorie auch die Darstellung von akademischen Einheiten oder von Personen umfasst. Administrative, soziale, technische oder der Navigation dienende Funktionen traten dagegen nur selten auf. Die Links dienten der Klassifikation zufolge ebenfalls häufig professionellen Zwecken (einschließlich Forschung und Lehre 60,2%) sowie der allgemeinen Information (11,4%). Auch hier sind andere Funktionen selten. Eine besondere Innovation dieser Untersuchung liegt in dem Anspruch, Ursprungs- und Zielseite nicht nur unabhängig voneinander zu klassifizieren, sondern auch den Bezug zwischen ihnen. Zu den häufigsten der insgesamt über sechzig Kategorien zählen hier u.a. Hinweise auf organisationale Einheiten („ist eine akademische Einheit“, „ist eine akademische Institution“), auf nützliche Informationen und Dienste („ist nützlich“, „bietet Informationen zum Thema“, „bietet Service“) oder auf weiterführende Informationen zu erwähnten Personen („arbeitet dort“, „studierte dort/hat dort studiert“, „ist RednerIn“, „ist Mitglied im Programmkomitee“, „ist Mitglied dieser Organisation“). In einer zweiten Studie entwickelte Bar-Ilan (2005c) ein etwas modifiziertes Klassifikationsschema, das eine noch differenziertere Betrachtung der Ursprungsseite sowie den Bezug von Ursprungs- und Zielseite erlaubt. Die Ergebnisse entsprechen in wesentlichen Teilen denen der ersten Untersuchung.

Neben den bereits beschriebenen, auf Links zwischen Universitäten fokussierten Analysen gibt es zwei Untersuchungen (Smith 2004; Chu 2005), die Verweise beliebiger Herkunft umfassen und die infolgedessen eine breitere Streuung von Linkfunktionen erwarten lassen. Die explorative neuseeländische Studie (Smith 2004: o.S.) befasste sich mit insgesamt fünfzehn Websites von Universitäten, Forschungseinrichtungen, individuellen Forschern, Fachgesellschaften und E-Journals. Ihrzufolge konnten 10 Prozent der Links als virtuelle Äquivalente von Zitationen betrachtet werden, da sie auf Artikel in E-Journals, Konferenzbeiträge oder Berichte verwiesen. Da lediglich die Ursprungs- und Zielseiten der Links kategorisiert wurden und nicht der Zweck der Links selbst, ist diese Schlussfolgerung fraglich. Im übrigen zeigten Links besonders häufig auf allgemeine Informationen sowie Informationen zu Forschung und Lehre (56%). Entsprechend stammten viele der Verweise aus fachspezifischen Linklisten (48%), wobei die Startseiten der untersuchten Internetauftritte das beliebteste Ziel waren (36%). Die Ergebnisse unterstützen die These, dass das Web vor allem als Medium zur Verbreitung von Informationen eine zentrale Rolle spielt und weniger als Publikationsorgan.

Eine Untersuchung der Websites von 54 bibliotheks- und informationswissenschaftlichen Universitätsinstituten in den USA (Chu 2005: 18ff) ergab, dass lediglich 13 bzw. 14 Prozent der Inlinks auf Seiten mit Forschungsinhalten bzw. Lehrinhalten verwiesen. Beliebtere Linkziele waren Serviceseiten mit nützlichen Ressourcen, Ankündigungen, Stellen- und Stipendienausschreibungen, Studierenden- und Fachvereinigungen etc. (46%) sowie die Homepages der Einrichtungen und ihrer MitarbeiterInnen (27%). Unter den häufig verlinkten Seiten befanden sich auch solche ohne informationswissenschaftlichen Bezug, wie z.B. eine Liste von Naturschutzgebieten, die entsprechend auch das Ziel von Links aus der Umwelt- und Tourismusbranche war.

Mit Links von Universitäten auf den Regierungssektor „gov.uk“ untersuchten Stuart und Thelwall (2005) eine spezielle Untergruppe von Links. Sie konstatierten, dass diese speziellen Verweise nur zu einem sehr geringen Teil forschungsbezogen waren, sondern vor allem auf akademische Ressourcen hinwiesen. Auch Forschungsförderung spielte hier keine Rolle, da die entsprechenden Einrichtungen in der Domain „.ac.uk“ angesiedelt sind. Entsprechend ließ sich keine Korrelation dieser Linkwerte mit der Forschungsproduktivität pro MitarbeiterIn nachweisen.

Aus der Übersicht über die bisher durchgeführten Linkklassifikationen wird deutlich, dass ihre Schlussfolgerungen sich zum Teil erheblich unterscheiden. Dies dürfte auch ein Resultat der heterogenen Datengrundlagen sein, die sich auf verschiedene Länder beziehen, mit unterschiedlichen Erhebungsinstrumenten gewonnen wurden und die in einigen Fällen außeruniversitäre Inlinks einbeziehen, in anderen dagegen nicht. Insgesamt wird deutlich, dass sich

auf akademischen Websites eine Reihe von Linkfunktionen finden, deren Diversität erheblich ist und weit über die Funktionen von Zitationen hinausreicht. Forschungsbezogene Verweise nehmen in den beschriebenen Studien je nach Datengrundlage und verwendeten Kategorien einen variierenden, jedoch durchweg hohen Anteil an allen Links ein. Es kann also eine gewisse Aussagekraft angenommen werden. Es bleibt jedoch festzuhalten, dass der Fokus von Linkanalysen nicht wie bei Zitationen auf der Publikation von Forschungsergebnissen, sondern auf einem breiteren Spektrum liegt, das alltägliche wissenschaftliche Aktivitäten ebenso einschließt wie die formalisierte Selbstdarstellung.

Aufgrund dieser im Vergleich zur Zitationsanalyse heterogeneren (und nach wie vor diffuseren) Datengrundlage wird Linkanalysen zum Teil eine grundsätzliche Skepsis entgegengebracht:

„I suppose that some day the people who run google and other search engines will figure out a way to separate true research citations from mere mentions of names, but in the meantime it is really not defensible to compare information retrieval via WOS [Web of Science] or STN [Scientific & Technical Information Network] or Dialog of ePsyche or whatever, to searches using google or other search engines over the Internet. That is why I don't waste my time trying to do so“ (Garfield 2002: o.S., Anmerkungen in eckigen Klammern: TR).

Trotz der Vielfalt der Linkfunktionen und ihrer geringen Überschneidung mit Zitationsfunktionen, sprechen inzwischen zahlreiche Studien dafür, dass ein Zusammenhang von Forschungsqualität und Linkdaten besteht. In Untersuchungen über britische, australische, neuseeländische, kanadische, taiwanesischen und chinesischen Hochschulen wurden statistische Korrelationen zwischen Inlinkmaßen – z.B. der absoluten Zahl der externen Inlinks oder Varianten des Web Impact Factors – und der Platzierung in Forschungsrankings wie dem britischen Research Assessment Exercise (RAE) oder anderen Produktivitätsmaßen gefunden (Thelwall 2001b; Smith/Thelwall 2002; Vaughan/Thelwall 2005; Thelwall/Tang 2003; Qiu/Chen/Wang 2004; Aguillo et al. 2006; Payne/Thelwall 2008). Sogar für Links, die von verschiedenen nicht-akademischen Top-Level-Domains aus auf britische Universitäten verweisen, konnte dieser Bezug bestätigt werden (Thelwall 2002a). Das Gleiche gilt auch für die Ebene von universitären Instituten der Disziplinen Biologie, Chemie und Physik in Australien, Kanada und Großbritannien (Li et al. 2005a). Eine vorausgehende Klassifikation der vorhandenen Links und eine Beschränkung auf forschungsbezogene Links verstärkt für britische Universitäten zwar die Korrelation, jedoch nur in schwachem Umfang (Thelwall/Harries 2003).

Ungeachtet der zum Teil nicht wissenschaftsbezogenen Funktionen von Links scheint es somit einen gewissen Zusammenhang von Forschungsqualität und der Vernetzung im World Wide Web zu geben. Ob es sich dabei um einen Effekt der allgemeinen Reputation von

Forschungseinrichtungen handelt, der sich auf die Sichtbarkeit der Internetauftritte auswirkt (vgl. Thelwall 2001b; Qiu/Chen/Wang 2004), oder ob Institute mit aktiver Forschungstätigkeit auch ihre Website besonders attraktiv und nützlich gestalten, ist noch zu klären. Die breite Streuung der Linkfunktionen erfordert in jedem Fall eine umsichtige Interpretation von Linkdaten, die von Aussagen zur Forschungsqualität Abstand nimmt und sich auf Schlussfolgerungen zu virtuellen Kommunikationsstrukturen innerhalb der Wissenschaft beschränkt (vgl. Abschnitt 2.5.2).

2.3.3 Disziplinäre Unterschiede in der Gestaltung von Internetauftritten

Bevor von den Internetauftritten wissenschaftlicher Einrichtungen Schlüsse auf Kommunikationsmuster gezogen werden, gilt es zu prüfen, welche Rolle disziplinäre Eigenheiten bei der Gestaltung der Websites spielen. Für disziplinäre Spezifika finden sich Anhaltspunkte in einigen webometrischen Studien:

Aufgrund der fachlichen Affinität und der vorhandenen Vorkenntnisse lässt sich erwarten, dass Internetauftritte in der Informatik und anderen technischen und naturwissenschaftlichen Disziplinen besonders stark genutzt werden. Ein Vergleich britischer Fachbereiche in Bezug auf Umfang ihrer Websites und ihrer Forschungsaktivitäten (basierend auf dem RAE) bestätigt die überproportionale Präsenz der Natur- und Ingenieurwissenschaften im Web im Vergleich zu Geistes-, Sozial- und Biowissenschaften (vgl. Thelwall/Price 2003). Eine disziplinäre Klassifikation der akademischen Domains in Australien und Taiwan, die besonders viele Links auf sich ziehen, spricht ebenfalls für die dominante Position technischer Fächer, insbesondere der Informatik (Thelwall et al. 2003). Dies gilt auch für einen Vergleich von insgesamt 89 Geschichts-, Psychologie- und Chemieinstituten an US-amerikanischen Universitäten: Der Median der Websitegröße liegt im Fach Geschichte mit 113 Seiten pro Institut weit unter den Medianen der Fächer Psychologie (603,5 Seiten pro Institut) und Chemie (1.786 Seiten pro Institut) (Tang/Thelwall 2003: 448). Auch zwischen verschiedenen naturwissenschaftlichen Disziplinen gibt es erhebliche Unterschiede. Eine Untersuchung von Biologie-, Chemie- und Physik-Instituten ergab, dass sowohl in Großbritannien als auch in Kanada und Australien die Physik die größten Internetauftritte aufweist (Li et al. 2005b). Es scheint also erhebliche Unterschiede zwischen verschiedenen Fachrichtungen zu geben.

Ferner gibt es Hinweise auf unterschiedliche Linkintensitäten: Beispielsweise zeigt die eben genannte Vergleichsstudie zu US-amerikanischen Universitätsinstituten (Tang/Thelwall 2003: 448) deutliche Unterschiede in der Vernetzung der Fachbereiche. Die Websites der untersuchten historischen Institute weisen im Durchschnitt Links zu einem Prozent aller US-amerikanischen Geschichtsinstitute auf. In der Psychologie liegen durchschnittlich Links zu 14 Prozent zu den übrigen Institute derselben Disziplin vor, in der Chemie 36 Prozent. Eine Anschlussstudie (Tang/Thelwall 2004: 481) offenbart Unterschiede im Hinblick auf die inter-

ationale Einbettung der drei Disziplinen in den USA: Während der Anteil der internationalen Inlinks bei den Chemie-Instituten 19 Prozent und bei den Psychologie-Instituten 16 Prozent beträgt, kommen bei den historischen Instituten nur 6 Prozent aller Inlinks aus anderen Staaten. Auch in anderen Ländern gibt es Anhaltspunkte für unterschiedliche Linkintensitäten. So weisen in Taiwan naturwissenschaftliche Fächer einen statistisch signifikanten Zusammenhang von der Anzahl der In- und Outlinks mit der Anzahl der Publikationen pro MitarbeiterIn auf, die in den Sozialwissenschaften bedeutend schwächer ausfällt (vgl. Thelwall/Tang 2003: 168f). Ein Vergleich naturwissenschaftlicher Universitätsinstitute ergab in Australien, Kanada und Großbritannien für die Disziplinen Biologie, Chemie und Physik einen signifikanten Zusammenhang zwischen der Zahl der Inlinks und verschiedenen Maßen der Forschungsproduktivität (Li et al. 2005a). Dennoch zeigten sich Variationen in den disziplinären Linkmustern: Die Physik-Institute waren am stärksten untereinander vernetzt. Außerdem wiesen alle Fachbereiche verschiedene Schwerpunkte in den Zielen der in ihren Websites enthaltenen Outlinks auf (Li et al. 2005b).

Eine besondere Position nehmen interdisziplinäre Links ein. Eine explorative Studie über Links zwischen britischen Universitäten führte zu dem Ergebnis, dass wiederum Informatik, Mathematik und Naturwissenschaften besonders häufig das Ziel disziplinenübergreifender Links sind. Stark vertreten ist jedoch auch die Pädagogik, was auf ein an allen Fachbereichen bestehendes Interesse an Konzepten und Materialien für die Lehre zurückgeführt wurde (Thelwall/Harries/Wilkinson 2003). Interessanterweise zeigen sich disziplinäre Unterschiede auch bei den Homepages von NobelpreisträgerInnen: Zwar nutzen sie alle ihre persönliche Seite in ähnlicher Weise, wobei insbesondere die Präsentation eines Photos (85,7%), einer Emailadresse (65,3%), eines Forschungsprofils (63,3%) und einer Bibliographie (57,1%) üblich sind. Bei den auf die Seiten verweisenden Links liegt der Durchschnitt in der Ökonomie bei 132,1 und in der Physik bei 105,5, in der Chemie jedoch mit 32,5 bei einem Bruchteil dieser Werte (Nelson 2005: 195). Dass die Chemie hier geringere Inlinkzahlen aufweist als die Physik, erstaunt nach den Befunden der übrigen disziplinären Studien weniger, als dass die Wirtschaftswissenschaften so viel populärer zu sein scheinen. Da die Inlinksuche in diesem Fall weder auf akademische Links noch auf externe Links eingeschränkt wurde, kann ihr starkes Abschneiden zum Teil auch auf die stärkere interne Verlinkung oder eine besondere Vermarktung der PreisträgerInnen zurückzuführen sein. Schließlich gibt es Anhaltspunkte dafür, dass nicht nur die quantitativen Linkintensitäten mit der Disziplin variieren, sondern auch die Linkfunktionen und die Art der Seiten, die sie verbinden. Eine Untersuchung der Mathematik-, Physik- und Soziologie-Institute an britischen Universitäten zeigte hier signifikante Differenzen (Harries et al. 2004).

Es stellt sich die Frage, ob die gefundenen disziplinären Unterschiede eine Folge der Vertrautheit mit den neuen Medien sind oder ob andere Gründe vorherrschen. Es erscheint

naheliegend, die starke Webnutzung in der Informatik auf die fachliche Nähe und die vorhandene Expertise zurückzuführen. Kling und McKim (2000) vertreten dagegen die These, dass die Intensität und die Art der Internetnutzung in erster Linie mit disziplinären Eigenschaften zu tun hat, die bereits vor der Entstehung des World Wide Webs bestanden. So gibt es in der Hochenergiephysik eine lange Tradition zur Herausgabe von gedruckten Preprints, die durch das Web lediglich vereinfacht und beschleunigt wurde. In den biologischen Disziplinen ist es hingegen unüblich, Artikel vor der Peer-Review öffentlich zu verbreiten. Stattdessen ist es gebräuchlich und z.T. sogar die Bedingung für Publikationen, erhobene Daten in gemeinsamen Webdatenbanken zu speichern. Die Autoren führen dies auf die unterschiedliche kommunikative Praxis der Disziplinen und ihre spezifischen Interessen zurück. Anpassungstendenzen sind folglich, wenn überhaupt, nur in geringem Maße zu erwarten.

Gewisse Nutzungsweisen des Webs scheinen jedoch universell zu sein: Hierzu zählen die Selbstdarstellung von Einrichtungen, Projekten und Personen sowie die Ankündigung und Vorbereitung von Konferenzen, über die sich auf elektronischem Weg jederzeit die aktuellsten Informationen verbreiten lassen. Auch bei Zeitschriften scheint es – bei Aufrechterhaltung des Peer-Review-Systems – einen generellen Trend zu elektronischen Ausgaben zu geben, wobei gegenwärtig vielfach noch eine zusätzliche Printausgabe erscheint. Häufig ersetzt das Web lediglich traditionelle, langsamere und aufwendigere Kommunikationswege in der Wissenschaft (vgl. Gläser 2003), die sich von Disziplin zu Disziplin unterschiedlich gestalten.

Es bleibt also festzuhalten, dass Faktoren wie die vorhandene Expertise und Ressourcen in Bezug auf Webgestaltung sowie die jeweiligen fachlichen Eigenheiten und Bedürfnisse die Gestaltung von Websites beeinflussen. Diese Variationen sind zunächst nichts Überraschendes. Auch in der Bibliometrie sind erhebliche disziplinäre Unterschiede im Publikationsverhalten bekannt (vgl. van Raan 2003), die sich auf die Strukturmerkmale, Publikations- und Zitationsraten von Fachartikeln erstrecken. Sofern jedoch Websites aus mehreren Disziplinen in eine empirische Studie involviert sind, gilt es, den disziplinären Unterschieden Rechnung zu tragen.

2.3.4 Die Identifikation wissenschaftlicher Websites im World Wide Web

Wenngleich das Web zur Unterstützung von Forschungsaktivitäten entwickelt wurde, wird es heute auf äußerst vielfältige Weise genutzt. Aufgrund der geringen Standardisierung des Webs ist es dabei nicht einfach, die Websites wissenschaftlicher Einrichtungen anhand äußerlicher Kriterien zu identifizieren. Ein Vorteil der Pluralität im Web ist es, dass hier möglicherweise die wechselseitigen Beziehungen von Wissenschaft und anderen sozialen Sektoren studiert werden können. Dies hängt jedoch davon ab, ob es gelingt, die Zugehörigkeit

eines Internetauftritts zu bestimmten Nutzergruppen – wie Forschungs- und Bildungseinrichtungen – eindeutig zu festzustellen.

Das einzige universelle Merkmal von Webseiten ist ihre Adresse, die URL. Durch die darin enthaltene Domain (wie www.uni-bielefeld.de) ist die Zuordnung einzelner Dokumente zu einem bestimmten Internetauftritt (dem der Universität Bielefeld) möglich. Durch die hierarchische Struktur der URLs lassen sich häufig auch Untereinheiten von Websites untersuchen. So enthält die Subdomain www.ub.uni-bielefeld.de die Seiten der Universitätsbibliothek. Eine Alternative ist die Ablage von Dokumenten in Unterverzeichnissen, sogenannten Directories, die durch einen Schrägstrich von dem Domainnamen getrennt sind (z.B. www.uni-bielefeld.de/iwt). URLs sind also in der Regel gut dazu geeignet, die zu einer ausgewählten Organisation oder einer ihrer Untereinheiten gehörenden Webseiten zu identifizieren, sofern bekannt ist, welche Adressbereiche sie umfassen.

Will man einen größeren Bereich des Webs (z.B. alle britischen Universitäten oder alle deutschen Unternehmen) untersuchen, ohne die darin enthaltenen Organisationen sowie ihre Internetauftritte zu kennen, bietet die URL nur wenig Ansatzpunkte: Mit der landesspezifischen oder allgemeinen Top-Level-Domain (wie „.de“ oder „.com“) enthält sie einen Hinweis auf die nationale bzw. sektorale Verortung des Internetauftritts. In vielen Ländern gibt es zusätzlich zur landesspezifischen TLD noch eine Second-Level-Domain, die Aufschluss über die sektorale Zugehörigkeit gibt: In Großbritannien stehen z.B. „.gov.uk“, „.mod.uk“ und „.ac.uk“ für Einrichtungen der Regierung, des Militärs und der höheren Bildung. Die TLDs „.gov“, „.mil“ und „.edu“ sind denselben Bereichen in den USA vorbehalten und beinhalten somit gleichzeitig einen Hinweis auf Land und Sektor. In zahlreichen Ländern ist Hochschulen eine spezifische Second-Level-Domain vorbehalten, so neben Großbritannien und den USA z.B. Australien, Neuseeland, China, Japan, Israel oder Südafrika. In anderen Ländern wie Deutschland, Frankreich, den Niederlanden, Russland etc. gibt es eine solche Differenzierung nicht. Folglich lassen sich Hochschul-Websites nur manuell identifizieren. Geht es nicht nur um die Untersuchung von Hochschulen, sondern auch um weitere Forschungseinrichtungen, ist zu beachten, dass sie häufig nicht unter der akademischen Top- oder Second-Level-Domain geführt werden. In den USA besitzen staatliche Forschungsinstitute häufig Domains mit der Endung „.gov“, private dagegen mit „.org“ und militärische mit „.mil“.

Zu beachten ist, dass die landesspezifischen Top-Level-Domains zwar für bestimmte Länder und die übrigen allgemeinen TLDs wie „.com“, „.org“ und „.net“ für bestimmte Verwendungszwecke eingerichtet wurden, dass ihre Vergabe aber nicht zweckgebunden erfolgt: Hinter einer Domain mit der Endung „.com“ verbirgt sich nicht notwendigerweise ein Unternehmen, hinter „.de“ nicht immer eine Person oder Einrichtung aus Deutschland (für eine detailliertere

Betrachtung des Zusammenhangs von landesspezifischen TLDs und nationaler Zugehörigkeit vgl. Abschnitt 3.3.4).

„The loose connection between top-level-domain specifications (e.g., .com, .edu and .org) and their actual content is also a frustrating problem. For example, many .com sites contain noncommercial content, although .com is ostensibly the main commercial top-level domain“ (Thelwall/Vaughan/Björneborn 2005: 81).

Sicherheit in Bezug auf die tatsächliche sektorale und geographische Verortung kann nur durch eine manuelle Prüfung erreicht werden. Da dieses Verfahren aufwendig und nicht in jedem Fall erfolgreich ist, werden sektorale Top-Level-Domains trotzdem als „rough deliniators“ (Snyder/Rosenbaum 1999: 377) bei der Kategorisierung von Websites nach thematischen Bezügen eingesetzt:

„TLDs are relatively simple to extract from hyperlinks yet are collectively providers of potentially valuable information because of the formal, if imperfect, association of most with countries or organization types“ (Thelwall 2004b: 126).

Das Problem der Aussagekraft von TLDs mildern zahlreiche webometrische Studien, indem sie sich auf die Untersuchung der Internetauftritte ausgewählter Universitäten und der Links zwischen ihnen bzw. auf Links innerhalb von landesspezifischen akademischen Second-Level-Domains beschränken (z.B. Smith/Thelwall 2002; Thelwall/Smith 2002; Thelwall/Tang 2003; Thelwall/Tang/Price 2003; Bar-Ilan 2004a; Bar-Ilan 2005c; Cothey 2004). Dies schließt zwar die Möglichkeit aus, Beziehungen zwischen Hochschulen und Wirtschaft, Politik oder Medien zu untersuchen; Vorteile dieser Vorgehensweise sind jedoch die einfache manuelle oder sogar automatisierte Auswahl universitärer Websites (letzteres anhand der akademischen Second-Level-Domain) sowie die vorwiegende Konzentration auf interakademische Links, bei denen der Wissenschaftsbezug ausgeprägter sein dürfte als bei der Berücksichtigung aller In- und Outlinks dieser Internetauftritte (vgl. Abschnitt 2.3.2). In der hier vorgestellten Untersuchung wird bewusst auf diese Beschränkung verzichtet, um zu prüfen, wie aufwendig eine manuelle Klassifikation der Link-URLs ist, wie nützlich die Informationen sind, die sich auf diese Weise aus ihnen ableiten lassen, und wie hoch der Anteil von Links ohne Bezug zur wissenschaftlichen Tätigkeit der untersuchten Einrichtungen ist.

Es bleibt festzuhalten, dass alle Websites einfach zu verarbeitende Merkmale in Form von Adressen aufweisen. Um ihr Potential ganz auszuschöpfen und Linkanalysen nicht auf ausgewählte Internetauftritte beschränken zu müssen, ist jedoch eine Prüfung der Aussagekraft der Top- und Second-Level-Domains unabdingbar. Sie ist daher Teil der folgenden empirischen Fallstudie. Neben den bereits beschriebenen Aspekten wissenschaftsbezogener Webstrukturen sind auch die Erhebungsverfahren im World Wide Web von Beginn an

Gegenstand kritischer Prüfungen. Im folgenden Abschnitt werden zwei Werkzeuge vorgestellt, mit denen Linkdaten gewonnen werden können.

2.4 Erhebungsinstrumente für Linkanalysen

Websites lassen sich mit Hilfe eines Browsers manuell durchsehen und auf Links absuchen. Dies ist jedoch mit einem äußerst hohen Zeitaufwand verbunden und erscheint deshalb allenfalls für Einzelfälle von sehr begrenztem Umfang praktikabel. Daneben besteht die Möglichkeit, Software zu nutzen, um die Links aus HTML-Seiten herauszufiltern. Dazu können zwei verschiedene Arten von Programmen verwendet werden, nämlich Webcrawler und kommerzielle Suchmaschinen.

2.4.1 Die Verwendung von Webcrawlern zur Erhebung von Links

Webcrawler sind spezielle, auf die Suche nach Hyperlinks ausgerichtete Programme. Übliche Bezeichnungen sind auch *spider*, *robot*, *bot*, *wanderer*, *downloader* oder kurz *crawler*. Solche Software speichert zunächst eine vorgegebene Startseite ab, folgt dann allen darin enthaltenen Links, speichert die so gefundenen Dokumente und extrahiert wiederum alle Verweise. Dieser Vorgang wird *crawl* genannt. Damit die Suche sich nicht auf das ganze World Wide Web erstreckt, verfolgt der Webcrawler nur Verknüpfungen innerhalb eines im Voraus definierten Bereichs (z.B. innerhalb der Domain „uni-bielefeld.de“). Darüber hinausgehende Links werden abgespeichert, aber nicht aufgerufen. Üblicherweise wird während der Abfrage auch eine Prüfung auf Duplikate durchgeführt, so dass Webseiten mit gleicher URL oder gleichen Inhalts nur einmal gespeichert werden (vgl. Thelwall 2004a: 9ff; Cothey 2004).

Zum Teil wird Webcrawler-Software kommerziell angeboten wie z.B. das Website-Management-Programm Microsoft Siteanalyzer, das Webmastern einen Überblick über ihren Internetauftritt liefert und helfen soll, darin enthaltene strukturelle Fehler zu finden. Es identifiziert beispielsweise *broken links*, also Links, deren Zieldokument aufgrund eines Tippfehlers, einer URL-Änderung oder einer Löschung nicht aufzufinden ist. In der Webometrie werden zum Teil selbst entwickelte Webcrawler verwendet. Sie werden auch als *academic crawler*, *research crawler* oder *specialist crawler* bezeichnet und sind an die jeweiligen individuellen Anforderungen angepasst. Ein in der akademischen Forschung populäres Exemplar ist der SocSciBot, den Mike Thelwall entwickelt und für Forschungszwecke

kostenfrei zur Verfügung gestellt hat.⁶ Die Version 3 dieses Webcrawlers wird in der folgenden empirischen Untersuchung benutzt.

Der SocSciBot sucht alle HTML-Links in einem definierten Bereich des Webs. Andere Formate (Flash, JavaScript etc.) kann er nicht lesen, er erfasst jedoch den Namen und den Speicherort eingebundener Dateien. Die Startseite ist frei wählbar (siehe Punkt 2 in Abb. 1).

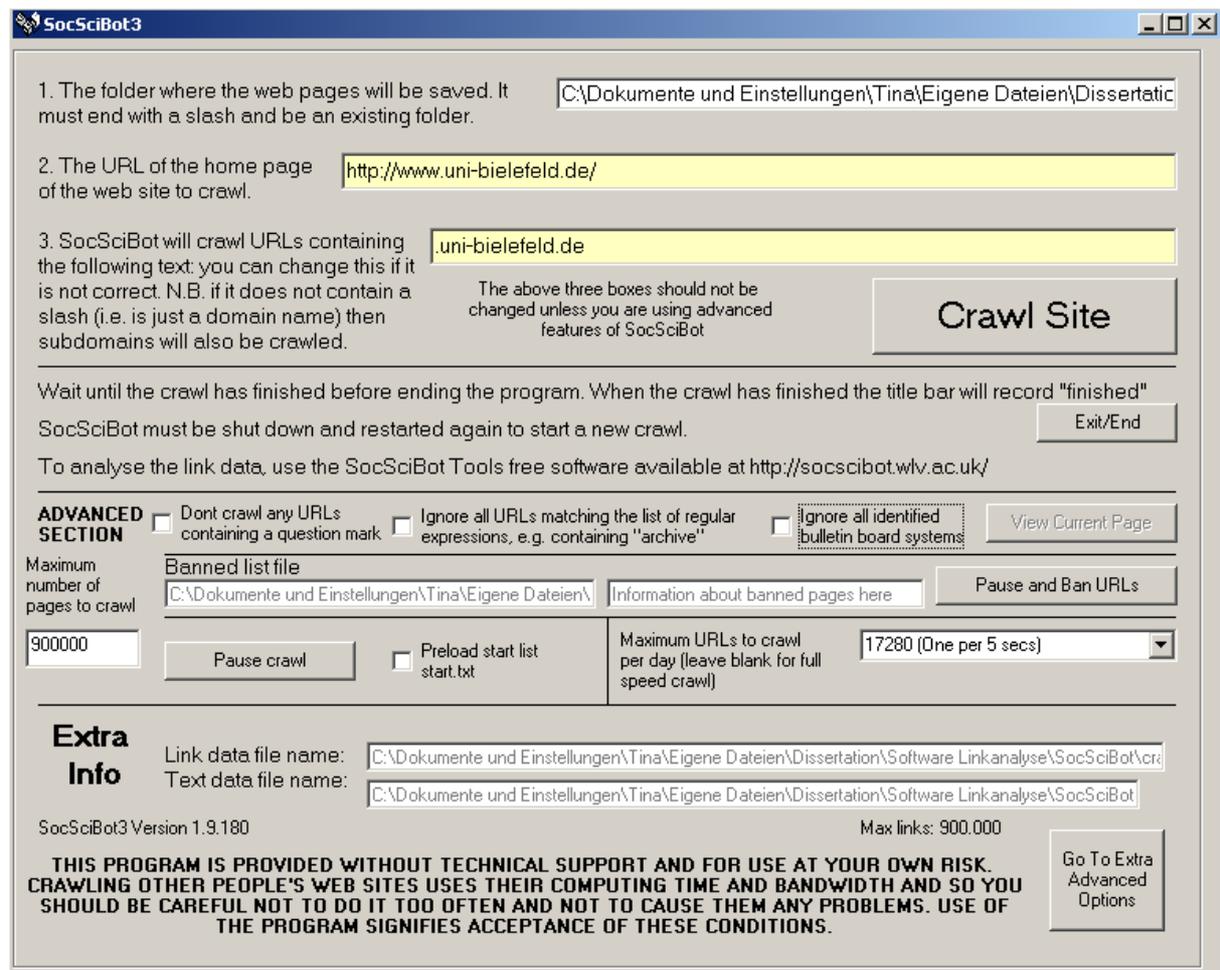


Abbildung 1: Oberfläche des Webcrawlers SocSciBot

Ferner ist der Bereich zu definieren, den der Webcrawler nicht verlassen soll (siehe Punkt 3 in Abb. 1). In der *advanced section* ist die Wahl mehrerer Startseiten über eine Liste möglich („preload start list“). Dies kann zu einer besseren Abdeckung schwach verlinkter oder nicht vollständig in HTML gestalteter Internetauftritte beitragen. Es besteht ferner die Option, dynamische Seiten („URLs containing question marks“) und Bulletin Boards („BBS“) zu ignorieren, da sie häufig keine externen Links enthalten, aber Linkschleifen generieren können.

⁶ Das Programm ist unter <http://socscibot.wlv.ac.uk> abrufbar (zuletzt abgerufen am 19.11.2009). Für eine Ansicht der Oberfläche vgl. Abb. 1.

Aus den gleichen Gründen kann auch auf eine Liste zurückgegriffen werden, die auf Erfahrungen des Autors zurückgeht („list of regular expressions“). Alternativ kann auch eine eigene Liste von URLs („banned list“) angegeben werden, um zielgerichtet auf voluminöse, aber irrelevante Bereiche einer Website oder Linkschleifen reagieren zu können. Diese *banned list* kann auch während eines Crawls aktualisiert werden. Der SocSciBot respektiert außerdem das Robots Exclusion Protocol (siehe Abschnitt 2.4.3).

Während des Crawls werden Textversionen von allen in dem definierten Webbereich gefundenen Seiten gespeichert. Anschließend liefert der SocSciBot eine Liste der erfassten URLs. Eine andere Liste spezifiziert gegebenenfalls alle URLs, die aufgrund des robots.txt-Protokolls ausgelassen wurden. In einer weiteren Ausgabedatei werden für jede durchsuchte Webseite alle darin enthaltenen URLs von internen Links, abrufbaren Dateien und externen Links, die über die angegebene Domain hinausweisen, genannt.

Die Auswertungssoftware Tools extrahiert eine Liste aller externen Links. Bei der Analyse mehrerer Internetauftritte berechnet sie außerdem die Zahl von Links zwischen den verschiedenen Websites, Directories, Subdomains oder Webseiten.

Die Version 3 des SocSciBot begrenzt die maximale Zahl der erfassten Links auf 50.000. Sie unterbricht außerdem das Crawlen nach jeder Seite für fünf Sekunden, um die abgesuchte Website nicht zu überlasten. Für das hier vorliegende Dissertationsprojekt stellte Mike Thelwall eine auf 900.000 Links limitierte und beliebig zu taktende Version zur Verfügung, um die Erfassung größerer Internetauftritte zu ermöglichen.

Webcrawler können auch, verbunden über eine Kontrolleinheit, auf mehreren PCs gleichzeitig laufen. Dieses Verteilungsprinzip, das auch von Suchmaschinenbetreibern verwendet wird, ermöglicht es, z.B. Computerräume in Universitäten nachts oder am Wochenende zu nutzen und so die Rechenkapazitäten zu erhöhen und die Laufzeiten zu verkürzen (vgl. Thelwall 2001a). Eine weniger ressourcen- und zeitaufwendige Alternative stellt die Linksuche mit Hilfe von Suchmaschinen dar.

2.4.2 Die Nutzung kommerzieller Suchmaschinen zur Linksuche

Mehrere kommerzielle Suchmaschinen erlauben nicht nur die Suche nach Stichworten, sondern auch die Abfrage von Links, die auf ausgewählte URLs (z.B. www.uni-bielefeld.de, www.ub.uni-bielefeld.de, www.uni-bielefeld.de/iwt/index.html) verweisen. Alternativ können Links in die entsprechenden Webbereiche recherchiert werden (z.B. Links auf eine beliebige Seite innerhalb der Domain www.uni-bielefeld.de, innerhalb der Subdomain www.ub.uni-bielefeld.de oder der Directory www.uni-bielefeld.de/iwt). Auch Verweise in bestimmte Top-Level-Domains (z.B. „.de“) lassen sich suchen. In den letzten Jahren kamen neue Möglich-

keiten hinzu, wie z.B. die sprach- oder länderweise Abfrage. Die verfügbaren Optionen und die entsprechenden Befehle unterscheiden sich dabei je nach Anbieter.

In den meisten frühen webometrischen Studien wurden Suchmaschinen als Zugang zu Linkdaten gewählt (u.a. Larson 1996; Almind/Ingwersen 1997; Rousseau 1997; Ingwersen 1998), vermutlich weil sie schnell, kostenlos und ohne technische Vorkenntnisse Ergebnisse liefern. Thelwall, Vaughan und Björneborn (2005) nehmen an, dass diese ab Mitte der 1990er Jahre zunächst von AltaVista und AllTheWeb angebotene Suchoption einen entscheidenden Anstoß für die Entwicklung der webometrischen Linkanalyse gegeben hat.

Kommerzielle Suchmaschinen nutzen ebenfalls Webcrawler, mit denen sie regelmäßig das Web nach neuen oder veränderten Seiten absuchen. Die gefundenen Dokumente werden indiziert, d.h. die darin enthaltenen Wörter werden – jeweils verknüpft mit der URL und anderen Merkmalen des Dokuments – in Datenbanken abgelegt (vgl. Bar-Ilan 2004b: 235). Spezielle Prozessoren helfen dabei, sie bei Anfragen in kurzer Zeit wieder abzurufen. Ranking-Algorithmen sortieren die gefundenen Treffer, so dass mit hoher Wahrscheinlichkeit relevante Ergebnisse zuerst erscheinen. Dabei werden je nach Suchmaschine verschiedene Faktoren berücksichtigt, wie z.B. die Häufigkeit und der Ort, an dem das gesuchte Wort im Dokument erscheint, oder die Anzahl der Links, die auf eine Seite verweisen (Thelwall 2004a: 19f).

Nicht alle Suchmaschinen bieten eine Abfragemöglichkeit von Links auf bestimmte und zwischen bestimmten Websites an. Im März 2007, als dieser Abschnitt verfasst wurde, konnten dafür Google, die Overture-Dienste Yahoo und Altavista sowie die aus MSN beta hervorgegangene Live Search von Microsoft genutzt werden. Diese Anbieter nehmen eine marktbeherrschende Position ein. Im Februar 2007 wurden in den USA 55,8 Prozent aller Suchanfragen über Google Search abgewickelt, weitere 20,7 Prozent über Yahoo Search sowie 9,6 Prozent über Live Search (Nielsen Netratings 2007). Allerdings sind die erweiterten Suchoptionen häufig nicht ordentlich dokumentiert, so dass man sich die benötigten Befehle aus der Literatur und durch Ausprobieren erschließen muss.

In dem abgebildeten Beispiel (Abbildung 2) wurden mit Live Search die Inlinks der gesamten Website der Universität ermittelt, wobei alle Links ausgeschlossen wurden, die von der Website selbst stammen (für eine genauere Erläuterung des Verfahrens vgl. Abschnitt 4.3.3). Dieses Verfahren wird in der Regel verwendet, um interne Links von der Erhebung auszuschließen, da sie primär Navigationszwecken dienen und da ihre Anzahl stark von dem Aufbau der Seite abhängt.

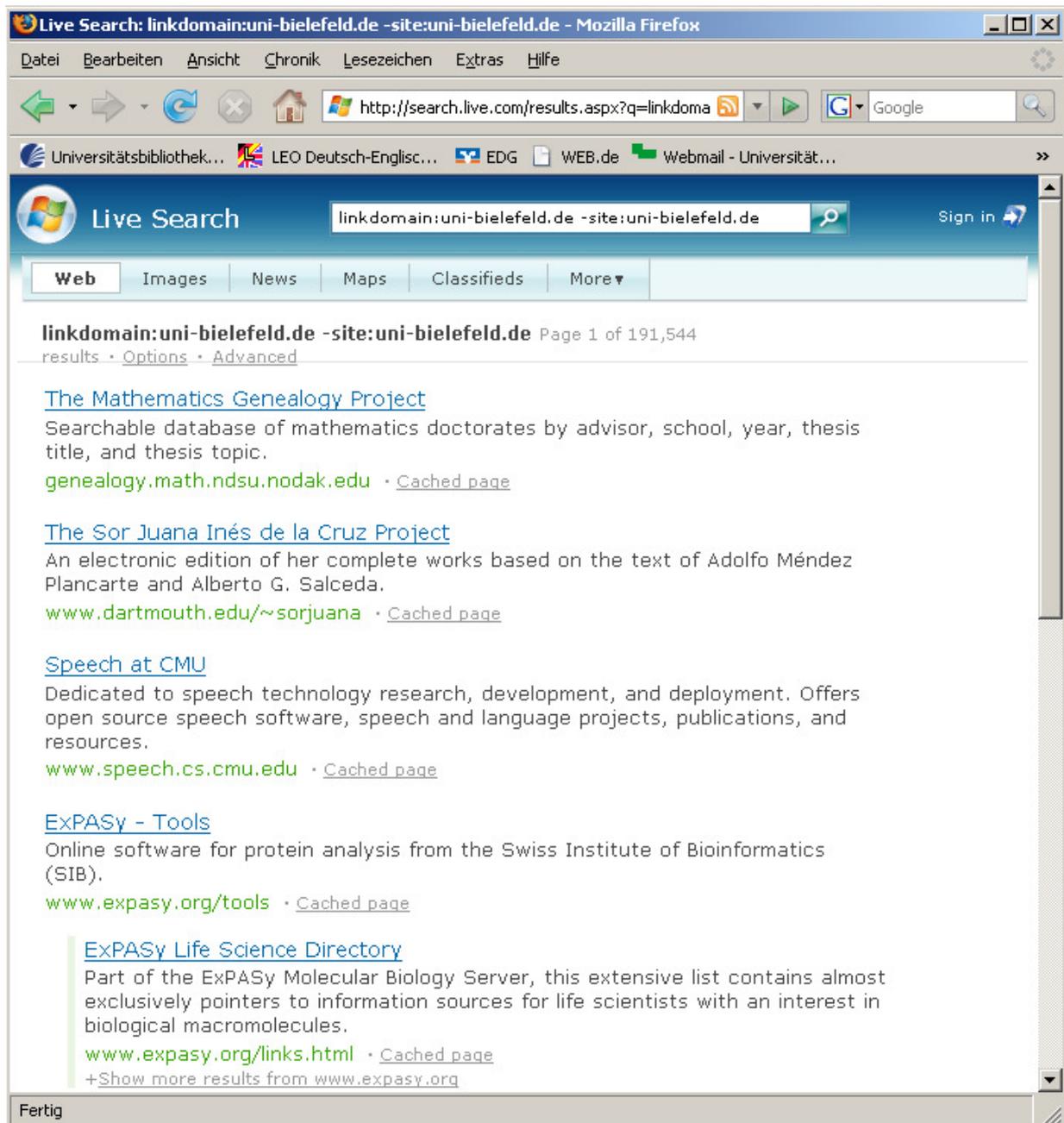


Abbildung 2: Ansicht der Ergebnisausgabe der kommerziellen Suchmaschine Live Search (Abfrage von Links auf die Website der Universität Bielefeld am 21.03.07)

2.4.3 Eigenheiten von Linkerhebungen mit Webcrawlern

Webcrawler suchen zielgenau und zu einem vorgegebenen Zeitpunkt einen definierten Bereich des Webs ab. Ihre Suchstrategien sind in der Regel nachvollziehbar und stabil, so dass sie präzise, aktuelle und über die Zeit vergleichbare Primärdaten liefern. Die Erhebung von Linkdaten mit Hilfe eines Webcrawlers erfordert je nach Umfang und Struktur der untersuchten Website zum Teil erhebliche Computer- und Zeitressourcen.

Abfragen von größeren Bereichen des Webs (wie z.B. einer ganzen Top-Level-Domain) sind daher auf diesem Weg nicht zu bewältigen. Aufgrund dieser Limitierung ist die Verwendung von Webcrawlern auf bestimmte Fragestellungen beschränkt: Zwar ist die offene Suche nach Inlinks, die vom ganzen Web aus auf eine bestimmte Website verweisen, nicht mit Crawlern zu bewältigen. Es können jedoch die in ausgewählten Websites enthaltenen Outlinks erhoben werden. Auf dieser Basis können auch die Links zwischen einer Auswahl von Internetauftritten ermittelt werden, indem nach dem Crawlen aus allen in einer Website enthaltenen Outlinks diejenigen herausgefiltert werden, die auf eine der anderen Websites im Sample zielen. In vielen Studien wird diese Vorgehensweise verwendet, beispielsweise um die virtuellen Verknüpfungen zwischen Universitäten in Großbritannien, Australien und Neuseeland (Smith/Thelwall 2002) oder in China und Taiwan (Thelwall/Tang 2003) zu analysieren.

Der SocSciBot ist ein Beispiel für ein transparentes und anpassungsfähiges Programm. Er ermöglicht es, eine Liste von auszuschließenden Webbereichen manuell einzugeben, um flexibel auf die zu untersuchenden Seiten eingehen zu können. Er erlaubt auch die flexible Einstellung der Ebene des Webs, die erhoben werden soll, so dass nicht notwendigerweise eine ganze Website, sondern auf Wunsch nur eine Subdomain, Directory oder ein einzelnes HTML-Dokument abgesucht wird. Das ebenfalls speziell für Forschungszwecke entwickelte Zusatzprogramm Tools erleichtert die Auswertung von Website-Netzwerken. Für Textanalysen steht ein weiteres Programm unter dem Namen Cyclist ebenfalls zur Verfügung.

Dennoch kann die Verwendung eines Webcrawlers keine vollständige Erfassung von Websites garantieren, da ausschließlich Dokumente durchsucht werden, die von der gewählten Startseite aus durch Links erreicht werden können und ein lesbares Format aufweisen. Bestimmte Bereiche des Webs sind somit von der Erhebung ausgeschlossen. Sie werden auch als „Invisible Web“ (Sherman/Price 2003) oder „deep Web“ (Bergman 2001) bezeichnet. Im einzelnen gehören die folgenden Punkte a) bis e) dazu:

a) Unverlinkte Bereiche

Da Crawler sich anhand von Links durch Websites arbeiten, erreichen sie nur jene Dokumente, zu denen von der Startseite aus direkt oder indirekt Links führen. Teile, die nicht

verlinkt sind, weil sie beispielsweise veraltet oder nicht für die Öffentlichkeit bestimmt sind, kann ein Crawler nicht finden. Der über Verweise erreichbare Bereich einer Websites wird auch als öffentlich indizierbares Web („publicly indexable web“, vgl. Lawrence/Giles 1999: 107) bezeichnet. Cothey (2004) hat die Erfassung einer vollständig bekannten Website simuliert und gezeigt, dass sich die Abdeckung durch die Wahl mehrerer Ausgangspunkte wesentlich erhöhen lässt. Besonders Directories, die ein Linkverzeichnis untergeordneter Dateien enthalten, haben sich als günstige Startseiten erwiesen. Sie selbst sind jedoch häufig nicht durch Links, sondern durch das Trunkieren bekannter Adressen zu erreichen (Cothey 2004: 1232ff).

b) Geschützte Bereiche

Ferner gibt es auf einigen Websites Bereiche, die für alle nicht autorisierten NutzerInnen oder speziell für Crawler gesperrt sind. So können passwortgeschützte Teile nicht ausgewertet werden. Das Gleiche gilt für Websites, deren Nutzung eine kostenlose oder gebührenpflichtige Registrierung erfordert. Darüber hinaus stellt das Robots Exclusion Protocol (vgl. Koster o.J.) Regeln auf, nach denen Webmaster in einer mit „robots.txt“ benannten Datei alle oder bestimmte Crawler auffordern können, eine Website oder definierte Teile davon nicht abzusuchen (vgl. Sherman/Price 2003). „Ethical robots“ (Thelwall 2004a: 17) wie der SocSciBot respektieren dieses Protokoll und ignorieren die angegebenen Bereiche.

c) Datenbanken

Häufig besitzen Datenbanken eine Oberfläche, die die Eingabe von Suchbegriffen oder -kommandos erfordert. Hierzu gehören z.B. Suchmaschinen oder Bibliothekskataloge. Üblicherweise sind Webcrawler nicht so eingerichtet, dass sie eigenständig Anfragen ausführen, weshalb in diesen Fällen lediglich die Maske der betreffenden Datenbank abgesucht wird (vgl. Sherman/Price 2003). In Bezug auf die Linkanalyse besitzt dieses Phänomen allerdings insofern geringe Relevanz, als Datenbanken häufig keine Links enthalten.

d) Nicht in HTML verfasste Dokumente

Eine weitere Begrenzung der Linksuche ergibt sich aus dem Dokumententyp. Ursprünglich bestand das World Wide Web im Wesentlichen aus HTML-Dokumenten. Deshalb werden Webcrawler dafür eingerichtet, die in den HTML-Code eingebetteten Links anhand ihrer spezifischen HTML-Tags zu identifizieren. Inzwischen werden im Web jedoch zunehmend Dateien anderer Formate abgelegt, wie z.B. PDF, Postscript, MS Word oder Excel. Hinzu kommen verschiedene Grafik- und Multimediaformate sowie komprimierte und ausführbare

(exe-)Dateien. Einige von ihnen können Links enthalten, die jedoch durch Webcrawler nicht entdeckt werden können.

Ein größeres Problem ist, dass auch die Webseiten selbst zunehmend in anderen Programmiersprachen als HTML geschrieben werden: Mit Java, JavaScript, Shockwave oder Flash lassen sich komplette Internetauftritte gestalten. Die darin enthaltenen Links sind jedoch nicht anhand von Tags zu erkennen und folglich nicht von auf HTML spezialisierten Webcrawlern zu finden. Betrifft dies nur einen Teil der Website, kann durch eine geschickte Wahl von einer oder mehreren Startseiten häufig dennoch eine gute Abdeckung erreicht werden (vgl. Thelwall 2004a). Die einzige Alternative ist das manuelle Verfolgen und Erfassen der Links, das jedoch nur in geringem Umfang praktikabel ist.

e) Dynamische Seiten

Große Teile des Webs bestehen aus statischen Webdokumenten, also fertig formatierten, unter festgelegten URLs abgelegten Seiten stabilen Inhalts. Daneben gibt es jedoch dynamische Seiten, die erst in dem Moment erzeugt werden, in dem eine Abfrage erfolgt. Sie werden deshalb auch als automatisch generierte Seiten bezeichnet. Diese Technik ermöglicht es, Datenbanken im Web effizient zu organisieren. Suchmaschinen, Onlinekaufhäuser oder Klimadatensammlungen ermitteln nach der Eingabe von Stichworten in ihrer Datenbank die passenden Ergebnisse und fügen sie unter Verwendung von Programmen wie PHP oder Microsoft ASP zu einer dynamischen Webseite zusammen, die nur für die Dauer der Abfrage existiert. Diese Seiten bestehen häufig aus HTML-Code, so dass Webcrawler sie lesen können; sie sind jedoch häufig darauf programmiert, sie zu ignorieren, da dynamische Seiten erstens nicht stabil und nahezu beliebig erzeugbar sind und da es zweitens zu Linkschleifen kommen kann, sogenannten *spider traps*. So generiert ein dynamisch gestalteter Veranstaltungskalender immer wieder eine neue Seite, wenn ein Link wie „nächster Tag“ oder „nächster Monat“ aufgerufen wird. Dieser Effekt kann auch durch Nachlässigkeit bei der Programmierung sowie bei der absichtlichen Einrichtung von Spam-Seiten entstehen (vgl. Sherman/Price 2003).

Da Datenbanken häufig nur sich wiederholende Links zur internen Navigation enthalten, hätte ihr Ausschluss meist keine weitreichenden Auswirkungen auf die Ergebnisse der Linksammlung. Problematischer ist der gegenwärtige Trend, Websites ganz oder in Teilen dynamisch zu organisieren, um Speicherplatz zu sparen. Wird ein interner Link auf einer dynamisch angelegten Website aufgerufen, wird eine damit verknüpfte Datenbankabfrage gestartet und die Ergebnisse als neue Webseite ausgegeben. Inhalte, die zu mehreren Seiten gehören (z.B. verschiedenen Projektprofile oder Mitarbeiterlisten), müssen so nur einmal in der Datenbank gespeichert werden (Thelwall 2004a: 15f). Es ist bei der Nutzung

eines Crawlers somit genau darauf zu achten, ob dynamische Seiten ausgeschlossen werden und inwiefern dies bei den untersuchten Websites ins Gewicht fällt.

Der Grad der Vollständigkeit hängt jedoch nicht nur von den beschriebenen Gründen ab. Wie Cothey (2004) durch die Simulation der Erfassung einer vollständig bekannten Website durch verschiedene Crawlertypen zeigte, gehört zu den positiven Einflussfaktoren auch die Zahl der Startpunkte oder die automatische Prüfung, ob sich durch Trunkierung bereits gefundener Adressen zusätzliche Dokumente finden lassen. Dagegen kann die Begrenzung der Crawltiefe (auf Adressen mit maximal x Schrägstrichen hinter der Domain) zu eingeschränkten Ergebnissen führen. Gleiches gilt, wenn zusätzlich zum nicht erfassbaren *Invisible Web* prinzipiell erreichbare Teile von Websites bewusst nicht auf Links untersucht werden. Bei einigen Webcrawlern können zu ignorierende Bereiche definiert werden, um Linkschleifen oder umfangreiche experimentelle Webseiten oder Webstatistiken zu meiden, die ausschließlich interne Funktion haben und zumeist keine externen Links enthalten, deren Erfassung aber einen großen Aufwand bedeuten würde (vgl. Thelwall 2001c). Folglich ist, um die Replizierbarkeit zu gewährleisten, eine genaue Spezifikation der Crawler-Konfiguration nötig:

„[...] to interpret or compare analyses of crawl-based data, one must know the crawl policy under which the data was selected. That is, Web crawling as a description of a data selection technique is inadequate in that it encompasses too many hidden variables [...] and in the absence of additional qualification leads to inconsistent results“ (Cothey 2004: 1235).

Aufgrund der beschriebenen Limitierungen von Webcrawlern ist davon auszugehen, dass nur in bestimmten Fällen – vor allem bei kleineren, gut verlinkten Websites – eine vollständige Erfassung einer Website erreicht werden kann (vgl. Thelwall 2004a: 11). Aus diesem Grund sollte bei webometrischen Studien das Wissen darum präsent sein, dass man nur Aussagen für den öffentlich indizierbaren Bereich einer Website treffen kann:

„The phrase ‚publicly indexable pages‘ can be applied to the collection of pages in a site that can be found by a crawler by following links from the home page and obeying ethical issues. Using this phrase in reporting crawler coverage gives a convenient reminder that no claim is made that all pages in the crawled sites have been downloaded“ (Thelwall 2004a: 18).

Mit Webcrawlern gesammelte Daten stellen folglich Annäherungen an Websitestrukturen dar, die zwar keinen Anspruch auf Vollständigkeit erheben können, die jedoch in den meisten Fällen den entscheidenden Teil der untersuchten Website wiedergeben dürften. Da Programme mit verschiedenen Konfigurationen zu unterschiedlichen Ergebnissen kommen, ist die Deklaration der verwendeten Parameter die Voraussetzung für die nötige Reliabilität.

2.4.4 Vor- und Nachteile von Linkerhebungen mit Suchmaschinen

Im Vergleich zu Webcrawlern verfügen kommerzielle Suchmaschinen über einige evidente Vorzüge: Sie sind vergleichsweise einfach zu handhaben, ihre Nutzung ist kostenlos und Abfragen führen unmittelbar zu Ergebnissen, da auf einen umfassenden Korpus von bereits erhobenen Daten zurückgegriffen werden kann. Bei Fragestellungen, die eine Auswertung größerer Teile des World Wide Webs erfordern, sind Suchmaschinen nach wie vor die einzige Option (vgl. Li 2003). Die Erhebung eines großen Webbereichs, wie z.B. einer Top-Level-Domain, erfordert Rechen- und Speicherkapazitäten, über die die Grundlagenforschung in diesem Bereich üblicherweise nicht verfügt. Auch hier gilt es jedoch, bei der Erhebung die Begrenzungen im Blick zu haben, die in den folgenden Punkten a) bis j) skizziert werden:

a-e) Keine Erfassung des „Invisible Web“

Da auch Suchmaschinen ihre Daten mit Hilfe von Crawlerprogrammen sammeln, unterliegen sie ähnlichen Limitationen wie spezielle Forschungscrawler in Bezug auf das *Invisible Web* (vgl. Thelwall 2004a, für eine ausführliche Darstellung der Aspekte a) bis e) vgl. Abschnitt 2.4.3). Links können auch hier nur in HTML-Dokumenten identifiziert werden, da sie nur hier durch eindeutige Tags gekennzeichnet sind. Unverlinkte oder passwort- bzw. robots.txt-geschützte Bereiche des Webs können von den Webcrawlern, die Daten für Suchmaschinen sammeln, nicht durchsucht werden. Ebenso werden dynamische Seiten von ihnen häufig ignoriert. Eines der zentralen Ziele der fortlaufenden Weiterentwicklung der Suchmaschinentechnologie ist die Erschließung von Webdatenbanken, die einen Großteil des *Invisible Web* ausmachen. Bislang stehen die entscheidenden Schritte jedoch noch aus (vgl. Ru/Horowitz 2005).

Allerdings gibt es bezüglich des erfassten Webraums bereits jetzt einige Vorteile von Suchmaschinen gegenüber gezielt eingesetzten Webcrawlern: Da Suchmaschinen einen großen Teil des Webs abdecken, können sie auch Seiten finden, die zwar nicht über Links von der eigenen Website, wohl aber von anderen Websites aus zu erreichen sind. Außerdem können Suchmaschinen auf einen Speicher mit älteren Daten zurückgreifen, so dass sie teilweise auch Seiten kennen, auf die früher Links verwiesen haben, welche zum Zeitpunkt der Suche nicht mehr existieren. Darüber hinaus werden Suchmaschinencrawler bei Sperrungen durch das robots.txt-Protokoll zum Teil ausgenommen, damit die betreffenden Dokumente bei Internetrecherchen zu finden sind. Somit sind mit Hilfe von Suchmaschinen nicht notwendigerweise eine kleinere Zahl von Links zu finden, die in einer bestimmten Website enthalten sind, als mit gezielt eingesetzten Crawlerprogrammen (vgl. Thelwall 2001b; Smith/Thelwall 2002).

f) Eingeschränkte Aktualität

Suchmaschinen aktualisieren und erweitern mit Hilfe ihrer Webcrawler permanent ihre Datenbestände. Deshalb ist ein Teil der Ergebnisse aktuell, während andere vor längerer Zeit gefunden wurden und zum Zeitpunkt der Suche eventuell nicht mehr im Web zu finden sind. In einer Studie von Lawrence und Giles (1999: 108f) zeigte sich, dass bei den elf verglichenen Suchmaschinen zwischen 2,2 Prozent und 14,0 Prozent der angezeigten Treffer nicht mehr aufrufbar waren. Auch Bar-Ilan (2002b: 313) fand bei einer Untersuchung von sechs großen Anbietern über das gesamte Jahr 2000 heraus, dass bei den monatlichen Abfragen zwischen 6,1 Prozent und 22,6 Prozent aller Treffer technisch unkorrekt waren, d.h. sie waren entweder nicht aufrufbar oder enthielten nicht (mehr) die gesuchten Begriffe.

Eine neuere Studie zu täglich aktualisierten Webseiten in Deutschland (Lewandowski/Wahlig/Meyer-Bautor 2006) belegte grundsätzlich eine hohe Aktualität. Während überregionale Nachrichtenportale mit einem Durchschnitt von drei Tagen sehr häufig in den Suchmaschinenindizes aktualisiert wurden, wurden Portale zu regionalen Nachrichten, Wissenschaft und anderen Themen immerhin durchschnittlich alle sechs Tage aktualisiert. Problematisch ist die Existenz extremer Update-Lücken, die die Verlässlichkeit von zwei der drei untersuchten Anbieter erheblich schmälerte. Es ist anzunehmen, dass diese Aktualisierungsdefizite Websites, die keine Nachrichtenportale sind, noch stärker betreffen.

g) Unvollständige und ungleichmäßige Abdeckung des Web

Keine einzelne Suchmaschine verfügt über die Kapazität, das gesamte Web erfassen zu können. Vergleicht man die Resultate gleicher Anfragen bei verschiedenen Suchmaschinen, wird deutlich, wie gering und wie unterschiedlich sie das World Wide Web abdecken. Häufig zeigen die Ergebnislisten nur geringe Überschneidungen (vgl. Cronin et al. 1998; Snyder/Rosenbaum 1999; Lawrence/Giles 1999; Björneborn/Ingwersen 2001; Bar-Ilan 2001). Lawrence und Giles (1999: 108) schätzen, dass die von ihnen untersuchten elf Suchmaschinen insgesamt etwa 42 Prozent aller Seiten im Netz abdecken, jede einzelne dagegen nur zwischen 2,2 Prozent und 16,0 Prozent. Bar-Ilan (2001: 20) fand bei einem ähnlichen Vergleich heraus, dass nur sechs Prozent aller bei einer bestimmten Abfrage gefundenen Dokumente von allen sechs untersuchten Suchmaschinen ausgegeben wurden. Aufgrund der geringen Abdeckungsanteile einzelner Anbieter empfiehlt sie die Kombination mehrerer Suchmaschinen: Da sich die Ergebnisse häufig nur wenig überschneiden, kann die Abdeckung so erheblich gesteigert werden. Eine Studie von Smith (2003) zeigt, dass lokale Suchmaschinen selbst bei regionalspezifischen Themen eine noch geringere Abdeckung aufweisen als globale Anbieter. Dies liegt vermutlich daran, dass ihnen eine kleinere Datenbankgrundlage zur Verfügung steht, die außerdem aufgrund eines geringen Angebots von

Suchoptionen (limitierte Anzahl von Suchbegriffen erlaubt, wenige Operanden verfügbar), sowie deren spärlicher Dokumentation nicht optimal durchsucht werden kann. Angesichts der andauernden Expansion des World Wide Webs ist es fraglich, ob die Abdeckung sich in Zukunft erhöhen wird.

Bei den von Suchmaschinen erfassten Webbereichen handelt es sich nicht um einen zufälligen Ausschnitt: Die zur Datengewinnung eingesetzten Crawler finden Dokumente, auf die viele Links verweisen, mit höherer Wahrscheinlichkeit als unverlinkte. Deshalb werden neue Seiten in Ermangelung von Verweisen oft erst nach Wochen oder Monaten gefunden. Es handelt sich hierbei also um eine Art Matthäus-Effekt, insofern als Webdokumente durch Suchmaschinen erst für ein breiteres Publikum sichtbar werden und dadurch mit höherer Wahrscheinlichkeit weitere Verweise auf sie eingerichtet werden (Lawrence/Giles 1999: 109). Inlinkwerte sind neben zahlreichen anderen Faktoren ein elementarer Bestandteil der Rankingalgorithmen, mit denen die gefundenen Ergebnisse geordnet werden. Dabei führt die spezifische Gewichtung dieser Faktoren ebenfalls dazu, dass die Ausgabelisten verschiedener Anbieter sich stark unterscheiden. Da die Zahl der gezeigten Ergebnisse inzwischen meist limitiert ist (vgl. den folgenden Punkt h), führen auch sie dazu, dass die von den einzelnen Suchmaschinen gezeigten Ausschnitte des Webs zum Teil sehr unterschiedlich ausfallen (Jepsen et al. 2004; Bar-Ilan 2005a).

Ferner gibt es Hinweise darauf, dass auch Länder ungleichmäßig abgedeckt sind: Vaughan und Thelwall (2004) haben die Erfassung kommerzieller Websites aus vier Top-Level-Domains durch drei große Suchmaschinen verglichen. Es zeigte sich, dass zufällig ausgewählte Internetauftritte aus den USA zu einem wesentlich höheren Teil, die aus Singapur jedoch zu einem geringeren Teil abgedeckt waren als solche aus China oder Taiwan. Da alle untersuchten Websites aus den USA und aus Singapur englischsprachig waren, diejenigen aus China und Taiwan dagegen chinesischen Text enthielten, wurde die Priorisierung der englischen Sprache als Ursache für die ungleiche Abdeckung ausgeschlossen. Die Gründe für den Bias sehen die AutorInnen in dem durchschnittlichen Alter der Websites sowie in dem Phänomen, dass Links innerhalb eines Landes häufiger sind als internationale: Da die USA eine vergleichsweise lange Tradition der Webnutzung aufweisen und Webcrawler hier seit entsprechend langer Zeit Daten für Suchmaschinen sammeln, sind Websites, die Links auf neue Internetauftritte einrichten, bereits mit hoher Wahrscheinlichkeit indiziert. Folglich finden Webcrawler bei einem Aktualisierungsbesuch auf den älteren Websites auch die Verweise auf neuere. In anderen Ländern, wie z.B. China, gibt es dagegen einen hohen Anteil neuer, nicht-indizierter Websites. Die Chance, dass die Inlinks auf diese Seiten von bereits durch Suchmaschinen erfassten Seiten stammen, ist folglich geringer. Obgleich es sich nicht um einen vorsätzlichen Bias handelt, ist er doch eng mit der Strategie zur Daten-

gewinnung verknüpft und wird sich aufgrund des kontinuierlichen, weltweiten Wachstums des Webs nicht abschwächen.

In einer neueren Studie haben Barjak und Thelwall (2008) die Zahl der Webseiten und Inlinks für ein internationales Sample aus Internetauftritten von 400 Forschungsgruppen mit Google, Yahoo und MSN erhoben. Der Vergleich dieser Suchmaschinen ergab deutliche Unterschiede in den Trefferzahlen insgesamt sowie in dem Stellenwert einzelner Länder, die sich nicht allein durch das Alter der Websites erklären ließen. Starke Unterschiede zwischen den Suchmaschinen Google, Yahoo, MSN Live und Exalead im Hinblick auf die Abdeckung einzelner Länder und Weltregionen fanden auch Aguillo et al. (2008).

Eine andere Art unvollständiger Abdeckung ist die partielle Indizierung von Dokumenten: Vermutlich um Zeit und Kapazitäten zu sparen, erfassen einige Suchmaschinen gefundene Seiten nicht komplett, sondern nur in Teilen. So ist beispielsweise die Stichwortsuche über Google offensichtlich nur erfolgreich, wenn die gesuchten Begriffe in den ersten 101kB eines von Google indizierten Webdokuments auftreten (vgl. Bar-Ilan 2005b).

Speziell für die Abfrage von Links hat sich gezeigt, dass Suchmaschinen einen Teil der Verweise, die in von ihnen indizierten Dokumenten enthalten sind, nicht finden. Die untersuchten Anbieter zeigten zwischen 3 und 70 Prozent aller Links auf das E-Journal Cybermetrics nicht an, obwohl die Dokumente, in denen sie enthalten waren, in der Suchmaschinen-datenbank vorhanden waren, wie eine gezielte Suche nach ihnen ergab. Offensichtlich gibt es verschiedene Algorithmen für die Abfrage von Link-URLs und von Dokumenten-URLs. Die Abdeckung ist bei der Linksuche schlechter als sie gemäß der Abdeckung des Webs sein müsste (vgl. Bar-Ilan 2002a: 459).

h) Limitierung von Suchoptionen und angezeigten Ergebnissen

Kommerzielle Suchmaschinen erlauben Boolesche Verknüpfungen häufig nur in begrenztem Umfang und mit ungenügender Zuverlässigkeit (vgl. Snyder/Rosenbaum 1999). Grund für diese Beschränkung ist vermutlich der begrenzte kommerzielle Nutzen, da Boolesche Operanden für viele AnwenderInnen kein wichtiges Angebot darstellen (vgl. Bar-Ilan 2005b). Aus webometrischer Perspektive ist vor allem die zum Teil fehlende Kombinationsmöglichkeit von Booleschen Operanden und Linkabfrage ein Manko, durch die beispielsweise ein Ausschluss interner Links unmöglich wird (vgl. Li 2003). Google begrenzt Anfragen außerdem auf maximal 10 Suchbegriffe, was komplexere Kombinationen erschwert (Bar-Ilan 2005b: 94).

Ein weiterer Aspekt ist, dass die Abfrage von Inlinks bei einigen Suchmaschinen auf ganze Domains (wie www.uni-bielefeld.de) oder Top-Level-Domains (wie „.de“) beschränkt ist (vgl. Li 2003). Nicht immer können Links, die auf eine komplette URL (wie [53](http://www.uni-</p></div><div data-bbox=)

bielefeld.de/iwt/forschung/programm.html) oder eine Directory (wie www.uni-bielefeld.de/iwt/) verweisen, gezielt abgefragt werden. Außerdem sind sie nicht notwendigerweise in den allgemeinen Domainabfragen enthalten. Auch Subdomains (wie www.ub.uni-bielefeld.de) sind nur dann eingeschlossen, wenn eine Suche nach URL-Teilen (wie uni-bielefeld.de statt www.uni-bielefeld.de) möglich ist.

Eines der größten praktischen Probleme ist, dass alle großen Suchmaschinen die Zahl der einsehbaren Ergebnisse begrenzen (vgl. Jepsen et al. 2004; Bar-Ilan 2005b; Thelwall 2008b). Google, Yahoo und MSN beta nennen zwar für jede Abfrage die Zahl der gefundenen Treffer, zeigen jedoch nur die ersten 1.000 bzw. 500 Ergebnisse an. Die angegebene Anzahl der Ergebnisse ist dabei nicht verlässlich (vgl. Bar-Ilan 2002b; Qiu/Chen/Wang 2004; Chu 2005; Thelwall 2008b). Ruft man alle Seiten einer Trefferliste auf, die unter diesem Limit liegt, ist festzustellen, dass die Zahl der tatsächlich gefundenen Ergebnisse meist deutlich geringer ist als die angegebene Gesamtzahl. Der Wert allein genügt jedoch für den Großteil webometrischer Untersuchungen nicht: Immer wenn es darum geht, die URLs oder den Inhalt der Suchergebnisse zu analysieren, werden die kompletten Adressen der Treffer benötigt, um die entsprechenden Seiten speichern zu können. Auch ist es aus Gründen der Replizierbarkeit notwendig, die vollständigen Rohdaten zu besitzen, um sie ggf. anderen ForscherInnen zur Verfügung stellen zu können (vgl. Rousseau 1997, Bar-Ilan 2002b). Die bloße Trefferzahl kann somit allenfalls als grober Näherungswert für Studien mit sehr allgemeiner Fragestellung dienen.

Beschränkt man sich auf die Auswertung der tatsächlich gefundenen Treffer, ist zu berücksichtigen, dass es sich dabei nicht um eine zufällige, sondern eine sortierte Auswahl handelt. Aufgrund dieses Rankings können somit nur Schlussfolgerungen auf Basis der Ergebnisse gezogen werden, die die Suchmaschine als nützlich eingestuft hat (vgl. Thelwall 2004a: 20). Eine Möglichkeit, eine höhere Anzahl sichtbarer URLs zu erzielen, besteht darin, die gewünschte Abfrage in mehrere, komplementäre Abfragen aufzusplitten (vgl. Thelwall 2008b). Überschreitet beispielsweise die Suche nach einem Begriff das Outputlimit, kann sie für jede existierende Top-Level-Domain einzeln abgefragt und nachträglich aggregiert werden. Dies reduziert das Problem, kann es jedoch nicht vollständig lösen, da auch hier einzelne Abfragen das Limit überschreiten können.

i) Instabilität

Aufgrund der volatilen Struktur des World Wide Webs ist es nicht verwunderlich, dass die gleiche Abfrage zu unterschiedlichen Resultaten führen kann, sofern sie nicht zum exakt selben Zeitpunkt ausgeführt wird. Darüber hinaus zeigte sich jedoch in mehreren Studien,

dass Suchmaschinen Instabilitäten aufweisen können, die die Zahl der Ergebnisse verzerrten.

Bar-Ilan (1998/1999) zeigte, dass einige Suchmaschinen in großem Umfang Informationen verlieren: Bei wiederholten Abfragen wurden Seiten, die bei einer früheren Erhebung gefunden wurden und die gesuchten Begriffe immer noch erhielten, nicht mehr angezeigt.

Rousseau (1999) fragte 21 Wochen lang täglich dieselben drei Stichworte bei AltaVista und Northern Light ab. Während die Zahl der Treffer bei Northern Light durch häufige, kleinere Updates kontinuierlich wuchs, schwankten die Ergebnisse bei AltaVista erheblich von Tag zu Tag, und zwar in unterschiedlicher Weise für die drei Suchbegriffe. Eine größere Umgestaltung der Suchmaschine im Untersuchungszeitraum verringerte die Schwankungen deutlich, beseitigte sie jedoch nicht vollständig. Ähnliche Unregelmäßigkeiten zeigten sich auch bei HotBot und Northern Light (vgl. Bar-Ilan 2001; Jepsen et al. 2004).

Ursachen für diese Fluktuationen können technische Probleme oder begrenzte Speicherkapazitäten sein, die dazu führen, dass ältere Einträge gelöscht werden, wenn neue Daten hinzukommen (Bar-Ilan 1998/1999). Außerdem arbeiten Suchmaschinenanbieter zum Teil mit mehreren Datenbanken, die nicht immer auf dem gleichen Stand sind, sowie mit themenspezifischen Indizes, die jedoch nicht kontinuierlich verfügbar sind. Schließlich wird bei hoher Auslastung einer Suchmaschine mitunter die Zahl der verwendeten Suchalgorithmen reduziert, um Rechenressourcen zu sparen und alle Anfragen bewältigen zu können (vgl. Mettrop/Nieuwenhuysen 2001). All dies kann zu variierenden Resultaten führen. Neuere Studien mit Messwiederholungen weisen auf eine größere Stabilität hin, die zwischen den verfügbaren Anbietern jedoch nach wie vor unterschiedlich hoch ausfällt (z.B. Vaughan 2004; Vaughan/Thelwall 2004). Konstanz kann, anders als bei geschlossenen Datenbanken, nicht vorausgesetzt werden.

j) Intransparenz und Diskontinuität aufgrund kommerzieller Ausrichtung

Suchmaschinen haben nicht das Ziel, ein möglichst vollständiges Set von Ergebnissen zu liefern. Vielmehr hängt ihr kommerzieller Erfolg davon ab, schnell die relevantesten Treffer aus einem möglichst großen Teil des Webs zu präsentieren. Eine Optimierung bzgl. Vollständigkeit und Aktualität lohnt sich nicht mehr ab dem Punkt, an dem die Kosten dafür höher sind als der Nutzen durch verbesserte Suchergebnisse. Hinzu kommt, dass sich die Ergebnisse in kleinen Datenbanken schneller ermitteln lassen als in umfassenderen, was die Ressourcen der Anbieter schont und auch den NutzerInnen entgegenkommt: Sie haben zumeist kein Interesse daran, längere Zeit zu warten oder eine umfangreiche Liste von Treffern durchzusehen (vgl. Li 2003: 409f; Bar-Ilan 2001; Thelwall 2001c). Die Gründe für die in verschiedener Hinsicht begrenzte Abdeckung des World Wide Webs (vgl. Punkt g)) sind

somit nicht nur bei den technischen Möglichkeiten zu suchen. Es handelt sich dabei weniger um einen Mangel, als um eine wirtschaftsstrategische Entscheidung der Suchmaschinenanbieter (vgl. Lawrence/Giles 1999). Entsprechend erfolgt eine Fokussierung auf interessante, d.h. allgemein gefragte Themen (Bar-Ilan 2005b).

Für die Webometrie bieten Suchmaschinen also lediglich Sekundärdaten, die nicht nach Forschungskriterien erhoben wurden (vgl. Björneborn 2004):

„Researchers are using a tool for web-link analysis that is not specifically designed for this task. Search engines are designed for subject and content retrieval and not the examination of web links. The fact that engines can be used in this way is a by-product of the engines' main purposes“ (Snyder/Rosenbaum 1999: 381).

Die Suchalgorithmen kommerzieller Suchmaschinen unterliegen – ebenso wie die Abdeckung des Webs, die Indizierungsregeln, die Rankingprozeduren und die Aktualisierungshäufigkeit – dem Geschäftsgeheimnis. Die beschriebenen empirischen Untersuchungen deuten darauf hin, dass diese Parameter sich zwischen verschiedenen Anbietern erheblich unterscheiden (vgl. die bisher behandelten Punkte a)-i)). Da also wesentliche Bedingungen, unter denen die Resultate ermittelt werden, verborgen sind, kann ihre Qualität kaum eingeschätzt werden:

„Particularly troubling in this context is the fact that the search methodology is opaque. It is unclear how and to what extent the web was searched, so no reliable criteria exist to determine the relative accuracy of the results“ (Snyder/Rosenbaum 1999: 379).

ForscherInnen bleibt nur die Möglichkeit, anhand der ausgegebenen Ergebnisse Schlüsse auf die verwendeten Verfahren zu ziehen (vgl. Thelwall 2001c; Björneborn 2004: 29). Ohne detaillierte Kenntnisse über die gewählten Einstellungen der verwendeten Crawler sowie über die übrigen Verarbeitungsschritte ist jedoch keine hinreichende Reliabilität zu erreichen, da die Ergebnisse – unabhängig von Veränderungen im Web – nicht replizierbar sind (vgl. Cothey 2004).

Ein anderer Aspekt mangelnder Kontrolle über das Erhebungsinstrument ist, dass Änderungen jederzeit stattfinden können und oft unbemerkt bleiben. So können beispielsweise neue Suchstrategien oder Rankingprozeduren eingeführt werden, durch die sich u.U. auch die Abdeckung des Webs ändert. Auch Umstrukturierungen des Angebots sind permanent und ohne Ankündigung möglich: Suchoptionen (z.B. nach Sprache, nach Inlinks) können hinzukommen oder verschwinden und Ausgabelimits eingeführt werden (vgl. Snyder/Rosenbaum 1999; Bar-Ilan 2005b). Im Extremfall kann es auch passieren, dass Angebote unerwartet eingestellt werden – so geschehen im Frühjahr 2004, als sich der Suchmaschinenmarkt neu ordnete und die bis dahin zur Linkanalyse herangezogenen AltaVista, AllTheWeb und

Inktomi vom Konkurrenten Yahoo aufgekauft wurden (vgl. Pandia Search Engine News 2004). Mit den eigenständigen Suchmaschinen verschwand – vorübergehend – die Option der Linksuche. Zwar gibt es früher oder später meist Ausweichmöglichkeiten, aber dies erschwert eine zuverlässige Planung von Studien, insbesondere von Zeitreihen, ebenso wie Vergleiche verschiedener Studien. Diese Diskontinuität des Suchmaschinenangebots ist neben den kontinuierlichen Veränderungen im World Wide Web ein wesentlicher Grund dafür, dass Linkdaten immer binnen möglichst kurzer Zeit erhoben und unter Angabe des genauen Datums publiziert werden sollten (vgl. Bar-Ilan 2001: 10f; Bar-Ilan 2002b: 309).

Zusammenfassend lässt sich feststellen, dass Suchmaschinen über die Limitierungen von Webcrawlern hinaus eine Reihe von Schwächen aufweisen, die im Wesentlichen daraus resultieren, dass sie nicht für akademische, sondern für kommerzielle Zwecke konstruiert sind:

„Search engines are complex machines designed to give good quality information quickly to surfers. They are not designed to give high quality, accurate information to researchers based upon uniform, unbiased coverage of the web. Despite this, they are still an attractive source of link information, as long as users are aware of their limitations. These limitations encompass, and extend, the limitations of web crawlers“ (Thelwall 2004a: 20).

Somit handelt es sich um ein unaufwendiges Erhebungsinstrument im Hinblick auf Bedienung und Zeit, dessen Ergebnisse jedoch mit Vorsicht zu interpretieren sind. Forderungen nach der Entwicklung einer speziell auf webometrische Zwecke zugeschnittenen Suchmaschine sind trotz des damit verbundenen Aufwands die logische Konsequenz (vgl. Snyder/Rosenbaum 1999; Bar-Ilan 2002a; Thelwall et al. 2002; Bar-Ilan 2005b). Ihr Ziel wäre es, ausschließlich den wissenschaftlichen Teil des Webs zu umfassen, dessen Größe unbekannt ist.⁷ Durch eine Kombination technischer und menschlicher Selektionsverfahren könnte sie einer Qualitätskontrolle unterliegen und wissenschaftsspezifische Abfrageoptionen enthalten (z.B. eine Gliederung nach Disziplinen). Solange ein solches speziell auf die Erhebung wissenschaftlicher Daten ausgelegtes Instrument fehlt, können kommerzielle Suchmaschinen eine Datengrundlage für „Trockenübungen“ zur Entwicklung und Erprobung webometrischer Verfahren liefern:

„Until these high quality tools become a reality, we can and should continue to experiment with informetric techniques and methods, try to develop new ones, while always keeping in mind that the validity of the results is influenced by the currently available data collection tools“ (vgl. Bar-Ilan 2001: 27).

Einige der beschriebenen Begrenzungen nehmen mit zunehmender Weiterentwicklung der Suchmaschinentechologie ab. So hat sich die Aktualität und Stabilität der Daten in den letzten Jahren verbessert. Auch die Erfassung von Dokumenten, die nicht in HTML verfasst sind, sowie von dynamischen Seiten wird üblicher (vgl. Sherman/Price 2003). Nichtsdestotrotz können kommerzielle Suchmaschinen gegenwärtig nur als Provisorium zur Methodentwicklung dienen, und die mit ihrer Hilfe gewonnenen Daten sind inhaltlich nur begrenzt valide und reliabel.

Abschließend lässt sich zusammenfassen, dass beide Erhebungsinstrumente über Anwendungsbereiche verfügen, für die sie speziell geeignet sind: Suchmaschinen erlauben die Erhebung von Inlinks aus einem großen Teil des Webs, während begrenzt eingesetzte Webcrawler recht zuverlässig die Outlinks eines kleineren Webbereichs extrahieren. Durch beide Verfahren lassen sich die Linknetzwerke zwischen einer Auswahl von Websites (z.B. von Universitäten) ermitteln: Bei Suchmaschinen gibt es oft die Möglichkeit, gezielt die Links zwischen zwei Domains abzufragen, während aus Webcrawlerdaten nachträglich die Verweise auf bestimmte Ziele isoliert werden können.

Beide Werkzeuge verfügen über spezifische Limitierungen, die die Vollständigkeit und zum Teil auch die Validität und Objektivität der Erhebungen beeinträchtigt. Bei Linkanalysen ist es essentiell, sich darüber im klaren zu sein, dass man weder die Links des gesamten World Wide Webs noch die einzelner Websites in ihrer Gänze erfassen kann. Mit den gegenwärtig verfügbaren Mitteln lassen sich nur Näherungen an die gegebenen Linkstrukturen erreichen:

„In summary, ‘crawling a website’ is a relative rather than absolute concept“ (Thelwall 2004a: 20).

Dies gilt für Suchmaschinen aufgrund der beschriebenen zahlreichen, nicht neutralisierbaren Einflussfaktoren in wesentlich stärkerem Maß als für Webcrawler.

2.5 Die Nutzung der Webometrie zur Evaluation

In den letzten Jahren erfuhren quantitative bibliometrische Indikatoren und ihre Verrechnung zu Rankings eine steigende Beliebtheit. Grund hierfür ist vor allem die vordergründige Einfachheit, mit der entsprechende Daten erhältlich und vergleichbar sind. Sie bedienen das angesichts der zunehmend verknappenden Forschungsmittel steigende Bedürfnis der Politik nach einer transparenten und nachvollziehbaren Grundlage für die Bewertung von Forschung, die Entscheidungen Legitimität verleiht (vgl. Weingart 2005).

⁷ In einer früheren Studie wurde er anhand zufällig ausgewählter Internetauftritte auf sechs Prozent aller Websites geschätzt (vgl. Lawrence/Giles 1999: 107).

Auch ein „Webometrics ranking of world universities“ existiert bereits (vgl. InternetLab 2006b). Dennoch ist die Webometrie bislang kaum für Evaluationen genutzt worden. Dies liegt sicher teilweise daran, dass sie aufgrund ihres geringen Alters noch nicht besonders verbreitet ist, vor allem aber daran, dass sich die Verfahren nach wie vor in der Erprobungsphase befinden. In der webometrischen Forschung gibt es, wie in den vorangegangenen Abschnitten beschrieben, eine Reihe von offenen Fragen und ungelösten Problemen. Deshalb ist die Qualität dieser Erhebungen bisher nicht mit der von Zitationsanalysen vergleichbar und somit für evaluative Zwecke keinesfalls hinreichend (vgl. Scharnhorst/Thelwall 2005).

Zunächst sind hier Eigenschaften des Webs zu nennen. Seine Flüchtigkeit verhindert die Replizierbarkeit von Daten, was ihre Glaubwürdigkeit speziell im Kontext von Bewertungen untergräbt:

„[...] the potential impermanence of some Web citations may introduce problems for bibliometric or webometric research and for use of citation measures in evaluation – what confidence can promotion committees, or even the readers of this paper have that the Web citation counts reported are accurate?“ (Vaughan/Shaw 2003: 1321).

Die Archivierung des kompletten Datensatzes ist bei jeder webometrischen Studie unabdingbar, bietet letztlich jedoch keine Garantie für die Authentizität und Korrektheit des empirischen Materials. Außerdem erlaubt es das flexible Design von Websites und die Abwesenheit jeglicher Qualitätskontrolle, Webseiten und Links in nahezu beliebiger Weise zu generieren. Die Gestaltung von Seiten hat somit Einfluss auf die Linkdaten, so dass Webindikatoren manipulationsanfällig sind (vgl. Thelwall 2001b).

Darüber hinaus ist nicht abschließend geklärt, welche Aussagen auf der Grundlage von linkbasierten Maßen getroffen werden können und wie sich forschungsbezogene Links von anderen trennen lassen. Einigkeit besteht – nachdem die anfängliche Euphorie verflogen ist – darin, dass Links nicht als moderner Ersatz für Zitationen anzusehen sind (Vaughan/Shaw 2003):

„Some people suggest that electronic document facilities such as hypertext-linking will replace largely the ‚classical‘ way of citing. This is very unlikely. Hypertext-linkage as new type of citation is too anonymous, too superficial. It does not provide a clear recording of the use of other’s work“ (van Raan 2001: 60).

Entsprechend erscheint die Vorstellung, Zitationsindikatoren durch Linkmaße zu ersetzen derzeit als „plainly ridiculous“ (Thelwall 2001b: 1166). Wenngleich die gefundenen Korrelationen von Webindikatoren mit umfassenden Forschungsrankings wie dem RAE (vgl. Abschnitt 2.3.2) darauf hindeuten, dass Linkstrukturen einen Zusammenhang mit der Forschungstätigkeit von Hochschulen aufweisen, bleibt die Vielfalt an Funktionen, denen

Links dienen können und die nicht notwendigerweise einen Wissenschaftsbezug aufweisen, eine Quelle von Unwägbarkeiten. Der Vorschlag, vor jeder Evaluation alle relevanten Links einer Klassifikation zu unterziehen (Chu 2005), um ihre Aussagekraft sicherzustellen, ist aufgrund des enormen Arbeitsaufwands nicht praktikabel (vgl. Thelwall/Harries 2003).

Ein weiterer Punkt sind die Erhebungsinstrumente: Mit kommerziellen Suchmaschinen lassen sich zwar schnell und kostengünstig Linkdaten ermitteln. Da ihre Reliabilität aufgrund zahlreicher Faktoren – u.a. begrenzte Abdeckung, Instabilität, Inaktualität und vor allem geheime und variable Suchalgorithmen – eingeschränkt ist, sind Suchmaschinendaten als Bewertungsgrundlage ungeeignet (vgl. Bar-Ilan 2002a; vgl. Abschnitt 2.4.4). Dies gilt insbesondere für Evaluationen einzelner Personen (vgl. Cronin 2001; Bar-Ilan 2004b). Webcrawler decken nur die öffentlich indizierbaren Teile von Websites ab, liefern aber nachvollziehbare Ergebnisse. Auch hier kann es zu Verzerrungen kommen, wenn Teile einzelner Websites aufgrund der Gestaltung (Verwendung von Flash etc.) oder aufgrund von Sperrungen (durch Passwörter oder das robots.txt-Protokoll) nicht auf Links absuchbar sind (vgl. Thelwall 2001b; vgl. Abschnitt 2.4.3). Manuelle Kontrollen sind hier unabdingbar.

Neben Problemen der Erhebungsqualität gibt es grundsätzliche Bedenken gegen eine evaluative Nutzung von Webindikatoren:

„Taking advantage of the partial analogy situation-citation to assess an institutions' performance in the traditional meaning would be deeply misleading. The picture can be as different as a city's nightscape of neon signs and its roadmap. The related indicators, Web visibility and academic performance, are completely different species“ (Prime/Bassecouard/Zitt 2002: 306).

Internetauftritte von WissenschaftlerInnen oder Forschungseinrichtungen und die damit verbundenen Links erlauben – anders als Publikationen und Zitationen – insofern keine direkten Rückschlüsse auf die Qualität der Forschung, als die Präsentation von Ergebnissen auf Websites nur einen geringen Stellenwert einnimmt.

„Counts of links [...] represent predominantly an amalgamation of many different causes loosely related to research and various forms of informal scholarly communication. One consequence of this finding is that their use to evaluate research impact fails the validity test“ (Tang/Thelwall 2003: 441).

Von einer Beurteilung der Forschungsgüte anhand webometrischer Daten sollte also Abstand genommen werden. Linkmaße können jedoch Aufschluss geben über die Sichtbarkeit und Popularität einer wissenschaftlichen Einheit in anderen gesellschaftlichen Sektoren (z.B. bei Medien, Politik, Wirtschaft, NGOs, interessierten Laien etc.; vgl. Thelwall 2002c). Gleiches gilt für die internationale Vernetzung innerhalb der Wissenschaft (vgl. Kap. 3). Insbesondere Forschungsk Kooperationen zwischen Industrie- und Entwicklungsländern schlagen sich nicht notwendigerweise in gemeinsamen Publikationen nieder (vgl. Dahdouh-

Guebas et al. 2003). Wechselseitige Verweise im Netz könnten hier einen niederschwelligeren Ansatz bieten, indem sie auch Kontakte und Zusammenarbeit unterhalb des Niveaus formeller Publikationen abbilden.

Auch dies kann Gegenstand einer Evaluation sein. Webindikatoren bieten außerdem den Vorteil, dass sie besonders günstig und schnell zu erheben sind. Webdokumente können gelesen werden, sobald sie online sind. Anders als bei Publikationen vergeht weder Zeit für die Veröffentlichung von Dokumenten noch von späteren Beiträgen, in denen sie zitiert werden. Auch die Erhebung ist zeitnah möglich, da Webcrawler Links ab dem Zeitpunkt ihrer Entstehung finden – eine Zeitlücke wie bei der Erfassung von Artikeln in Datenbanken tritt nicht auf. Somit kann also auch die Arbeit von jüngeren WissenschaftlerInnen, die noch nicht über eine umfangreiche Publikationsliste verfügen, angemessen erfasst werden (Vaughan/Shaw 2003; Vaughan/Shaw 2005).

Zusammenfassend lässt sich feststellen, dass Webindikatoren sicher keinen Ersatz für andere szientometrische Indikatoren darstellen, da Weblinks nicht unmittelbar wissenschaftliche Qualität repräsentieren. Aufgrund der beschriebenen methodischen Mängel sollten sie außerdem mit äußerster Vorsicht eingesetzt werden. Insbesondere für Mittelzuweisungen oder individuelle Bewertungen stellen sie keine geeignete Grundlage dar. Denkbar ist jedoch, dass in den nächsten Jahren genauere Erkenntnisse über die Aussagekraft von Links gewonnen werden sowie zuverlässige Erhebungsstandards entwickelt werden, die Qualitätsanforderungen an die Erhebungsinstrumente sowie Ausschlussregeln für bestimmte Links oder Website-Bereiche umfassen.

Auf dieser Basis könnten Linkanalysen eine sinnvolle Ergänzung anderer Evaluationsindikatoren darstellen. Bei bestimmten Fragestellungen – z.B. bei der gesellschaftlichen Einbettung oder der internationalen Vernetzung – können Webindikatoren eine interessante *zusätzliche* Perspektive liefern, indem sie über die Sichtbarkeit und Attraktivität der jeweiligen Einrichtung im Web Aufschluss geben (vgl. Holmes/Oppenheim 2001; Thelwall 2001b; Thelwall/Tang 2003). Um in der Metaphorik von Prime, Bassecouard und Zitt (2002) zu bleiben: Nächtlige Leuchtreklamen zeichnen sicher ein spezielles Bild von einer Stadt, aber es ist durchaus lebensnah und eine gute Ergänzung zu einer nützlichen, aber zweidimensionalen Straßenkarte. Ebenso wie bibliometrische können auch webometrische Indikatoren in jedem Fall nur die Grundlage für eine informiert Peer-Review, also eine durch quantitative Daten gestützte Einschätzung durch ExpertInnen, bilden (van Raan 2003; Weingart 2005).

2.6 Zwischenfazit

Abschließend werden einige grundlegende Eigenschaften und Einsatzmöglichkeiten der Webometrie resümiert (Abschnitt 2.6.1) und klärenswerte Aspekte im Hinblick auf internationale Fragestellungen herausgestellt (Abschnitt 2.6.2).

2.6.1 Grundlegende Eigenschaften der Webometrie

Das World Wide Web lässt sich als spezifische Datenquelle begreifen, mit deren Hilfe sich neue Erkenntnisse über Wissenschaftsstrukturen gewinnen lassen. Seine einführende Beschreibung (Abschnitt 2.1) hat zahlreiche positive Aspekte aufgezeigt, die es zu einer wertvollen Datenquelle für die Wissenschaftsforschung machen. Zu ihnen zählen die kostenlose Nutzung des Webs, seine Aktualität, seine vielfältige Verwendung innerhalb der Wissenschaft, seine Offenheit gegenüber anderen sozialen Sektoren und der einfache Datenzugang. Die Vorstellung der Webometrie (Abschnitt 2.2) und die Eigenschaften der Internetauftritte wissenschaftlicher Einrichtungen (Abschnitt 2.3) zeigen jedoch zahlreiche Beschränkungen webometrischer Analysen auf, wie die mangelnde Qualitätskontrolle, die schwierige Unterscheidung von Wissenschaft und Nicht-Wissenschaft, die fehlende Replizierbarkeit, die Heterogenität von Dokumenten sowie die Diversität an Linkmotivationen. Hinzu kommen Defizite bei den derzeit verfügbaren Erhebungsverfahren (Abschnitt 2.4). Insbesondere im Hinblick auf die Nutzung der Webometrie zu evaluativen Zwecken ist somit Zurückhaltung geboten (Abschnitt 2.5).

Die entscheidende konzeptionelle Beschränkung der Webometrie liegt darin, dass das Erstellen von Webdokumenten – anders als das Verfassen von Publikationen – nicht zum Kern wissenschaftlichen Arbeitens zählt. Während die Nutzung des Webs inzwischen zum Forschungsalltag gehört, trifft dies nicht für die Gestaltung von Webseiten und das Veröffentlichen im Netz zu. Das Einrichten von Links ist – anders als das Zitieren – nach wie vor keine zentrale wissenschaftliche Aktivität. Scharnhorst und Thelwall (2005: 1521) prophezeien webbasierten Link- und Netzwerkanalysen deshalb eine Zukunft als „supporting actors to mainstream citation analysis“ und nehmen an, dass sie dieser Nebenrolle erst entwachsen können, wenn sich das Web weiterentwickelt und einen bedeutend höheren Stellenwert einnimmt, wenn z.B. digitale Bibliotheken üblicher werden als solche mit Schwerpunkt auf Printmedien.

Unter den gegenwärtigen Bedingungen ist die Webometrie deshalb in erster Linie für explorative Zwecke geeignet. Wenngleich die Linkanalyse keinen futuristischen Ersatz für die Zitationsanalyse darstellt, bietet sie eine besondere Perspektive auf die Wissenschaft als soziales System. In der Erschließung dieses spezifischen Zugangs zur Wissenschaft liegt

das Potential der Webometrie. Um es entfalten zu können, gilt es, die Grundlagen webometrischer Untersuchungen sorgfältig zu untersuchen und eventuelle Unsicherheiten zu klären. Nachdem in diesem Kapitel die zentralen Eigenschaften und Verfahren der Linkanalyse beschrieben wurden, geht es daher in den folgenden Teilen darum zu klären, welche besonderen Anforderungen mit internationalen Linkanalysen verbunden sind und inwiefern die Webometrie ihnen Rechnung tragen kann.

2.6.2 Relevante Aspekte im Hinblick auf internationale Fragestellungen

Abschließend für Kapitel 2 gilt es nun, vor dem Hintergrund der folgenden empirischen Erhebung relevante Aspekte internationaler Fragestellungen zu benennen (Abschnitt 2.5.2).

Zunächst stellt sich die Frage, inwiefern die Erhebungswerkzeuge für internationale Vergleiche geeignet sind. Webcrawler durchsuchen jede Website mit Hilfe einer stabilen Prozedur, die unabhängig von dem Land oder der Sprache abläuft, in denen der Internetauftritt gestaltet wurde. Es führt jedoch zu ungleichmäßigen Abdeckungsraten, wenn in einigen Regionen größere Teile des Webs nicht zu erfassen sind. So gibt es beispielsweise Hinweise darauf, dass in China ein substantieller Anteil der Websites mit modernen, nicht crawlbaren Flash-Applikationen gestaltet ist (Thelwall/Tang 2003). Suchmaschinen decken anscheinend Länder mit einer langen Tradition der Webnutzung, wie z.B. die USA, besser ab als solche mit einem hohen Anteil neuer Internetauftritte, wie z.B. China (vgl. Vaughan/Thelwall 2004). Zudem gibt es Anhaltspunkte dafür, dass die Suchalgorithmen großer Anbieter auf die englische Sprache zugeschnitten sind und den Eigenheiten anderer Sprachen (z.B. durch Berücksichtigung von Akzenten) nicht in angemessener Weise Rechnung tragen, was die Qualität der Ergebnisse beeinträchtigt (vgl. Bar-Ilan/Gutman 2005). Letzteres hat vermutlich keinen Effekt auf die Linksuche. Eine genaue Prüfung möglicher Einflüsse auf internationale Fragestellungen ist im Rahmen dieser Studie nicht möglich, da die Prüfung der Erhebungsinstrumente ein spezifisches Design erfordert. Es ist jedoch zu vermuten, dass sie die ohnehin begrenzte Qualität von Linkdaten – insbesondere bei der Verwendung von Suchmaschinen – weiter mindern. Die speziellen Optionen, die Suchmaschinen für die Untersuchung internationaler Fragestellungen anbieten und deren Qualität daher äußerst relevant ist, wird im folgenden Kapitel im Zusammenhang mit möglichen Indikatoren für internationale Vernetzung behandelt. Gleiches gilt für die Nutzungsmöglichkeiten von landesspezifischen Top-Level-Domains (vgl. Abschnitt 3.3).

Genaueres Wissen über die Eigenschaften von Websites und Links im wissenschaftlichen Kontext ist die unbedingte Voraussetzung für eine szientometrische Linkanalyse. Schließlich fehlte ansonsten die Kenntnis darüber, was mit Hilfe der gesammelten Linkdaten überhaupt gemessen werden kann. Auf Basis der allgemeinen Beschreibung der Webometrie in diesem

Kapitel lassen sich die folgenden kritischen Aspekte identifizieren, die im Hinblick auf internationale Fragestellungen besonderes Gewicht haben und der Klärung bedürfen:

- Eine Bedingung für die Analyse der internationalen Vernetzung wissenschaftlicher Einrichtungen im Web ist die Vergleichbarkeit der Nutzung von Websites durch wissenschaftliche Einrichtungen in verschiedenen Ländern, über die bisher keine systematischen Ergebnisse vorliegen.
- Daneben ist zu klären, inwiefern Links im nationalen und im internationalen Kontext ähnliche Funktionen aufweisen.
- Im Zusammenhang mit offenen internationalen Linkabfragen, die ohne Beschränkung auf Verweise zwischen wissenschaftlichen Einrichtungen durchgeführt werden, ist eine relevante Frage, wie häufig Links ohne Wissenschaftsbezug auftreten.
- Eng verknüpft damit ist die Frage, welche Möglichkeiten zur Abgrenzung von Wissenschaft in verschiedenen nationalen Kontexten gegeben sind, z.B. über entsprechende Top- oder Second-Level-Domains.

Diese Aspekte werden in Kapitel 4 in die Zusammenstellung von Leitfragen für die Fallstudie eingehen. Zuvor wird in Kapitel 3 analysiert, welche Rolle internationale Kommunikation in der Wissenschaft spielt und wie die Webometrie zu ihrer Untersuchung beitragen kann.

3. Die webometrische Untersuchung internationaler Zusammenhänge in der Wissenschaft

Kommunikation und Kooperation mit anderen ForscherInnen sind essentielle Bestandteile wissenschaftlichen Arbeitens. Innerhalb der *scientific community* werden aktuelle Erkenntnisse verbreitet, neue Ideen entwickelt, Fragen und strittige Punkte diskutiert. Jedes ihrer Mitglieder profitiert von den Befunden, den Anregungen und der Kritik der Gemeinschaft. Die dabei genutzten Kommunikationswege unterscheiden sich im Grad ihrer Formalisierung und der Öffentlichkeit: Während sich Zeitschriften an die gesamte Fachöffentlichkeit richten und durch einheitliche Gestaltung der Artikel und institutionalisierte Qualitätskontrolle einen stark formalisierten Modus der wissenschaftlichen Kommunikation darstellen, kann informelle Kommunikation in Gestalt von Gesprächen oder Korrespondenz auf individueller Ebene erfolgen, sie kann sich aber auch – beispielsweise durch Mailing-Listen – auf größere Gruppen erstrecken. Konferenzen schaffen eine Verbindung zwischen der öffentlichen, eher formellen Präsentation neuer Erkenntnisse – z.T. mit anschließender Veröffentlichung – und der Möglichkeit, informelle Gespräche zu führen, Anregungen auszutauschen und Kontakte zu knüpfen.

Die Varianten wissenschaftlicher Kommunikation sind vielfältig, fest steht jedoch, dass sie insgesamt einen hohen Stellenwert innehaben und als fruchtbar für die wissenschaftliche Arbeit gelten. Die zentrale Rolle, die die Kommunikation von Ideen und Befunden in der Wissenschaft spielt, rührt auch daher, dass wissenschaftliches Wissen erst durch die (fach-) öffentliche Äußerung zu existieren beginnt: Eine brillante, aber nur im Kopf einer einzelnen Person bestehende Erkenntnis ist in der Gesellschaft nicht vorhanden (Schott 1993). Das Kommunizieren neuer Ergebnisse und Theorien, z.B. durch Publikation, dient somit der Weitergabe von Information; es besitzt aber auch soziale Relevanz für die *scientific community*:

„Kommunikation bildet die Basis jeglicher Tätigkeit im wissenschaftlichen Bereich, da sie drei grundlegende Aufgaben wahrnimmt: die Erzeugung und Verbreitung von Wissen und die Schaffung von Reputation. Damit ist der Akt des Kommunizierens sowohl auf der *kognitiven* als auch auf der *sozialen* Ebene von zentraler Bedeutung. Einerseits existiert Wissen [...] ohne die Weitergabe an andere ForscherInnen im wissenschaftlichen Sinn eigentlich nicht. Andererseits kommt dadurch der Sicherung von intellektuellem Eigentum eine wesentliche Funktion zu, da Reputation auf diesem aufbaut [...]“ (Felt/Nowotny/Taschwer 1995: 66f; Hervorhebungen im Original).

Erst durch die öffentliche Kommunikation kann ein wissenschaftlicher Diskurs – einschließlich der kritischen Überprüfung von Erkenntnissen, der Replikation empirischer Daten und der Zuschreibung von Reputation – einsetzen.

Auch Kooperation ist ein weiter Begriff: Sie umfasst die wechselseitige Vermittlung von theoretischem, empirischen, methodologischen oder technischem Wissen ebenso wie die Weiterentwicklung von Ideen, das Herstellen von Kontakten, die Bereitstellung von Ausrüstung oder Daten, das gemeinsame oder arbeitsteilige Bearbeiten von Projekten, die Veröffentlichung von Ergebnissen in Koautorenschaft sowie das kritische Kommentieren dieser Produkte (Laudel 2002). Auch der Austausch von Studierenden und ForscherInnen, die Koordination wissenschaftlicher Aktivitäten in einem Forschungsprogramm sowie die Organisation von Konferenzen sind im weiteren Sinn unter dem Begriff Forschungskooperation zu subsumieren (van Raan 1997). Zu den Vorteilen gemeinschaftlichen Arbeitens zählt, dass sich so Projekte mit komplexen Fragestellungen und in geringer Zeit bearbeiten lassen. Mehrere Personen können außerdem ihre Expertise und ihre Kontakte bündeln, einander neue Fertigkeiten beibringen, Fehler mit höherer Wahrscheinlichkeit entdecken und mehr Möglichkeiten wahrnehmen, ihre Ergebnisse zu präsentieren. Ein gut zusammengesetztes Team trägt ferner dazu bei, die Motivation aufrecht zu erhalten (Bordons/Gómez 2000; Beaver 2001). Auch wer keine arbeitsteiligen Projekte durchführt, ist auf den Austausch mit anderen angewiesen: JedeR einzelne ForscherIn besitzt ein mehr oder weniger großes und dichtes Netzwerk aus *peers*. Sie bieten nicht nur fachlichen Rat, sondern dienen auch als Referenzrahmen zur Einschätzung der Originalität und Qualität der eigenen Ideen und Ergebnisse (Schott 1991; Schott 1993). Der hohe Stellenwert von Kooperation innerhalb der *scientific community* zeigt sich auch darin, dass WissenschaftlerInnen durch ihre Erkenntnisse zwar Anerkennung für ihre Leistung in Form persönlicher Reputation gewinnen, jedoch keine über die intellektuelle Urhebererschaft hinausgehenden Eigentumsrechte daran erwerben. Dem Ethos der Wissenschaft gemäß gelten Forschungsergebnisse – anders als technologische Entwicklungen – als „Produkt sozialer Zusammenarbeit und werden der Gemeinschaft zugeschrieben“ (Merton 1972: 51).

Der Austausch zwischen WissenschaftlerInnen kann viele Formen annehmen: Er kann sich in informellen persönlichen Kontakten ebenso manifestieren wie in einem formellen *Memorandum of Understanding* zwischen Forschungseinrichtungen. Er kann flüchtig sein oder die Grundlage für eine stabile Arbeitspartnerschaft bilden. Er kann zwischen gleichen oder ungleichen PartnerInnen erfolgen. Kommunikation und Kooperation können innerhalb einer Forschungsgruppe stattfinden, sie können aber auch Disziplinen- und Ländergrenzen überschreiten.

In dieser Dissertation geht es speziell um die empirische Erforschung internationaler Vernetzung innerhalb der Wissenschaft und um die Frage, inwiefern webometrischer Verfahren hierzu einen Beitrag leisten können. Zunächst ist daher zu klären, welche Rolle grenzüberschreitende Kontakte in der Forschung spielen (3.1). Darüber hinaus wird in diesem Kapitel diskutiert, welchen Einfluss das Internet auf die internationale Forschungskommunikation hat

(3.2). Gegenstand des darauf folgenden Abschnitts (3.3) ist eine Auswertung der bisher erschienenen, länderübergreifenden webometrischen Studien, die sich auf die jeweiligen Ziele sowie die verwendeten Indikatoren erstreckt. Das Zwischenfazit (3.4) enthält Schlussfolgerungen für die Gestaltung der empirischen Fallstudie, die in den Kapiteln 4, 5 und 6 vorgestellt wird.

3.1 Der Stellenwert von Internationalität in der Wissenschaft

Wissenschaft gilt als genuin international. Ein oft angeführtes Argument ist, dass die gemeinsame Suche nach universeller Wahrheit eine globale *scientific community* eint (Storer 1970; Crane 1972; Elzinga 1996). Es spiegelt die Mertonsche Universalitätsnorm wider, derzufolge wissenschaftliche Erkenntnisse unabhängig von dem persönlichen oder kulturellen Hintergrund der forschenden Personen zu prüfen und zu beurteilen sind (Merton 1972). Die Wissenschaftsforschung hat jedoch in den letzten Jahrzehnten gezeigt, dass die Forschung variierenden sozialen Bedingungen unterliegt und von ihnen geprägt wird. Die Vorstellung einer einheitlichen, globalen Wissenschaft, in der sich Zusammenarbeit allein an fachlichen Interessen ausrichtet, scheint vor diesem Hintergrund fraglich. Beide Auffassungen werden von diversen Anhaltspunkten gestützt, wie in den Abschnitten 3.1.1. und 3.1.2 ausgeführt wird.

3.1.1 Die internationale Ausrichtung der Wissenschaft

Der Austausch unter WissenschaftlerInnen über Länder- und Sprachgrenzen hinweg ist seit Jahrhunderten üblich, sowohl durch Veröffentlichungen als auch durch persönliche Korrespondenz. Bereits zu Beginn der modernen Naturwissenschaft im 16. Jahrhundert erlaubte die lateinische Sprache als *lingua franca* die weitreichende Kommunikation und Diskussion wissenschaftlicher Thesen und Befunde innerhalb Europas (Stichweh 1996). Des Weiteren umfasst der Themenbereich der Natur-, Sozial- und Geisteswissenschaften den gesamten Globus: In Disziplinen wie Biologie, Geologie, den Sprachwissenschaften oder Ethnologie haben Feldforschung und Forschungsreisen in alle Regionen der Welt eine jahrhundertelange Tradition, in der Astronomie ist die ganze Erde als Planet Gegenstand der Forschung. Auch internationale Kopublikationen gibt es bereits annähernd ebenso lange wie moderne Fachzeitschriften: Bibliometrische Studien zeigen, dass die ersten gemeinschaftlich veröffentlichten Artikel in den 1660er Jahren erschienen; Publikationen mit Autoren aus verschiedenen Ländern folgten wenig später (Beaver/Rosen 1978). Die Autoren fassen dies als Zeichen der Professionalisierung von Wissenschaft im Rahmen der Wissenschaftlichen Revolution auf. Im 18. und 19. Jahrhundert nahmen gemeinschaftliche Veröffentlichungen jedoch nur einen geringen Anteil an allen Publikationen ein (Beaver/Rosen 1978; Wagner-

Döbler 2001). Um die Wende vom 19. zum 20. Jahrhundert entstanden zahlreiche internationale Fachgesellschaften und erste internationale Konferenzen fanden statt. Dies betraf insbesondere feldforschung-intensive Disziplinen, die einen starken Bedarf an Abstimmung in Bezug auf Nomenklatur, Methoden, Erhebungseinheiten etc. aufwiesen (Crawford/Shinn/Sörlin 1993).

In der jüngeren Vergangenheit finden sich Hinweise auf eine deutliche Intensivierung sowohl nationaler als auch internationaler Kooperation: Während des gesamten 20. Jahrhunderts stieg der Anteil an Publikationen mit mehreren AutorInnen in verschiedenen europäischen Zeitschriften deutlich an (Beaver/Rosen 1979; Wagner-Döbler 2001). Zwischen 1988 und 2003 sank der Anteil von Publikationen mit nur einer einzigen institutionellen Adresse weltweit stark (von 60,3% auf 41,3%), während zugleich der Anteil von Publikationen mit mehreren Adressen innerhalb eines Landes (von 31,5% auf 39,1%) sowie insbesondere der Anteil von Publikationen mit Adressen in mehreren Ländern deutlich anstiegen (von 8,2% auf 19,5%) (NSB 2006: Abb. 5-39). Dieser Trend zu mehr internationalen Kopublikationen zeigte sich in den letzten Jahrzehnten in vielen Weltregionen und Fachgebieten (Luukkonen/Persson/Sivertsen 1992; Glänzel 2001; Engels/Ruschenburg/Weingart 2005; Wagner/Leydesdorff 2005), wobei sich die einzelnen Länder und Disziplinen deutlich im Anteil internationaler Kopublikationen an allen nationalen bzw. disziplinären Veröffentlichungen unterschieden (Frame/Carpenter 1979; Luukkonen/Persson/Sivertsen 1992; Leclerc/Gagné 1994; Bordons et al. 1996; Glänzel 2001; EC 2003: 303f; Glänzel/Schubert 2005; NSB 2006: Abb. 0-19). Die Zunahme an internationalen Kopublikationen geht mit einer Tendenz zu Veröffentlichungen in englischer Sprache einher (van Raan 1997). So waren im Jahr 2005 98,7 Prozent der im Science Citation Index (SCI) erfassten Publikationen auf Englisch erschienen (Leydesdorff 2006).

Verschiedene Gründe für diesen Trend zur Internationalisierung lassen sich identifizieren: Die zunehmende (sub-)disziplinäre Differenzierung in der Wissenschaft führt zur Ausweitung der Suche nach FachkollegInnen über das eigene Institut und die eigene Nation hinaus (Stichweh 1996; Bordons/Gómez 2000). Internationale Fachgesellschaften unterstützen den länderübergreifenden Zusammenhalt einzelner Disziplinen. Auch von politischer Seite verfolgen diverse Programme auf nationaler und supranationaler Ebene das Ziel, die internationale Zusammenarbeit in der Wissenschaft zu stärken. Viele Nationen, die mit steigenden Forschungskosten ohne äquivalente Steigerung der Forschungsgelder konfrontiert sind, erhoffen sich von Kooperationsprojekten höhere Produktivität und sparsameren Einsatz von Mitteln durch gemeinschaftliche Finanzierung. Dies betrifft insbesondere, aber nicht ausschließlich, Forschungsfelder, die der *big science* zuzurechnen sind, wie Teile der Physik oder der Erd- und Weltraumwissenschaften (van Raan 1997). Einen regionalen Kooperationsanreiz bietet die Forschungsförderung der Europäischen Union, die die Beteiligung

mehrerer Länder voraussetzt. Politisch gefördert wird in der Folge der globalen Umweltkonventionen auch *capacity building* in Entwicklungsländern. Insbesondere durch die Biodiversitätskonvention wurden die Rechte von Entwicklungsländern gestärkt, über ihre Artenvielfalt selbst zu verfügen. Auf dieser Grundlage kann beispielsweise die Erteilung von Forschungsgenehmigungen an die Bedingung der Kooperation mit einheimischen Einrichtungen geknüpft werden. Dies hat insbesondere Schwellenländern zu einer stärkeren Beteiligung an internationalen Forschungsprojekten verholfen (Engels/Ruschenburg 2006).

Schließlich erleichtern wissenschaftsexterne Faktoren die internationale Kooperation, wie die technische und ökonomische Entwicklung: Das heute vergleichsweise unaufwendige und günstige Reisen führt dazu, dass Forschungsreisen oder der Besuch internationaler Tagungen keine Seltenheit mehr sind (Katz/Martin 1997). Die modernen Kommunikationsmittel tragen ebenfalls dazu bei, dass räumliche Entfernung an Bedeutung verloren hat, indem sie eine Intensivierung der internationalen Vernetzung ermöglichen (Stichweh 1996; Bordons/Gómez 2000). Dieses Argument wurde interessanterweise bereits mit Bezug auf frühere Entwicklungen angewendet: So wurden verbesserte Reise- und Kommunikationsmöglichkeiten durch die Ausbreitung des Eisenbahnnetzes in Europa und den Einsatz von Dampfschiffen als Voraussetzung für die Entstehung der ersten internationalen Fachgesellschaften und der Veranstaltung der ersten internationalen Konferenzen gegen Ende des 19. Jahrhunderts angeführt (vgl. Crawford/Shinn/Sörilin 1993: 13). Mit Fax und E-Mail wurden auch jüngere Entwicklungen als Grund für zunehmende internationale Kooperation angeführt, bevor das Web verfügbar war und die Kommunikationsmöglichkeiten noch einmal deutlich erweitert hat (Schott 1991; Luukkonen/Persson/Sivertsen 1992).

3.1.2 Die nationale Ausrichtung der Wissenschaft

Trotz dieser Anhaltspunkte stellt sich die Frage, inwiefern grundsätzlich die Universalität und Einheitlichkeit der Wissenschaft angenommen werden kann. Schließlich muss nicht nur der kulturell variierenden Sozialisation ein gewisser Einfluss zugestanden werden, sondern auch den politischen Prioritäten der einzelnen Staaten (Crawford/Shinn/Sörilin 1993). Denn nahezu alle bedeutenden Organisationen der Grundlagenforschung, wie Universitäten und andere Forschungseinrichtungen, sind strukturell in einer einzelnen Nation verankert (Stichweh 1996). Für internationale Homogenität sorgt dabei die Übernahme des Modells westlicher Forschungs- und Bildungssysteme durch die meisten Länder der Erde (Stichweh 1996; Drori et al. 2003; Meyer/Ramirez 2005), das im Hinblick auf soziale Normen eine kosmopolitische Einstellung der ForscherInnen beinhaltet, die sich aus der angenommenen Universalität des Wissens ergibt und die die Aufnahme kollegialer Kontakte über große Entfernungen hinweg begünstigt (Schott 1993).

Dennoch besteht in der nationalen Forschungspolitik ein großer Spielraum bei der Ausgestaltung des Wissenschaftssystems. Er entsteht nicht zuletzt dadurch, dass auch die Forschungsförderung überwiegend national organisiert ist, wenn man von Ausnahmen wie der übernationalen EU-Förderung absieht, die nur einen geringen Anteil der weltweiten Forschungsmittel ausmachen. Die nationale Forschungspolitik bestimmt nicht nur den Umfang der möglichen Forschung über die zur Verfügung gestellten finanziellen Mittel, sondern beeinflusst durch Evaluationskriterien und Schwerpunktprogramme auch die Ausrichtung der Wissenschaft.

An dieser Stelle wirkt sich auch das globale Ungleichgewicht enorm aus, das in der Wissenschaft noch stärker ausfällt als in der Wirtschaft (Schott 1991): Hochentwickelte, ressourcenstarke Forschungssysteme finden sich nur in wenigen Ländern, die entsprechend den Output an Veröffentlichungen dominieren. Allein aus den G8-Staaten stammten im Jahr 2003 62,5 Prozent⁸ aller in Science Citation Index und Social Science Citation Index (SSCI) enthaltenen Publikationen (eigene Berechnung auf Grundlage von NSB 2006, Tabelle 5-42). Stärker als bei der Anzahl der Publikationen drückt sich die Dominanz der Industrieländer in der Verteilung der Zitationen aus: Von allen Zitationen, die in den ISI-Datenbanken bis zum Jahr 2002 auf Publikationen der Jahrgänge 1997 bis 2001 verweisen, entfallen allein 49,43 Prozent resp. 39,3 Prozent⁹ auf Artikel von Forschungseinrichtungen in den USA und in der EU-15 (King 2004, Tabelle 1). Auch internationale Koautorenschaftsnetzwerke konzentrieren sich auf wenige Länder, wie eine Analyse der Publikationen aus dem Jahr 2000 im SCI zeigt: Den Kern internationaler Koautorenschaftsnetzwerke bilden acht – ausschließlich westliche – eng untereinander vernetzte Staaten (Wagner/Leydesdorff 2005).

Einzelne Schwellenländer schließen zu diesen Industrienationen auf. So kamen im Jahr 2003 10,5 Prozent aller in SCI und SSCI enthaltenen Artikel von Forschungseinrichtungen in China, Südkorea, Indien, Taiwan und Brasilien (eigene Berechnung auf Grundlage von NSB 2006, Tabelle 5-42). In diesen Ländern lagen die Ausgaben für Forschung und Entwicklung 2003 im Verhältnis zum Bruttoinlandsprodukt (BIP) zwischen 1,0 Prozent (Brasilien) und 2,5 Prozent (Südkorea). Ihre relativen FuE-Investitionen bewegen sich somit in einem ähnlichen Spektrum wie die der Industrieländer, die zwischen 0,6 Prozent (Griechenland) und 4,0 Prozent (Schweden) ausmachen (vgl. UNESCO 2006; keine aktuellen Daten für Indien).

⁸ Diese Daten aus NSB 2006 sind durch *fractional counting* berechnet worden. Dies bedeutet, dass Artikel, an denen Forschungseinrichtungen in mehreren Ländern beteiligt waren, den jeweiligen Staaten nur anteilig zugeschrieben, so dass sich die Anteile aller Nationen zu 100 Prozent aufsummieren.

⁹ Die hier genannten Daten aus King 2004 basieren auf *whole counts*. Dabei werden Zitationen von Artikeln, die Adressen von Forschungseinrichtungen in mehreren Ländern aufweisen, allen beteiligten Nationen voll angerechnet. Alle nationalen Anteile zusammen ergeben daher einen Wert von mehr als 100 Prozent ergeben. Eine Addition der Prozentwerte mehrerer Länder ist nicht sinnvoll, da der Anteil gemeinsamer – und somit mehrfach gezählter – Kopublikationen unbekannt ist.

Für viele Entwicklungsländer scheint sich die Lage dagegen nicht deutlich zu verbessern: Sieht man von dem Grenzfall Israel ab, das nach der Länderklassifikation der Vereinten Nationen zu den Entwicklungsländern gehört, wenngleich es in internationalen Handelsstatistiken als Industrieland geführt wird und 2003 4,7 Prozent seines BIP für Forschung und Entwicklung aufwandte, investieren sie alle deutlich unter ein Prozent ihres BIP in die Forschung (vgl. United Nations Statistics Division 2006; UNESCO 2006). Sie verfügen somit absolut und relativ über einen Bruchteil der Forschungsressourcen von Industrieländern, was sich auf die Qualität der Ausbildungsmöglichkeiten auswirkt. Zudem leiden sie unter dem sogenannten *brain drain*, wenn im Ausland ausgebildete ExpertInnen nicht ins Heimatland zurückkehren (Schott 1998). Insbesondere die Sichtbarkeit von *least developing countries* in den Publikationsdaten scheint sich kaum zu erhöhen, da selbst kooperative Projekte – bei Durchführung der Feldforschung im Land durch einheimische ForscherInnen – nicht notwendigerweise in Kopublikationen münden, sondern KooperationspartnerInnen aus Industrieländern die Ergebnisse häufig exklusiv in internationalen Zeitschriften publizieren (Dahdouh-Guebas et al. 2003). Hochwertige Forschung setzt gewisse finanzielle Mittel voraus, nicht nur wenn es um die Beschaffung kostspieliger Forschungsgeräte geht. Bereits der Zugang zu internationalen Zeitschriften oder die Teilnahme an Kongressen kann in einigen Ländern nicht durch die Haushalte der Forschungseinrichtungen gedeckt werden. Die globale *scientific community* ist somit keine homogene Gemeinschaft. Sie ist vielmehr gekennzeichnet von regionalen Verdichtungen und internationalen Hierarchien (Schott 1993).

Neben der Forschungspolitik existieren weitere, eher mittelbare nationale Einflussfaktoren wie die Außen- und Sicherheitspolitik. Sie wirken sich indirekt auf die Bewilligung von gemeinsamen Projekten sowie über die Einreise-, Ein- und Ausfuhrbestimmungen auf die Möglichkeiten zum Besuch von ProjektpartnerInnen oder Konferenzen und zur Feldforschung aus. So hemmten die beiden Weltkriege im letzten Jahrhundert vor allem die Forschungsk Kooperationen in Europa. Der nachfolgende Kalte Krieg erlaubte den Austausch zwischen Ost und West nur in sehr eingeschränktem Umfang. Ein jüngeres Beispiel für die Behinderung von Forschungsaktivitäten ist die *homeland security*-Politik in USA nach den Terroranschlägen vom 11. September 2001. Die folgende Verschärfung der Einreisebestimmungen machte es ForscherInnen und Studierenden aus einigen Nationen (Kuba, China etc.) nicht einfach, Gastaufenthalte zu absolvieren. Der bürokratische Aufwand für die Beschaffung von Visa, die nur in einem Teil der Fälle erfolgreich verläuft, hat erhebliche administrative Kapazitäten in Forschungseinrichtungen gebunden und zahlreiche Gaststudierende von US-amerikanischen Hochschulen ferngehalten, was in diversen Fachzeitschriften kritisiert wurde (Greenberg 2002; Kennedy 2002; Brumfiel 2004). Die erhöhten Sicherheitsanforderungen bei Flugreisen erschweren zudem das Mitführen von Forschungsgeräten und -materialien für im Ausland forschende US-WissenschaftlerInnen. Nach wie vor

ist also davon auszugehen, dass die Rahmenbedingungen für Forschung – trotz des unter WissenschaftlerInnen nach wie vor verbreiteten universalistischen Ethos' – einem nicht zu unterschätzenden staatlichen Einfluss unterliegen, der sich auch auf Art und Ausmaß der internationalen Kooperation auswirkt.

3.1.3 Weiterer Forschungsbedarf

Forschungsbedarf ist somit in mehrfacher Hinsicht gegeben: Für die Wissenschaftsforschung ist es erstens von Interesse, das tatsächliche Ausmaß der grenzüberschreitenden Vernetzung empirisch zu erforschen. Zweitens sollten die Unterschiede zwischen verschiedenen Ländern und Disziplinen weiter analysiert werden. Drittens gilt es, die Art der internationalen Kontakte genauer zu spezifizieren. Ihr Spektrum reicht von Wahrnehmung und Rezeption von Forschungsaktivitäten über persönliche Bekanntschaft und direkten Austausch bis hin zur kooperativen Bearbeitung von Projekten.

Bisher gibt es erstaunlich wenige empirische Studien in diesem Bereich, die nicht ausschließlich auf Daten zur Koautorenschaft beruhen. Eines der seltenen Beispiele ist ein Artikel von Glänzel und Schubert (2005), die im Web of Science die Publikationen des Jahres 2000 und ihre Zitationen bis zum Jahr 2002 untersuchten und für 36 Länder die Verteilung von Referenzen und Zitationen auf das In- und Ausland ermittelten. Eine Schwierigkeit, die bei Kalkulationen dieser Art auftritt, ist, dass sowohl die zugrundegelegten Artikel selbst, als auch die zitierten und zitierenden Publikationen Adressen aus mehreren Ländern enthalten können, so dass die nationale Zuordnung recht komplex ist und grundlegenden Entscheidungen der Autoren unterliegt. Im Gegensatz dazu bietet Koautorenschaft eine vergleichsweise einfache Berechnung und gute Vergleichbarkeit. Sie gilt als "most obvious form of international collaboration – and the most easily measured" (Frame/Carpenter 1979: 481), deckt jedoch nur einen minimalen Bruchteil der tatsächlichen Bandbreite der internationalen Zusammenarbeit ab. Sie verengt die Perspektive auf ein bestimmtes, relativ hohes Niveau der Kooperation und bleibt gleichzeitig diffus. Denn einerseits führt Zusammenarbeit nicht notwendigerweise zu gemeinsamen Veröffentlichungen; andererseits repräsentiert Koautorenschaft nicht in jedem Fall Kooperation (Katz/Martin 1997; Laudel 2002).

Hinzu kommt bei bibliometrischen Untersuchungen allgemein eine Abhängigkeit von Literaturdatenbanken, die häufig regionale Schwerpunkte aufweisen und das weltweite Publikationsaufkommen nicht gleichmäßig abdecken. Für den Science Citation Index ist eine Konzentration auf den *mainstream* an Publikationen der zentralen, vorwiegend westlichen Wissenschaftsnationen und eine unterdurchschnittliche Erfassung von Zeitschriften, die in nicht-lateinischer Schrift erscheinen, belegt (Carpenter/Narin 1981). Zwar ist anzunehmen, dass dieser Bias durch die Ausweitung der ISI-Zeitschriftengrundlage in den letzten Jahr-

zehnten zumindest abgeschwächt wurde, allerdings erscheint nach wie vor der Großteil der Zeitschriften der ISI-Datenbanken in den USA und anderen englischsprachigen Ländern (Paasi 2005). Ob Entwicklungsländer, für die bibliometrische Datenbanken meist nur einen sehr geringen Teil aller erschienenen Publikationen ausweisen (Shrum 1997), inzwischen stärker repräsentiert sind, ist fraglich.

Das World Wide Web bietet vielfältige Möglichkeiten, die auf kommunikativem Austausch basierende wissenschaftliche Arbeit zu unterstützen, und entsprechend umfassend wird es heute in der Wissenschaft eingesetzt (vgl. Abschnitt 2.1.3). Webstrukturen beinhalten nicht nur Informationen über Formen des wissenschaftlichen Austauschs, die das spezielle und relativ fortgeschrittene Niveau von Kopublikationen haben, sondern auch über andere, niederschwelligere Formen des Kontakts bis hin zur ein- oder wechselseitigen Wahrnehmung. Für die Wissenschaftsforschung erscheint es daher als vielversprechende empirische Grundlage, um ein breiteres Bild wissenschaftlicher Kommunikation zu gewinnen. Dies gilt insbesondere für internationale Zusammenhänge: Ebenso wie E-Mail die Kommunikation über große Entfernungen erleichtert, vereinfacht das World Wide Web speziell die weltweite Verbreitung von Informationen, weshalb beide Dienste häufig zur grenzüberschreitenden Kommunikation genutzt werden. Im folgenden Abschnitt wird die spezifische Rolle des Internets im Kontext wissenschaftlicher Kooperation genauer beleuchtet.

3.2 Die Rolle des Internets in der internationalen wissenschaftlichen Kommunikation

Den neuen Medien wird häufig das Potenzial zugeschrieben, Forschungsk Kooperationen stark zu erleichtern. Durch E-Mail wird der kommunikative Austausch zwischen WissenschaftlerInnen erheblich beschleunigt. Das World Wide Web ermöglicht die Präsentation aktueller Informationen vor der weltweiten Öffentlichkeit ebenso wie die Recherche von unmittelbar zugänglichen Dokumenten. Die Vorbereitung einer Konferenz oder die institutsübergreifende Nutzung von Datenbanken sind Beispiele für Aktivitäten, die sich online sehr viel einfacher, aktueller, komfortabler und kostengünstiger gestalten lassen als offline. Dies gilt insbesondere für grenzüberschreitende Kooperationen: Sowohl E-Mail als auch das World Wide Web tragen dazu bei, die für den Austausch von Informationen aufgewendete Zeit und die Kosten unabhängig von der überbrückten Distanz zu reduzieren. Ihre Auswirkungen auf den wissenschaftlichen Arbeitsalltag können ohne Zweifel als bedeutend eingestuft werden (vgl. Abschnitt 2.1.3). Es stellt sich jedoch die Frage, ob alle Teile der Welt in gleichem Maße von dieser Entwicklung profitieren.

Im Folgenden werden zunächst zwei gegensätzliche Auffassungen zum Einfluss des Internets auf die internationale Ungleichheit vorgestellt (Abschnitt 3.2.1). Anschließend werden empirische Befunde zur internationalen Verteilung des Zugangs zum Internet (Abschnitt

3.2.2) sowie der Präsenz im Web (3.2.3) analysiert, um die beiden Auffassungen gegeneinander abwägen zu können (Abschnitt 3.2.4).

3.2.1 Der Einfluss des Internets auf die internationale Ungleichheit – zwei konträre Einschätzungen

In den hochentwickelten Forschungssystemen der ersten Welt haben die neuen Medien inzwischen einen festen Platz. Ihr Nutzen ist hier unumstritten. Was die globale Perspektive betrifft, lassen sich allerdings zwei gegensätzliche Einschätzungen formulieren:

These 1: Das Internet trägt dazu bei, räumliche Distanzen und Ungleichheit zu überwinden.

Die erste These fußt auf der Feststellung, dass das Internet ein weltumspannendes Medium ist, das unabhängig vom Standort einen breiten Zugang zu Informationen bietet und dadurch zum Bedeutungsverlust räumlicher Distanz beiträgt:

„Physical location is no longer a barrier to the free and easy exchange of information“ (Beaver 2001: 376).

Informationen können von beliebigen Servern abgerufen werden. Ist man nicht mehr auf lokale Quellen angewiesen, lassen sich geeignete KooperationspartnerInnen, relevante Literatur und nützliche Datenbanken besser identifizieren. Gleichzeitig können Forschungseinrichtungen Informationen über sich, ihre WissenschaftlerInnen, Projekte und Ergebnisse ins Netz stellen und so ihren Bekanntheitsgrad erhöhen (de Roy 1997). Anders als bei Telefonaten oder Post haben Entfernungen hier keinen Einfluss auf die Höhe der anfallenden Kosten, die ohnehin als gering einzustufen sind.

„The Internet’s ability to span continents at a single bound has reduced the constraints of time and distance. Global communication is as easy, rapid, flexible, and low cost as chats between neighbors“ (Koku/Nazer/Wellman 2001: 1752).

Mit dieser These ist die Hoffnung verknüpft, dass insbesondere Entwicklungsländer von den neuen Medien profitieren. Ein Problem, mit dem ForscherInnen dort häufig konfrontiert sind, ist die Isolation, die aus dem geringen Umfang der lokalen *scientific community*, den begrenzten Kommunikationsmöglichkeiten mit KollegInnen im Ausland und dem schwierigen Zugang zu internationaler Literatur resultiert (Dedijer 1963; Salam 1966; de Roy 1997; Davidson/Sooryamoorthy/Shrum 2002). Vor diesem Hintergrund erscheint das Internet als Sprungbrett zum kommunikativen Anschluss an die Forschungsnationen der ersten Welt, der nie schneller und günstiger zu haben war.

„It can easily be seen that the availability of computer-mediated communication (email, file transfer, bulletin boards, online publication) improves the possibilities of long-distance collaboration in science. This is supported by other telecommunicative media (cheaper long-distance calls, fax machines). The differentiation of science in centres and peripheries is weakened by these developments“ (Stichweh 1996: 335f).

Der optimistische Gedanke, dass sich die Differenz zwischen Zentrum und Peripherie durch die neuen Medien abschwächt, manifestiert sich in der Metapher vom *global village* sowie in der speziell auf die Wissenschaft zugeschnittenen Adaption *global research village*, die die OECD als Titel für eine Reihe von Konferenzen und damit zusammenhängenden Berichten zur Vernetzung der internationalen Forschungslandschaft durch IuK-Technologien verwendet (vgl. z.B. OECD 1998). Dieser Position zufolge verbindet das Internet die verstreuten Mitglieder der unterschiedlichen nationalen *scientific communities* zu einer dörflichen Nachbarschaft, in der alle miteinander bekannt sind und in der jede Frage rasch geklärt werden kann.

These 2: Das Internet bildet lediglich bereits bestehende Strukturen ab und verfestigt dadurch die vorhandene Ungleichheit.

Die zweite These drückt die skeptische Position aus, dass die Nutzung des Internets soziale Ungleichheiten nicht nivelliert, sondern vielmehr reproduziert und eventuell sogar verstärkt. Die Option einer weltweiten Vernetzung, die die Internettechnologie bietet, beinhaltet nicht die faktische Universalität der Nutzungsmöglichkeiten. Begrenzter Zugang zum Internet, mangelnde Fähigkeiten im Umgang mit den neuen IuK-Technologien oder fehlende Sprachkenntnisse können Gründe sein, aus denen ForscherInnen in Entwicklungsländern weniger von den Vorteilen des Internets profitieren als ihre westlichen KollegInnen (Brooks/Donovan/Rumble 2005). Trifft es zu, dass mit Hilfe der neuen Medien in erster Linie Forschungsprozesse in entwickelten Ländern effizienter gestaltet werden, wäre eine weitere Marginalisierung von ForscherInnen in der Dritten Welt die Folge. Sie resultiert aus der Verstärkung der bereits bestehenden internationalen Ungleichheit in Bezug auf den Zugang zu Literatur und anderen Datenquellen, in Bezug auf die Zahl der wissenschaftlichen Kontakte und in der Folge auch in Bezug auf die Qualität und den Umfang der Forschungsergebnisse:

„While the rapid expansion of the Internet in the West and the speedy transition to electronic publishing will help scientists and scholars in the developed countries to interact with colleagues and members of their invisible colleges, the already isolated scientists of the developing world will be reduced to mere spectators. The already existing gulf in the levels of science and technology performed in the developed and the poorer countries will be widened further“ (Arunachalam 1999: 470).

Der Begriff *digital divide*, der für ungleich verteilte Möglichkeiten zur Nutzung moderner Informationstechnologie steht und sich nicht exklusiv auf die Wissenschaft bezieht, illustriert diese Einschätzung. Sein genauer Ursprung ist unbekannt, jedoch ist er seit seiner Verwendung in mehreren Berichten der National Telecommunications and Information Administration (NTIA) der USA Ende der 1990er Jahre verbreitet (vgl. u.a. NTIA 1998 und 1999). Sofern speziell die Ungleichheit verschiedener Weltregionen – im Gegensatz zur Ungleichheit entlang verschiedener sozialer Dimensionen innerhalb einzelner Regionen oder Nationen – im Mittelpunkt steht, wird auch der Terminus *global digital divide* verwendet (vgl. z.B. Drori/Jang 2003; Castells 2005).

Eine Abwägung beider Auffassungen – der These des *global research village* und der These der *global digital divide* – ist aufgrund des begrenzten empirischen Materials schwierig. Auf Grundlage der verfügbaren Studien werden in den folgenden Abschnitten einige zentrale Aspekte der globalen Verteilung von Internetzugang und Webpräsenz beleuchtet, um anschließend eine begründete Einschätzung der Rolle des Internets für die internationale Kommunikation in der Wissenschaft geben zu können.

3.2.2 Globale Ungleichheit im Internetzugang

Der Zugang zum Internet ist nicht ubiquitär. Um den Informationsfundus des Webs und die Möglichkeit zur schnellen Kommunikation per E-Mail nutzen zu können, ist ein geeigneter Computer mit Internetanschluss erforderlich. Darüber hinaus sind Informationen über die Optionen, die das Internet bietet, sowie Grundkenntnisse im Umgang mit den neuen Medien die Voraussetzung dafür, dass Menschen Interesse an der Internetnutzung entwickeln und in befriedigender Weise mit der Technologie umgehen können. In zahlreichen entwickelten Ländern ist es heute eine Selbstverständlichkeit, dass wissenschaftliche Arbeitsplätze mit Rechnern und einem schnellen Internetanschluss ausgestattet sind. Einer Studie aus Israel, das ein hochentwickeltes Forschungssystem besitzt, zufolge verfügten im Jahr 2001 92% der befragten ForscherInnen über einen Internetzugang in ihrem eigenen Büro (Bar-Ilan/Peritz/Wolman 2003). Die Anbindung von Rechenzentren und Bibliotheken an das Internet bietet häufig auch kostenlose, hochwertige Zugangsmöglichkeiten für Studierende. Als umfassend und vielfältig beschreibt Al-Ansari (2006) die Ausstattung und Nutzung von Computern und Internet an Universitäten in Kuwait, das nach den Länderkategorien der Vereinten Nationen zu den Entwicklungsländern zählt (vgl. United Nations Statistics Division 2006), aber eines der höchsten Pro-Kopf-Einkommen der Welt aufweist.¹⁰

¹⁰ Vgl. <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html>, zuletzt abgerufen am 19.11.2009.

In vielen anderen Entwicklungsländern ist die Verfügbarkeit von Computern dagegen begrenzt. Einer Befragung von ForscherInnen in Indien, Ghana und Kenia zufolge wurde zwischen 2000 und 2002 jeder PC im Durchschnitt von sieben Personen genutzt, weshalb die Autoren folgenden Schluss zogen: „PCs are not so much ‘personal computers‘ as ‘public computers‘“ (Ynalvez et al. 2005: 61). Unter diesen Nutzungsbedingungen ist die Zeit, sich mit Computer- und Internetanwendungen vertraut zu machen, knapp.

Eine Studie zum Internetzugang in Bangladesch wies ebenfalls auf beschränkte Möglichkeiten der akademischen Internetnutzung hin (Uddin 2003). Generell gab es in diesem Land im Jahr 2002 keine Breitbandzugänge an Universitäten, stattdessen verfügten die meisten Einrichtungen über ein bis drei Standleitungen. Fast alle befragten Lehrenden der dortigen Rajshahi Universität nutzten das Internet, wobei E-Mail von den meisten Personen (88,1%) verwendet wurde. Andere Anwendungen wie Recherchen oder das Downloaden von Dateien waren weniger beliebt, was nicht zuletzt auf die geringe Geschwindigkeit der Verbindungen von 3-4 Kilobyte pro Sekunde zurückzuführen war, die 82,7 Prozent der Befragten bemängelten. Zwei Drittel (67,3%) hielten die Zugangsmöglichkeiten für nicht ausreichend. Studierende hatten an der Universität – ebenso wie in den meisten Fällen auch privat – keine Möglichkeit, ins Internet zu gehen, weswegen speziell auf die Lehre bezogene Anwendungen keine Rolle spielten. Hinzu kommt, dass einige nützliche Dienste (Online-Zugang zu Bibliothekskatalogen, Stellenausschreibungen etc.) noch nicht verbreitet waren. Insofern wundert es nicht, dass das Internet für die meisten Befragten eher ein Kommunikations- als ein Informationsmedium darstellte. Insbesondere die internationale Kommunikation schien im Vordergrund zu stehen: Ein Großteil korrespondierte per E-Mail mit Ausbildungs- bzw. Forschungseinrichtungen (70,6%) oder FreundInnen und KollegInnen (63,5%) im Ausland. Inwiefern das Internet zur Erweiterung oder Intensivierung der früheren Kontakte beigetragen hatte, wurde nicht erfragt.

Eine Befragung an nigerianischen Universitäten zeigte, dass im Jahr 2001 64,4 Prozent der dortigen WissenschaftlerInnen ein Computer zur Verfügung stand (Ehikhamenor 2003a). Die Hälfte der Befragten (50,4%) nutzte das Internet, allerdings verfügte lediglich eine Minderheit von 5,3 Prozent über einen Anschluss in ihrem Büro. Während weitere 5,0 Prozent andere Zugangsmöglichkeiten in der Universität nutzten, griff der Großteil auf außeruniversitäre Optionen zurück, und zwar insbesondere auf kommerzielle Internet-Cafés (36,9%). Dies impliziert, dass vier von fünf InternetnutzerInnen die Kosten, die im Verhältnis zur Kaufkraft wesentlich höher sind als in entwickelten Ländern, privat trugen. Entsprechend wurden in der Studie der begrenzte Computerzugang und die hohen Kosten als größte Hindernisse einer intensiveren Nutzung der neuen Medien identifiziert. Als weiteres Problem wurde die mäßige Qualität der Verbindungen genannt, die die Informationssuche erschwerte und Geduld von den NutzerInnen forderte. Das generelle Interesse war dennoch groß: 84,3 Prozent der

Befragten erwarteten, dass das Internet in Zukunft unabdingbar für ihre wissenschaftliche Arbeit sein würde.

Andere Erhebungen in Afrika führten zu ähnlichen Ergebnissen: Auch für Ghana und Kenia wurden neben den Kosten für Internetnutzung und gebührenpflichtige Seiten vor allem der Zeitaufwand für das Verbinden und Laden von Seiten sowie die Instabilität der Verbindung als verbreitete Probleme benannt (Ynalvez et al. 2005). An der Universität von Ibadan in Nigeria wurde das Internet weniger intensiv genutzt als von den Befragten gewünscht, obgleich der Großteil es als Informationsquelle (96,8%) und Mittel zur informellen Kommunikation (94,5%) schätzte. Als häufigste Ursache wurde noch vor den fehlenden finanziellen Mitteln der Universitäten für die Nutzung von Internet und kostenpflichtigen Internetquellen (76,5%) und den hohen Kosten für die ersatzweise genutzten kommerziellen Zugangsmöglichkeiten (31,3%) die instabile Stromversorgung (83,5%) genannt (Sangowusi 2003: 132). In einer qualitativen, nicht auf Universitäten begrenzten Untersuchung in verschiedenen Ländern Afrikas südlich der Sahara wurden neben der mangelhaften Infrastruktur mit unzureichenden Telefonverbindungen, unzuverlässigen Internetanschlüssen und Stromausfällen auch Wissensdefizite als Ursachen für eine schwache Internetnutzung identifiziert (Adam/Wood 1999). Mangelnde technische Kenntnisse der AnwenderInnen, fehlende Wartungsmöglichkeiten und die vorrangige Auslegung der Hardware auf die Verwendung in trockenem, gemäßigttem Klima führten dazu, dass die Ausstattung häufig nicht dauerhaft in Betrieb genommen werden konnte (de Roy 1997; Adam/Wood 1999).

Zwei auf derselben Befragung beruhende Studien (Duque et al. 2005; Ynalvez et al. 2005) wiesen auf deutliche Unterschiede im Internetzugang zwischen verschiedenen Entwicklungsländern hin. Zwischen 2000 und 2002 verfügten deutlich mehr ForscherInnen in der indischen Provinz Kerala (86%) über Zugang zu E-Mail als in Ghana (65%) oder Kenia (51%). Ob es sich dabei um Zugangsmöglichkeiten in wissenschaftlichen Einrichtungen handelt, wurde nicht ausgeführt. Die gute Ausstattung in Indien, die in annähernd dem westlichen Niveau entspricht, wurde von den Autoren auf hohe öffentliche Investitionen in IuK-Technologien zurückgeführt. Interessant ist die Position Ghanas, das zwar in wirtschaftlicher Hinsicht schwächer entwickelt ist als Kenia, aber eine regionale Vorreiterrolle in der Liberalisierung seines Telekommunikationssektors einnimmt, die zu einer Ausbreitung der Zugangsmöglichkeiten geführt hat (Duque et al. 2005).

Besonders stark fällt ein fehlender oder mangelhafter Internetzugang ins Gewicht, wenn wissenschaftliche Aktivitäten komplett ins Internet verlagert werden. Ein Beispiel ist die Praxis einiger Zeitschriften, Manuskripte ausschließlich über Online-Formulare zu akzeptieren (Arunachalam 1999; Azzi 2005), ein weiteres die zunehmende Substitution gedruckter Zeitschriften durch E-Journals. Während Preprintforen tatsächlich eine webspezifische und –

abgesehen vom Internetzugang – kostenlose Informationsquelle sind, von der ForscherInnen weltweit profitieren können, bringt die Einrichtung von E-Journals anstelle gedruckter Zeitschriften keinen nivellierenden Effekt, da üblicherweise ein kostenpflichtiges Abonnement erforderlich ist. Gleiches gilt für wissenschaftliche Datenbanken:

„Due to the evolution of the information systems, libraries will be substituted by online access journals. Without connectivity, access to published information, already very difficult and expensive in developing countries, will become almost impossible. Paradoxically, the progress offered by electronic communication to the development of world science, if not shared with less industrialized countries, will result in the elimination of the already difficult and limited scientific contribution from a large part of the world“ (Azzi 2005: 261).

Folglich ist zu beachten, dass die Anbindung an das Internet allein noch keinen Zugang zu zentralen wissenschaftlichen Informationen gewährleistet, der aufgrund der schwachen finanziellen Ausstattung vieler Forschungseinrichtungen in Entwicklungsländern nach wie vor begrenzt ist.

Insgesamt zeigt sich, dass in der Forschung die Zugangsmöglichkeiten zum Internet von Land zu Land variieren, wobei entwickelte Länder und Schwellenländer einen erheblichen Vorsprung im Hinblick auf Breite und Qualität des Zugangs und den individuell zu tragenden Kosten besitzen. Wie groß diese Lücke ist, hängt auch davon ab, inwiefern der nationale Telekommunikationsmarkt unter öffentlicher Regulierung für private Anbieter geöffnet ist (World Bank 2005: 18ff) und wieviel Unterstützung einzelne Regierungen für die Ausstattung von Universitäten mit IuK-Technologien und spezifischen Anwendungen (z.B. Abonnements von E-Journals oder Online-Datenbanken) bieten (Brooks/Donovan/Rumble 2005). Häufig geht eine schwache technische Ausstattung der Universitäten mit generellen Defiziten im Hinblick auf den Literaturzugang etc. einher, wie eine der nigerianischen Fallstudien unterstreicht (Ehikhamenor 2003a).

Es ist möglich, dass die Lücke in den nächsten Jahren kleiner wird, sich der Internetzugang in Entwicklungsländern in qualitativer und quantitativer Hinsicht verbessert und sich die gegenwärtigen Probleme als „teething troubles“ (Davidson/Sooryamoorthy/Shrum 2002: 500) erweisen. In technischer Hinsicht gibt es Möglichkeiten, auch großflächige Regionen mit geringer telekommunikativer Infrastruktur ans Netz anzubinden, wie beispielsweise die satellitenbasierten Very Small Aperture Terminals (VSAT), die in einigen Ländern Afrikas bereits seit längerem eingesetzt werden, aber auch hohe Kosten verursachen (vgl. de Roy 1997; Azzi 2005; Duque et al. 2005). Ostafrika hat kürzlich als eine der letzten Weltregionen eine terrestrische Verbindung an das Internet-Backbone erhalten, weshalb dort mittelfristig ein hochwertigeres und günstigeres Internetangebot zu erwartet ist (vgl. Perras 2009).

Allerdings ist anzunehmen, dass eine Verbesserung der Infrastruktur allein nicht zu einer Optimierung des Internetzugangs an wissenschaftlichen Einrichtungen in Entwicklungsländern führt. Zwar werden mittelfristig sinkende Nutzungsgebühren erwartet, eine Angleichung an die niedrigen Tarife der Industrieländer jedoch nicht, da die Infrastrukturkosten wie z.B. für den Backbone-Anschluss überwiegend privat finanziert werden (vgl. Perras 2009). Zu dem ist ein gewisser Fahrstuhleffekt zu beobachten: Mit der Verbreitung hochwertiger Breitbandanschlüsse in Industrieländern werden dort auch immer komplexere Websites veröffentlicht, die auf Netzanschlüsse mit hohen Datenübertragungsraten ausgerichtet sind. Auch wenn die Anschlussqualität weltweit steigt, bleiben bestehende Diskrepanzen erhalten.

3.2.3 Internationale Unterschiede in der Webpräsenz

Zur Klärung des Stellenwerts des World Wide Webs in der internationalen wissenschaftlichen Kommunikation ist neben der globalen Verteilung von Zugangsmöglichkeiten zum Internet auch die Präsenz verschiedener Weltregionen im Web von Belang. Es ist anzunehmen, dass Umfang und Güte des Internetzugangs Faktoren sind, die die Neigung beeinflussen, Websites einzurichten und zu pflegen. Eine systematische Untersuchung dieses Zusammenhangs gibt es bisher nicht, jedoch deuten die Ergebnisse von Befragungen in Ländern mit begrenztem, instabilem und teurem Internetzugang darauf hin, dass das Design von Webseiten nicht zu den Internetanwendungen gehört, die die dortigen ForscherInnen nutzen. So hat keiner der 128 befragten NaturwissenschaftlerInnen in der nigerianischen Erhebung von Sangowusi (2003: 131) jemals eine Webseite eingerichtet.

Bisher gibt es kaum vergleichende Untersuchungen zur Präsenz von Forschungseinrichtungen in Industrie- und Entwicklungsländern im Web. Allerdings zeigen bereits internationale Vergleiche unter Nationen, die sich in der Ausstattung mit IuK-Technologien ähneln, deutliche Unterschiede in Gestaltung, Umfang und Verlinkung von Websites. Generelle qualitative Unterschiede im Webdesign konstatierte eine Studie zu 90 kommunalen Websites in Deutschland, Japan und den USA. Das Layout, die Verwendung von Grafiken, Karten und Linklisten sowie die Verfügbarkeit von Übersetzungen und Suchoptionen variierten signifikant über die drei Länder (vgl. Cyr/Trevor-Smith 2004).

Im wissenschaftlichen Bereich zeigen sich außerdem diverse quantitative Differenzen zwischen verschiedenen Ländern. Onyancha und Ocholla (2007) haben die Websites aller Universitäten von Kenia und Südafrika verglichen. Aufgrund von substantiellen Unterschieden diagnostizieren sie eine *content divide*. Dieser Begriff bezieht sich nicht auf die Qualität der Website-Inhalte, sondern auf die Menge der Webseiten, die in Südafrika wesentlich höher ausfällt als in Kenia. Dies ist bisher die einzige Studie, die die Größe von universitären Internetauftritten in Entwicklungsländern zum Gegenstand hat. Ihre Aussagekraft wird allerdings dadurch begrenzt, dass der Umfang der Websites lediglich mit Suchmaschinen erho-

ben wurde. Bei der Erhebung von Inlinkdaten haben die Autoren darauf verzichtet, interne Links auszuschließen, so dass die Ergebnisse nicht sinnvoll interpretierbar sind.

Eine Untersuchung naturwissenschaftlicher Universitätsinstitute in Großbritannien, Kanada und Australien ergab, dass die britischen Websites größer waren als die der beiden anderen Länder. Gleichzeitig waren die kanadischen Fachbereiche am stärksten untereinander vernetzt. Sie wiesen außerdem den höchsten Anteil internationaler Inlinks auf, Australien dagegen einen sehr niedrigen (Li et al. 2005b). Wenngleich sich die gefundenen landesspezifischen Muster in dieser Untersuchung zum Teil mit disziplinären überschneiden (vgl. Abschnitt 2.3.3), bietet sie einige Hinweise auf nationale Eigenheiten, die zwischen verschiedenen sprachigen Ländern vermutlich noch deutlicher ausfallen.

Einen Hinweis darauf liefert eine Studie zu 670 Internetauftritten europäischer Universitäten, derzufolge sich der durchschnittliche Umfang der Websites von Land zu Land erheblich unterscheidet (Thelwall et al. 2002). Dabei zeigte sich ein deutlicher Zusammenhang mit dem durchschnittlichen Pro-Kopf-Einkommen: Je reicher das Land, desto umfangreicher waren die universitären Internetauftritte. Die beschriebenen Unterschiede sind durchaus erheblich: Eine durchschnittliche akademische Website in den Niederlanden war 2000mal größer als eine in Bulgarien, dessen Universitäten im Web kaum präsent waren. Die AutorInnen vermuten einen direkten Einfluss der materiellen Ausstattung der Hochschulen, von der die gegenwärtigen Rechner- und Netzwerkressourcen sowie die in den letzten Jahrzehnten erworbene Expertise und die Breite der NutzerInnenbasis abhängt. In einer weiteren Studie (Thelwall/Zuccala 2008) wurde eine starke Korrelation zwischen dem nationalen Brutto-Inlandsprodukt europäischer Staaten und der Zahl der Inlinks, die ihre Universitäten im Durchschnitt erhalten, gefunden. Ob ein Zusammenhang mit der Websitegröße gegeben war, wurde nicht erhoben.

Auch ohne weltweite Studien kann angenommen werden, dass sich der materielle Einfluss nicht nur in einem europäischen Ost-West-Gefälle, sondern ebenfalls in einem globalen Nord-Süd-Gefälle zeigt, bei dem der Umfang universitärer Websites und damit die Vielfalt der Anwendungen und die Sichtbarkeit der Forschungs- und Lehraktivitäten stark variieren. Diese These wird von einer Studie zu akademischen Websites in dreizehn Ländern im pazifischen Raum gestützt, die extreme Unterschiede in der Größe der Internetauftritte aufdeckt: Während auf den akademischen Second-Level-Domains von Australien und Japan im Januar 2002 jeweils gut über 3,4 bzw. 3,7 Millionen Seiten lagerten, verfügten Polynesien (55 Seiten), die Fiji-Inseln (582), aber auch größere Länder wie Indonesien (ca. 36.000) nur über einen Bruchteil davon (Thelwall/Smith 2002: 368). Kritisch anzumerken ist an dieser Stelle, dass der Einfluss der unter Umständen lokal variierenden Abdeckung der verwendeten Suchmaschine AltaVista unklar bleibt und dass die Autoren keinen Versuch unternom-

men haben, die Größe des akademischen Webraums in Beziehung zu der tatsächlichen Größe der jeweiligen nationalen Universitätssysteme zu setzen, die sich zwischen den genannten Ländern ebenfalls stark unterscheidet. Die Studie unterstützt jedoch die Vermutung, dass akademische Internetauftritte in industrialisierten Ländern wie Japan und Australien intensiver genutzt werden als in Entwicklungsländern.

Thelwall et al. (2002) weisen ferner darauf hin, dass es kulturelle Unterschiede in der Gestaltung von Internetauftritten geben kann. Frankreichs Universitäten verfügen laut ihrer Erhebung über deutlich kleinere Websites, als aufgrund des Pro-Kopf-Einkommens zu erwarten wäre. Ein Grund hierfür liegt darin, dass Inhalte nicht selten auf Domains außerhalb der eigentlichen Universitätsadresse abgelegt und folglich nicht erfasst werden. Eine weitere Ursache für ihre geringe Präsenz sehen die AutorInnen jedoch darin, dass – während die offiziellen Rahmenseiten professionell gestaltet sind – die einzelnen Universitätsangehörigen vergleichsweise wenig beitragen. In diesem Fall ist davon auszugehen, dass es sich nicht um eine Frage der vorhandenen Ressourcen handelt, sondern vielmehr um eine Frage der Hochschulkultur: die regulative Frage, ob die Website in erster Linie ein Aushängeschild darstellt, das vorwiegend der Eigenwerbung und der allgemeinen Information dient, oder ob sie den einzelnen HochschulmitarbeiterInnen oder auch Studierenden als Werkzeug zur Verfügung stehen soll. Vermutlich handelt es sich meist um Mischformen, wobei extreme Schwerpunkte wie im Fall Frankreichs vorkommen können.

Auch die Verwendung unterschiedlicher Sprachen auf akademischen Websites ist bereits untersucht worden: Im Allgemeinen ist Englisch die vorherrschende Sprache im Web. Dies gilt auch für wissenschaftliche Einrichtungen in Europa. In den sechzehn untersuchten westeuropäischen Ländern (Norwegen, Schweiz sowie die EU der 15 Länder ohne Luxemburg) bestehen universitäre Internetauftritte im Durchschnitt zur Hälfte aus englischsprachigen Dokumenten. Während irische und britische Universitäten nahezu ausschließlich englische Seiten präsentieren, machen sie in den übrigen Ländern mit Ausnahme von Spanien und Griechenland ein bis zwei Drittel aus. Den geringsten Stellenwert hat die englische Sprache auf den Websites griechischer Hochschulen (ca. 15%), die gleichzeitig selten Ziel und Ursprung internationaler Links sind (vgl. Thelwall/Tang/Price 2003: 422f). Auch für Kanada zeigte sich, dass die Websites der englischsprachigen Universitäten deutlich mehr Links auf sich ziehen als die der französischsprachigen Hochschulen, bei denen eine englische Version häufig schwer zu finden war oder gar nicht vorlag (vgl. Vaughan/Thelwall 2005). Ein weiterer Hinweis auf die Bedeutung der Landessprache findet sich im pazifischen Raum: 2002 erhielten die Websites der akademischen Domain in Australien im Vergleich zu Japan mehr als doppelt so viele Inlinks aus Ostasien, Südostasien und Ozeanien, was durch andere Faktoren wie den Umfang der Forschungsaktivitäten oder den Zugang zu IuK-Technologien nicht zu erklären ist (Thelwall/Smith 2002). Insgesamt zeigt sich, dass im akademi-

schen Web wie in wissenschaftlichen Publikationen die Verwendung der englischen Sprache üblich und ein entscheidendes Kriterium für internationale Sichtbarkeit ist. Sprachliche Hürden, im Fall Griechenlands verstärkt durch die spezifische Schrift, können dagegen zur Isolation im Web führen.

In diesem Zusammenhang stellt sich auch die Frage, welchen Einfluss räumliche Distanzen auf die Vernetzung von Websites haben. Sofern die wenigen verfügbaren Studien die vorherrschenden Tendenzen abbilden, zeigen sich zwei verschiedene Muster. Innerhalb von Nationen scheint sich räumliche Nähe positiv auf die wechselseitige Verlinkung von Universitäten auszuwirken, wie eine britische Studie ergab (Thelwall 2002b). Dieser Befund könnte ein weiterer Hinweis auf die Bedeutung persönlicher Begegnungen für die wechselseitige Wahrnehmung und die Kooperationsneigung sein, die sich im Netz abbildet. Eine bibliometrische Studie auf der Grundlage von nationalen Koautorenschaftsdaten zu Großbritannien, Kanada und Australien führte ebenfalls zu diesem Schluss (Katz 1994).

Auf der internationalen Ebene ist der Zusammenhang zwischen der Entfernung und der Intensität der Verlinkung weniger eindeutig ausgeprägt. Stattdessen lassen sich regionale Schwerpunkte identifizieren, die unabhängig von der Distanz besonders viele internationale Inlinks auf sich ziehen und von denen viele internationale Outlinks ausgehen. Die Untersuchung der akademischen Domains in Ostasien, Südostasien und Ozeanien zeigte, dass Australien und Japan innerhalb dieser Region die bedeutendsten Linkziele waren, gefolgt von Neuseeland und den Schwellenländern Taiwan, Korea, Hong Kong und Singapur (Thelwall/Smith 2002). Die Entwicklungsländer Malaysia, Philippinen, Indonesien und Vietnam waren dagegen nur selten das Ziel von Links. Wurden die Inlinks jedoch anhand der Zahl der im Zielland vorhandenen akademischen Webseiten standardisiert, erhielten diese Länder mehr Links als die Industrie- und Schwellenländer. Die Ursache für den gefundenen „inverse Matthew effect“ (Katz 2000: 35) kann darin liegen, dass sich kleine Websites häufig auf ein Kernangebot beschränken, das besonders beliebte Linkziele enthält wie Startseiten der Universitäten und Institute, während große Websites auch große, selten verlinkte Bereiche von interner oder lokaler Relevanz enthalten. Ähnliche positive oder negative Skaleneffekte, also nicht-lineare Zusammenhänge, wurden bereits im Zusammenhang mit Publikationsmaßen häufig gefunden (Katz 2000) und durch verschiedene *power laws* wie Lotkas oder Zipfs Gesetz beschrieben. Insofern ist die überdurchschnittliche Zahl von Inlinks auf die akademischen Domains der genannten Entwicklungsländer weniger ein Zeichen außergewöhnlicher Attraktivität, als vielmehr eine Auswirkung eines generellen exponentiellen Zusammenhangs zwischen dem Umfang von Webpräsenzen und der Zahl der Inlinks. Hinweise auf einen positiven Effekt räumlicher Nähe fanden sich kaum, sieht man von der engen Linkbeziehung zwischen Australien und Neuseeland ab. Sie wird nicht allein durch die

geringe Entfernung, sondern auch durch die sprachliche, kulturelle, politische und wirtschaftliche Nähe beider Länder und ihrer Wissenschaftssysteme begünstigt.

Nicht nur bei den Linkzielen, sondern auch bei den Linkquellen sind Websites in Entwicklungsländern schwach repräsentiert. Tang und Thelwall (2004) fanden bei der Analyse der Websites US-amerikanischer Geschichts-, Chemie- und Psychologie-Institute heraus, dass in allen drei Disziplinen über die Hälfte der internationalen Inlinks aus Europa, 15 bis 25 Prozent aus Asien sowie 5 bis 10 Prozent aus Ozeanien stammen. Nicht nur von Websites in Afrika und dem mittleren Osten, sondern auch von denen Süd- und Mittelamerikas oder des unmittelbaren Nachbarlands Mexiko ging nur ein geringer Anteil der gefundenen Inlinks aus (jeweils 0-3%). Gleichzeitig hatten zwischen 10 und 20 Prozent der Inlinks ihren Ursprung in Kanada. Dies spricht für die Interpretation der Autoren, dass in Ländern mit verbreiteter Internetnutzung auch mehr Links eingerichtet werden als in anderen Nationen. Eine naheliegende Ursache stellen die bereits beschriebenen Unterschiede im Internetzugang und der Präsenz im Web dar, die mit Umfang und Ausstattung der nationalen Wissenschaftssysteme zusammenhängen.

Die Chance, Forschungsaktivitäten im Web weltweit sichtbar zu machen, wird in verschiedenen Ländern in unterschiedlichem Maß genutzt. Weltweit, aber selbst innerhalb Europas, zeigen sich deutliche Unterschiede. Während die meisten ressourcenstarken westlichen Wissenschaftsnationen gut vertreten sind, sind Entwicklungsländer sowie die früheren Ostblockstaaten häufig weit weniger sichtbar. Die Möglichkeit, sich im Web zu präsentieren, wird also bisher insbesondere in den Staaten genutzt, die auch in Fachzeitschriften stark vertreten sind. Da die materielle Ausstattung der nationalen Wissenschaftssysteme und die der Internetnutzung häufig korrelieren, lässt sich der Einfluss beider Faktoren auf die Webpräsenz auf der Grundlage des vorhandenen empirischen Materials nicht trennen. Offensichtlich ist jedoch, dass Entwicklungsländer, die in Journals kaum sichtbar sind, dies in der Regel nicht durch starke Webpräsenz kompensieren können.

Es stellt sich die Frage, ob ein solches Aufholen von Entwicklungsländern nicht durch ungebremstes Wachstum der Webpräsenz von Hochschulen und Forschungseinrichtungen in Industrieländern ausgeschlossen scheint. Allerdings zeigte eine Längsschnittstudie über die Webpräsenz der Universitäten Großbritanniens, Australiens und Neuseelands in den Jahren 2000 und 2005, dass in diesen Ländern nach starkem Wachstum akademischer Webbereiche eine Stagnationsphase begonnen hat (Payne/Thelwall 2007).

Der gezielte Einsatz des Webs als Arbeitswerkzeug und als Mittel, um die eigene internationale Sichtbarkeit zu erhöhen, dürfte neben den zentralen westlichen Wissenschaftsnationen in erster Linie Schwellenländern gelingen, deren Forschungssystem eine kritische Größe überschreitet und denen eine umfassende Ausstattung mit den nötigen technischen und

ökonomischen Ressourcen möglich ist. Diese Einschätzung wird – trotz gewisser Vorbehalte bzgl. der Validität der mit Suchmaschinen erhobenen und nur grob sortierten Daten – von dem weltweiten webometrischen Universitätsranking gestützt, das auf Maßen der Websitegröße und der Zahl der Inlinks beruht und bei dem im Januar 2006 beispielsweise Brasilien oder Taiwan weit besser abschnitten als Spanien, Frankreich oder Dänemark (vgl. Aguillo et al. 2005; InternetLab 2006a).

3.2.4 Global research village oder global digital divide?

Abgesehen von der globalen Verteilung der Zugangsmöglichkeiten und der Präsenz im Web bleibt die grundsätzliche Frage, ob die neuen Medien die mit ihnen verknüpften Hoffnungen auf mehr Kontakte und mehr Kooperationen im In- und Ausland tatsächlich erfüllen können, insbesondere was die Integration von Entwicklungsländern in die weltweite *scientific community* betrifft.

Einer kanadische Untersuchung zur Kommunikation in Forschungsnetzwerken zufolge ersetzen E-Mails persönliche Treffen nicht (Koku/Nazer/Wellman 2001). Vielmehr dienen sie vorwiegend als kommunikative Ergänzung regelmäßiger *face-to-face*-Kontakte. Solange jedoch die persönliche Begegnung eine wesentliche Voraussetzung für einen stabilen E-Mail-Kontakt ist, ist auch ein hochwertiger Internetzugang ohne bestehende Forschungskontakte von begrenztem Nutzen.

„In short, e-mail creates opportunities for all to communicate, but it does not necessarily create reasons to communicate or force people to communicate. [...] Ubiquitous connectivity does not mean universal communication“ (Koku/Nazer/Wellman 2001: 1768).

Diese Einschätzung ist konsistent mit den Ergebnissen einer Feinauswertung (Ehikhamenor 2003b) der bereits erwähnten Studie zur Internetnutzung an nigerianischen Universitäten (Ehikhamenor 2003a). Sie führte zu dem Ergebnis, dass die Zahl der Forschungskontakte unter den befragten InternetnutzerInnen nicht höher war als die der Nicht-NutzerInnen: Personen, die das Internet nutzten, besaßen im Durchschnitt fünf, die übrigen Personen über zwölf Kontakte innerhalb und außerhalb Nigerias. In diesem Zusammenhang wäre es interessant, etwas über die Verteilung von Alter bzw. Rang im Verhältnis zur Internetnutzung bzw. Nicht-Nutzung zu erfahren, da die Zahl der Forschungskontakte üblicherweise mit dem Alter steigt. Der Artikel enthält diese Informationen nicht. E-Mails waren nach E-Journals der zweitbeliebteste Verwendungszweck und wurden von 63,2 Prozent der Befragten mit wichtig oder sehr wichtig bewertet. Vermutlich nutzten sie das Internet also eher zur Pflege bestehender als zur Aufnahme neuer Kontakte. Entsprechend gaben nur 4,2 Prozent der NutzerInnen an, aufgrund des Internets mit mehr ForscherInnen zu kommunizieren.

Die beiden Vergleichsstudien zu Indien, Kenia und Ghana (Duque et al. 2005; Ynalvez et al. 2005) zeigten ebenfalls keinen eindeutigen Zusammenhang zwischen Internetnutzung und Kooperationsintensität. In der Gruppe der indischen Befragten war der Internetzugang annähernd selbstverständlich, die Zahl der Kooperationen jedoch geringer als in Ghana und Kenia. Ein Grund dafür könnte darin bestehen, dass Indien über ein gut entwickeltes und umfassendes Forschungssystem verfügt, was sich auch in der vergleichsweise hohen Publikationsaktivität der indischen Befragten zeigte. Die Autoren nahmen ferner an, dass die höhere Zahl der internationalen Kooperationen in Ghana und Kenia weniger das Ergebnis von umfassendem Internetzugang noch von reinem Forschungsinteresse, sondern vielmehr der schlechten Ausstattung und dem sozialen und politischen Kontext geschuldet sei. Entsprechend wirkten sich diese Kooperationen teilweise sogar negativ auf die Forschungsproduktivität aus (Duque et al. 2005):

„Given the continual economic difficulties faced by professionals who are not politically connected, they are encouraged to take up ‚collaborations for development‘, regardless of their direct connection with personal scientific interests. They search for consulting projects and teach additional courses for needed family income. As they brave the deadlines, hazards, and reporting requirements of increased collaboration, they undermine gains in productivity by incurring additional research problems“ (Duque et al. 2005: 777).

Folglich hat das Internet eine paradoxe Wirkung: Es kann, insbesondere durch die Verwendung von E-Mail, Kommunikationsabläufe erleichtern und bestehende Kooperationen vereinfachen. In vielen Entwicklungsländern mangelt es jedoch an den nötigen technischen Ressourcen sowie an „*personal time and space for gaining the sophistication that might transform Internet connectivity into a collaborative research tool [...]*“ (Ynalvez et al. 2005: 64; Hervorhebung im Original). Wird es im Kontext schwach entwickelter Länder genutzt, um die Zahl der Kooperationen aus wissenschaftsexternen Gründen gezielt zu maximieren, bringt es aufgrund struktureller Probleme unter Umständen sogar mehr Aufwand als Entlastung:

„The research institutions of sub-Saharan Africa, for which collaboration has seemed to hold the greatest promise, are the least equipped to benefit, since the very conditions that problematize the relationship between collaboration and productivity also undermine the benefits of new information and communication technologies. It is not collaboration [...] that causes research problems. It can also be attributed to poverty, corruption, and family obligations – in short, the routine problems of everyday life in many less-developed regions“ (Duque et al. 2005: 777).

Der vielfach unterstellte Nutzen des Internets ist somit nicht universell, sondern kommt insbesondere stabilen, stärker entwickelten Nationen mit gut ausgestatteten Forschungssystemen zugute.

Insgesamt unterstützt das empirische Bild trotz des Fehlens groß angelegter Vergleichsstudien in erster Linie die zweite, skeptische These von der *global digital divide*. Die vorhan-

denen Untersuchungen werfen Zweifel daran auf, dass die Verbreitung des Internets grundsätzlich dazu beiträgt, Ungleichheit im Forschungssystem zu nivellieren. Während die gelungene elektronische Vernetzung entscheidend zur Integration in die internationale *scientific community* beitragen kann, verstärkt die kommunikative Isolation aufgrund langsamer und teurer Internetanschlüsse, fehlender Anwendungskennntnisse oder geringer Webpräsenz bestehende Marginalität.

„The potential research benefits of internationalisation through the Internet have been recognised before the age of the web but it is, conversely, equally a threat to those countries that are not using it effectively“ (Thelwall et al. 2002: 96).

Verschiedene Einflussfaktoren auf das Ausmaß und das Niveau der Internetnutzung wurden bereits empirisch ermittelt: So begünstigen Wettbewerb und Deregulierung des Telekommunikationssektors die Verbreitung preiswerter und breiter Internetzugänge (Guillén/Suárez 2005; Wallsten 2005). Stärker als die wirtschaftliche Situation wirkt sich außerdem der Grad der Einbindung eines Landes in die Weltgesellschaft – im Sinne von *world polity* – auf den Umfang und die Vielfältigkeit der Internetnutzung aus. Er äußert sich u.a. in der Mitgliedschaft in übernationalen Organisationen sowie in der Existenz eines umfassenden Bildungs- und Wissenschaftssystems (Drori/Jang 2003).

Die verfügbaren empirischen Studien deuten darauf hin, dass der Ausstattung von Universitäten und Forschungseinrichtungen mit Internetzugängen insbesondere in den – gemessen am Forschungsoutput – peripheren Nationen der globalen *scientific community* geringe Priorität zugemessen wird. Der Versuch, die bestehende Marginalität durch strategische Förderung der Internetnutzung zumindest teilweise zu kompensieren, ist in den untersuchten Entwicklungsländern nicht zu beobachten. Es ist zu vermuten, dass Phänomene wie begrenzte Sichtbarkeit der Forschung zahlreicher Entwicklungsländer in Fachzeitschriften und ihre geringe Internetnutzung sowie Webpräsenz auf ähnliche Ursachen zurückgehen: Alltägliche Unwägbarkeiten (wie z.B. die unsichere Stromversorgung oder soziale Probleme), mangelhafte Ausstattung (z.B. in Bezug auf Rechnerzugang, Internetverbindungen, Forschungsgeräte und Literatur) und begrenzte Fortbildungsmöglichkeiten (z.B. in Webdesign, der Nutzung neuer Software oder der englischen Sprache) behindern die Präsenz in hochwertigen Fachjournalen ebenso wie die Sichtbarkeit im Web. Die geringe Nutzung der neuen Medien ist somit häufig nur ein Symptom von vielen, die auf materielle, politische und soziale Defizite in Forschung und Lehre hinweisen. Die Erwartung, dass die modernen IuK-Technologien eine generelle Nivellierung der Ungleichheit im globalen Wissenschaftssystem herbeiführen können, erscheint vor diesem Hintergrund überhöht.

3.3 Die webometrische Untersuchung internationaler Vernetzung

Bevor in Kapitel 4 Leitfragen zur webometrischen Analyse internationaler Vernetzung für die Fallstudie zusammengestellt werden, bleibt zu prüfen, welche Indikatoren genutzt werden können, um internationale Beziehungen im Web zu untersuchen. Eine Auswahl der hier präsentierten Internationalitätsmaße wird in der Fallstudie auf ihre Aussagekraft und Zuverlässigkeit hin geprüft, indem sie mit manuell klassifizierten Daten abgeglichen werden.

Internationale Studien wurden in der webometrischen Forschung bisher recht selten durchgeführt, wofür das geringe Alter der Webometrie und die andauernde Beschäftigung mit methodologischen Fragen eine Erklärung bieten kann. In einigen Studien wurden Daten über die Wissenschaftssysteme verschiedener Länder gesammelt, um sie statistisch zu beschreiben und untereinander zu vergleichen, ohne jedoch ihre internationale Einbettung oder ihre Vernetzung zu untersuchen. So suchen Thelwall et al. (2002) methodologisch relevante Zusammenhänge zwischen nationalen Unterschieden in der Beschaffenheit universitärer Websites, ihrer Abdeckung durch Suchmaschinen und dem Brutto-Inlands-Produkt. Eine andere Fragestellung verfolgen Thelwall et al. (2003), die zwei Länder vergleichen, um Differenzen in der virtuellen Sichtbarkeit verschiedener Disziplinen zu identifizieren.

Derzeit sind neun webometrische Untersuchungen bekannt, in denen die internationale Einbettung wissenschaftlicher Einrichtungen im World Wide Web mit Hilfe verschiedener Indikatoren analysiert wird. An jeder dieser Studien war die Statistical Cybermetrics Research Group der University of Wolverhampton in England beteiligt, die der Mathematiker und Informatiker Mike Thelwall leitet. Nicht in allen Beiträgen steht die Analyse der internationalen Vernetzung im Vordergrund. Vielmehr geht es zum Teil auch um methodologische Fragen (z.B. Thelwall 2004b), so dass nicht notwendigerweise ein differenziertes Set internationaler Indikatoren eingesetzt wird. Vier von ihnen (Smith/Thelwall 2002; Thelwall/Smith 2002; Thelwall/Tang/Price 2003; Cothey 2005) sind auf ein Netzwerk aus mehreren nationalen Webbereichen und die Links zwischen ihnen beschränkt. Zwei weitere Studien (Tang/Thelwall 2004; Thelwall 2004b) verzichten auf die Beschränkung und analysieren die allgemeine internationale Einbettung wissenschaftlicher Websites. In den drei übrigen Fällen (Thelwall/Tang 2003; Li et al. 2005b; Thelwall/Zuccala 2008) wurde eine Mischform aus offener Erhebung und Netzwerkanalyse gewählt. Die genannten Untersuchungen unterscheiden sich außerdem in der Definition ihrer landesspezifischen Webbereiche: Sie sind zum Teil aus einzelnen Universitätswebsites zusammengesetzt, indem die erhobenen Daten entweder länderweise zusammengefasst oder in Form deskriptivstatistischer Kennzahlen wie Durchschnittswerten oder Medianen aggregiert wurden. Eine Alternative, die allerdings nur bei der Nutzung von Suchmaschinen Anwendung findet, ist die Abgrenzung durch akademische Second-Level-Domains wie „ac.uk“. Eine weitere Differenz liegt darin, dass die meisten

Studien auf der Erhebung vollständiger Universitätswebsites beruhen, während zwei (Tang/Theilwall 2004; Li et al. 2005b) auf bestimmte Disziplinen fokussieren und entsprechend die Internetauftritte von Universitätsinstituten und ähnlichen organisationalen Einheiten analysieren.

Um der Vielfalt der möglichen Bezugsebenen Rechnung zu tragen, wird in diesem Abschnitt der allgemeine Ausdruck „Webbereich“ als Überbegriff für einzelne oder akkumulierte Websites sowie anhand akademischer Second-Level-Domains eingegrenzter Bereiche etc. verwendet. Handelt es sich um einen Bereich, von dem Outlinks ausgehen, wird auch von „Quellbereich“ gesprochen, während es im Zusammenhang mit Inlinks um „Zielbereiche“ geht.

In den folgenden Abschnitten werden potentiell nützliche webometrische Internationalitätsindikatoren vorgestellt. Der Großteil der präsentierten Maße wurde eigens für diese Dissertation definiert und benannt. Sofern sie jedoch – in der hier beschriebenen Rohform oder in einer standardisierten Version – bereits empirisch genutzt wurden, wird dies in Überblickstabellen ausgewiesen. Ob den einzelnen Studien Daten auf der Ebene von Domains zugrundeliegen oder ob Daten von Instituts- oder Universitätswebsites gepoolt oder statistisch aggregiert wurden, wird nicht spezifiziert, da es in diesem Zusammenhang irrelevant ist. Gleiches gilt für die Unterscheidung, ob Suchmaschinen oder Webcrawler zur Erhebung eingesetzt wurden (vgl. Abschnitt 2.4). Sind einzelne Indikatoren nur mit Hilfe eines bestimmten Erhebungswerkzeugs berechenbar, wird darauf jeweils in der Beschreibung hingewiesen.

Die Internationalitätsmaße unterscheiden sich im Hinblick darauf, ob sie auf Inlink- oder Outlinkdaten basieren. In beiden Fällen beziehen sie sich vorwiegend auf die nationale Verortung von Inlinkquellen bzw. Outlinkzielen. Bei der Verwendung von Suchmaschinen wird gelegentlich auch ihre Verteilung auf bestimmte Sprachen einbezogen. Eine weitere Kategorie sind Maße, die speziell zur Untersuchung internationaler Netzwerke geeignet sind. Die meisten Erhebungen stützen sich auf ähnliche Grunddaten, wie beispielsweise die Anzahl der Seiten in einem Webbereich, die Zahl aller In- und Outlinks sowie aller In- und Outlinks aus bzw. in bestimmte Top-Level-Domains.

Die absoluten In- und Outlinkwerte sind nicht nur als Basis für die Berechnung von Indikatoren wertvoll. Sie stellen eine sinnvolle Ergänzung zu relativen Maßen dar, indem sie Aufschluss über die Größenverhältnisse der untersuchten Entitäten geben (Luukkonen et al. 1993). Sofern die Gesamtzahl aller In- oder Outlinks herangezogen wird, handelt es sich grundsätzlich um externe Links. Interne Links werden in webometrischen Studien üblicherweise ausgeschlossen, da sie spezifischen Zwecken wie z.B. der Navigation dienen und keine Aussagekraft in Bezug auf die äußere Wirkung eines Internetauftritts besitzen (Li

2003). Zur Standardisierung werden neben webbezogenen Daten wie der Anzahl der Webseiten im Referenzbereich auch webexterne Daten herangezogen, etwa die Zahl der MitarbeiterInnen der untersuchten Einrichtung. Auch Maße der Forschungsproduktivität sind hier prinzipiell einsetzbar. Abbildung 3 gibt eine Übersicht über häufig verwendete Standardisierungsgrößen, die unvollständig bleiben muss, da insbesondere die Standardisierungsgrößen nahezu beliebige Möglichkeiten der Variation bieten.

- Anzahl der Webseiten
- Anzahl der Outlinks resp. Inlinks
- Anzahl der TLDs/Länder/Regionen/Kontinente/Sprachen
- Personalumfang der untersuchten Einrichtungen
- Publikationsoutput der untersuchten Einrichtungen

Abbildung 3: Webometrische Standardisierungsgrößen

Im Folgenden werden mögliche sowie bereits empirisch genutzte inlink- und outlinkbasierte Maße (Abschnitt 3.3.1 und 3.3.2) sowie Netzwerkmaße (Abschnitt 3.3.3) vorgestellt. Aufgrund der großen Auswahl an Standardisierungsgrößen besteht auch hier kein Anspruch auf Vollständigkeit. Vielmehr geht es darum, einen Überblick über die bestehenden Möglichkeiten zu geben, um feststellen zu können, welche Aspekte einer genaueren Prüfung bedürfen (Abschnitt 3.3.4).

3.3.1 Inlinkbasierte Internationalitätsmaße

Auf der Grundlage von Inlinks lassen sich drei verschiedene Arten von Indikatoren berechnen, die sich auf den Stellenwert internationaler Inlinks, die Streuung der Inlinks sowie das Spektrum internationaler Inlinkquellen beziehen. Von ihnen wird Aufschluss über die internationale Sichtbarkeit (Li et al. 2005b: 192) oder „international recognition“ (Cothey 2005: 213) erwartet. Zu jedem Aspekt werden in diesem Abschnitt mögliche Maße in der unstandardisierten Rohform vorgestellt. Sie werden anschließend in Tabelle 1 zusammengefasst, die gegebenenfalls auch die Studien ausweist, in denen sie oder eine standardisierte Variante bereits zur Anwendung gekommen sind. Eine offene, das ganze Web umfassende Ermittlung von Inlinks ist lediglich mit Suchmaschinen möglich, da nur sie die Kapazitäten besitzen, einen großen Teil des Webs zu erfassen (vgl. Abschnitt 2.4.4). Bei Webcrawlern besteht jedoch zumindest die Möglichkeit, geschlossene länderübergreifende Linknetzwerke zu untersuchen (z.B. alle Verweise zwischen Websites von Universitäten in der EU) und die gefundenen Linkbeziehungen in diesem begrenzten Rahmen zu analysieren.

Alle inlinkbasierten Internationalitätsmaße setzen die Unterscheidung von einheimischen und internationalen Inlinks voraus. Wenngleich im Web keine eindeutigen Kennzeichen für die nationale Verortung zu finden sind, gibt es Anhaltspunkte. Beliebte ist die Auswertung von Top-Level-Domains (TLDs) als „proxy for country“ (Cothey 2005: 213). Dabei werden die gefundenen Inlinks auf eine Website anhand des Länderkürzels nach Herkunftsländern klassifiziert und in einheimische und ausländische Inlinks aufgeteilt. Einige Suchmaschinen bieten die Option, die Suche auf Links aus einer einzigen TLD zu begrenzen, so dass eine Klassifikation der Linkquellen nicht nötig ist. Die Zahl der internationalen Inlinks lässt sich in diesem Fall durch Subtraktion der einheimischen Inlinks von der Gesamtsumme der Inlinks ermitteln. Problematisch an der Auswertung von Top-Level-Domains ist, dass URLs mit sektoralen TLDs keine Information zur nationalen Anbindung enthalten und dass URLs mit einer landesspezifischen TLD nicht in jedem Fall einen Bezug zu dem jeweiligen Land implizieren (vgl. Abschnitt 3.3.4). Ist die Erhebung auf ein Netzwerk von Hochschulen und anderen staatlichen Forschungseinrichtungen beschränkt, dürfte das Landes Kürzel jedoch – anders als bei der offenen Suche nach Inlinks – eine recht hohe Zuverlässigkeit besitzen.

Eine Alternative, die von einigen Suchmaschinen angeboten wird, ist die Suche nach Links aus bestimmten Ländern. Analog zur Vorgehensweise beim Rückgriff auf TLDs werden hierbei die Inlinks aus dem eigenen Land von allen übrigen unterschieden. Der Anbieter MSN erlaubt es zur Zeit sogar, die absolute Zahl der internationalen Inlinks direkt zu ermitteln, indem Webseiten aus einem bestimmten Land oder einer TLD ausgeschlossen werden. Vorteilhaft an der länderweisen Suche nach Inlinks ist, dass auch Webseiten mit sektoralen TLDs einbezogen und nach Nationen sortiert werden. Nachteilig ist, dass die Klassifikationsprozedur unbekannt ist und ihre Zuverlässigkeit nicht eingeschätzt werden kann.

Von besonderer Bedeutung für die Analyse inlinkbasierter Internationalitätsmaße sind dabei drei im Folgenden näher ausgeführte Aspekte: der Stellenwert internationaler Inlinks sowie die Streuung und das Spektrum der Inlinkquellen. Tabelle 1 gibt einen Überblick über alle inlinkbasierten Internationalitätsmaße. Es ist jeweils angegeben, ob und in welcher der bisherigen Studien sie bereits Anwendung gefunden haben.

		Smith/Theilwall 2002	Theilwall/Smith 2002	Theilwall/Tang 2003	Theilwall/Tang/Price 2003	Tang/Theilwall 2004	Theilwall 2004b	Cothey 2005	Li et al. 2005b	Theilwall/Zuccala 2008
<i>a) Der Stellenwert internationaler Inlinks</i>										
I.	Inlinks aus der eigenen TLD								X	
II.	Inlinks aus anderen TLDs		X			X		X		X
III.	Inlinks aus dem eigenen Land									
IV.	Inlinks aus anderen Ländern									
<i>b) Die Streuung der Inlinkquellen</i>										
I.	Inlinks aus bestimmten ccTLDs									
II.	Inlinks aus bestimmten gTLDs					X				
III.	Inlinks aus bestimmten Ländern									
IV.	Inlinks aus bestimmten Weltregionen					X				
V.	Inlinks aus bestimmten Kontinenten									
VI.	Inlinks aus bestimmten Ländergruppen (Industrieländer/Übergangsländer/Entwicklungsländer)									
VII.	Inlinks mit Quellseiten in bestimmten Sprachen			X						
<i>c) Das Spektrum der Inlinkquellen</i>										
I.	Anzahl der unter allen Inlinkquellen vertretenen ccTLDs									
II.	Anzahl der unter allen Inlinkquellen vertretenen gTLDs									
III.	Anzahl der unter allen Inlinkquellen vertretenen Länder									
IV.	Anzahl der unter allen Inlinkquellen vertretenen Weltregionen									
V.	Anzahl der unter allen Inlinkquellen vertretenen Kontinente									
VI.	Anzahl der unter allen Inlinkquellen vertretenen Industrieländer, Übergangsländer und Entwicklungsländer									
VII.	Anzahl der unter allen Inlinkquellen vertretenen Sprachen									

Tabelle 1: Inlinkbasierte Internationalitätsmaße und ihre bisherige Anwendung in webometrischen Studien

a) Der Stellenwert internationaler Inlinks

Ein Weg, um verschiedene Webbereiche im Hinblick auf ihre internationale Sichtbarkeit zu vergleichen, ist die Bestimmung des Stellenwerts internationaler Inlinks (I.). Hierzu kann beispielsweise der Anteil der internationalen Inlinks – definiert als Links aus anderen TLDs oder von Suchmaschinen klassifizierten Ländern – als prozentualer Anteil an der Gesamtzahl der Inlinks angewiesen werden. Als internationale Version des WIF ist auch eine

Standardisierung an der Websitegröße der untersuchten Einrichtung möglich (Cothey 2005: 213). Eine weitere Referenzgröße, die bereits einer Variante des allgemeinen WIF zugrundegelegt wurde, ist der Personalbestand der untersuchten Einrichtungen (Thelwall 2001b). Dieses Maß eignet sich für den Vergleich einzelner Organisationen, kann aber durch deskriptivstatistische Aggregation zu Medianwerten o.ä. auch für den Vergleich ganzer Länder eingesetzt werden.

Da die Möglichkeit, eine Suchmaschinenanfrage wie bei MSN direkt auf internationale Inlinks zu beschränken, recht neu ist, wurde bisher zum Teil auf indirekte Vorgehensweisen zurückgegriffen. Beispielsweise ermittelten Li et al. (2005b) den Stellenwert internationaler Inlinks anhand des Anteils der Inlinks aus dem eigenen Land bzw. aus der eigenen TLD, was eine einfache und transparente Lösung ist (II).

Da Suchmaschinen zum Teil auch die Abfrage nach Inlinks aus bestimmten Ländern erlauben, könnten auch Inlinks aus anderen (III.) und aus dem eigenen Land (IV.) die Basis helfen, den Stellenwert internationaler Inlinks zu bestimmen.

b) Die Streuung der Inlinkquellen

Die Analyse der Streuung der Inlinkquellen dient der Identifikation nationaler oder regionaler Schwerpunkte in der internationalen Sichtbarkeit.

Häufig werden dazu die Inlinks aus einer bestimmten landesspezifischen Top-Level-Domain (ccTLD) ins Verhältnis zur Gesamtzahl der Inlinks gesetzt oder auf die Größe des untersuchten Webbereichs bezogen (I.). Auch die Anzahl der Inlinks in einzelne sektorspezifische Top-Level-Domains (gTLD) kann bestimmt werden, um einen Eindruck vom gesamten Linkprofil und von der Größenordnung des nicht lokalisierbaren Linkanteils zu bekommen (II.).

Bei der Verwendung von Suchmaschinen bietet sich auch hier die Nutzung der Inlinkabfrage nach Ländern an, um den relativen Stellenwert einzelner Nationen zu bestimmen (III.). Um einen Überblick über die weltweite Verteilung der Inlinks zu bekommen, können die Inlinks aus bestimmten Weltregionen ermittelt werden (IV.). Während entsprechende Abfragen früher bei der Suchmaschine Altavista möglich waren (vgl. Tang/Thelwall 2004: 477), erfordert dieses Maß heute eine nachträgliche manuelle Aggregation. Da Suchmaschinen die länderweise Abfrage meist nur für eine begrenzte Auswahl von Staaten anbieten, erscheint die Abfrage einzelner Top-Level-Domains hier als die geeignetere Grundlage, um globale Zusammenhänge abzubilden. Auf derselben Datengrundlage können Inlinks nach Kontinenten zusammengefasst werden (V.).

Ebenso lassen sich die Inlinks aus einzelnen TLDs nach ihrer Zugehörigkeit zu Ländergruppen von unterschiedlichem Entwicklungsstand gruppieren (VI.). Indikatoren zur Vertei-

lung der Inlinkquellen auf Industrieländer, Entwicklungsländer und die Übergangsländer des früheren Ostblocks können eine gute Ergänzung zu Maßen der weltweiten Streuung von Inlinks sein.

Untersuchungen die auf Suchmaschinendaten beruhen, können außerdem auf Informationen zur Sprache der Quelldokumente von Inlinks zurückgreifen (VII.). Auch hier ist unklar, wie die Bestimmung der Sprache erfolgt und entsprechend wie zuverlässig die Klassifikation ist (vgl. Abschnitt 3.3.4). Die Streuung der Quelldokumente nach Sprachen kann jedoch weiter dazu beitragen, regionale Schwerpunkte zu identifizieren und das Gesamtprofil zu abzubilden.

c) Das Spektrum der Inlinkquellen

Die dritte Kategorie von inlinkbasierten Indikatoren dient der Untersuchung der Frage, wie breit die weltweite Einbettung bestimmter Webbereiche ist. In der Bibliometrie werden ähnliche Maße, wie die Zahl der Staaten, mit denen ein Land in einem bestimmten Zeitraum Koautorenschaftsbeziehungen aufweist, gelegentlich verwendet, um das Spektrum der internationalen Kontakte zu illustrieren (NSB 2004: Tab. 5.22; Engels/Ruschenburg/Weingart 2005). In der Webometrie wurden sie dagegen bisher nicht genutzt.

Um abzubilden, wie breit die Quellen der Inlinks weltweit streuen, kann die Zahl der landesspezifischen Quell-TLDs herangezogen werden (I.). Auch hier gibt die Abdeckung der sektoralen TLDs zusätzlich Auskunft über die Breite der Verankerung (II.). Eine Alternative bei der Verwendung von Suchmaschinen besteht auch hier darin, die Anzahl der Länder zugrunde zu legen (III.).

Darüber hinaus kann die Anzahl der Regionen (IV.) und Kontinente (V.) unter den Inlinkquellen einen weiteren Anhaltspunkt in Bezug die Abdeckung der Welt geben. Wie bei den Streuungsmaßen sind vermutlich auch hier landesspezifische TLDs eine bessere Datengrundlage als länderweise Suchmaschinendaten. Ebenso ist eine Differenzierung nach Ländergruppen von unterschiedlichem Entwicklungsstand möglich (VI.). Auch die Anzahl der Sprachen, in denen die Quelldokumente der Inlinks verfasst sind, ist ein mögliches Maß für die Breite des Spektrums der Inlinkquellen (VII.).

Neben den üblichen Standardisierungsmöglichkeiten steht hier eine weitere zur Verfügung: Die auftretenden Größen können als Anteil an allen Elementen ihrer Art ausgedrückt werden. Ein Beispiel wäre der Anteil aller existierenden Staaten, der durch die Inlinks des untersuchten Webbereichs abgedeckt wird. Allerdings ist hier mit deutlichen Skaleneffekten zu rechnen, da die Abdeckung von Ländern, Weltregionen oder Kontinenten vermutlich überproportional zur Anzahl der Inlinks steigt.

3.3.2 Outlinkbasierte Internationalitätsmaße

Outlinks können ein Mittel sein, um die internationale Orientierung von Organisationen und anderen Entitäten zu analysieren. Sie lassen sich grundsätzlich wie Inlinks klassifizieren. Daher bestehen hier analoge Grundkategorien in Bezug auf den Stellenwert internationaler Outlinks, die Streuung der Outlinkziele sowie die Breite des Spektrums von Outlinkzielen. Eine Übersicht über die outlinkbasierten Maße bietet Tabelle 2, wobei jeweils ausgewiesen wird, in welchen Studien sie bereits eingesetzt wurden.

		Smith/Thelwall 2002	Thelwall/Smith 2002	Thelwall/Tang 2003	Thelwall/Tang/Price 2003	Tang/Thelwall 2004	Thelwall 2004b	Cothey 2005	Li et al. 2005b	Thelwall/Zuccala 2008
<i>a) Der Stellenwert internationaler Outlinks</i>										
I.	Outlinks in dieselbe TLD					X				
II.	Outlinks in andere TLDs		X							X
<i>b) Die Streuung der Outlinkziele</i>										
I.	Outlinks in bestimmte ccTLDs							X		
II.	Outlinks in bestimmte gTLDs					X		X		
III.	Outlinks in bestimmte Weltregionen									
IV.	Outlinks in bestimmte Kontinente									
V.	Outlinks in bestimmte Ländergruppen (Industrieländer/Übergangsländer/ Entwicklungsländer)									
<i>c) Das Spektrum der Outlinkziele</i>										
I.	Anzahl der unter allen Outlinkzielen vertretenen ccTLDs									
II.	Anzahl der unter allen Outlinkzielen vertretenen gTLDs									
III.	Anzahl der unter allen Outlinkzielen vertretenen Weltregionen									
IV.	Anzahl der unter allen Outlinkzielen vertretenen Kontinenten									
V.	Anzahl der unter allen Inlinkquellen vertretenen Industrieländer, Übergangsländer und Entwicklungsländer									

Tabelle 2: Outlinkbasierte Internationalitätsmaße und ihre bisherige Anwendung in webometrischen Studien

Allerdings ergeben sich Unterschiede, da die Ziele der Outlinks eines Webbereichs nur durch Webcrawler zu ermitteln sind. Zwar erlauben einige Suchmaschinen die Suche nach Links aus einem Webbereich in einen bestimmten anderen – z.B. eine Website oder eine Top-Level-Domain. Dabei werden jedoch nur die Seiten ausgegeben, von denen die Links ausgehen, nicht die Outlinkziele. Die Ergebnisse sind somit nicht direkt nachvollziehbar,

weshalb die Verwendung von Suchmaschinen zur Bestimmung eines vollständigen Outlinkprofils methodisch nicht sinnvoll ist. Mit Hilfe von Suchmaschinen lassen sich Outlinkdaten lediglich erheben, indem die Suche auf Linkmuster zwischen Paaren von Websites oder Second-Level-Domains begrenzt wird. Diese Abfragen zeigen jedoch nur einen Bruchteil der gesamten Outlinks und enthalten keine Angaben zu Sprachen oder Ländern (vgl. auch Abschnitt 2.4).

In der Folge stehen spezifische Optionen von Suchmaschinen – wie länderweise oder sprachweise Suche – bei der Auswertung von Outlinkdaten nicht zur Verfügung, was die Bandbreite der möglichen Indikatoren deutlich schmälert. Als primäres Kriterium für die Unterscheidung von Ländern bleiben lediglich die landesspezifischen Top-Level-Domains. Es stehen die gleichen Möglichkeiten der Standardisierung zur Verfügung wie bei den Inlinks. Ebenso sind auch bei den internationalen Outlinks Maße zu unterscheiden, die sich auf ihren Stellenwert, ihre Streuung und ihr Spektrum beziehen.

a) Der Stellenwert internationaler Outlinks

Um die internationale Ausrichtung eines Webbereichs zu untersuchen, können die Outlinks in andere TLDs als Grundmaß dienen (I.). Wie bei den Inlinks kann auch hier auf ein Komplementärmaß, nämlich den Anteil der Outlinks in die eigene TLDs zurückgegriffen werden (II.), wenn Suchmaschinen die direkte Suche nach internationalen Outlinks nicht erlauben und es erwünscht ist, den Ausschluss sektoraler TLDs transparent zu machen. Beide Maße wurden bereits angewandt, wobei sowohl die Anzahl aller Outlinks als auch die der Webseiten im untersuchten Webbereich zur Standardisierung herangezogen wurde (Thelwall 2004b und Thelwall/Zuccala 2008 resp. Thelwall/Smith 2002). Sofern einzelne Organisationen untersucht werden, scheint auch hier die alternative Standardisierungsgröße des Personalumfangs der untersuchten Einrichtung sinnvoll.

b) Die Streuung der Outlinkziele

Zur Bestimmung des internationalen Profils der Outlinks ist es nützlich, die Verteilung der Outlinkziele nach Top-Level-Domains zu bestimmen. Neben den landesspezifischen TLDs geben die sektoralen TLDs Hinweise auf den Stellenwert der nicht national lokalisierbaren Outlinks und der Orientierung gegenüber außerwissenschaftlichen Gesellschaftsbereichen (I. und II.).

Die Aggregation der Outlinkziele nach Weltregionen und Kontinenten auf der Grundlage der Anteile einzelner landesspezifischer TLDs trägt dazu bei, regionale Schwerpunkte in der internationalen Orientierung zu identifizieren (III. und IV). Ebenso lässt sich die Verteilung der Outlinks auf Industrie-, Übergangs- und Entwicklungsländer bestimmen (V).

c) Das Spektrum von Outlinkzielen

Um die Breite des Spektrums der Outlinkziele zu bestimmen, kann die Anzahl der vertretenen landesspezifischen und sektoralen Top-Level-Domains herangezogen werden (I. und II).

Ein weiterer Indikator in diesem Zusammenhang ist die Anzahl der Weltregionen (III.) und Kontinente (IV.), die durch Outlinks adressiert werden. Er ergibt sich aus den unter den Outlinkzielen vertretenen landesspezifischen TLDs.

Durch eine entsprechende Gruppierung der länderweisen Daten lässt sich auch die spezifische Abdeckung der Industrie-, der Übergangs- und der Entwicklungsländer bestimmen (V.).

3.3.3 Netzwerkmaße

Studien zur gezielten Untersuchung von Ländernetzwerken steht grundsätzlich eine breitere Auswahl an Maßen zur Verfügung als offenen Erhebungen. Sie können neben den bereits beschriebenen Inlink- und Outlinkmaßen auch netzwerkspezifische Indikatoren nutzen. Hierbei bestehen zahlreiche Optionen zur Standardisierung, da – anders als im offenen Erhebungsdesign – sowohl Informationen über den Quellbereich als auch den Zielbereich der Links vorliegen. So kann beispielsweise bei der Untersuchung chinesischer und taiwanesischer Universitäten der Umfang der Webseiten oder der Personalumfang in beiden Ländern – also im Quell- *und* im Zielbereich – bestimmt werden, während dies bei der offenen Suche nach Inlinks auf chinesische Universitätsseiten nicht möglich ist.

Untersuchungen internationaler Netzwerke bauen in der Regel auf Varianten der bereits oben angeführten grundlegenden Inlink- und Outlinkdaten auf (vgl. Tabelle 1 und Tabelle 2). Sie werden in unterschiedlicher Weise standardisiert, die u.a. von den Möglichkeiten der Erhebungsinstrumente und der Verfügbarkeit von Daten abhängig ist. Daher wird an dieser Stelle lediglich eine Übersicht über die bereits in anderen Studien verwendeten Netzwerkmaße gegeben (vgl. Tabelle 3).

Eine Gruppe von Netzwerkmaßen besteht aus Variationen standardisierter In- und Outlinkdaten. Sie beziehen sich zum Teil auf die Links zwischen zwei Ländern. Das Gesamtbild aller bilateralen Beziehungen zeigt in diesem Fall Schwerpunkte und Lücken des internationalen Netzwerks. Zur Standardisierung können verschiedene Größen genutzt werden. Um die Anzahl der Inlinks zu vergleichen, die ein einzelnes Land aus einem bestimmten anderen Land erhält, können beispielsweise die Zahl der untersuchten Webseiten oder der Personalumfang in beiden Ländern herangezogen werden. Wie bei den Inlink- und Outlinkmaßen geht es auch hier weniger um die Bestimmung der idealen Standardisierungsgröße als darum, Möglichkeiten der Indikatorenbildung auf der Grundlage von Linkdaten zusammenzustellen. Daher werden die folgenden Maße am Beispiel der Standardisierung anhand der Webseitenanzahl von Quell- und Zielbereich vorgestellt. Es sei jedoch ausdrücklich darauf

hingewiesen, dass Variationen der Bezugsgrößen und ihrer Verrechnung möglich und durchaus verbreitet sind.

		Smith/Theilwall 2002	Theilwall/Smith 2002	Theilwall/Tang 2003	Theilwall/Tang/Price 2003	Tang/Theilwall 2004	Theilwall 2004b	Cothey 2005	Li et al. 2005b	Theilwall/Zuccala 2008
I.	Anzahl der Inlinks aus einem bestimmten Webbereich	X	X						X	X
II.	Anzahl der Outlinks, die auf einen bestimmten Webbereich verweisen	X	X						X	X
III.	Anteil der Quellseiten in einer bestimmten Sprache an allen Seiten mit Links von Webbereich A nach Webbereich B			X	X					
IV.	Anteil der verlinkten Netzwerkknoten an allen möglichen Paaren								X	

Tabelle 3: Netzwerkmaße und ihre bisherige Anwendung in webometrischen Analysen

Im Übrigen müssen die verglichenen Einheiten nicht notwendigerweise Länder sein – die Netzwerkindikatoren könnten beispielsweise auch auf die Universitäten eines Landes angewendet werden. Um die Websites wissenschaftlicher Einrichtungen in verschiedenen Ländern abzugrenzen, stehen auch hier mehrere Verfahren zur Verfügung. Sofern akademische Second-Level-Domains existieren, können sie dazu genutzt werden, um den gesamten Universitätssektor eines Landes zu umreißen. Andernfalls können die Websites von Hochschulen, Universitätsinstituten oder außeruniversitären Einrichtungen und die Links zwischen ihnen auch einzeln erhoben und akkumuliert werden. Da in beiden Fällen ausschließlich Links zwischen wissenschaftlichen Einrichtungen erhoben werden, die ihre Website in aller Regel unter der jeweiligen nationalen Top-Level-Domain betreiben, kann hier eine hohe Aussagekraft der TLD unterstellt werden. Insofern sind bei der nationalen Verortung Alternativen – wie die länderweisen Suchoptionen – zu vernachlässigen.

Basale Maße für die Untersuchung der Linkbeziehungen zwischen zwei Ländern sind die Anzahl der Inlinks bzw. Outlinks zwischen ihnen (I. und II.), die beispielsweise anhand der Zahl der Webseiten im Zielbereich oder im Quellbereich standardisiert werden kann. Hierbei handelt es sich um bilaterale Nebenformen allgemeiner In- und Outlinkindikatoren zum Stellenwert internationaler Links. Auch sie unterliegen Skaleneffekten: So betont ein Indikator, dem die Zahl der Inlinks aus einem bestimmten Webbereich pro Webseite zugrunde liegen, kleinere Webbereiche, da sie im Verhältnis zu größeren Sektionen üblicherweise mehr Inlinks pro Seite erhalten. Dagegen erlaubt der zweite Indikator, bei dem die Zahl der

Outlinks in einen definierten Bereich pro Quellseite berechnet wird, die Identifikation besonders beliebter Ziele innerhalb einer Ländergruppe. Allerdings sind große Webbereiche hier möglicherweise unterrepräsentiert, sofern sie zu größeren Teilen Dokumente ohne externe Links enthalten (Thelwall/Smith 2002).

Ein verbreitetes Verfahren ist die Standardisierung bilateraler Linkbeziehungen anhand von Kennwerten zu Quell- und Zielbereich. Eine ähnliche Vorgehensweise ist auch in der Bibliometrie üblich. Ein zentrales bilaterales Maß zur Analyse von Koautorenschaftsbeziehungen ist *Salton's measure*, das die Anzahl der gemeinsamen Kopublikationen zweier Entitäten anhand der Wurzel des Produkts aller ihrer Publikationen standardisiert. Eine Alternative stellt *Jaccard's measure* dar, bei dem die gemeinsamen Kopublikationen zweier Entitäten durch die Summe ihrer übrigen Publikationen geteilt werden. Beide Standardisierungsformen weisen spezifische Vor- und Nachteile auf, da sie bestimmte Koautorenschaftsbeziehungen besonders stark hervorheben oder besondere Visualisierungsmöglichkeiten bieten (für eine genauere Darstellung vgl. Luukkonen et al. 1993; Wagner/Leydesdorff 2005: Fußnote 8). Fest steht jedoch, dass sie dazu beitragen, Effekte zu reduzieren, die auf die Größe der verglichenen Forschungssysteme zurückgehen.

Bilaterale Linkwerte können zu Standardisierungszwecken beispielsweise durch das Produkt der Webseitenzahl von Quell- und Zielbereich geteilt werden. Die Variante, bei der das Produkt des Personalumfangs als Referenzgröße verwendet wird, ist auch als *propensity to link* bekannt (Smith/Thelwall 2002: 374). Allerdings sind Links – anders als Koautorenschaftsdaten – gerichtet und charakterisieren eine asymmetrische Beziehung. Insofern lässt sich dieser Indikator für Links von Webbereich A auf Webbereich B ebenso berechnen wie für Links von B nach A. Die Unterscheidung von In- und Outlinks spielt bei diesem Indikator insofern keine Rolle, als weder eine Fokussierung auf Quell- und Zielbereich erfolgt, sondern das Verhältnis zwischen ihnen im Mittelpunkt steht.

Eine multilaterale Version dieses Indikators ist die *mean number of international peer inlinks* (vgl. Li et al. 2005b). Dabei liegt der Fokus jeweils auf den Outlinks, die ein Webbereich in andere Webbereiche aufweist. Um Größenunterschiede zu neutralisieren, verwenden Li et al. (2005b) als Standardisierungsgröße die Anzahl der Einrichtungen im Quellbereich, die mit der Summe der Einrichtungen in allen Zielbereichen multipliziert wird.

Bei der Verwendung von Suchmaschinen als Erhebungswerkzeug ist eine Berücksichtigung der Sprache der Outlinkquellen möglich (III.). Dies kann Aufschluss über den Stellenwert verschiedener Sprachen in einem Netzwerk geben, z.B. innerhalb der EU (vgl. Thelwall/Tang/Price 2003). Die Anzahl von bilateral verlinkten Seiten in einer bestimmten Sprache wurde empirisch bereits als Anteil an allen (Thelwall/Tang 2003) Webseiten eines

Landes ausgedrückt als auch einer Standardisierung durch das Produkt der Webseitenzahl im Quell- und Zielbereich unterzogen (Thelwall/Tang/Price 2003).

Ferner lässt sich ein Netzwerk als Ganzes charakterisieren. Seine Dichte lässt sich durch den Anteil der Knoten, die durch Links in einer beliebigen Richtung verbunden sind, an allen möglichen Knotenpaaren beschreiben (IV.). Sinnvoll ist seine Anwendung insbesondere im Vergleich mehrerer Gruppen von Internetauftritten. Li et al. (2005b) verwenden diesen Indikator, um die Vernetzung universitärer Institute mehrerer Disziplinen in verschiedenen Ländern zu vergleichen, so dass die internationale Einbettung unberücksichtigt bleibt. Sie kann jedoch einbezogen werden, indem disziplinäre Strukturen länderübergreifend analysiert werden, beispielsweise anhand der Zahl der grenzüberschreitend verlinkten Mathematik- und Physik-Institute innerhalb der EU. Auch zum Vergleich regionaler Netzwerke kann dieser Indikator dienen, indem z.B. der Anteil der durch Links verbundenen Universitäten in Amerika, Europa und Asien bestimmt wird. In jedem Fall ist für diese Art der Untersuchung eine Bezugsgröße unterhalb der Länderebene – wie Universitäts- oder Instituts-Websites – zu wählen. Da vermutlich der Großteil aller nationalen Wissenschaftssysteme Linkbeziehungen untereinander aufweist, würde die Analyse akademischer Second-Level-Domains hier zu unspezifischen Ergebnissen führen.

3.3.4 Offene Fragen bei der Untersuchung internationaler Zusammenhänge im World Wide Web

Wie sich gezeigt hat, gibt es zahlreiche Möglichkeiten, auf der Basis von Inlinks und Outlinks webometrische Internationalitätsindikatoren zu bilden, die sich für die offene Anwendung oder die Verwendung zur Netzwerkanalyse eignen. Der Großteil der genannten Maße wurde erst für die vorliegende Dissertation konstruiert und ist, wie die tabellarischen Übersichten zeigen, bisher nicht empirisch eingesetzt worden. Die folgende Fallstudie wird dazu beitragen zu prüfen, inwiefern sie zu sinnvollen – also aussagekräftigen und zuverlässigen – Ergebnissen führen.

Bereits an dieser Stelle sind einige Aspekte hervorgetreten, die im Zusammenhang mit webometrischen Analysen von internationalen Dimensionen klärungsbedürftig sind. Ein Grundproblem besteht darin, dass es im Web nur wenige Ansatzpunkte für die nationale Verortung der URLs von Linkquellen oder -zielen gibt. Gleiches gilt für Zuordnung zu gesellschaftlichen Sektoren (vgl. Abschnitt 2.3.4). Die Top-Level-Domain stellt diesbezüglich den einzigen unmittelbar und ohne weiteren Aufwand zugänglichen Hinweis dar:

„The target TLD is the most general information obtainable from the text of an URL without accessing other contents of its hosting page“ (Thelwall 2004b: 130).

Insgesamt existieren derzeit spezifische Top-Level-Domains für knapp 250 Länder (vgl. IANA 2005). Ihre Vergabe erfolgt auf Grundlage der Länderliste der Internationalen Standardisierungsorganisation (ISO), die ausschließlich von den Vereinten Nationen anerkannte Staaten und jeweils zweibuchstabile Abkürzungen umfasst (Wass 2003b). Daneben bestehen zwei übernationale Kürzel: Die sektorale TLD „int“ ist der Nutzung durch internationale, auf überstaatlichen Abkommen basierende Organisationen (von den Vereinten Nationen über die Europäische Zentralbank bis hin zu Interpol) vorbehalten. Sie können jedoch auch andere Adressen wählen. So nutzt z.B. die Vereinten Nationen die Adresse www.un.org, die World Meteorological Organization entsprechend ihrer geographischen Verortung www.wmo.ch. 2005 wurde eine weitere übernationale TLD eingeführt – „eu“ für Europa –, die allen Nutzergruppen offensteht. Übernationale Aktivitäten sind keine Bedingung für den Zugang zu einer Domain in diesem Bereich, so dass unklar ist, inwiefern die NutzerInnen in einem einzelnen Mitgliedsstaat oder auf multi- bzw. supranationaler Ebene zu verorten sind. Andere nicht-länderspezifische Top-Level-Domains wie „com“ oder „org“ sind für spezifische Organisationstypen – z.B. Unternehmen oder NGOs – vorgesehen, die nicht notwendigerweise in mehreren Staaten präsent sind. Diese sektoralen TLDs enthalten somit keinen Hinweis auf den nationalen oder übernationalen Status.

Aufgrund des bequemen Zugangs ist der Einsatz von TLDs als Indikator nationaler Zugehörigkeit verbreitet (vgl. u.a. Rousseau 1997; Almind/Ingwersen 1997; Wormell 2001; Vaughan/Shaw 2003; Li et al. 2005b). Bei Erhebungen, die auf Links zwischen Websites mit bekannter Länderzugehörigkeit – z.B. Links zwischen den Internetauftritten der Universitäten mehrerer Staaten – beschränkt sind, kann die TLD in aller Regel zuverlässig Auskunft geben. Die Websites von Hochschulen und öffentliche Forschungseinrichtungen werden üblicherweise unter den entsprechenden Länderkürzeln eingerichtet. Einen Sonderfall stellen die USA dar, in der für tertiäre Bildungseinrichtungen die spezifische Endung „edu“ und für staatliche Forschungseinrichtungen üblicherweise „gov“ verwendet wird, so dass auch hier eine eindeutige nationale Zuordnung möglich ist. Anders verhält es sich bei offenen, nicht auf Netzwerke beschränkte Forschungsdesigns, in denen die Verteilung aller Linkquellen oder -ziele untersucht wird.

Es gibt keine weltweiten Richtlinien für die Vergabe von Webadressen. Jede TLD wird von einer nationalen – nicht notwendigerweise staatlichen – Stelle verwaltet. In einigen Staaten wird die Vergabepaxis eher restriktiv gehandhabt, was zur Folge hat, dass ein großer Teil der Websites unter anderen, leichter erhältlichen sektoralen oder landesspezifischen TLDs eingerichtet wird. So wurden beispielsweise in Schweden bis 2003 nur Unternehmen mit eingetragenen entsprechenden Namensrechten Adressen unter „se“ zugebilligt, um zu gewährleisten, dass Angebote innerhalb dieser Domain seriös sind, und um die Rechte öffentlicher und privater Körperschaften zu schützen. Um das Ausweichen von Privat-

personen und nicht entsprechend registrierter Firmen auf andere TLDs zu stoppen, werden inzwischen Anstrengungen unternommen, das Länderkürzel attraktiver zu machen (Lindén 2003). Bei der landesspezifischen TLD der USA („.us“) hat sich die verbindliche Nutzung von zusätzlichen Second-Level-Domains für den jeweiligen Staat und Third-Level-Domains für den Bezirk als unattraktiv erwiesen: Der eigentliche Domainname kam erst an vierter Stelle, und die gesamte Adresse war wenig einprägsam (z.B. www.amazon.seattle.wa.us) und hatte daher – nicht nur für Unternehmen – einen deutlich geringen Wert als eine kurze „.com“-Adresse (www.amazon.com). Diese Regelung, die die Verfügbarkeit von Adressen gewährleisten sollte, wurde 2001 aufgehoben (Wass 2003c). Adressen unter Indiens „.in“ wurden bis 2001 erst nach aufwendigen bürokratischen Prozeduren, nach Prüfung der Namensrechte und nur bei Wahl des monopolistisch-staatlichen Internetanbieters freigegeben. Trotz einiger Erleichterungen wurden selbst staatliche Bildungseinrichtungen und Behörden auch nach 2001 durch hohe Gebühren für die Domainnutzung abgeschreckt (Gandhi 2003).

In anderen Ländern sind Webadressen dagegen ganz ohne besondere Anforderungen erhältlich, so dass nicht notwendigerweise ein Bezug zu dem in der URL angegebenen Land oder Sektor gegeben ist. So setzte die südpazifische Inselgruppe Niue gezielt auf die internationale Vermarktung der eigenen TLD „.nu“, um den Aufbau der eigenen Internetanbindung zu finanzieren. Heute kann der geographisch isolierte Kleinstaat seiner Bevölkerung einen hochwertigen kostenlosen Internetzugang bieten. Interessanterweise werden Domains unter „.nu“ primär in Schweden genutzt, da „.nu“ auf schwedisch die Bedeutung „jetzt“ hat und das skandinavische Land daher zu den wichtigsten Adressaten der Vermarktungsinitiative Niues gehörte (StClair 2003). Auch TLDs, die für gängige Abkürzungen insbesondere in der englischen Sprache stehen, werden gezielt außerhalb des bezeichneten Landes vermarktet. Beispiele sind „.md“ (Moldavien), das eine attraktive Adress-Endung für US-amerikanische Ärzte (*medical doctors*) darstellt, oder „.tv“ (Tuvalu), „.fm“ (Mikronesien) und „.am“ (Armenien), die unter Radio- und Fernsehsendern beliebt sind (Gallup 2003). In der Konsequenz erscheint eine exemplarische Analyse der Aussagekraft von landesspezifischen Top-Level-Domains im Zusammenhang mit den Links in und auf die Websites wissenschaftlicher Einrichtungen geboten.

Bei der Verwendung von Suchmaschinen stellt die Nutzung spezieller Funktionen zur Beschränkung auf ausgewählte Länder eine Alternative zur Unterscheidung anhand von Top-Level-Domains dar. Jedoch lässt nicht jede Suchmaschine die Kombination von Linksuche und umfassender Länder- bzw. Sprachauswahl zu. Wie die Algorithmen kommerzieller Suchmaschinen an sich (vgl. Abschnitt 2.4.4), sind auch die Verfahren hinter diesen Optionen geheim. Daher ist zu prüfen, wie zuverlässig die Bestimmung von Nationalität oder Sprache einer Webseite funktioniert. Ein weiterer zu untersuchender Aspekt ist die Konsistenz der verschiedenen Suchfunktionen: Da die generelle Linksuche, die sprach- und die

länderweise Recherche möglicherweise auf unterschiedlichen Algorithmen und Datenbanken basieren, ist zu hinterfragen, inwiefern Ergebnisse untereinander stimmig sind. Ein Kriterium wäre beispielsweise, ob sich die Resultate der Länderabfrage zu einem ähnlichen Wert aufsummieren wie die allgemeine Linkabfrage oder ob sie nur einen begrenzten Ausschnitt wiedergeben. Sowohl die Unterscheidung der Länderzugehörigkeit von Webadressen anhand der Top-Level-Domain als auch die Verwendung spezifischer Suchmaschinenoptionen werfen folglich klärungsbedürftige Fragen auf.

3.4 Zwischenfazit

Verbunden mit einer Zusammenfassung der wesentlichen Ergebnisse dieses Kapitels werden abschließend im Hinblick auf internationale webometrische Studien relevante Aspekte identifiziert, die in der folgenden Fallstudie geprüft werden.

Wissenschaft ist in vielfacher Hinsicht kommunikationsbasiert und universell ausgerichtet (vgl. Abschnitt 3.1). Staatliche Einheiten spielen insofern eine Rolle, als im Wesentlichen die einzelnen Nationen die Rahmenbedingungen für Forschung schaffen. Kontakte über Ländergrenzen hinweg finden sich von Beginn der modernen Wissenschaft an und haben sich in den letzten Jahrzehnten aufgrund sich ändernder politischer, wirtschaftlicher und technologischer Rahmenbedingungen verstärkt. Die weitreichenden Kommunikationsmöglichkeiten, die das Internet bietet, haben diese Entwicklung erheblich erleichtert. Folglich ist das World Wide Web ein interessanter Ansatzpunkt zur empirischen Untersuchung internationaler Kommunikation in der Wissenschaft. Da es eine Vielzahl wissenschaftlicher Aktivitäten unterstützt, bildet es vermutlich ein breiteres Spektrum von direkten und indirekten Kontakten ab als der gängige Indikator der Kopublikation. Da es auch niederschwelligere Bezüge, wie z.B. einseitige oder wechselseitige Wahrnehmung, einschließt, lassen sich anhand des Webs Beziehungen zwischen Forschungseinrichtungen in entwickelten und Entwicklungsländern unter Umständen umfassender nachvollziehen, als anhand der relativ komplexen Kooperationsform der gemeinschaftlichen Veröffentlichung. Um zu prüfen, inwiefern webometrische Indikatoren ein bzgl. Umfang, Detailreichtum und Aussagekraft vergleichbares oder deutlich anderes Bild liefern als klassische bibliometrische Indikatoren, werden in der empirischen Fallstudie webometrische Daten mit Publikationsdaten über dieselben Fälle verglichen.

Bevor das World Wide Web als empirische Datengrundlage für die Wissenschaftsforschung herangezogen wird, ist zu bedenken, dass es sich um ein ambivalentes Medium handelt, das – anders als sein Name nahelegt – keineswegs weltumspannend ist und dessen Auswirkungen auf die internationale wissenschaftliche Kommunikation nicht abschließend geklärt sind (vgl. Abschnitt 3.2). Es kann als sicher gelten, dass einige Länder in Abhängigkeit von

den Zugangsmöglichkeiten mehr von seinen Vorzügen profitieren als andere. Ob dies die Ungleichheit innerhalb der globalen *scientific community* dauerhaft stabilisiert, sie noch verstärkt oder sie langfristig doch abschwächt, muss separat erforscht werden. Eine Konsequenz für die folgende Fallstudie ist jedoch, dass Forschungseinrichtungen aus Ländern mit unterschiedlichem Entwicklungsstand in die Fallauswahl einbezogen werden, um Hinweise auf mögliche Unterschiede in Umfang und Nutzungsweise der Internetauftritte sowie ihrer internationalen Einbettung durch Links zu sammeln. Ferner wird besonderes Augenmerk darauf gerichtet, in welchem Ausmaß und in welchem Zusammenhang Links auf und von Seiten in Entwicklungsländern auf den untersuchten Websites auftreten.

In der Webometrie sind Untersuchungen der internationalen Einbettung und Vernetzung bisher selten, wofür die Unsicherheit in Bezug auf die Datengrundlage sicherlich eine Ursache ist (vgl. Abschnitt 3.3). Entsprechend ist die Zahl der bereits verwendeten Indikatoren begrenzt, so dass eine weitere Differenzierung geboten scheint. Die hier vorgenommene systematische Definition und Zusammenstellung möglicher Maße belegt die Vielfalt der Optionen. Mit inlink- und outlinkbezogenen Maßen kann der Stellenwert der internationalen Links eines Webbereichs bestimmt werden. Ferner lässt sich mit ihnen die Streuung und das Spektrum von Inlinkquellen bzw. Outlinkzielen spezifizieren. Daneben erlaubt eine Reihe von Netzwerkmaßen die Analyse der Beziehungen und des Vernetzungsgrades innerhalb einer Gruppe von Webbereichen.

Vielversprechende Ansatzpunkte für webbasierte internationale Untersuchungen sind somit vorhanden, allerdings besteht auch hier Klärungsbedarf. Sofern er die weltweite Abdeckung von Suchmaschinen oder die optimale Standardisierung von Indikatoren betrifft, kann ihm im Rahmen dieser Arbeit nicht nachgegangen werden, da dies ein spezielles, wesentlich breiteres Forschungsdesign erfordert. In der folgenden Fallstudie wird jedoch geprüft, wie zuverlässig die in diesem Kontext äußerst nützlichen TLD- und landesspezifischen Suchmaschinenoptionen sind. Außerdem wird der Zusammenhang zwischen den in Inlinkquellen und Outlinkzielen verwendeten TLDs und der tatsächlichen nationalen Verortung näher betrachtet.

4. Leitfragen, Fallauswahl und Datenerhebung der Fallstudie zu internationalen Linkstrukturen in der Meeresforschung

Im Folgenden werden die Eigenheiten internationaler Linkstrukturen in einer Fallstudie detailliert untersucht. In diesem Kapitel wird der methodische Rahmen dieser Analyse umrissen, bevor in weiteren Abschnitten die empirischen Ergebnisse präsentiert (Kapitel 5) und im Hinblick auf die allgemeinen Forschungsfragen dieser Dissertation diskutiert werden (Kapitel 6).

Zunächst werden die auf Basis der beiden vorangegangenen Kapitel zusammengestellten Leitfragen für die Fallstudie vorgestellt (4.1). Danach wird ein geeignetes Forschungsfeld zur Klärung der Leitfragen identifiziert (4.2) und eine Auswahl von Fällen getroffen (4.3). Eine Beschreibung der Erhebung und Verarbeitung der herangezogenen Daten schließt das Kapitel ab (4.4).

4.1 Leitfragen für die empirische Fallstudie zur webometrischen Untersuchung internationaler Zusammenhänge

Unter Einbeziehung der Schlussfolgerungen aus den Kapiteln 2 und 3 lässt sich eine Reihe von Leitfragen zusammenstellen, die sich auf die besonderen Anforderungen an die Webometrie im Kontext internationaler Fragestellungen beziehen (vgl. Abbildung 4). Zusammengefasst tragen sie zur umfassenden Beantwortung der beiden Forschungsfragen dieser Dissertation bei, die die Möglichkeiten und Grenzen webometrischer Verfahren zur Erforschung internationaler Strukturen der Wissenschaft fokussieren (vgl. Kapitel 1.2).

In technischer Hinsicht gilt es, geeignete Ansatzpunkte für die nationale Verortung von Linkzielen und -quellen zu ermitteln. Dazu wird die generelle Aussagekraft von Top-Level-Domains als Indikator für Nationalität geprüft (Leitfrage 1).

Grundlegend ist ferner die Frage, inwieweit sich die durch webometrische Analyse gewonnene Einschätzung der ausgewählten Forschungseinrichtungen mit dem Bild überschneidet, das bibliometrische Indikatoren auf der Grundlage von Publikations-, Koautorenschafts- oder Zitationsdaten liefern (Leitfrage 2). So lässt sich Klarheit darüber gewinnen, ob die Webometrie – wie aufgrund der Vielfalt der Nutzungsmöglichkeiten des Internets gelegentlich vermutet – eine andere, möglicherweise breitere und vielfältigere Perspektive bietet. Davon hängt entscheidend ab, ob sie eine wertvolle Ergänzung klassischer, publikationsbasierter Verfahren sein kann.

Bei der dritten Leitfrage steht die Suche nach globalen Unterschieden im Mittelpunkt. Insbesondere stellt sich die Frage, ob das World Wide Web eine Möglichkeit bietet, die generelle

sozio-ökonomische Ungleichheit im Sinne der Idee vom *global research village* zu überwinden oder sich die ungleiche Verteilung der Ressourcen im Sinne der *digital divide*-These auch im virtuellen Raum abbildet. Aus diesem Grund wird für das weltweite Sample von Instituten aus Industrie- und Entwicklungsländern untersucht, inwiefern Umfang und Vernetzung der Internetauftritte mit dem Entwicklungsstand der jeweiligen Sitzländer zusammenhängt (Leitfrage 3). Im Rahmen dieser Leitfrage wird ferner geprüft, ob web- oder publikationsbasierte Indikatoren stärker mit der sozio-ökonomischen Entwicklung der Sitzländer wissenschaftlicher Einrichtungen zusammenhängen.

1. Wie aussagekräftig sind Top-Level-Domains als Indikatoren der nationalen Verortung?
2. Inwieweit ähneln sich die Bilder, die webometrische und bibliometrische Indikatoren von der internationalen Einbettung von Forschungseinrichtungen zeigen?
3. Hängen Präsenz und Vernetzung wissenschaftlicher Einrichtungen im World Wide Web (bzw. im Web of Science) mit dem Entwicklungsstand ihrer Sitzländer zusammen?

Abbildung 4: Leitfragen für die empirische Fallstudie

Interessant wäre außerdem die Frage, wie sich die Verteilung von Inlinks zu der faktischen Nutzung der Websites verhält. Die Annahme, dass es hier – gerade vor dem Hintergrund globaler Disparitäten – Unterschiede gibt, ist nicht unplausibel, da es wesentlich voraussetzungsvoller ist, eine eigene Website aufzubauen und sie mit anderen zu verlinken, als einfach andere Internetauftritte zu betrachten. Um die Frage zu klären, ob Inlinkwerte tatsächlich einen geeigneten Indikator für die Sichtbarkeit einer Seite im World Wide Web darstellen, wäre es notwendig, die Inlinkprofile der ausgewählten Forschungseinrichtungen mit Daten zur faktischen Nutzung ihrer Websites zu vergleichen. Anders als geplant, kann dies hier nicht geleistet werden, da keine geeigneten Daten erhältlich sind. Details dazu finden sich in Abschnitt 4.3.5.

In den vorangegangenen Kapiteln wurden noch einige weitere Aspekte hervorgehoben, deren empirische Prüfung lohnenswert erscheint (vgl. Kap. 2.6.2 und 3.4). Sie würden jedoch den Rahmen dieser Dissertation sprengen und werden daher nicht weiterverfolgt. Zur Klärung der oben genannten Leitfragen wird in den folgenden Kapiteln eine Auswahl von Internetauftritten wissenschaftlicher Einrichtungen empirisch untersucht.

4.2 Die Auswahl der Fälle

In der folgenden Fallstudie wurde eine Auswahl von Meeresforschungseinrichtungen einschließlich ihrer kompletten Websites sowie ihrer rezenten Publikationen untersucht. Soll eine Studie repräsentative Schlussfolgerungen erlauben, bedarf es einer Stichprobe, die ein strukturgeleiches, verkleinertes Abbild der Grundgesamtheit liefert – also möglichst einer Zufallsstichprobe. Eine alternative Vorgehensweise ist die Auswahl typischer Fälle. Die Ergebnisse der Untersuchung sind nicht als allgemeingültig anzusehen, sondern gelten nur für vergleichbare Fälle. In dem hier behandelten Fall ist die Grundgesamtheit – nämlich sämtliche Internetauftritte wissenschaftlicher Einrichtungen im World Wide Web – nicht abzugrenzen. Folglich kann keine Zufallsauswahl getroffen werden. Da es in dieser Studie um die Prüfung einer Methode geht und nicht um repräsentative Befunde zum weltweiten Wissenschaftssystem, ist hier eine gezielte Selektion angemessen. Möglichen Unterschieden zwischen Industrie- und Entwicklungsländern wurde durch die Auswahl von Fällen Rechnung getragen, die in Ländern mit unterschiedlichem Entwicklungsstand angesiedelt sind. Um ein Mindestmaß an Vergleichbarkeit sicherzustellen, wurde die Untersuchung auf wissenschaftliche Einrichtungen aus demselben Forschungsfeld beschränkt. Der Fragestellung entsprechend wurde ein Gebiet gesucht, das besonders international ausgerichtet ist, um zu gewährleisten, dass die betreffenden Internetauftritte ein hohes Maß an internationalen In- und Outlinks aufweisen.

4.2.1 Ein Fachgebiet mit starken internationalen Bezügen: die Meeresforschung

Ein Gebiet, das zahlreiche internationale Bezüge aufweist, ist die Meeresforschung. Sie ist ein interdisziplinäres Feld, das die physikalischen, geologischen, biologischen und chemischen Gegebenheiten und Prozesse in den Meeren und Ozeanen erforscht. Sie wird größtenteils den Erdwissenschaften zugerechnet, von denen viele ein überdurchschnittliches Maß an internationaler Kooperation und Koordination aufweisen:

„There are several fields within earth and space science that require an international coordination of research resources and facilities, for example, astronomy, oceanography, and atmospheric and space science“ (Luukkonen et al. 1992: 119).

Wenngleich in den letzten Jahren keine weltweiten feldspezifischen Statistiken zur internationalen Koautorenschaft erschienen sind, unterstreichen Daten aus dem „Third European Report on Science & Technology Indicators 2003“ der Europäischen Kommission die außergewöhnlich starke Neigung der Erd- und Weltraumwissenschaften zur grenzüberschreitenden Kooperation: Zwischen 1995 und 1999 wurden in diesem Fachgebiet in den USA (28,1%), in der EU der 15 Staaten (53,1%) und in Japan (37,0%) ein großer Anteil aller Publikationen in internationaler Koautorenschaft veröffentlicht, der in den USA und in der EU

der 15 Staaten lediglich von der Physik übertroffen wurde und der in Japan weit über allen anderen Disziplinen rangierte (EC 2003: Abb 5.4.3). Im gleichen Zeitraum lag der Anteil der internationalen Kopublikationen am Publikationsaufkommen in den USA (18,0%) und Japan (15,2%) deutlich niedriger. Für die EU-15 liegen keine aggregierten Daten vor, die einzelnen Mitgliedstaaten wiesen jedoch in den Jahren 1995 bis 1999 zwischen 29,3 Prozent (Großbritannien) und 50,8 Prozent (Portugal) an internationalen Kopublikationen auf (EC 2003: Abb. 5.4.2). Unter den Disziplinen, die die Erde als Ganzes untersuchen, zeichnen sich insbesondere Felder mit einer systemischen Perspektive durch eine starke Neigung zu internationaler Kooperation aus. So stellten in der Ozeanographie sowie im Feld Meteorologie und Atmosphärenforschung in den Jahren 2002 und 2003 24,3 Prozent resp. 25,8 Prozent aller Einträge im Science Citation Index internationale Kopublikationen dar, während der gesamte SCI einen Durchschnitt von lediglich 18 Prozent aufwies. In Fachgebieten, in denen das globale Gesamtbild vorwiegend durch die Akkumulation regionaler Befunde generiert wird, wie z.B. in der Ökologie oder der Erforschung von Süßwasserhaushalten, fiel der Anteil internationaler Kopublikationen deutlich geringer (21,3% resp. 18,7%), wenngleich noch überdurchschnittlich aus (Jappe 2007: 383).

Die Gründe für die internationale Orientierung der Erdwissenschaften sind zahlreich, wie das Beispiel der Meeresforschung zeigt: Ozeane sind allein aufgrund ihrer physischen Ausmaße ein außergewöhnlicher Forschungsgegenstand. Das Ziel, die Weltmeere zu vermessen und ihre weltweite Dynamik, ihre komplexen biologischen, chemischen und physikalischen Zusammenhänge und ihren Austausch mit der Atmosphäre zu verstehen, impliziert, dass Forschungsaktivitäten auf der ganzen Welt stattfinden. Dabei können einzelne Forschungsprojekte lediglich einen äußerst begrenzten Ausschnitt der Ozeansysteme dokumentieren, so dass koordiniertes, arbeitsteiliges Vorgehen mit einheitlichen Standards in der Datenerhebung notwendig scheint (vgl. Wooster 1969).

„The need for international cooperation between governments and scientists engaged in oceanographic research is great. This is especially true for projects of rather large dimensions [...] which require participation by many ships of diverse nationalities to make standardized and sometimes simultaneous observations and measurements“ (Miles 1972: 264f).

„[The] substantial level of scientific internationalism presents itself with a strong element of necessity. In order to reach consensus scientists need to agree upon calibration of instruments, standards for observation protocols, units of measurements and research methods. [...] In large scale projects in astronomy, oceanography, meteorology and other areas there enters in also systematic division of labour across observatories, amongst national expeditions, or laboratories“ (Elzinga 1996: 15f).

Erst durch internationale Abstimmung im Hinblick darauf, welche Art von Daten gesammelt wird, welche Forschungseinrichtungen die Erhebung in bestimmten Regionen übernehmen

und welche Standards für die Verarbeitung der Daten gelten, wird es möglich, ein umfassendes Gesamtbild zu generieren.

„[...] the physical aspects of [...] the oceans demand a high level of international cooperation if certain problems are to be investigated. The areas are vast and therefore require standardized observations and measurements together with support systems providing for the exchange of information derived from the investigations“ (Miles 1972: 267).

Die koordinierte Erhebung und der weltweite Austausch von Daten wird auf übernationaler Ebene durch zwischenstaatliche Einrichtungen wie die Intergovernmental Oceanographic Commission (IOC) der UNESCO, die World Meteorological Organization (WMO), das UN-Umweltprogramm (UNEP), andere UN-Organisationen oder das International Council of Scientific Unions (ICSU) unterstützt. So hat das 1961 gegründete Programm International Oceanographic Data and Information Exchange (IODE) des IOC inzwischen über 60 Datenzentren hervorgebracht, die über das Internet das Einspeisen und Abrufen ozeanographischer Daten erlauben (IODE 2009). Weitere Beispiele für weltweite Programme, von denen einige Unterprogramme mit regionalen Schwerpunkten enthalten, sind das Global Ocean Observing System (GOOS), das Global Sea-Level Observing System (GLOSS) oder das Global Coral Reef Monitoring Network (GCRMN). Auch Nichtregierungsorganisationen sind zum Teil in die internationalen Programme involviert und tragen auch selbst zur internationalen Abstimmung zwischen wissenschaftlichen und forschungspolitischen Akteuren bei. So besitzt ICSU seit 1957 ein Scientific Committee on Oceanic Research (SCOR), das u.a. mit der Identifizierung von Themen befasst ist, die für die internationale ozeanographische *scientific community* von hoher Aktualität und Relevanz sind und den gemeinschaftlichen Einsatz von Kompetenzen und Ressourcen erfordern.

ICSU organisierte auch das International Geophysical Year (IGY) in den Jahren 1957 und 1958, das – anders als die vorangegangenen International Polar Years von 1882/1883 und 1932/1933 – nicht nur die Polregionen, sondern die Ozeane und die Atmosphäre als Ganzes in den Mittelpunkt verstärkter internationaler Forschungsbemühungen stellte. Trotz des Kalten Krieges beteiligten sich ForscherInnen aus über 60 Nationen daran. Obwohl das IGY durchaus auch genutzt wurde, um nationale politische Interessen voranzutreiben (Elzinga 1993), kann es folglich als Hinweis darauf dienen, dass internationale Koordination und Kooperation in der Ozeanographie eine lange Tradition besitzen. In den Jahren um das IGY expandierte die *scientific community* dieses Fachgebiets deutlich (Wooster 1969). Meeresforschung wird heute im Hinblick auf den Einsatz von wissenschaftlichem und technischem Personal, Schiffen und anderer Ausrüstung in großem Stil betrieben und kann als *big science* eingeordnet werden. Da die Erforschung ganzer Erdsysteme erhebliche Ressourcen erfordert, ist die Bündelung der finanziellen Mittel mehrerer Nationalstaaten und die

länderübergreifende Nutzung von Ausrüstung – z.B. von Forschungsschiffen oder Bojen – verbreitet (Crane 1972).

Die räumlichen Dimensionen des Forschungsgegenstands Erde führen zu einer hohen wechselseitigen Abhängigkeit (Whitley 1984) von Forschungseinrichtungen in zahlreichen erd- und umweltwissenschaftlichen Disziplinen (vgl. Jappe 2007). Aus der beschriebenen Notwendigkeit, Forschungsaktivitäten zu koordinieren und zu standardisieren, aus der länderübergreifenden Finanzierung zahlreicher Projekte und Geräte sowie aus der Existenz unterstützender übernationaler Strukturen ergibt sich die starke Internationalität in der Meeresforschung. Insofern erscheint die Vermutung, hier ein besonders hohes Maß an grenzüberschreitenden Linkstrukturen zu finden, berechtigt. Vor dem Hintergrund der Kapitalintensität der Meeresforschung stellt sich jedoch die Frage, welchen Stellenwert Forschung in Schwellen- und Entwicklungsländern in diesem Fachgebiet einnimmt. Die SCI-Einträge im Bereich Ozeanographie in den Jahren 2002 und 2003 stammten zu 91,3 Prozent aus den zwanzig publikationsstärksten Ländern, während es über alle Forschungsgebiete nur 86,8 Prozent waren (Jappe 2007, Tab. 3). Dies legt den Schluss nahe, dass sich die Meeresforschung mehr als andere Disziplinen auf Länder mit gut ausgestatteten Forschungssystemen konzentriert. In eine ähnliche Richtung weisen die Ergebnisse einer Untersuchung der Publikationszahlen auf Basis des Science Citation Index des Jahres 2000 (Dastidar 2004). Sie ergab, dass unter den zehn – gemessen am Publikationsaufkommen – aktivsten Nationen im Fachgebiet Ozeanographie sämtliche G-8-Staaten sowie Australien und die Niederlande zu finden waren, wobei die USA eine dominante Position einnahmen. Dort befanden sich auch 23 der produktivsten 32 Meeresforschungseinrichtungen. Beides überrascht in Anbetracht des Kostenaufwands in dieser Großforschungsdisziplin wenig. Allerdings wies das Bruttoinlandsprodukt im Vergleich zu anderen Disziplinen nur eine schwache Korrelation mit dem Publikationsvolumen auf. Offensichtlich haben andere Faktoren – wie möglicherweise die geographische Lage oder die Existenz internationaler Programme mit regionalem Fokus – Einfluss auf das Engagement in diesem Forschungsfeld. Somit ist zu erwarten, dass die Aktivität von Entwicklungsländern in diesem Gebiet nicht ungewöhnlich gering ausfällt. Dafür spricht auch, dass sich im Jahr 2000 unter den 25 produktivsten Nationen in der Ozeanographie immerhin sechs Schwellenländer befanden, und zwar China, Taiwan, Indien, Mexiko, Südafrika und Brasilien (Dastidar 2004).

An 7,6 Prozent der SCI-Einträge im Fachgebiet Ozeanographie waren in den Jahren 2002 und 2003 AutorInnen in Entwicklungs- und Schwellenländern sowie in Ländern des früheren Ostblocks beteiligt. Dieser Anteil ist in Disziplinen mit starker Feldforschung in äquatornahen Klimazonen wie der Ökologie (8,2%) oder Gebieten mit starkem Anwendungsbezug und unmittelbarer ökonomischer Relevanz wie der Forschung zu Süßwasserhaushalten (9,3%) höher (Jappe 2007, Tab. 4). Vergleichbare Werte für die übrigen Kategorien des SCI liegen

nicht vor, so dass nicht geprüft werden kann, wie sich die Ozeanographie in dieser Hinsicht zu anderen Disziplinen verhält. Jedoch wird der Meeresforschung eine hohe Bedeutung im Zusammenhang mit globalen Umweltveränderungen zugemessen. Ihre weltweiten Dimensionen sowie die potentielle Betroffenheit vieler küstenreicher, niedrig gelegener Entwicklungsländer legen eine Zusammenarbeit von Nationen verschiedenen Entwicklungsstands nahe, die auch im Rahmen verschiedener internationaler Umweltabkommen wie der Klimarahmenkonvention der Vereinten Nationen unterstützt wird (Engels/Ruschenburg/Weingart 2005).

Insofern kann angenommen werden, dass die Meeresforschung sich nicht stärker auf die gut ausgestatteten Forschungsnationen in Nordamerika und der EU sowie auf Japan, Russland und Australien konzentriert als die meisten übrigen Disziplinen. Folglich ist sie ein geeignetes Feld für eine Fallstudie zu internationalen Linkstrukturen.

4.2.2 Kriterien für die Auswahl der Fälle

Im Folgenden werden sechs Kriterien beschrieben, anhand derer die Auswahl von Meeresforschungseinrichtungen getroffen wurde. Ziel war, Institute zu finden, die im Hinblick auf ihre Organisationsform, ihre fachliche Ausrichtung sowie ihre Größe vergleichbar sind und eine Analyse ihrer Websites und ihrer Publikationen erlauben.

Fachliche Ausrichtung

Es werden ausschließlich Institute berücksichtigt, deren fachlicher Schwerpunkt eindeutig auf der Meeresforschung liegt. Einrichtungen, die innerhalb dieses Forschungsfelds eine große Bandbreite an Themen bearbeiten, werden solchen mit starker Spezialisierung (z.B. auf Marine Mikroorganismen oder Marine Tropenökologie) vorgezogen. Reine Datenplattformen wie die nationalen Datenzentren des IODE-Programms der Vereinten Nationen werden nicht berücksichtigt.

Organisationstyp

Hinsichtlich der Organisationsform sind im Wesentlichen universitäre und außeruniversitäre Forschungseinrichtungen zu unterscheiden. Wie andere *big science*-Bereiche auch, ist die Meeresforschung aufgrund ihres interdisziplinären Charakters, ihrer Ressourcenintensität und dem nationalen Interesse an den Forschungsergebnissen – z.B. im Hinblick auf Fischereiwesen, Küstenschutz, Klimawandel etc. – häufig außerhalb von Hochschulen angesiedelt. In dieser Fallstudie wird außeruniversitären Instituten der Vorzug gegeben, was mit methodologischen Anforderungen an die Internetauftritte zusammenhängt (siehe nächster Punkt).

Art der Webpräsenz

Ein fixes Auswahlkriterium ist der Besitz einer Website, die in einer eigenen Domain abgelegt ist (wie z.B. <http://www.ifm-geomar.de>). Internetauftritte, die lediglich in einer eigenen *directory* gebündelt werden (wie z.B. <http://www.uni-bielefeld.de/iwt>), können nicht berücksichtigt werden, da Suchmaschinen hier keine Inlinkabfrage für den gesamten institutionellen Webbereich erlauben. Bei Internetauftritten, die Teil einer größeren Website sind, ist außerdem ungewiss, welche Inhalte sie selbst abdecken und welche in andere Bereiche ausgelagert sind (Fachbibliothek, EDV, Verwaltung etc.). Dies schränkt die Vergleichbarkeit ein. Anders als Institute an Hochschulen verfügen außeruniversitäre Einrichtungen in der Regel über eine eigene Domain. Außerdem sollte die Website nicht zu wesentlichen Teilen in Flash programmiert sein, da der Webcrawler die enthaltenen Links darin nicht extrahieren kann.

Sprache

Die Inhalte der Websites stellen eine wesentliche Grundlage der Untersuchung dar. Um umfassendes Verständnis sicherzustellen, beschränkt sich die Auswahl auf Fälle, deren Internetauftritte größtenteils in englischer oder deutscher Sprache verfasst sind.

Größe der Website

Aus Gründen der Vergleichbarkeit wurden Institute von ähnlicher Websitegröße zusammengestellt. Sehr große Internetauftritte, die aus zehntausenden oder mehr Seiten bestehen, wurden aus zwei Gründen ausgeschlossen: Sie sind extrem aufwendig zu erfassen und sie finden sich nur in Industrieländern. Daher erfolgte eine Beschränkung auf kleinere und mittlere Fälle. Zur pragmatischen Größenabschätzung wurde am 26. Januar 2007 bei der Suchmaschine Windows Live die Zahl der Seiten abgefragt. Zu berücksichtigen ist, dass die von der Suchmaschine angegebene Seitenzahl nur einen Näherungswert darstellt und erheblich von den durch einen Webcrawler ermittelten Daten abweichen kann. Sie erlaubt eine schnelle, aber grobe Einschätzung. Als Ausschlusskriterium wurde eine Websitegröße von 15.000 Seiten bei Windows Live gewählt. Stellte sich bei einer Probeerhebung mit dem SocSciBot heraus, dass die tatsächliche Größe 50.000 Seiten überschritt, wurden die Institute ebenfalls ausgeschlossen. Dies war zweimal der Fall (vgl. Abschnitt 4.3.2).

Alter der Einrichtungen

Da auch ein Abgleich von Web- und Publikationsdaten ausgeführt werden soll, können kürzlich neugegründete, umstrukturierte oder umbenannte Einrichtungen nicht in die Auswahl einbezogen werden. Die Präsenz im Web of Science setzt mehrjähriges Bestehen voraus,

da die Publikation von Artikeln sowie ihre Erfassung in der Datenbank einige Zeit in Anspruch nimmt.

4.2.3 Auswahlverfahren

In einer weltweiten webbasierten Recherche wurden mögliche Fälle zusammengestellt. Dabei wurden Datenbanken der Intergovernmental Oceanic Commission (wie <http://oceanportal.org/>, <http://www.oceanexpert.org>, <http://www.iode.org/contents.php?id=97> oder <http://www.africanoceans.net/>), Portale der *scientific community* (wie <http://www.indian-ocean.org>, <http://www.ioiscience.org>, <http://www.deutsche-meeresforschung.de> oder <http://www.meeresforschung.de>.) sowie von Forschungseinrichtungen erstellte weltweite Linklisten (wie <http://www.researchvessels.org>, <http://www.phys.ocean.dal.ca/other-sites.html>, oder <http://www.mth.uea.ac.uk/ocean/vl/>) herangezogen (alle Websites wurden zuletzt am 11.04.2007 aufgerufen).

Um die Variation zu begrenzen und etwaige nationale Spezifika hervorzuheben, wurde versucht, die Zahl der Länder zu begrenzen und jeweils mehrere Einrichtungen aus einem Land zu erheben. Mit den USA und Deutschland wurden zwei Industrienationen ausgewählt, die eine zentrale Stellung im globalen Wissenschaftssystem einnehmen und über eine umfangreiche Forschungslandschaft im Bereich Meeresforschung verfügen. Anhand der oben ausgeführten Auswahlkriterien wurden drei geeignete US-amerikanische und zwei deutsche Forschungseinrichtungen identifiziert. Für Entwicklungs- und Schwellenländer erwies sich die Konzentration auf ein einzelnes Land als unpraktikabel. Tabelle 4 und Tabelle 5 geben eine Übersicht über die Ergebnisse der Recherche und weisen den Kriterien entsprechende bzw. nicht entsprechende Fälle aus.

Meeresforschungseinrichtungen in den USA

Neben zahlreichen Instituten an Universitäten weisen die USA auch eine Reihe von außer-universitären Meeresforschungseinrichtungen in privater und öffentlicher Trägerschaft auf. Die Websites der Woods Hole Oceanographic Institution (WHOI) sowie des Marine Biological Laboratory (MBL) umfassten laut Windows Live mehr als 15.000 Seiten. Sie wurden ebenso ausgeschlossen wie mit PMEL und AOML zwei Institute der National Oceanic and Atmospheric Administration (NOAA), deren Websites bei einer Probeerhebung mit dem SocSciBot das oben gesetzte Limit von 50.000 Seiten weit überschritten.

Endgültig in die Auswahl einbezogen wurden das Bigelow Laboratory for Ocean Sciences (BLOS), die Harbor Branch Oceanographic Institution (HBOI) sowie das Monterey Bay Aquarium Research Institute (MBARI).

Akronym	Institut	fachlicher Fokus	Träger	Website	Seiten
<i>USA</i>					
WHOI	Woods Hole Oceanographic Institution, Woods Hole, Massachusetts	Meeresforschung	privat	www.whoi.edu	57.986
MBL	Marine Biological Laboratory, Woods Hole, Massachusetts	Meeresbiologie	privat	www.mbl.edu	23.346
PMEL	Pacific Marine Environmental Lab, Seattle, Washington	Meeresforschung	staatlich	www.pmel.noaa.gov	13.295¹¹
AOML	Atlantic Oceanographic and Meteorological Lab, Miami, Florida	Meeresforschung	staatlich	www.aoml.noaa.gov	3.965¹¹
<i>Deutschland</i>					
AWI	Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven	Meeres- und Polarforschung	staatlich	www.awi-bremerhaven.de	28.197
---	Max-Planck-Institut für Marine Mikrobiologie, Bremen	Marine Mikrobiologie	staatlich	www.mpi-bremen.de	855
---	Max-Planck-Institut für Meteorologie, Hamburg	globales Klimasystem	staatlich	www.mpimet.mpg.de	1.254
Marum	Zentrum für Marine Umweltwissenschaften, Bremen	Meeresforschung	diverse	www.rcom.marum.de	2.670
<i>Indien</i>					
NCOAR	National Centre for Antarctic and Ocean Research, Goa	Indian Antarctic Programme	staatlich	ncaor.nic.in	299
NIOT	National Institute of Ocean Technology, Chennai	Meerestechnologie	staatlich	www.niot.res.in	1.297
<i>Bermuda</i>					
BUEI	Bermuda Underwater Exploration Institute	Meeresforschung	staatlich	www.buei.bm www.buei.org	239 502
<i>Bangladesh</i>					
NOAMI	National Oceanographic and Maritime Institute, Dhaka	Meeresforschung	staatlich	keine Website	---
<i>Namibia</i>					
NatMirc	National Marine Information and Research Centre, Swakopmund	Meeresforschung	staatlich	keine Website	---

Tabelle 4: ausgeschlossen **außeruniversitäre Meeresforschungsinstitute in Deutschland, den USA und englischsprachigen Entwicklungsländern (Eigenschaften, die die Auswahlbedingungen verletzen, jeweils fett hervorgehoben)**

¹¹ PMEL und AOML wurden ausgeschlossen, nachdem der Webcrawler – anders nach den Daten von Windows Live erwartbar - in einer Probeerhebung für beide Websites mehr als 50.000 Webseiten gefunden hatte.

Akronym	Institut	fachlicher Fokus	Träger	Website	Seiten
<i>USA</i>					
[BLOS]	Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor (ME)	Meeresforschung	privat	www.bigelow.org	1.300
HBOI	Harbor Branch Oceanographic Institution, Harbor Branch (FL)	Meeresforschung und -technologie, marine Ressourcen	privat	www.hboi.edu	3.054
MBARI	Monterey Bay Aquarium Research Institute, Moss Landing (CA)	Meeresforschung und -technologie	privat	www.mbari.org	12.582
<i>Deutschland</i>					
[IFM]	Leibniz-Institut für Meereswissenschaften IfM-Geomar, Kiel	Meeresforschung	staatlich	www.ifm-geomar.de	2.894
IOW	Institut für Ostseeforschung, Warnemünde	Meeresforschung	staatlich	www.io-warnemuende.de	6.122
<i>Indien</i>					
NIO	National Institute of Oceanography, Goa	Meeresforschung	staatlich	www.nio.org	7.197
<i>Pakistan</i>					
[NIOPK]	National Institute of Oceanography, Karachi	Meeresforschung	staatlich	www.niopk.gov.pk	164
<i>Mauritius</i>					
MOI	Mauritius Oceanography Institute, Quatre Bornes	Meeresforschung, marine Ressourcen	staatlich	moi.gov.mu	83
<i>Kenia</i>					
KMFRI	Kenya Marine and Fisheries Research Institute, Mombasa	Meeres- und Süßwasserforschung, marine Ressourcen	staatlich	www.kmfri.co.ke	8
<i>Südafrika</i>					
ORI	Oceanographic Research Institute, Durban	Meeresforschung, marine Ressourcen, Biodiversität	privat	www.ori.org.za	27

Tabelle 5: geeignete Meeresforschungsinstitute in Deutschland, den USA und englischsprachigen Entwicklungsländern (von der Promovendin eingesetzte Akronyme in eckigen Klammern)

Meeresforschungseinrichtungen in Deutschland

Vor allem in Norddeutschland ist die Meeresforschung an den Hochschulen verankert. Daneben gibt es eine Reihe außeruniversitärer Forschungseinrichtungen. Einige von ihnen sind Teile größerer Organisationen, wie das Deutsche Zentrum für Marine Biodiversitätsforschung des Senckenberg-Instituts oder das Institut für Küstenforschung des GKSS-Forschungszentrums. Andere wie das Marum in Bremen sind eine Gemeinschaftseinrichtung mehrerer Universitäten und Forschungseinrichtungen. Zwei Institute der Max-Planck-Gesellschaft arbeiten ebenfalls in diesem Bereich, kommen aber wegen ihrer Spezialisierung auf marine Mikrobiologie resp. Klimaforschung nicht für diese Auswahl in Frage (vgl. Tabelle 4).

Als weitere eigenständige Institute außerhalb von Universitäten existieren das Alfred-Wegener-Institut für Meeres- und Polarforschung (AWI), das Max-Planck-Institut für Marine Mikrobiologie, das Leibniz-Institut für Ostseeforschung (IOW) und das Leibniz-Institut für Meereswissenschaften IfM-Geomar (IFM).

Während der Internetauftritt des AWI die Obergrenze von 15.000 Seiten überschreitet, stellen das IOW und das IFM geeignete Fälle für die weitere Untersuchung dar. Das IOW fällt hier insofern auf, als es auf ein einzelnes Meer, nämlich die Ostsee, ausgerichtet ist. Da das disziplinäre Forschungsspektrum des IOW sehr breit ist und grundsätzlich auch Forschung in Ozeanen stattfindet, beinhaltet diese regionale Spezialisierung keine generelle Einschränkung. Der lokale Bezug und die Konzentration auf ein Weltmeer findet sich auch bei Meeresforschungseinrichtungen in anderen Ländern.

Meeresforschungseinrichtungen in Entwicklungsländern

Zum Zeitpunkt der Auswahl, im Januar 2007, war die Zahl außeruniversitärer Meeresforschungsinstitute in englischsprachigen Entwicklungsländern stark begrenzt. In zwei Fällen verfügten die vorhandenen Institute nicht über einen eigenen Internetauftritt, namentlich das National Oceanographic and Maritime Institute (NOAMI) in Bangladesh und das National Marine Information and Research Centre (NatMIRC) in Namibia. Dagegen unterhielt das Bermuda Underwater Exploration Institute (BUEI) gleich zwei offizielle Domains, deren Inhalt nicht identisch war und deren spezifische Funktionen nicht eindeutig zu klären waren. Beide wiesen außerdem übermäßige Flashanteile auf, so dass sie nicht in die Erhebung einbezogen werden konnten.

Lediglich in Indien gab es mehrere Websites wissenschaftlicher Einrichtungen, von denen jedoch nur eine die oben genannten Anforderungen erfüllte. Aus diesem Grund wurden sämtliche Meeresforschungseinrichtungen in englischsprachigen Entwicklungsländern, die sich bei der Prüfung als geeignet erwiesen, in die Untersuchung einbezogen. Es handelt sich

hierbei um das Kenya Marine and Fisheries Research Institute (KMFRI), das Maritius Oceanography Institute (MOI), das Oceanographic Research Institute (ORI) in Südafrika, sowie das National Institute of Oceanography (NIO) Indiens und eine gleichnamige Einrichtung in Pakistan (NIOPK).

Einige Meeresforschungseinrichtungen in Entwicklungsländern haben laut ihrer Website einen starken Anwendungsbezug, der auf der Bewahrung von Biodiversität (ORI) oder der Nutzung von Meeresressourcen (MOI, KMFRI, ORI) liegen kann. Dies wurde akzeptiert, da zusätzlich allgemeine Forschungsaktivitäten stattfanden und die Institute gut in die internationale *scientific community* integriert zu sein schienen (z.B. als nationaler IOC-Ansprechpartner oder durch die Teilnahme an internationalen wissenschaftlichen Konferenzen). In deutschen und US-amerikanischen Meeresforschungseinrichtungen sind diese Schwerpunkte weniger ausgeprägt, kommen jedoch auch vor. So verfügt die HBOI über eine Abteilung für Aquakultur.

4.3 Datengrundlage und -erhebung

4.3.1 Struktur der Websites

Die Grundlage für die gesamte Auswertung bildet die Struktur der Websites der Forschungseinrichtungen. Mit dem Webcrawler SocSciBot (vgl. Abschnitt 2.4.1) wurden bereits vor der endgültigen Auswahl der Fälle alle Websites mehrfach erfasst, um Umfang und Qualität der Seite zu prüfen, geeignete Startseiten zu identifizieren und mögliche Hindernisse wie z.B. Spidertraps auszumachen. Dynamische Seiten wurden grundsätzlich zugelassen. Am 2. Juli 2007 wurden die Daten gesammelt, die im Folgenden verwendet werden (vgl. Tabelle 6 für die genauen Zeiten und die Startseiten).

Institut	Startseite	Beginn	Ende
BLOS	www.bigelow.org	02.07.2007 09:48	02.07.2007 10:35
HBOI	www.hboi.edu	02.07.2007 10:36	02.07.2007 11:05
IFM	www.ifm-geomar.de	02.07.2007 11:16	02.07.2007 13:20
IOW	www.io-warnemuende.de	02.07.2007 13:20	02.07.2007 16:37
KMFRI	www.kmfri.co.ke	02.07.2007 11:06	02.07.2007 11:06
MBARI	www.mbari.org	02.07.2007 10:37	02.07.2007 19:01
MOI	moi.gov.mu	02.07.2007 11:07	02.07.2007 11:08
NIO	www.nio.org/jsp/indexNew.jsp	02.07.2007 16:44	02.07.2007 18:48
NIOPK	www.niopk.gov.pk	02.07.2007 17:05	02.07.2007 17:06
ORI	www.ori.org.za	02.07.2007 11:09	02.07.2007 11:16

Tabelle 6: Startseiten und Zeiten des Crawlens der Websites

In Tabelle 7 sind für vier Institutswebsites Bereiche, die nicht erhoben wurden, mit der jeweiligen Begründung für ihren Ausschluss angeführt.

<i>Institut</i>	<i>gebannte Adresse</i>	<i>Grund für den Ausschluss</i>
HBOI	http://www.mtg.hboi.edu	Der Server mtg.hboi.edu war permanent inaktiv.
IFM	http://www.sfb574.ifm-geomar.de	Diese Domain enthielt zum Zeitpunkt der Erhebung eine veraltete Website des SFB 574, die nicht länger gepflegt wurde. Sie wurde gebannt, da sie sehr umfangreich war und da explizit auf die aktualisierte Version unter www.sfb574.uni-kiel.de hingewiesen wurde.
IOW	http://www.io-warnemuende.de/research/helcom_zp/specieslist/index.php?	Die Website des IOW enthielt eine dynamische Seite, bei der Teile ausgeblendet werden konnten, so dass sich zahlreiche Variationen ergaben. Durch das Bannen der nebenstehenden URL wurde nur die vollständige Ansicht erfasst.
NIO	http://drs.nio.org/drs/browse-date? http://drs.nio.org/drs/browse-title? http://drs.nio.org/drs/browse-author? http://drs.nio.org/drs/items-by-author?	Die Website des NIO enthielt eine umfangreiche Literaturdatenbank, die nach mehreren Kriterien durchgegangen werden konnte. Durch das Bannen der nebenstehenden Adressen wurden alle Einträge genau einmal erfasst.

Tabelle 7: Von der Erhebung ausgeschlossene Webbereiche

Der SocSciBot hält für jede Website die Zahl der Seiten und sonstiger enthaltener Dateien fest. Er dokumentiert ferner alle Weiterleitungen (*redirects*) und Fehler beim Download. Bei letzterem kann es sich um temporäre Zugriffsschwierigkeiten handeln, wenn beispielsweise ein Server überlastet oder abgeschaltet ist. Größtenteils handelt es sich jedoch um dauerhafte Fehler aufgrund von Links, die nicht mehr aktuell sind oder Rechtschreibfehler enthalten und die deshalb nicht aufgerufen werden können. Ferner weist der Crawler Dateien gleichen Inhalts, die unter unterschiedlichen URLs abgelegt sind, als Dubletten aus. Diese Dublettenprüfung erwies sich allerdings als unvollständig, so dass in einem manuellen Verfahren weitere Dubletten identifiziert, überprüft und ausgeschlossen wurden. Dabei handelte es sich um URLs, die sich von einer anderen nur durch ein eingefügtes "amp;" oder eine Session-ID unterschieden. Eine besondere Form von Doppelung war beim IFM zu finden: Deutsch- und englischsprachige Versionen der gleichen Seite unterschieden sich bei dieser Website durch ein "&L=0" bzw. "&L=1", das an die URL angehängt wurde. Die gleiche URL ohne diesen Anhang entsprach jedoch ebenfalls der deutschen Fassung, so dass eine Version ausgeschlossen wurde. Tabelle 8 gibt eine Übersicht über die Website-Daten. Für die folgende Analyse werden lediglich die Zahl der Webseiten und der übrigen, nicht in HTML verfassten Dateien verwendet.

<i>Institut</i>	<i>Webseiten</i>	<i>Dateien (nicht html)</i>	<i>Weiterleitungen</i>	<i>Download-Fehler</i>	<i>Dubletten</i>
BLOS	972	1.378	27	27	44
HBOI	905	786	5	27	2
IFM	3.626	1.396	23	39	1.046
IOW	3.960	3.248	46	212	779
KMFRI	27	1	0	1	0
MBARI	5.310	13.372	3.924	1.064	136
MOI	27	13	0	1	1
NIO	2.915	2.150	4	217	181
NIOPK	27	11	0	5	1
ORI	11	67	0	1	1

Tabelle 8: Umfang der erhobenen Websites

4.3.2 Outlinkdaten

Neben der Liste der Webseiten und der Dateien umfasst der Output des Webcrawlers für jede Website eine Datei, die für jede HTML-Seite die enthaltenen Links protokolliert. Aus dieser Datei wurden mit Visual-Basic-Makros die URLs der externen Outlinks extrahiert. Der Internetauftritt des KMFRI enthielt eine Linkliste in Flash, deren Inhalt manuell ergänzt wurde. Für das IFM wurden 393 Links auf die Domains der Vorgängereinrichtungen – "ifm.uni-kiel.de" und "geomar.de" – als interne Links bewertet und von der folgenden Analyse ausgeschlossen. Gleiches gilt für zehn Links des NIO, die auf die IP-Adresse der eigenen Website verwiesen und somit als intern anzusehen sind. Des Weiteren wurden 13 unvollständige URLs (z.B.: <http://edv-59/cis>) aussortiert.

Da auch Outlinks mit unterschiedlicher URL auf die gleiche Zielseite verweisen können, wurden Adressen mit typischen Variationen zusammengefasst. Die zugrundegelegten Kriterien sind in Abbildung 5 aufgeführt.

Domains wurden als gleich betrachtet, wenn sie sich ausschließlich in folgender Hinsicht unterschieden:

- ein vorangestelltes "www." (z.B. "www.google.com" = "google.com")
- ein angehängtes "/" (z.B. "www.noaa.gov/" = "www.noaa.gov")
- eine auf die Domain folgende Port-Nummer (z.B. "www.uni-kiel.de:8080" = "www.uni-kiel.de")
- eine an den Servernamen angehängte Nummer (z.B. "www1.unep.org" = "www.unep.org")
- eine geringfügige Variation des Domainnamens bei gleichem Inhalt (z.B. "www.bfa-fish.de" = "www.bfa-fisch.de"; "www.oceanexpert.net" = "www.oceanexperts.net"; "www.metoffice.gov.uk" und "www.met-office.gov.uk")

Links wurden zusammengefasst, wenn sich ihre URLs in den bereits genannten und/oder in folgenden Kriterien unterschieden:

- ein an beliebiger Stelle eingefügtes "amp;"
- eine Erweiterung um eine Startseiten-Suffix wie "index.html", "home.html", "default.html" oder entsprechende Variationen wie "index.asp", "index.htm", "index.jsp", "index.shtml", "index.php" etc.

Seiten in verschiedenen Sprachen wurden grundsätzlich nicht als identisch betrachtet. In allen Zweifelsfällen wurde die Inhalte der URLs bzw. Domains manuell überprüft.

Abbildung 5: Kriterien, nach denen Outlink-URLs und -domains zu einem Ziel zusammengefasst wurden

Darüber hinaus wurde für jede Domain die Grunddomain bestimmt, anhand derer sich die Links nach Sites aggregieren lassen. In der Regel war die Second-Level-Domain die Grundlage für die Gruppierung. So liegen die Domains "www.elnino.noaa.gov" und "www.pmel.noaa.gov" auf der Site "noaa.gov". Bei Domains, die eine sektorspezifische SLD enthielten, wurde entsprechend die Third-Level-Domain herangezogen. In diesem Fall teilen beispielsweise Outlinks auf "www.botany.uwc.ac.za" und "www.science.uwc.ac.za" die Ziel-site "uwc.ac.za".

Sämtliche Daten wurden zur weiteren Auswertung in eine MySQL-Datenbank überführt. Für jede Institutswebsite wurden die distinkten Outlinkziele, -domains und -sites ermittelt (vgl. Tabelle 9). Alle Outlinkziele wurden außerdem aufgerufen und abgespeichert, um im Rahmen der ersten Leitfrage ihre nationale Anbindung zu untersuchen.

<i>Institut</i>	<i>Anzahl der Outlinks</i>	<i>distinkte Outlink-URLs</i>	<i>distinkte Outlinkziele</i>	<i>distinkte Outlink-Domains</i>	<i>distinkte Outlink-Sites</i>
BLOS	1.105	686	677	422	280
HBOI	507	342	339	210	189
IFM	7.190	2.039	2.007	897	720
IOW	2.386	827	813	476	422
KMFRI	17	15	15	15	15
MBARI	4.359	2.053	2.006	1.018	690
MOI	23	23	23	20	19
NIO	3.252	88	87	66	61
NIOPK	2	1	1	1	1
ORI	38	38	38	38	37
gesamt	18.879	6.112	6.006	3.163	2.434
distinkt	---	6.009	5.852	2.742	1.983

Tabelle 9: Outlinkwerte nach Instituten

4.3.3 Inlinkdaten

Die Erhebung der Inlinkdaten erforderte verschiedene Entscheidungen und beinhaltete aufwendige Datenbereinigungen, die in den folgenden Abschnitten ausgeführt werden.

Die Inlinkdaten für die Websites aller Forschungseinrichtungen wurden mit Yahoo erhoben. Diese Suchmaschine verfügte im Sommer 2007 über die umfassendsten Möglichkeiten der Linkrecherche, nachdem Windows Live seine weitreichende Inlinksuche zum 30. März 2007 eingestellt hatte. Die französische Suchmaschine Exalead stellt für weltweite Fragestellungen keine geeignete Alternative dar, da sie einen europäischen Bias aufweist (Aguillo et al. 2008). Anders als Google erlaubte es Yahoo nicht nur, Links auf einzelne Seiten, sondern auf ganze Domainbereiche zu suchen (ersteres mit dem Operator "link:", letzteres mit "linkdomain:"). Ferner ließen sich die Abfragen bei Yahoo auf einzelne Inlinkquellen oder auf Top-Level-Domains beschränken (mit "site:").

Wie andere Anbieter beschränkt Yahoo die Zahl der angezeigten Treffer (s.u.). Um das Limit von 1.000 Inlinks überschreiten und die weltweite Verteilung der URLs analysieren zu können, wurde gezielt nach Inlinks aus einzelnen Top-Level-Domains gesucht. Zum Zeitpunkt der Erhebung waren laut Internet Assigned Numbers Authority zwanzig generische und 250 landesspezifische Top-Level-Domains zugelassen.¹² Es wurden Inlinks aus jeder dieser TLDs auf jede Institutswebsite abgefragt, nach dem Muster "linkdomain:kmfri.co.ke

¹² Vgl. <http://iana.org/domains/root/db/>, Information abgerufen am 24.09.2007.

site:de". Eine Übersicht über sämtliche Abfragemuster, die an die einzelnen Institutswebsites angepasst wurden, findet sich im Anhang (vgl. Abschnitt 8.1). Die Erhebung aller 2.700 Abfragen erfolgte am 25. September 2007 mit der Software LexiURL Searcher¹³, die sie automatisiert an Yahoos Application Programming Interface (API) schickte.

Fünf der von IANA geführten TLDs (".ad", ".ax", ".cat", ".do", ".name") wurden von Yahoo offensichtlich nicht als Teil der Domain, sondern als normales Schlagwort behandelt. Entsprechend führten die Abfragen nicht zu korrekten Ergebnissen und wurden nachträglich ausgeschlossen. Nur ein kleiner Teil der übrigen Abfragen führte zu Treffern (15,5%, vgl. Tabelle 10). Für die einzelnen Websites liegt dieser Anteil zwischen 4,9 Prozent (KMFRI) und 26,8 Prozent (IFM).

<i>Institut</i>	<i>1</i>	<i>2-10</i>	<i>11-100</i>	<i>101-1.000</i>	<i>1.001-10.000</i>	<i>>10.000</i>	<i>Σ Abfragen mit Treffern</i>	<i>Anteil an allen Abfragen</i>
BLOS	13	19	13	7	1	0	53	20,0%
HBOI	7	28	16	4	1	0	56	21,1%
IFM	7	27	19	14	3	1	71	26,8%
IOW	11	23	17	2	1	0	54	20,4%
KMFRI	4	8	1	0	0	0	13	4,9%
MBARI	8	20	24	9	3	0	64	24,2%
MOI	9	5	1	0	0	0	15	5,7%
NIO	11	18	16	4	0	0	49	18,5%
NIOPK	12	4	3	0	0	0	19	7,2%
ORI	8	7	3	0	0	0	18	6,8%
gesamt	90	159	113	40	9	1	412	15,5%

Tabelle 10: Inlinkabfragen bei Yahoo nach Zahl der Treffer

Bei zehn Abfragen überschritt die Trefferzahl Yahoos Outputlimit von maximal tausend angezeigten URLs (vgl. Tabelle 11). Bei ihnen ist unklar, inwiefern die von Yahoo ausgegebene Anzahl von Webseiten mit Inlinks auf einen vorgegebenen Webbereich tatsächlich die Anzahl der in der Datenbank der Suchmaschine enthaltenen passenden URLs wiedergibt. Die Tatsache, dass alle Werte durch zehn teilbar sind, deutet darauf hin, dass die Suchmaschine die Ergebnisse zumindest rundet.

¹³ Die LexiURL-Programme wurde von Mike Thelwall erstellt und sind kostenlos abrufbar unter <http://lexiurl.wlv.ac.uk/> (zuletzt abgerufen am 19.11.2009).

Institut	Treffer	Abfrage
IFM	10.200	linkdomain:ifm-geomar.de site:org
IFM	7.190	linkdomain:ifm-geomar.de -site:ifm-geomar.de site:de
IFM	3.900	linkdomain:ifm-geomar.de site:com
MBARI	2.890	linkdomain:mbari.org site:com
IOW	2.590	linkdomain:io-warnemuende.de -site:io-warnemuende.de site:de
MBARI	1.650	linkdomain:mbari.org -site:mbari.org site:org
MBARI	1.620	linkdomain:mbari.org site:edu
BLOS	1.580	linkdomain:bigelow.org -site:bigelow.org site:org
HBOI	1.120	linkdomain:hboi.edu site:org
IFM	1.080	linkdomain:ifm-geomar.de site:net

Tabelle 11: Inlinkabfragen bei Yahoo mit mehr als 1.000 Treffern

Ein Vergleich der ausgewiesenen Treffer mit den tatsächlich gezeigten und um Dubletten bereinigten URLs zeigt, dass die Trefferliste häufig nur einen Bruchteil der von Yahoo genannten Anzahl enthält. Wie aus Abbildung 6 zu erkennen, ist der durchschnittliche Anteil der gezeigten URLs an der ausgewiesenen Trefferzahl umso geringer, je höher die genannte Trefferzahl ist.

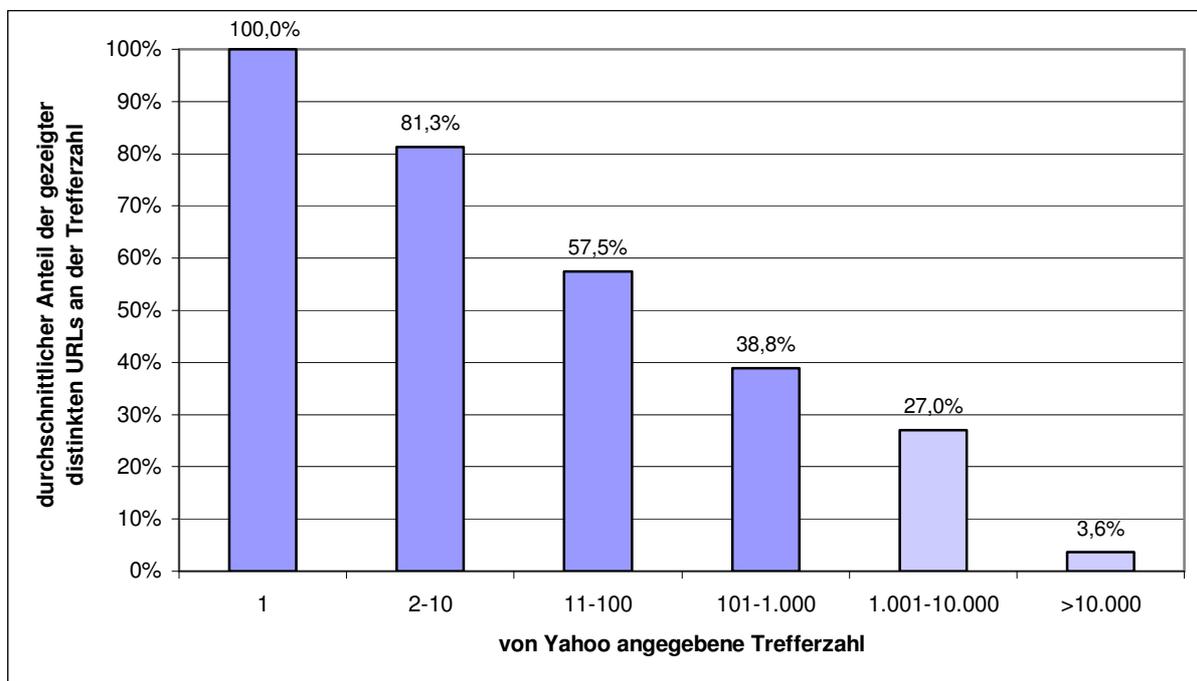


Abbildung 6: Abfragen nach dem durchschnittlichen Anteil der gezeigten distinkten URLs an der von Yahoo angegebenen Gesamttrefferzahl (gruppiert nach Gesamttrefferzahl, N = 412)

Bei Abfragen mit einem Treffer wurde in allen Fällen genau eine URL angezeigt. Bei zwei bis zehn Treffern wurden durchschnittlich 81,3 Prozent der angekündigten Treffer ausgegeben. Wenn Yahoo zwischen elf und hundert Treffer ankündigte, deckten die gezeigten URLs gut die Hälfte (57,5%) der Trefferzahl ab, bei 101 bis tausend Treffern war es ein starkes Drittel (38,8%). In einem einzigen Fall wurden zwei URLs mehr ausgewiesen als durch die Trefferzahl angekündigt. Bei den neun Abfragen, für die Yahoo zwischen 1.001 und 10.000 Treffer meldete, wurden im Durchschnitt 27,0 Prozent der ausgegebenen Werte als URLs angezeigt; bei der einzigen 10.000er Abfrage machten die aufgelisteten URLs nur 3,6 Prozent der angegebenen Trefferzahl aus.

Die von Yahoo angegebenen Werte stimmten bei Abfragen mit niedrigen Trefferzahlen relativ gut mit der Zahl der angezeigten URLs überein. Bei höheren Trefferzahlen lag der Anteil der tatsächlich ausgegebenen Treffer deutlich niedriger. Dies gilt bereits unterhalb der Schwelle von tausend Treffern. Es ist anzunehmen, dass dieser Effekt auch bei höheren Trefferzahlen auftritt; dort lässt er sich aufgrund von Yahoos Outputlimit jedoch nicht quantifizieren. Ein Vergleich der Suchmaschinen Yahoo, Google und Windows Live (Thelwall 2008d), dessen Datengrundlage aus dem Sommer 2007 stammt, zeigte, dass eine Überschätzung der tatsächlichen Trefferzahl zu dieser Zeit üblich war und dass Yahoo stärker als die anderen beiden Suchmaschinen dazu tendierte. Dagegen enthielten die von Yahoo ausgegeben URLs die kleinste Zahl an falschen Treffern und bei weitem die größte Bandbreite verschiedener Domains, Sites und Top-Level-Domains, weshalb Thelwall (2008d) sie für webometrische Zwecke empfiehlt.

<i>Institut</i>	<i>Trefferzahl (Yahoo)</i>	<i>gezeigte URLs</i>	<i>distinkte URLs</i>	<i>distinkte Inlinkquellen</i>	<i>distinkte Domains</i>	<i>distinkte Sites</i>
BLOS	4.171	2.484	1.682	1.674	1.326	1.166
HBOI	3.387	2.354	1.847	1.830	1.432	1.271
IFM	25.863	5.179	3.805	3.746	2.763	2.622
IOW	4.086	1.520	1.381	1.362	1.050	971
KMFRI	64	47	47	45	30	29
MBARI	10.472	5.167	3.686	3.652	2.858	2.439
MOI	62	46	46	43	33	31
NIO	2.654	1.469	889	880	701	662
NIOPK	110	88	88	88	76	75
ORI	128	90	90	90	79	76
gesamt	50.997	18.444	13.561	13.410	10.348	9.342
distinkt	---	---	13.025	12.876	8.285	7.101

Tabelle 12: Inlinkwerte nach Instituten

Tabelle 12 zeigt für jedes Institut die Zahl der von Yahoo angekündigten Treffer sowie die Zahl der tatsächlich gezeigten URLs, aus denen durch das Entfernen von Dubletten eine Liste distinkter URLs erstellt wurde. Ähnlich wie bei den Outlinks (vgl. Abbildung 5) wurden diejenigen Adressen zu Inlinkquellen zusammengefasst, die nicht identisch waren, aber auf die gleiche Seite bzw. Domain verwiesen.

Es wird deutlich, dass für alle Institute die Zahl der distinkten URLs die Zahl der Yahoo-Treffer deutlich unterschreitet. Wie oben gezeigt (vgl. Abbildung 6), hängen diese Ausfälle mit der Trefferzahl pro Abfrage zusammen. Um die Effekte der rohen Anzahl von Treffern auf die übrigen Inlinkwerte zu untersuchen, wurden die Institute in drei Gruppen zusammengefasst: Die Websites von MOI, KMFRI, NIOPK und ORI erzielten insgesamt weniger als 150 Treffer bei Yahoo. NIO, IOW, HBOI und BLOS rangierten mit 2.500 bis 4.500 Treffern im Mittelfeld. MBARI und IFM erreichten jeweils über 10.000 Treffer. Für jede Gruppe wurden die durchschnittlichen Anteile ermittelt, die die gezeigten URLs, die distinkten URLs und die Inlinkquellen an allen Yahoo-Treffern haben. Es zeigt sich ein deutliches Gefälle (vgl. Tabelle 13):

<i>Institute</i>	<i>Yahoo-Treffer</i>	<i>Trefferzahl (Yahoo)</i>	<i>gezeigte URLs</i>	<i>distinkte URLs</i>	<i>distinkte Inlinkquellen</i>
MBARI, IFM	>10.000	100,0%	34,7%	25,0%	24,7%
NIO, IOW, HBOI, BLOS	2.500-4.500	100,0%	55,4%	40,5%	40,2%
MOI, KMFRI, NIOPK, ORI	<150	100,0%	74,5%	74,5%	72,5%

Tabelle 13: Durchschnittlicher Anteil der Inlinkwerte an allen Yahoo-Treffern, gruppiert nach der Trefferzahl der Institutswebsites

Die Zahl der distinkten Inlinkquellen machten bei MBARI und IFM durchschnittlich nur ein knappes Viertel (24,7%) der Gesamttrefferzahl aus. Auch bei den Instituten mit mittlerer und kleinerer Trefferzahl erreichte die Zahl der distinkten Inlinkquellen im Durchschnitt nur 40,2 bzw. 72,5 Prozent der von Yahoo angekündigten Trefferzahl.

Um genauer zu analysieren, wie stark jeder einzelne Bereinigungsschritt die Zahl der Inlinks reduziert, wurde jeder Inlinkwert durch den der vorangehenden Kategorie geteilt. Es wurde wiederum der Durchschnitt dieser Anteilswerte für jede Websitegruppe ermittelt (vgl. Tabelle 14). Die Zahl der gezeigten URLs an allen Yahoo-Treffern macht bei den Instituten mit hoher Trefferzahl im Durchschnitt nur etwa ein Drittel aus (34,7%), was insofern nicht verwundert, als das Outputlimit bei drei bzw. vier Abfragen für MBARI und IFM überschritten wurde. Bei den Instituten mit mittlerer Trefferzahl, von denen drei das Outputlimit jeweils einmal überschritten, wurden nur 55,4 Prozent der angekündigten Yahoo-Treffer tatsächlich angezeigt. Selbst bei den Instituten mit geringer Trefferzahl, die das Outputlimit nicht

tangierten, machten die ausgegebenen URLs nur drei Viertel der Trefferzahl (74,5%) aus. Es zeigt sich somit ein Effekt des Outputlimits, jedoch liegt die von Yahoo angegebene Trefferzahl auch bei wenig und gar nicht betroffenen Websites deutlich zu hoch. Ferner ist festzustellen, dass Websites mit hohen Yahoo-Werten auch ein hohes Maß an Dubletten aufwiesen. Bei den Instituten mit hoher und mittlerer Trefferzahl blieben nach der Elimination mehrfach ausgegebener URLs nur jeweils knapp drei Viertel (72,4% bzw. 74,4%) aller ausgegebenen Treffer übrig. Dagegen enthielten die Trefferlisten von KMFRI, MOI, NIOPK und ORI keinerlei Dubletten: Hier gab Yahoo nur distinkte Inlink-URLs aus.

<i>Institute</i>	<i>Yahoo-Treffer</i>	<i>Anteil von</i>		
		<i>gezeigten URLs</i>	<i>distinkten URLs</i>	<i>distinkten Inlinkquellen</i>
		<i>an</i>		
		<i>Trefferzahl (Yahoo)</i>	<i>gezeigten URLs</i>	<i>distinkten URLs</i>
MBARI, IFM	>10.000	34,7%	72,4%	98,8%
NIO, IOW, HBOI, BLOS	2.500-4.500	55,4%	74,4%	99,1%
MOI, KMFRI, NIOPK, ORI	<150	74,5%	100,0%	97,3%

Tabelle 14: Durchschnittlicher Anteil der Inlinkwerte an der jeweils vorausgehenden Bereinigungsstufe, gruppiert nach der Gesamtrefferzahl der Institutswebsites

Dagegen zeigt die Bündelung distinkter Inlink-URLs zu inhaltlich identischen Inlinkquellen insgesamt nur einen geringen Effekt: Die Beseitigung von Adressvariationen und Spiegelseiten (für die zugrundegelegten Kriterien vgl. Abbildung 5) reduzierte keine der drei Kategorien um mehr als drei Prozent (vgl. Tabelle 14).

Zusammenfassend bedeutet dies, dass die von Yahoo genannte Trefferzahl keine geeignete Grundlage für die statistische Auswertung ist, da Yahoo die Zahl der Inlinkquellen generell überschätzte und für Websites mit mittleren und größeren Trefferzahlen außerdem zahlreiche Dubletten ausgab. Dagegen hatte die Zusammenfassung von Inlinks und Outlinks mit verschiedenen URLs, aber identischer Quelle bzw. Ziel nur einen geringen Effekt.

Für die weitere Untersuchung werden folglich die bereinigten Daten verwendet. Hierdurch wird eine starke Überschätzung der Linkzahl, insbesondere für Abfragen mit mehr als zehn Treffern, vermieden. Allerdings ist aufgrund des Outputlimits davon auszugehen, dass stattdessen die Inlinkwerte für die zehn Abfragen mit den höchsten Trefferzahlen leicht unterschätzt werden. Dies dürfte sich vorwiegend auf das IFM auswirken, zu dem die drei Abfragen mit den höchsten Werten gehören. Bei der Interpretation der Ergebnisse gilt es, dies zu berücksichtigen.

Zu beachten ist außerdem, dass Yahoo maximal zwei Links aus der gleichen Domain ausweist, so dass die tatsächliche Gesamtzahl aller Inlinkseiten auch bei kleineren Abfragen unbekannt bleibt. Als geeignete Ebene für die statistische Untersuchung der Inlinkmuster wurden deshalb Domains und nicht einzelne Seiten mit Inlinks gewählt (vgl. Abschnitt 5.2. und 5.3). Mehrfache Inlinks aus derselben Domain wurden somit nur einfach gezählt. Aus Gründen der Konsistenz wurde für die Outlinks die gleiche Ebene gewählt. Eine Ausnahme ist der inhaltsanalytische Teil der Untersuchung (vgl. Abschnitt 5.1), der sich konsequenterweise auf den Inhalt konkreter Inlinkquellen und Outlinkziele bezieht.

Nach der entsprechenden Bereinigung der Inlinkdaten mit Visual Basic wurden sie zur weiteren Verarbeitung in eine mySQL-Datenbank überführt.

4.3.4 Publikationsdaten

Die zweite Leitfrage (vgl. Abbildung 4) beinhaltet für jede Forschungseinrichtung einen Vergleich des webometrischen und des bibliometrischen Profils. Dazu wurden am 1. Oktober 2008 alle Einträge der zehn Institute für die Jahre 2005 bis 2007 aus dem Web of Science (WoS) heruntergeladen.

<i>Institut</i>	<i>Standardadresse(n) im WoS</i>	<i>Abfrage im WoS</i>	<i>Einträge 2005-07</i>
BLOS	Bigelow Lab Ocean Sci, W Boothbay Harbor, ME 04575 USA	bigelow lab	62
HBOI	Harbor Branch Oceanog Inst Inc, Ft Pierce, FL 34946 USA	harbor branch same oceanog same inst	108
IFM	Leibniz Inst Marine Sci / Leibniz Inst Meeresforsch / IFM-GEOMAR, D-24105 Kiel, Germany	kiel same (geomar or ifm or ctr mar* geosci or inst mar* or inst meer* or mar* res* inst*)	695
IOW	Inst Ostseeforsch Warnemuende / Balt Sea Res Inst Warnemunde / IOW D-18119 Rostock, Germany	((rost* or warnem*) same (iow or bal* or ostseeforsch* or ostseeforsch* or istseeforsch*)) or (warnem* same leibniz)	234
KMFRI	Kenya Marine & Fisheries Res Inst / KMFRI, Mombasa, Kenya (mehrere Standorte)	kenya* same (kmfri or (marin* same fish*))	43
MBARI	Monterey Bay Aquarium Res Inst / MBARI, Moss Landing, CA 95039 USA	mbari or monterey bay aquarium res inst or moss landing same bay aqua*	224
MOI	Mauritius Oceanog Inst, Quatre Bornes, Mauritius	mauritius SAME ocean*	8
NIO	Natl Inst Oceanog, Panaji 403004, Goa, India (mehrere Standorte)	india same (natl inst oceanog or nio)	398
NIOPK	Natl Inst Oceanog, Karachi, Pakistan	pakistan SAME ocean*	5
ORI	Oceanog Res Inst, ZA-4056 Durban / Marine Parade, South Africa	south africa same ocean* res* inst*	18

Tabelle 15: Abfrage der Publikationen aus dem Web of Science 2005-2007

Die verwendeten Abfragen sowie die jeweilige Anzahl der Publikationen finden sich in Tabelle 15. Die heruntergeladenen Datensätze wurden anhand des Unique Article Identifier (UT) auf Dubletten geprüft; sie enthielten ausschließlich distinkte Einträge.

Für einige Fragestellungen werden auch Zitationsdaten verwendet. Da die erhobenen Publikationen sich auf drei Jahrgänge verteilen, die noch nicht lange zurückliegen, ist die Zitation eines älteren Beitrages deutlich wahrscheinlicher, als die eines neueren. Die Zitationsdaten wurden daher mit Hinblick auf den Erscheinungszeitpunkt standardisiert. Es wurde angenommen, dass die Publikationen im Mittel am 1. Juli ihres Jahrgangs erschienen. Da die Daten am 1. Oktober 2008 aus dem Web of Science abgerufen wurden, ergab sich somit für Publikationen aus 2005 eine durchschnittliche Dauer der Existenz von 3,25 Jahren (bzw. für 2006 2,25 Jahre und für 2007 1,25 Jahre). Für jeden Jahrgang wurde die Anzahl der Zitationen durch diesen Wert geteilt, bevor die Daten aller Jahrgänge zu einem standardisierten Gesamtwert aufsummiert wurden (Tabelle 16). Wie für alle übrigen Maße wird hier sowohl diese standardisierte Summe, als auch die standardisierte Zahl der Zitationen pro Publikation resp. pro MitarbeiterIn verwendet.

<i>Institute</i>	<i>Summe der Zitationen</i>			<i>Summe der rohen Zitationen</i>	<i>Summe der standardisierten Zitationen</i>
	<i>2005</i>	<i>2006</i>	<i>2007</i>		
BLOS	323	78	41	442	166,9
HBOI	224	119	43	386	156,2
IFM	1888	1068	422	3378	1.393,2
IOW	404	287	115	806	343,9
KMFRI	107	23	6	136	47,9
MBARI	445	546	95	1086	455,6
MOI	21	9	0	30	10,5
NIO	326	256	52	634	255,7
NIOPK	0	2	1	3	1,7
ORI	8	15	2	25	10,7

Tabelle 16: Rohe und standardisierte Zitationen im Web of Science bis zum 1.10. 2008

Zusätzlich wurde von jeder Forschungseinrichtung ein Publikationsverzeichnis für das Jahr 2006 erbeten bzw. aus dem Internet heruntergeladen, um dessen Abdeckung durch das Web of Science prüfen zu können.

4.3.5 Weitere Daten

Drei zusätzliche Datenquellen wurden herangezogen, um Informationen zur Institutsgröße sowie zum Entwicklungsstand und zur geographischen Einbettung beteiligter Länder zu gewinnen.

Als Maß für die Größe einer Einrichtung wurde die Zahl der wissenschaftlichen MitarbeiterInnen Ende des Jahres 2007 der Website entnommen, sofern dort hinreichend detaillierte und aktuelle Angaben zu finden waren. Andernfalls wurde sie direkt beim Institut erfragt. Eine Übersicht bietet Tabelle 17.

<i>Institut</i>	<i>wiss. Personal</i>	<i>Datenquelle</i>
BLOS	37	Website
HBOI	69	erfragt
IFM	238	erfragt
IOW	89	erfragt
KMFRI	90	erfragt
MBARI	99	Website
MOI	12	Website
NIO	177	Website
NIOPK	29	Website
ORI	15	Website

Tabelle 17: Anzahl der wissenschaftlichen MitarbeiterInnen an den ausgewählten Meeresforschungseinrichtungen

Um den Entwicklungsstand der Länder, in denen die Meeresforschungsinstitute angesiedelt sind, zu bekommen, wurde der Human Development Index (HDI) der Vereinten Nationen herangezogen. In ihm sind grundlegende Indikatoren zum Pro-Kopf-Einkommen mit Kaufkraftparität (GDP PPP), zur Lebenserwartung und zum Bildungsstand in der Bevölkerung gebündelt, so dass er ein umfassenderes Bild von der sozio-ökonomischen Situation liefert als das Bruttoinlandsprodukt pro Kopf. Der verwendete HDI für die Jahre 2007/2008 basiert auf Daten des Jahres 2005 (vgl. UNDP 2007: 225ff). Tabelle 18 enthält die HDI-Werte der Sitzländer der ausgewählten Meeresforschungseinrichtungen.

<i>Land</i>	<i>HDI</i>
USA	0,951
Deutschland	0,935
Mauritius	0,804
Südafrika	0,674
Indien	0,619
Pakistan	0,551
Kenia	0,521

Tabelle 18: Human Development Index 2007/08 für die Sitzländer der ausgewählten Meeresforschungseinrichtungen

Die geographische Klassifikation der Sitz- und Kooperationsländer wurde von den Vereinten Nationen übernommen (vgl. United Nations Statistics Division 2006). Sie fassen die einzelnen Staaten nach Kontinent sowie – unterhalb der kontinentalen Ebene – nach Weltregionen wie z.B. Südeuropa oder Ostafrika zusammen. In sozio-ökonomischer Hinsicht unterscheiden sie Entwicklungsländer, Industrieländer und Übergangsländer (englisch: *countries in transition*). Zu letzteren zählen neben Albanien die Nachfolgestaaten Jugoslawiens und die Gemeinschaft Unabhängiger Staaten (GUS). Schwellenländer werden nicht als eigene Gruppe geführt. Das Verzeichnis „Abkürzungen und Lage der untersuchten Meeresforschungseinrichtungen“ (S. XI) zeigt die Sitzländer der zehn Institute sowie deren geographische und sozio-ökonomische Einordnung nach der UN-Klassifikation.

4.3.6 Exkurs: Daten zur Rezeption von Websites

Wie Zitationen nichts über die allgemeine Rezeption von Publikationen aussagen, sind Links nicht identisch mit der tatsächlichen Nutzung von Webseiten. Links einzurichten ist zwar nicht schwierig, jedoch doch ungleich aufwendiger als eine Seite zu besuchen – es setzt die Existenz einer eigenen Website voraus. Daher erscheint es sinnvoll, auch die Besuche von Websites zu analysieren. Dies gilt besonders vor dem Hintergrund globaler Disparitäten: In Regionen mit begrenztem Internetzugang ist es vermutlich eher möglich, E-Mails zu schreiben und Internetauftritte anzusehen, als selbst eine Website zu gestalten.

Grundsätzlich ist die Analyse der Rezeption einzelner Websites möglich: Alle Zugriffe auf eine Website werden von dem zugehörigen Server protokolliert. Das dazu angelegte Logfile dokumentiert u.a. Zeitpunkt und Dauer des Besuchs, aufgerufene Dateien und IP-Adressen der NutzerInnen. Eine Analyse der Logfiles erlaubt Rückschlüsse auf die Häufigkeit und – über die Server-Adressen – auch über die Verteilung auf Top-Level-Domains. Viele Webmaster lassen mit entsprechender Software aus den Logfiles Nutzungsstatistiken

erstellen. So sehen sie, wie häufig ihre Website aufgerufen wird, welche Seiten besonders beliebt sind etc. Im Unterschied zu den bisher behandelten Webdaten sind weder Logfiles noch Serverstatistiken öffentlich verfügbar.

Da die Analyse von Logfiles sehr aufwendig ist, wurde im Rahmen dieser Dissertation geprüft, ob die Auswertung von Nutzungsstatistiken eine Möglichkeit sein könnte, unaufwendig Informationen über die internationale Verteilung der Website-RezipientInnen zu gewinnen. Bei allen zehn Instituten wurde angefragt, ob sie Zugriffsdaten sammeln und bereit seien, sie für Forschungszwecke zur Verfügung zu stellen. Nur zwei Einrichtungen (BLOS und MBARI) stellten die Daten für den gewünschten Zeitraum von Oktober 2006 bis September 2007 bereit. NIOPK und HBOI reagierten auf keine der drei Anfragen. Drei Institute gaben an, keine Webdaten zu sammeln (MOI, KMFRI und IOW). Die übrigen drei (IFM, NIO und ORI) stellten Nutzungsdaten für kürzere Zeiträume zur Verfügung (vgl. Tabelle 19). Begründet wurde dies damit, dass die Daten nicht archiviert werden, dass die Datensammlung gerade erst aufgenommen worden sei bzw. dass man gerade den Server gewechselt und die Daten nicht übertragen habe. Für das ORI kommt hinzu, dass die Nutzungsdaten auch die Website des zugehörigen Aquariums umfassen. Eine weitere Beschränkung besteht darin, dass die in den Einrichtungen verwendeten Programme zur Dokumentation der Website-Nutzung in einem Fall *requests*, in den übrigen vier *hits* gezählt haben. Diese Werte sind nicht vergleichbar, da bei ersteren das Aufrufen ganzer Seiten gezählt wird, bei letzteren das Aufrufen von Seitenteilen wie Textkörper, Rahmen, eingelagerten Bilddateien etc.

<i>Institut</i>	<i>Datenstatus</i>	<i>Zeitraum</i>	<i>Zähleinheit</i>
BLOS	Daten komplett	10/2006-09/2007	requests
HBOI	keine Antwort	---	---
IFM	Daten teilweise gesammelt	08-11/2007	hits
IOW	Daten nicht gesammelt	---	---
KMFRI	Daten nicht gesammelt	---	---
MBARI	Daten komplett	10/2006-09/2007	hits
MOI	Daten nicht gesammelt	---	---
NIO	Daten teilweise gesammelt	07/2007 (teilweise)	hits
NIOPK	keine Antwort	---	---
ORI	Daten teilweise gesammelt	11/2007	hits

Tabelle 19: Auskünfte der Institute zur Website-Nutzung

Insgesamt waren die vorliegenden Daten so heterogen im Hinblick auf Erhebungszeitpunkt und Zählheit (*requests* vs. *hits*), dass sie keine aussagekräftigen Schlussfolgerungen zuließen. Von einer Untersuchung der Website-Nutzung wurde daher Abstand genommen.

Die Beschaffung aggregierter Daten zur Rezeption der Website hat sich somit als schwierig erwiesen. Neben Kooperationsbereitschaft setzt sie voraus, dass Forschungseinrichtungen die Nutzung ihrer Website kontinuierlich dokumentieren, was offensichtlich weder in Industrie- noch in Entwicklungsländern Standard ist. Hinzu kommt die Variation in der verwendeten Software. Mögliche Alternativen werden im Zwischenfazit (vgl. Abschnitt 4.4.3) kurz diskutiert.

4.4 Zwischenfazit

Schon bevor in den nächsten beiden Kapiteln die Daten analysiert und die Leitfragen beantwortet werden, erlauben die oben ausgeführten Erfahrungen bei der Fallauswahl und der Datenerhebung erste Schlussfolgerungen zu den Möglichkeiten und Grenzen der Webometrie im Hinblick auf internationale Fragestellungen.

4.4.1 Schlussfolgerungen aus der Fallauswahl

Obwohl mit der Meeresforschung ein hoch internationales Fachgebiet gewählt wurde, verfügten 2007 zwei einschlägige Forschungseinrichtungen in englischsprachigen Entwicklungsländern nicht über eine eigene Website (vgl. Abschnitt 4.2.3). Offensichtlich ist die vergleichsweise niederschwellige Möglichkeit, das Profil der eigenen Organisation im World Wide Web einer breiten Öffentlichkeit zu präsentieren, für Forschungseinrichtungen in einigen Ländern noch zu aufwendig.

Beide Institute dienen jeweils als nationale Kontaktstelle für ozeanographische UN-Organisationen und unterhalten auch internationale Kooperationen mit anderen Einrichtungen im Sample. Es handelt sich somit nicht um Institute mit ausgewiesener lokaler Ausrichtung. Somit bleibt festzuhalten, dass die Reichweite der Webometrie zum Erhebungszeitpunkt noch sehr begrenzt war, wenn es darum geht, speziell die Forschungsaktivitäten von Einrichtungen in Entwicklungsländern nachzuzeichnen.

4.4.2 Schlussfolgerungen aus der Linkdatenerhebung

Zu den Vorzügen webometrischer Untersuchungen wird häufig gezählt, dass sich schnell und einfach Daten beschaffen lassen, insbesondere wenn Suchmaschinen verwendet werden. Die Erhebung hat gezeigt, dass das bloße Sammeln der Daten tatsächlich schnell geht, dass die Vorbereitung der Erhebung und die Aufbereitung der Daten jedoch große Sorgfalt und viel Zeit erfordert:

Die *Datenerhebung per Webcrawler* braucht eine akkurate Vorbereitung und eine genaue Prüfung der Fälle, damit die endgültige Erhebung schnell und reibungslos läuft. Eine Datenbereinigung ist notwendig und zeitaufwendig (vgl. Abschnitt 4.3.1 und 4.3.2).

Suchmaschinen sind sehr schnell zu erheben und brauchen wenig Vorbereitung (vgl. Abschnitt 4.3.3). Nicht zu umgehen ist jedoch der Unsicherheitsfaktor Suchmaschine. In dieser Erhebung ist deutlich geworden, wie Yahoo die Zahl der Treffer überschätzt. Je mehr Treffer angekündigt werden desto größer im Durchschnitt die Überschätzung. Dies ist konsistent mit den Ergebnissen älterer Studien (vgl. Bar-Ilan 2002b; Qiu/Chen/Wang 2004; Chu 2005). Auch die rohe Zahl der ausgegebenen URLs enthält einen hohen Anteil von Dubletten. Somit bietet weder die angegebene Trefferzahl noch die Anzahl der ausgegebenen URLs eine geeignete Untersuchungsgrundlage. Eine sorgfältige Bereinigung samt Dublettenkontrolle ist unabdingbar. Daraus folgt, dass Inlinkdaten von hinreichender Qualität nur für Webbereiche gewonnen werden können, deren Inlinkabfragen das Outputlimit nicht überschreiten und vollständig bereinigt werden können.

Bei offenen Erhebungsdesigns, die zu großen Trefferzahlen führen, sind die Effekte des Outputlimits, der Überschätzung und der Doppelung von Treffern besonders stark. Daher ist neben sorgfältiger Datenbereinigung generell zu empfehlen, eher begrenzte Abfragen (z.B. auf Links zwischen Universitäten bzw. anderen relevanten Webbereichen oder auf Links aus einzelnen Top-Level-Domains) auszuführen.

Auch ein Beispiel für die Diskontinuität von Suchmaschinen (vgl. Abschnitt 2.4.4) bietet diese Erhebung: Windows Live stellte während der Pretests sein umfassendes Angebot zur TLD-, länder- oder sprachspezifischen Suche nach Inlinks unangekündigt ein, so dass kurzfristig eine Alternative gesucht werden musste. Würde ein solcher Suchmaschinenwechsel während Längsschnittuntersuchungen zur internationalen Sichtbarkeit im World Wide Web notwendig, hätte dies gravierende Folgen für das Forschungsdesign und die Vergleichbarkeit der Daten.

Sowohl Inlink- als auch Outlinkerhebung haben gezeigt, wie essentiell eine sorgfältige Vorbereitung der Erhebung und die Bereinigung von Webdaten ist. Der damit verbundene Aufwand ist somit größer als häufig beschrieben: Mit ein paar Klicks gelangt man keinesfalls zu zuverlässigen Daten.

4.4.3 Schlussfolgerungen zu Rezeptionsdaten

Daten zur Rezeption von Websites – z.B. dazu, wie oft eine Seite aufgerufen wurde – sind eine wertvolle Ergänzung zu Inlinkdaten. Letztere geben lediglich die durch Suchmaschinen vermittelte Sichtbarkeit in Form von Links wieder, die als Zeichen der Anerkennung der Qualität von Websiteinhalten interpretiert werden. Hier zeigt sich eine weitere Parallele zur Bibliometrie, die Zitationen als Indikator für die Qualität wissenschaftlicher Publikationen nutzt, wenngleich diese Annahme kontrovers diskutiert wurde und die Gründe des Zitierens äußerst heterogen sind (vgl. Hornbostel 1997: 283ff).

Indem Daten zur faktischen Nutzung der Websites einbezogen werden, erweitert sich die webometrische Analyse um die Dimension der Rezeption. Das bibliometrische Pendant wäre die Information, wie viele Personen ein Dokument gelesen haben, über die die etablierten Zitationsdatenbanken keine Auskunft geben. Der Trend von gedruckten zu elektronischen Zeitschriften könnte hier neue Möglichkeiten eröffnen, da sich hier neben der Zahl der Zitationen – sofern zugänglich – auch die der Downloads auswerten ließe.

Der Hauptgrund dafür, dass Analysen der Websitenutzung bisher nur selten durchgeführt werden, ist der begrenzte Zugang zu Daten, die nur über die BetreiberInnen der untersuchten Internetauftritte erhältlich sind. Die meisten der in dieser Fallstudie untersuchten Meeresforschungseinrichtungen haben die gewünschten Nutzungsstatistiken nicht für den gewünschten Zeitraum zur Verfügung gestellt. Zudem erwies sich die Heterogenität der Software, die die auskunftsbereiten Institute verwendeten, als Problem (vgl. Abschnitt 4.3.6).

Eine Alternative zu den hier erbetenen Nutzungsstatistiken wären die Server-Logfiles. Sie liefern präzisere und besser vergleichbare Daten, erfordern jedoch ebenfalls die Kooperation der Website-BetreiberInnen. Außerdem ist die Datenbereinigung und -analyse hier voraussetzungsvoller und aufwendiger, so dass eine sorgfältige Einarbeitung notwendig ist und deutlich mehr Zeit eingeplant werden muss.

Eine weitere Möglichkeit bietet der Dienst „Alexa Internet“ unter www.alexa.com. Er stellt kostenlos Nutzungsstatistiken für Websites zur Verfügung. Datenquelle sind hier nicht die Websiteserver, sondern WebnutzerInnen, die die Alexa-Suchleiste in ihrem Browser nutzen. Sie dokumentiert, welche Seiten aufgesucht werden, und sendet diese Daten an Alexa. Zu den Vorzügen dieser Datenquelle gehört der geringe Zeitaufwand. Gleichzeitig bestehen wie bei kommerziellen Suchmaschinen Validitäts- und Reliabilitätsprobleme, da weder die Datengrundlage noch ihre Verarbeitung und Gewichtung bekannt sind. Es ist zu vermuten, dass zusätzliche Tools wie die Alexa-Suchleiste eher in Regionen mit gutem Internetzugang genutzt wird, so dass von einem Bias gegenüber Entwicklungsländern auszugehen ist. Hinzu kommt, dass die NutzerInnen dieser Suchleiste häufig nicht darüber informiert sind, dass ihre

Webaktivitäten durch Alexa protokolliert und ausgewertet werden. Die Suchleiste kann daher auch als Spyware klassifiziert werden (vgl. Der Landesbeauftragte für Datenschutz und Informationsfreiheit Bremen 2007). Wenngleich es in der Webometrie Stimmen gibt, die sich für eine Nutzung von Alexa und ähnlichen Diensten aussprechen (Vaughan 2008), ist davon aus forschungsethischen und methodologischen Gründen abzuraten.

Eine gut zugängliche Datenquelle, die über die Nutzung einzelner Websites Auskunft gibt, eine pragmatische Auswertung erlaubt und gleichzeitig vollständige und vergleichbare Daten liefert, existiert zur Zeit nicht. Grundsätzlich erscheint ein internationaler Abgleich von Link- und Nutzungsmustern, insbesondere vor dem Hintergrund der ungleichen globalen Verteilung von Ressourcen und Internetzugängen, sinnvoll.

5. Die Ergebnisse der empirischen Fallstudie

In diesem Kapitel werden die fünf im vorangegangenen Kapitel angeführten Leitfragen (vgl. Abschnitt 4.1) empirisch geklärt. Für jede Leitfrage wird zunächst die Datengrundlage beschrieben, bevor die Ergebnisse präsentiert und diskutiert werden.

5.1 Wie aussagekräftig sind Top-Level-Domains als Indikatoren der nationalen Anbindung?

Die Leitfrage, ob sich Top-Level-Domains als Indikatoren für die nationale Verortung der Inhalte eignen, die Webseiten repräsentieren, enthält mehrere Teilaspekte: Zu klären ist für Inlink- wie Outlinkdaten beispielsweise, ob unter URLs mit länderspezifischen TLDs tatsächlich Seiten mit inhaltlicher Anbindung an das jeweilige Land abgelegt werden. Eine weitere zentrale Frage ist, ob Seiten aus einzelnen Ländern gleichermaßen unter TLDs mit und ohne Länderhinweis zu finden sind. Andernfalls wären bei einer Analyse, die sich auf die Auswertung von TLDs mit Länderhinweis beschränkt, einzelne Länder über- bzw. unterrepräsentiert.

Die hier verwendete Unterscheidung zwischen TLDs mit und ohne Länderhinweis ist nicht deckungsgleich mit den Kategorien country-code Top-Level-Domain (ccTLD) und generische Top-Level-Domain (gTLD) der IANA, von denen erstere alle TLDs mit zwei Buchstaben (wie „de“ oder „eu“) und letztere alle TLDs mit mehr Buchstaben (wie „edu“, „info“ oder „museum“) umfasst. Die folgende Analyse weicht insofern von dieser Einteilung ab, als die gTLDs „edu“, „gov“ und „mil“ als Hinweis auf die USA gewertet werden, während die ccTLD „eu“ nicht als Verweis auf eine einzelne Nation akzeptiert wird. Für TLDs mit Länderhinweis werden auch die Begriffe „nationale TLD“ und „landesspezifische TLD“ verwendet, für TLDs ohne Länderhinweis wird das Synonym „sektorale TLD“ oder „länderneutrale TLD“ gebraucht.

Zur Prüfung der ersten Leitfrage wurde ein Sample von Inlinkquellen und Outlinkzielen einer Inhaltsanalyse unterzogen, bei der die gespeicherten Seiten auf deutliche Bezüge zu einer Nation untersucht wurden. Ein solcher Bezug konnte beispielsweise die Anschrift einer repräsentierten Person oder Organisation sein, der Ort einer Veranstaltung oder ein räumlicher Bezug im Namen der vorgestellten Einrichtung (wie z.B. in „Universität Bielefeld“). Thematische Bezüge – z.B. zu einer Region, in der ein Forschungsprojekt durchgeführt wird – wurden nicht berücksichtigt. Sofern die repräsentierte Person oder Organisation bekannt war und der Inhalt der gespeicherten Seite sowie die verwendete Domain mit diesem Vorwissen konsistent waren, war auch die Codierung von Seiten zugelassen, die keine Hinweise auf ihre nationale Anbindung enthielten. In analoger Vorgehensweise wurde vermerkt, ob die Seiten binationale, multinationale oder zwischenstaatliche Bezüge enthiel-

ten. Grundsätzlich war bei der Codierung auf Konsistenz und Plausibilität der Informationen zu achten. Ein detailliertes Codebuch findet sich im Anhang.

5.1.1 Verwendete Daten

Outlinkdaten

Unter den am 2. Juli 2007 erhobenen Websitedaten der Meeresforschungseinrichtungen wurden institutsübergreifend 5.852 distinkte Outlinkziele (OLZ) identifiziert und vom 5. bis 14. Juli 2007 durch ein Visual-Basic-Makro abgespeichert. Aufgrund der großen Anzahl wurde nur eine mit SPSS generierte Stichprobe inhaltsanalytisch untersucht. Sie umfasst 1.463 Outlinkziele (25,0%) (vgl. Tabelle 20).

	<i>OLZ</i>	<i>Anteil</i>	<i>ILQ</i>	<i>Anteil</i>
distinkte OLZ / ILQ	5.852	---	12.876	---
Stichprobe	1.463	25,0% der distinkten OLZ	1.288	10,0% der distinkten ILQ
<i>Art der Codierung</i>				
Erstcodierung	1.450	99,1% der Stichprobe	1.196	92,9% der Stichprobe
mit Sprachkundigen	13	0,9% der Stichprobe	92	7,1% der Stichprobe
<i>Ausschluss von Fehlern</i>				
inaktive Seiten	74	5,1% der Stichprobe	58	4,5% der Stichprobe
leere Seiten	3	0,2% der Stichprobe	0	0,0% der Stichprobe
gültige OLZ / ILQ	1.386	94,7% der Stichprobe	1.230	95,5% der Stichprobe

Tabelle 20: Inlink- und Outlinkstichprobe für die Inhaltsanalyse

13 Seiten in diesem Sample waren in Sprachen verfasst, die keine zuverlässige Codierung durch die Promovendin erlaubten. Sie wurden gemeinsam mit sprachkundigen Personen klassifiziert. 77 OLZ wurden von der weiteren Analyse ausgeschlossen, weil die Domain, der Server oder die Seite zum Zeitpunkt des Downloads inaktiv waren (74 Fälle bzw. 5,1%) oder weil die aufgerufene Seite keinen Inhalt enthielt (3 Fälle bzw. 0,2%). Die folgende Analyse bezieht sich somit auf 1.386 gültige Outlinkziele.

Inlinkdaten

Aufgrund der großen Zahl von 12.876 Inlinkquellen (ILQ) wurde im Anschluss an die Erhebung vom 25. September 2007 mit SPSS eine zehnpromzentige Zufallsstichprobe gezogen und zwischen dem 30. September und dem 2. Oktober 2007 automatisiert aufgerufen und abgespeichert.

92 dieser 1.288 ILQ (7,1%) waren in Sprachen erfasst, die es erforderten, eine sprachkundige Person hinzuzuziehen. 58 ILQ (4,5%) waren zum Zeitpunkt des Downloads nicht aktiv, so dass sich die weitere Analyse auf 1.230 Seiten beschränkt (vgl. Tabelle 20).

Reliabilitätsprüfung

Die bei Inhaltsanalysen übliche Reliabilitätsprüfung hat das Ziel zu klären, ob die Codierungspraxis der codierenden Person im Zeitverlauf stabil geblieben ist (Intracoder-Reliabilität) und ob eine andere Person anhand des Codebuchs zu einem vergleichbaren Ergebnis kommt (Intercoder-Reliabilität) (Rössler 2005: 185ff). Um die Intracoder-Reliabilität zu prüfen, wurden nach Abschluss der Codierung die zuerst klassifizierten OLZ und ILQ erneut durch die Promovendin codiert (jeweils 5%). Außerdem wurde eine Zufallsstichprobe im Umfang von zehn Prozent der Datensätze durch eine zweite Person codiert, um die Intercoder-Reliabilität zu bestimmen. Die geforderte Mindestzahl von dreißig bis fünfzig wiederholten bzw. gemeinsamen Codierungen (Früh 1998: 166) wurde sowohl bei der Recodierung als auch bei der Zweitcodierung der OLZ und ILQ überschritten (vgl. Tabelle 21 und Tabelle 22). Die Reliabilitätsprüfung erstreckt sich auf die Kategorien der nationalen Anbindung sowie eines übernationalen Bezugs. Die Top-Level-Domains der In- und Outlinks wurden mit Hilfe eines Visual-Basic-Scripts aus den URLs extrahiert, dessen Präzision manuell geprüft wurde.

<i>Outlinkziele</i>	<i>Anzahl</i>	<i>nationaler Bezug</i>	<i>übernationaler Bezug</i>
Recodierung durch die Promovendin (5% der Erstcodierung)	73	0,99	0,99
Codierung durch zweite Person (10% der Erstcodierung)	145	0,89	0,97

Tabelle 21: Übereinstimmungsquotienten nach Holsti für die Outlinkziele

Für die Reliabilitätsprüfung nominalskaliertter Variablen gilt der Überschneidungsquotient nach Holsti als geeignetes Maß (vgl. Merten 1983: 303f; Rössler 2005: 190). Beim Vergleich zweier Codierungen entspricht er dem Anteil der übereinstimmenden Ergebnisse. Für die Recodierung der Bezüge zu einzelnen Nationen der Outlinkziele lag er bei 99 Prozent, für die Codierung durch eine zweite Person bei 89 Prozent. Die Klassifikation von übernationalen Bezügen ergab denselben Wert für die Recodierung (99%) sowie 97 Prozent für die Sekundärcodierung (vgl. Tabelle 21). Eine Ursache für die unterschiedlich hohen Werte bei der Zweitcodierung liegt darin, dass übernationale Bezüge generell selten sind und hier deutlich weniger Kategorien vorgegeben waren als bei den nationalen Bezügen. Dies

begünstigt die Wahrscheinlichkeit übereinstimmender Urteile (vgl. Merten 1983: 306; Rössler 2005: 193).

<i>Inlinkquellen</i>	<i>Anzahl</i>	<i>nationaler Bezug</i>	<i>übernationaler Bezug</i>
Recodierung durch die Promovendin (5% der Erstcodierung)	60	0,98	0,98
Codierung durch zweite Person (10% der Erstcodierung)	120	0,83	0,98

Tabelle 22: Übereinstimmungsquotienten nach Holsti für die Inlinkquellen

Die Recodierung der Inlinkquellen ergab mit je 98 Prozent bei den nationalen und übernationalen Bezügen ein ähnliches Bild (vgl. Tabelle 22). Das gleiche gilt für die Codierung durch eine zweite Person, die zu 83 bzw. 98 Prozent mit der Erstcodierung übereinstimmte. Nach Rössler (2005: 192) gelten für inhaltliche (d.h. nicht rein formale) Kategorien Reliabilitätskoeffizienten ab 0,80 oder mehr als hinreichend, so dass die gesamte Codierung als reliabel bewertet werden kann.

Prüfung der Genauigkeit von Yahoos TLD-spezifischer Suche

Für alle mit Yahoo erhobenen Inlinks auf die Websites der zehn Meeresforschungseinrichtungen wurde geprüft, inwiefern die in den Abfragen spezifizierten TLDs mit denen der ausgegebenen Inlinks übereinstimmen. Im Information Retrieval wird hier von der Genauigkeit bzw. der *precision* einer Datenbank gesprochen (vgl. Baeza-Yates/Ribeiro-Neto 1999: 75f).

Die instituts- und TLD-spezifischen Yahoo-Abfragen nach dem Muster „linkdomain:ifm-geomar.de site:org“ führten zu 18.444 Treffern (vgl. Tabelle 10). Nach der Eliminierung von Dubletten innerhalb der einzelnen Abfragen blieben 13.562 Treffer. Für jeden dieser Inlinks wurde abgeglichen, ob seine TLD der Abfrage entspricht.

Nur ein einziger Treffer trug nicht die TLD, die in der Abfrage vorgegeben war: In den Ergebnissen zu „linkdomain:ifm-geomar.de site:be“ war ein Inlink mit der TLD „.com“ enthalten.¹⁴ Der gleiche Inlink wurde zusätzlich in der passenden Abfrage „linkdomain:ifm-geomar.de site:com“ ausgegeben. 13.561 von 13.562 gefundenen Inlinks trugen somit die gesuchte TLD, so dass die Genauigkeit bei 99,9 Prozent lag.

¹⁴ Die vollständige URL des Inlinks lautet: http://cgi.ebay.com/METEOROLOGY-CLIMATE-AVIATION-EARTH-SCIENCE-PHYSICS_W0QQitemZ290154682576QqcmdZViewItem.

Zum Zeitpunkt der Erhebung erlaubte nur Yahoo weitreichende Möglichkeiten der Inlinksuche nach Top-Level-Domains (vgl. Abschnitt 4.3.3). Daher konnte dieses Ergebnis nicht mit der Qualität anderer Anbieter verglichen werden.

Auswahl der Korrelationskoeffizienten

Da es in diesem Abschnitt ausschließlich um bivariate Zusammenhänge zwischen nominalskalierten Daten geht, wird Chi-Quadrat χ^2 verwendet. Für den Spezialfall zweier dichotomer Variablen wird der Phi-Koeffizient ϕ gewählt. Beide Maße erfordern für jedes Feld einer Kreuztabelle eine *erwartete* Häufigkeit von mindestens fünf, die sich aus den Randsummen ergibt. Gegebenfalls werden daher einzelne Ausprägungen der Variablen entsprechend zusammengefasst.

5.1.2 Ergebnisse

Um zu prüfen, ob Top-Level-Domains als Indikator für die nationale Anbindung von Webseiten dienen können, wird die Leitfrage in sieben Teilaspekte aufgesplittet. In diesem Abschnitt wird jeweils für Outlinkziele und für Inlinkquellen untersucht, wie viele der verwendeten TLDs einen Länderhinweis enthalten (Abschnitt 5.1.2.1) und wie viele Seiten sich inhaltlich einem bestimmten Land zuordnen lassen (Abschnitt 5.1.2.2). Um Unterschiede zwischen Outlinkzielen und Inlinkquellen verstehen zu können, wird anschließend der Stellenwert von Spam näher beleuchtet (Abschnitt 5.1.2.3).

Einen zentralen Punkt dieses Kapitels stellt die Frage dar, wie gut der äußerliche Landesbezug durch die TLD und der inhaltliche Landesbezug übereinstimmen (Abschnitt 5.1.2.4). Eine hohe Übereinstimmung ist eine notwendige Voraussetzung für die Nutzung von TLDs als Nationalitätsindikator. Eine weitere essentielle Bedingung ist, dass unter den Seiten einzelner Länder TLDs mit und ohne Länderhinweis in ähnlichem Umfang vertreten sind (Abschnitt 5.1.2.5).

Aufmerksamkeit wird auch übernationalen Inlinkquellen und Outlinkzielen gewidmet. Hier geht es speziell um die Fragen, welche TLDs für Seiten gewählt werden, die sich auf mehrere Länder gleichermaßen oder auf eine zwischenstaatliche Ebene beziehen (Abschnitt 5.1.2.6) und welche Inhalte unter den supranationalen Top-Level-Domains „int“ und „eu“ abgelegt werden (Abschnitt 5.1.2.7).

Anschließend werden die Ergebnisse in Abschnitt 5.1.3 zusammenfassend diskutiert.

5.1.2.1 Wie viele TLDs von Outlinkzielen und Inlinkquellen enthalten einen Länderhinweis?

Outlinkziele

745 der 1.386 gültigen OLZ (53,8%) weisen eine TLD mit Länderhinweis auf, 641 eine andere TLD (46,2%) (vgl. Tabelle 23). Bei einer TLD-bezogenen Untersuchung der Outlinkziele nach Ländern könnte folglich nur eine Aussage über gut die Hälfte der OLZ getroffen werden.

<i>Inhaltlicher Länderbezug</i>	<i>TLD mit Länderhinweis</i>	<i>TLD ohne Länderhinweis</i>	<i>gesamt</i>
ja	637	360	997
nein	108	281	389
gesamt	745	641	1.386

Tabelle 23: Outlinkziele nach inhaltlichem und äußerlichem Länderbezug

Inlinkquellen

Von den 1.230 ILQ wiesen 453 (36,8%) eine TLD mit Länderhinweis auf (vgl. Tabelle 24). Bei den übrigen 777 ILQ (63,2%) enthielt die URL eine länderneutrale TLD.

<i>Inhaltlicher Länderbezug</i>	<i>TLD mit Länderhinweis</i>	<i>TLD ohne Länderhinweis</i>	<i>gesamt</i>
ja	211	317	528
nein	242	460	702
gesamt	453	777	1.230

Tabelle 24: Inlinkquellen nach inhaltlichem und äußerlichem Länderbezug

Nur für ein gutes Drittel der Inlinkquellen kann somit untersucht werden, wie gut inhaltlicher und äußerlicher Länderbezug durch die TLD übereinstimmen.

5.1.2.2 Wie viele Outlinkziele und Inlinkquellen haben einen inhaltlichen Hinweis auf die Landesbindung?

Outlinkziele

Der Großteil aller gültigen OLZ enthält einen inhaltlichen Länderhinweis (997 von 1.386 OLZ bzw. 71,9%; vgl. Tabelle 23). Allerdings ist dieser Anteil unter den OLZ, die TLDs mit Länder-

hinweis aufweisen (637 von 745 OLZ bzw. 85,5%), wesentlich höher als bei übrigen (360 von 641 OLZ bzw. 56,2%). Dieser Unterschied ist hoch signifikant ($\varphi = 0,326^{**}$; $p = 0,000$; $N = 1386$; $m.v. = 77$).

Ein möglicher Grund ist, dass TLDs ohne Länderhinweis häufiger für Internetauftritte gewählt werden, die multinationale Organisationen (z.B. multinationale Unternehmen oder NGOs unter „.com“ und „.org“) repräsentieren (vgl. Abschnitt 5.1.2.6) oder bei denen aus anderen Gründen die nationale Anbindung der AutorInnen im Hintergrund steht. So liegen themenzentrierte Seiten häufig unter „.info“ oder Open-Source-Software unter „.org“ oder „.net“. Für Internetauftritte, die eine einzelne Organisation, Person oder eine Veranstaltung repräsentieren, wird offensichtlich überdurchschnittlich häufig eine nationale TLD gewählt.

Inlinkquellen

Weniger als die Hälfte aller Inlinkquellen (528 von 1.230 bzw. 42,9%) enthält einen Bezug zu einem bestimmten Land (vgl. Tabelle 24). Die ILQ mit TLD mit und ohne Länderhinweis unterscheiden sich in dieser Hinsicht gering, aber signifikant ($\varphi = 0,056^{*}$; $p = 0,048$; $N = 1230$; $m.v. = 58$): 46,6 Prozent der ILQ mit landesspezifischer TLD und 40,8 Prozent der ILQ mit sektoraler TLD lassen sich inhaltlich einer Nation zuordnen.

Eine wahrscheinliche Erklärung für die Diskrepanz zwischen OLZ und ILQ liegt darin, dass erstere häufig auf die Startseiten von Internetauftritten verweisen, die häufig grundlegende Informationen wie eben Ortsangaben enthalten, aber selbst nur selten die Quelle von Links sind (vgl. Smith 2004; Chu 2005).

Ein weiterer Grund kann sein, dass unter den Inlinkquellen eher Spamseiten zu erwarten sind als bei den Outlinkzielen, die von den Forschungseinrichtungen selbst ausgewählt wurden. Internetauftritte, die obskuren Zwecken dienen, dürften in der Regel keine (authentischen) Angaben zur Urheberschaft enthalten. Der Stellenwert von Spam unter den verlinkten Seiten wird unter dem folgenden Punkt 5.1.2.3 näher betrachtet.

Für die weitere Untersuchung bedeutet dies, dass sich die Befunde nur auf eine Minderheit der Inlinkquellen beziehen und somit weniger belastbar sind als die Ergebnisse zu den Outlinkzielen.

5.1.2.3 Welche Rolle spielt Spam unter den Inlinkquellen und Outlinkzielen?

Bei den codierten Inlinkquellen und Outlinkzielen wurde ohne Reliabilitätsprüfung zusätzlich klassifiziert, ob die Seiten vermutlich Spam enthielten. Der Großteil der als Spam codierten Seiten ist drei Typen zuzuordnen (vgl. Abbildung 7).

Klone

Klone sind Kopien bestehender Seiten. Oft handelt es sich um Seiten aus Wikipedia oder dem Open Directory Project, bei denen das Kopieren von Inhalt und Layout nicht prinzipiell illegal ist. Diese beliebten Seiten werden oft unter ähnlichen Domainnamen wie z.B. www.wikipedia.net.pl ins Netz gestellt. So lässt sich mit wenig Aufwand eine seriös erscheinende Website kreieren und mit eingefügter Werbung Geld verdienen.

sinnlose Linksammlungen

Einige Seiten enthielten Linksammlungen, die meist einen professionellen äußeren Eindruck machten, aber nicht über einen inhaltlichen Zusammenhang verfügten. Eine Teilgruppe dieses Typs stellen „geparkte Domains“ dar, die zum Verkauf stehen und in der Zwischenzeit mit beliebigen Links und Werbung gefüllt werden, um von potentiellen BesucherInnen zu profitieren. Bei einem anderen Teil bestanden die Domainnamen aus Kombinationen von Webschlagwörtern wie "site", "bay", "library", "links", "super", "inter", "micro", "net", "dir" etc. Auch sie enthielten – bei identischem Layout – Werbung und Linksammlungen, die offensichtlich mit Hilfe von Suchmaschinen anhand von Schlagworten zusammengestellt worden waren. Möglicherweise spielen hier auch "Linkfarmen"¹⁵ eine Rolle, die dazu dienen, die Position bestimmter Seiten im Output von Suchmaschinen zu verbessern.

c) Verkauf pharmazeutischer Präparate

Auch die aus Spam-Mails bekannten Angebote zum Kauf diverser pharmazeutischer Produkte, beispielsweise zur Brust- oder Penisvergrößerung, waren unter den verlinkten Seiten zu finden.

Abbildung 7: Unter der Inlinkquellen auftretende Typen von Spam

Einige weitere Seiten wurden als Spam klassifiziert, konnten sich allerdings keiner der drei Kategorien eindeutig zuordnen lassen und wurden daher in einer Restkategorie zusammengefasst.

Outlinkziele

Die Outlinkziele wurden von den Meeresforschungseinrichtungen selbst gewählt. Daher war zu erwarten, dass Spamseiten unter ihnen keinen großen Stellenwert einnehmen. In der Tat gab es nur einen Fall von Spam unter den OLZ (Tabelle 25): Unter der URL <http://www.joi-odp.org/joi/SEISMIC/ocnsnofc.htm> waren Werbung für Antivirensoftware sowie vermutlich

pornographische Videos abgelegt. Der Domainname bezieht sich auf das bereits beendete Ocean Drilling Program (ODP), das von dem Konsortium Joint Oceanographic Institutions (JOI) koordiniert wurde. Somit handelt es sich um eine Adresse, die möglicherweise früher in einem meereswissenschaftlichen Zusammenhang stand und nun anders genutzt wird. Die Seite enthält keinen Hinweis zur nationalen Anbindung. Da es sich um einen Einzelfall handelt, ist der Einfluss auf die Verteilung aller 1.386 Outlinkziele marginal.

<i>Spamtyp</i>	<i>Outlinkziele</i>	<i>Inlinkquellen</i>
Klone	0	114
sinnlose Linksammlungen	0	99
Verkauf pharmazeutischer Produkte	0	5
Sonstige Spam	1	9
gesamt	1	227

Tabelle 25: Outlinkziele und Inlinkquellen mit Spaminhalten

Inlinkquellen

Unter den 1.230 Inlinkquellen finden sich 227 Spamseiten, die somit einen Anteil von 18,5 Prozent ausmachen. Der Großteil davon entfällt auf Klone (50,0%) und sinnlose Linksammlungen (43,6%) (vgl. Tabelle 25).

Spamseiten verteilen sich ähnlich wie die übrigen Inlinkquellen auf Top-Level-Domains mit und ohne Länderhinweis: 79 von 227 ILQ mit Spaminhalt (34,8%) weisen eine Top-Level-Domain mit Länderhinweis auf, während es bei den übrigen ILQ 37,3 Prozent sind (vgl. Tabelle 26). Dieser Unterschied ist nicht signifikant ($\phi = -0,20$ n.s.; $p = 0,483$; $N = 1230$; m.v. = 58).

<i>Typ der Inlinkquelle</i>	<i>ILQ mit Spaminhalt</i>	<i>ILQ ohne Spaminhalt</i>	<i>ILQ gesamt</i>
TLD mit Länderhinweis	79	374	453
TLD ohne Länderhinweis	148	629	777
gesamt	227	1.003	1.230

Tabelle 26: Inlinkquellen mit und ohne Spaminhalt nach TLD-Typ

Außerdem enthalten nur 14 von 227 ILQ mit Spaminhalt (6,2%) einen Bezug zu einem Land; bei den übrigen ILQ sind es dagegen 51,2 Prozent (vgl. Tabelle 27). Hier weichen die Werte hoch signifikant von den erwarteten ab ($\phi = -0,353^{**}$, $p = 0,000$, $N = 1230$, m.v. = 58).

¹⁵ Für eine genauere Erklärung vgl. <http://de.wikipedia.org/wiki/Linkfarm> (zuletzt abgerufen am 19.11.2009).

<i>Typ der Inlinkquelle</i>	<i>ILQ mit Spaminhalt</i>	<i>ILQ ohne Spaminhalt</i>	<i>ILQ gesamt</i>
mit inhaltlichem Länderbezug	14	514	528
ohne inhaltlichen Länderbezug	213	489	702
gesamt	227	1.003	1.230

Tabelle 27: Inlinkquellen mit und ohne Spaminhalt nach inhaltlichem Länderbezug

Somit erhöhen die Spamseiten unter den Inlinkquellen zwar nicht die Anzahl von TLDs ohne Länderhinweis, aber sie beeinflussen den Stellenwert von Seiten ohne inhaltlichen Bezug zu einem Land. Da Spamseiten unter den Outlinkzielen nur eine marginale Rolle spielen, tragen die Spamseiten deutlich zu dem Gefälle zwischen OLZ und ILQ im Hinblick auf den Anteil von Seiten mit inhaltlichem Länderbezug bei.

5.1.2.4 Wie hoch ist die Übereinstimmung von Seiteninhalt und der verwendeten Top-Level-Domain?

Outlinkziele

Bei den 637 Outlinkzielen, die sowohl eine TLD mit Länderhinweis als auch einen inhaltlichen Landesbezug aufwiesen, stimmten äußerlicher und inhaltlicher Bezug in 632 Fällen (99,2%) überein. Darin enthalten sind auch acht Seiten mit gleichwertigen Bezügen zu zwei Nationen, für die jeweils die TLD eines der beiden Länder gewählt worden war, was hier als Übereinstimmung gewertet wurde.

<i>Outlinkziel</i>	<i>TLD-Land</i>	<i>Inhaltlicher Länderbezug</i>
www.bremen.edu	USA	Deutschland
www.leiden.edu	USA	Niederlande
www.bbsr.edu	USA	Bermuda
www.coastalguide.to	Tonga	Niederlande
http://www.nbc4.tv/news/9957420/detail.html	Tuvalu	USA

Tabelle 28: Outlinkziele mit widersprüchlichen nationalen Bezügen in Seiteninhalt und Top-Level-Domain

Nur in fünf Fällen (0,8%) widersprechen sich inhaltlicher und TLD-Länderbezug. Hierbei handelte es sich dreimal um „.edu“-Adressen, die von Universitäten und Forschungseinrichtungen in Deutschland, den Niederlanden und den Bermudas genutzt wurden, sowie ein niederländischer Küsten-Guide unter „.to“ und ein US-Fernsehsender unter „.tv“ (vgl. Tabelle 28).

Inlinkquellen

201 von 211 Inlinkquellen mit Landesbezug und landesspezifischer TLD (95,3%) verweisen inhaltlich und durch die TLD auf dasselbe Land. Darunter sind sechs Seiten mit gleichwertigen Bezügen zu zwei Nationen, die als übereinstimmend betrachtet wurden, wenn die TLD von einem der beiden Länder genutzt wurde. Von den übrigen zehn ILQ (4,7%) stammen zwei Inlinks von Seiten, die die Top-Level-Domains „.ag“ und „.tv“ nutzen, um sich selbst als Aktiengesellschaft bzw. Fernsehsender auszuweisen (vgl. Tabelle 29).

<i>Inlinkquelle</i>	<i>TLD-Land</i>	<i>Inhaltlicher Länderbezug</i>
http://www.sohnix.ag/it-netzwerke_referenzen.html	Antigua	Deutschland
http://www.arte.tv/fr/histoire-societe/le-dessous-des-cartes/392,CmC=551470,CmPage=70.251900.392,CmPart=com.arte-tv.www,CmStyle=98674,view=links.html	Tuvalu	Deutschland und Frankreich
http://bgbm3.bgbm.fu-berlin.de/idb/botherb.html	Deutschland	Finnland
http://www.biologie.uni-hamburg.de/b-online/library/bioluminescence/Dino_web.html	Deutschland	USA
http://www.iasfbo.inaf.it/Services/Local/ScoutSci/02/scoutsci-990512.html	Italien	USA
http://www.jergym.hiedu.cz/~canovm/vyhledav/varianty/usa.html	Tschechien	USA
http://www.ebay.ph/viltem?ItemId=170142533257	Philippinen	USA
http://directory.google.ch/alpha/Top/World/Deutsch/Wissenschaft/Geowissenschaften/Ozeanographie/	Schweiz	USA
http://directory.google.ch/Top/World/Deutsch/Wissenschaft/Geowissenschaften/Geochemie/	Schweiz	USA
http://www.sincos.de.vu/	Vanuatu	Deutschland

Tabelle 29: Inlinkquellen mit widersprüchlichen nationalen Bezügen in Seiteninhalt und Top-Level-Domain

Die nächsten vier Inlinkquellen, die unter deutschen, italienischen und tschechischen TLDs abgelegt sind, enthalten Spiegel von Seiten aus Finnland und den USA. Drei weitere Inlinkquellen stammen aus Internetauftritten der US-Unternehmen Ebay und Google, die auf die Philippinen bzw. die Schweiz zugeschnitten sind und daher die TLDs dieser Länder tragen. Die letzte ILQ ist in der TLD Vanuatus abgelegt, deren Subdomain „.de.vu“ kostenlos und besonders in Deutschland vermarktet wird.¹⁶

Insgesamt ist die Übereinstimmung von Top-Level-Domains und inhaltlichen nationalen Bezügen sowohl bei Outlinkzielen als auch bei Inlinkquellen hoch. Durch eine gezielte Suche nach TLDs wie „.tv“, „.ag“ oder „.to“, die wie andere Top-Level-Domains außerhalb ihres

Nennlandes aufgrund übertragener Bedeutungen beliebt sind (vgl. Gallup 2003), ließe sich ein Teil der Abweichungen leicht identifizieren. Dagegen sind insbesondere Spiegelseiten anhand äußerlicher Kriterien kaum zu erkennen.

Die Seiten, die mit den ausgewählten Meeresforschungseinrichtungen verlinkt und unter einer landesspezifischen Top-Level-Domain abgelegt sind, sind somit nur selten Ländern zuzuordnen, die nicht ihrem äußeren Landesbezug entsprechen. Knapp die Hälfte der Outlinkziele und fast zwei Drittel der Inlinkquellen wiesen jedoch eine länderneutrale TLD auf (vgl. Abschnitt 5.1.2.2). Deshalb ist zu klären, ob sich die inhaltlichen Länderbezüge von Links mit TLDs mit und ohne Länderhinweis in vergleichbarer Weise auf die einzelnen Nationen verteilen.

5.1.2.5 Verteilen sich Seiten mit Bezug zu einzelnen Ländern gleichmäßig auf TLDs mit und ohne Länderhinweis?

Mit einem Chi-Quadrat-Unabhängigkeitstest wurde geprüft, ob bestimmte Länder einen ähnlichen Stellenwert unter den Links mit und ohne landesspezifischen TLDs einnehmen.

Outlinkziele

Dazu wurden die 997 Outlinkquellen mit inhaltlichem Bezug zu einem Land genauer untersucht. 13 dieser URLs mit TLDs mit und ohne Länderhinweis hatten eine gleichwertige inhaltliche Anbindung an zwei Länder. Beispiele hierfür wären ein deutsch-russischer Studiengang oder eine Themenseite, die gemeinsam von einem Deutschen und einem Kanadier unterhalten wird. In diesen Fällen wurden beide Bezüge berücksichtigt, so dass insgesamt 1.010 nationale Bezüge untersucht wurden.

Chi-Quadrat-basierte Maße erfordern eine minimale *erwartete* Häufigkeit von fünf in jedem Feld der zugrundeliegenden Kreuztabelle, die sich aus den Randsummen ergeben. Um diese Bedingung zu erfüllen, wurden weniger häufig vertretene Nationen in einer Restkategorie zusammengefasst (vgl. Tabelle 30).

Der hochsignifikante χ^2 -Wert von 223,5** ($p = 0,000$; $N = 1010$; m.v. = 389) deutet auf ungleichmäßige Verteilung beider Variablen hin. Eine genauere Auswertung der Kreuztabelle zeigt, dass australische, deutsche und südafrikanische Seiten zu einem außergewöhnlich hohen Anteil unter nationalen TLDs firmieren (Australien: 13 von 15 bzw. 86,7%, Deutschland: 277 von 308 bzw. 89,4%, Südafrika: 18 von 19 oder 94,7%), während indische und US-amerikanische Seiten überwiegend unter Top-Level-Domains ohne Länderhinweis abgelegt werden (Indien: 10 von 19 bzw. 52,6%, USA: 266 von 449 bzw. 59,2%). Für das vorliegende Sample würden Top-Level-Domains als Nationalitätsindikator somit den Stellenwert Indiens

¹⁶ Vgl. <http://www.nic.de.vu> (zuletzt abgerufen am 10.06.2009).

und der USA unter den Outlinkzielen unterschätzen, den der erstgenannten Länder dagegen deutlich überschätzen.

<i>Inhaltlicher Länderbezug</i>	<i>TLD mit Länderhinweis</i>	<i>TLD ohne Länderhinweis</i>	<i>gesamt</i>
Australien	13	2	15
Deutschland	278	33	311
Frankreich	9	5	14
Indien	9	10	19
Kanada	13	7	18
Südafrika	18	1	19
UK	21	18	39
USA	183	266	449
andere	101	23	124
gesamt	645	365	1.010

Tabelle 30: Outlinkziele nach inhaltlichem Länderbezug und Top-Level-Domain mit und ohne Länderhinweis

Inlinkquellen

Auch für die 528 Inlinkquellen mit inhaltlichem Bezug zu einem Land wurde ein Chi-Quadrat-Unabhängigkeitstest durchgeführt. Neun Inlinkquellen wiesen binationale Bezüge auf, die beide in die Analyse einbezogen wurden, so dass insgesamt die Verteilung von 537 Länderbezügen ausgewertet wurde. Wie bei den Outlinkzielen wurden Felder mit einer erwarteten Häufigkeit von weniger als fünf Inlinkquellen zu einer Restkategorie zusammengefasst (vgl. Tabelle 31).

<i>Inhaltlicher Länderbezug</i>	<i>TLD mit Länderhinweis</i>	<i>TLD ohne Länderhinweis</i>	<i>gesamt</i>
Deutschland	86	15	101
Spanien	8	7	15
UK	20	4	24
USA	22	260	282
andere	81	34	115
gesamt	217	320	537

Tabelle 31: Inlinkquellen nach inhaltlichem Länderbezug und Top-Level-Domain mit und ohne Länderhinweis

Auch die Verteilung der Inlinkquellen auf einzelne Länder und Top-Level-Domains mit und ohne Länderhinweis ist nicht unabhängig, wie der χ^2 -Wert von 270,9** ($p = 0,000$; $N = 537$; $m.v. = 702$) zeigt. Noch deutlicher als bei den Outlinkzielen finden sich Unterschiede zwischen einzelnen Nationen: Die Inlinkquellen mit einem Bezug zu Deutschland und Großbritannien sind weit überwiegend in der jeweiligen nationalen Top-Level-Domain zu finden (85,1% bzw. 83,3%). Dagegen liegen Seiten mit Anbindung an die USA mehrheitlich in TLDs ohne Länderhinweis (92,2%).

Während für die deutschen Inlinkquellen und Outlinkziele gleichermaßen häufig die TLD „de“ gewählt wurde, zeigt sich für die USA ein deutliches Gefälle: Während bei den US-Outlinkzielen knapp 60 Prozent in länderneutralen TLDs abgelegt waren, galt dies für die Inlinkquellen fast ausschließlich (s.o.). Der wesentliche Grund ist vermutlich, dass die Outlinks häufiger auf Seiten höherer Bildungseinrichtungen sowie staatlicher Institutionen und Programme verweisen, die berechtigt sind, die TLDs „edu“ bzw. „gov“ zu nutzen. Von Inlinks kann angenommen werden, dass sie heterogenere Quellen haben, die neben den bereits genannten Organisationen vermehrt Unternehmen, Medien, Privatleute, Spammer u.a. umfassen. Für diese Gruppen ist jedoch nur die US-spezifische TLD „us“ zugänglich, die in den USA jedoch nicht die Popularität von „com“ oder „org“ erreicht hat (vgl. Wass 2003c).

5.1.2.6 Welche TLDs werden für Seiten mit überstaatlichem Bezug gewählt?

In diesem Abschnitt wird geprüft, wie viele Outlinkziele und Inlinkquellen einen zwischenstaatlichen oder multinationalen Bezug aufweisen und unter welcher Art von Top-Level-Domain diese Seiten abgelegt wurden. Dabei wurden alle Seiten entsprechend klassifiziert, die sich auf mehrere einzelne Staaten (z.B. Forschungsprojekte mit TeilnehmerInnen aus mehreren Staaten oder Unternehmen mit Standorten in mehreren Ländern) oder auf zwischenstaatliche Organisationen wie die EU oder die UN bezogen. Die Seiten können gleichzeitig einen Hinweis auf einen Sitz in einem bestimmten Staat besitzen. So ist beispielsweise die World Meteorological Organisation der Vereinten Nationen eine zwischenstaatliche Einrichtung mit Sitz in Genf.

Outlinkziele

Bei hundert OLZ (7,2% der gültigen 1.386 Fälle) wurde ein multinationaler oder zwischenstaatlicher Bezug festgestellt (vgl. Tabelle 32). Für diese Seiten war zu 75,0 Prozent eine TLD ohne Länderhinweis gewählt worden. Dagegen trugen die übrigen URLs überwiegend nationale TLDs (720 von 1.286 OLZ bzw. 56,0%). Diese Verschiebung ist signifikant ($\varphi = 0,161$ **; $p = 0,000$; $N = 1386$; $m.v. = 77$).

<i>überstaatlicher Bezug</i>	<i>TLD mit Länderhinweis</i>	<i>TLD ohne Länderhinweis</i>	<i>gesamt</i>
ja	25	75	100
nein	720	566	1.286
gesamt	745	641	1.386

Tabelle 32: Outlinkziele mit übernationalem inhaltlichem Bezug

Inlinkquellen

50 von 1.230 gültigen ILQ (4,1%) bezogen sich auf eine überstaatliche Ebene. Sie verteilten sich ähnlich wie die übrigen ILQ auf TLDs mit und ohne Länderhinweis, so dass es keinen signifikanten Unterschied gab ($\varphi = 0,044^{**}$; $p = 0,308$; $N = 1230$; $m.v. = 58$).

<i>überstaatlicher Bezug</i>	<i>TLD mit Länderhinweis</i>	<i>TLD ohne Länderhinweis</i>	<i>gesamt</i>
ja	14	36	50
nein	439	741	1.180
gesamt	453	777	1.230

Tabelle 33: Inlinkquellen mit übernationalem inhaltlichem Bezug

Während Outlinks, die auf überstaatliche Inhalte verwiesen, außergewöhnlich häufig länderneutrale TLDs trugen, galt dies nicht für die Inlinkquellen.

5.1.2.7 Welchen Inhalt haben Seiten mit zwischenstaatlichen Domains?

Derzeit gibt es zwei Top-Level-Domains, die auf eine supranationale Ebene verweisen. Die Top-Level-Domain „.int“ ist speziell für zwischenstaatliche Einrichtungen wie z.B. UN-Organisationen, Interpol etc. vorgesehen und im Zugang kontrolliert. Die erst 2006 eingeführte TLD „.eu“ ist für den EU-Raum vorgesehen (vgl. EC 2009). Da sie frei vergeben wird, stellt sich die Frage, ob sich die darin abgelegten Webseiten auf die EU-Ebene beziehen oder nicht.

Outlinkziele

Nur sechs OLZ waren unter den supranationalen TLDs „.eu“ und „.int“ abgelegt (vgl. Tabelle 34). Die geringen Fallzahlen erlauben keine weitreichenden Schlussfolgerungen. Einige Hinweise bietet das Sample dennoch.

<i>Outlinkziele</i>	<i>TLD übernationaler Bezug</i>
ec.europa.eu/research/descartes/communication_en.htm	eu offizielle EU-Seite
www.baltex-research.eu/	eu multinationales Forschungsprogramm
www.bio-antifouling.eu/	eu kein Hinweis
www.eur-oceans.eu/	eu EU-Projekt
europa.eu.int/comm/fisheries/doc_et_publ/pub1_en.htm	int offizielle EU-Seite
www.iai.int/	int zwischenstaatliches Forschungsinstitut

Tabelle 34: Outlinkziele mit supranationalen Top-Level-Domains

Die frei zugängliche TLD „eu“ wurde bei vier OLZ im Sample genutzt, von denen zwei einen inhaltlichen EU-Bezug (eine offizielle Seite der EU sowie ein EU-Projekt) aufwiesen. Eine weitere Seite repräsentierte ein übernationales Forschungsprogramm zum Ostseeraum, dessen Teilnehmerschaft nicht eindeutig auf EU-Staaten begrenzt ist. Das vierte Outlinkziel enthielt keinen Bezug zur Europäischen Union.

Es wurden somit in den Outlinkzielen der TLD „eu“ keine starken Abweichungen vom EU-Raum gefunden. Allerdings wiesen zwanzig weitere OLZ einen Bezug zur Europäischen Union auf – vorwiegend multinationale Forschungsprojekte, die durch ein EU-Rahmenprogramm gefördert wurden, die jedoch nicht anhand der TLD der EU zugeordnet werden konnten. Sie waren größtenteils unter „org“ und anderen TLDs ohne Länderbezug abgelegt (vgl. Tabelle 35).

Die TLD „int“ nutzten in diesem Sample ein zwischenstaatliches Forschungsinstitut sowie eine offizielle Seite der Europäischen Union (vgl. Tabelle 34). Somit waren tatsächlich nur übernationale Inhalte zu finden, was der Erwartung an eine TLD mit kontrolliertem Zugang entspricht. Offizielle EU-Seiten waren sowohl unter „int“ als auch unter „eu“ zu finden. Beides erscheint insofern sinnvoll, als die EU eine zwischenstaatliche Einrichtung ist und die TLD „eu“ speziell auf den EU-Raum zugeschnitten wurde. Es irritiert dennoch, dass keine einheitliche Wahl getroffen wurde.

Auch viele zwischenstaatliche Einrichtungen bevorzugen offensichtlich sektorale Top-Level-Domains wie „org“ oder die TLD ihres Sitzlandes gegenüber ihrer spezifischen TLD „int“: 23 der 25 als zwischenstaatlich klassifizierten Seiten nutzten andere TLDs (vgl. Tabelle 35).

<i>Outlinkziele</i>	<i>TLD</i>	<i>zwischenstaatl. Bezug</i>	<i>EU-Bezug</i>
www.redfish.de/	de	nein	ja
www.jrc.it/	it	nein	ja
www.codyssey.co.uk/	uk	nein	ja
www.ecasa.org.uk/	uk	nein	ja
www.meramed.com/	com	nein	ja
www.eu-hermes.net/	net	nein	ja
www.coastalpractice.net	net	nein	ja
www.intermarec.net/	net	nein	ja
www.sea-search.net	net	nein	ja
www.biomareweb.org/	org	nein	ja
www.marbef.org	org	nein	ja
biocombe.org	org	nein	ja
www.carboocean.org	org	nein	ja
www.crimea-info.org/	org	nein	ja
www.elme-eu.org/	org	nein	ja
www.ensembles-eu.org/	org	nein	ja
www.eur-oceans.org/	org	nein	ja
www.carboschools.org/	org	nein	ja
www.eurosion.org	org	nein	ja
www.eur-oceans.eu/	eu	nein	ja
ec.europa.eu/research/descartes/communication_en.htm	eu	ja	ja
europa.eu.int/comm/fisheries/doc_et_publ/pub1_en.htm	int	ja	ja
www.iai.int/	int	ja	nein
www.wmo.ch/indexflash.html	ch	ja	nein
www.ipcc.ch	ch	ja	nein
www.unep.ch/	ch	ja	nein
www.wmo.ch	ch	ja	nein
www.ices.dk/iceswork/wgdetail.asp?wg=SGEH	dk	ja	nein
www.ices.dk/products/fiche/Plankton/SHEET158.PDF	dk	ja	nein
www.ices.dk	dk	ja	nein
www.helcom.fi/environment/indicators2003/sst.html	fi	ja	nein
www.helcom.fi	fi	ja	nein
www.acmad.ne/	ne	ja	nein
www.africanoceans.net	net	ja	nein
www.oceanexpert.net/	net	ja	nein
www.earthobservations.org/index.html	org	ja	nein
www.eurocean.org	org	ja	nein
www.ioc-goos.org/	org	ja	nein
www.iode.org	org	ja	nein
www.loicz.org	org	ja	nein
www.millenniumassessment.org/en/index.aspx	org	ja	nein
www.oceansatlas.org	org	ja	nein
www.odinafrica.org	org	ja	nein
www.thecommonwealth.org	org	ja	nein
www.worldbank.org	org	ja	nein

Tabelle 35: Outlinkziele mit inhaltlichem Bezug zur Europäischen Union und anderen zwischenstaatlichen Organisationen und Programmen

Inlinkquellen

Sieben Inlinkquellen waren unter der TLD „.eu“ abgelegt (vgl. Tabelle 36). Eine dieser Seiten repräsentierte die Europäische Kommission, zwei weitere von der EU geförderte, internationale Forschungsprojekte. Die übrigen vier enthielten keinen Hinweis auf übernationale Bezüge, so dass unklar ist, inwiefern sie eine inhaltliche Anbindung an die EU besitzen.

Die einzige Inlinkquelle mit der TLD „.int“ stammt ebenfalls von der europäischen Kommission, so dass offizielle Seiten der Europäischen Union unter beiden supranationalen TLDs vertreten waren.

<i>Inlinkquellen</i>	<i>TLD</i>	<i>übernationaler Bezug</i>
ec.europa.eu/fisheries/tools/links_fr.htm	eu	offizielle EU-Seite
www.alticore.eu/index.php?option=com_content&task=view&id=51&Itemid=1	eu	kein Hinweis
www.alzakera.eu/music/vetenskap/Biologia/bio-0050.htm	eu	kein Hinweis
www.gowarsaw.eu/warsaw-info/index.php?title=Ophir	eu	kein Hinweis
www.novac-project.eu/partners-es.htm	eu	EU-Forschungsprojekt
www.rendy.eu/novinky/samice-kladivouna-se-muze-rozmnoziti-bez-samce	eu	kein Hinweis
www.uncover.eu/index.php?id=86	eu	EU-Forschungsprojekt
www.europa.eu.int/comm/research/endocrine/projects_completed_en.html	int	offizielle EU-Seite

Tabelle 36: Inlinkquellen mit supranationalen Top-Level-Domains

Eine Auswertung aller 15 Inlinkquellen mit supranationalen Bezügen (Tabelle 37) zeigt, dass insgesamt neun ILQ einen Bezug zur EU aufweisen, von denen nur die aus Tabelle 36 bekannten vier unter den TLDs „.eu“ bzw. „.int“ abgelegt waren. Von den fünf übrigen Seiten stammten zwei aus „.org“ und drei weitere aus nationalen TLDs.

Außer den beiden Seiten der EU-Kommission hatten noch sechs Inlinkquellen zwischenstaatliche Urheber wie die Food and Agriculture Organization (FAO) der Vereinten Nationen oder International Oceanographic Data and Information Exchange (IODE). Sie wählten durchgängig die länderneutrale TLD „.org“, wobei das IODE-Portal „Ocean Experts“ zusätzlich mit der Spiegelseite in „.net“ vertreten war.

Unter den Outlinkzielen und Inlinkquellen, die aus den supranationalen Top-Level-Domains „.eu“ und „.int“ stammten, gab es keine Fälle, in denen diese TLDs eindeutig ohne entsprechenden Bezug verwendet wurden. Allerdings überwogen bei OLZ wie ILQ die Fälle, in denen EU-bezogene oder zwischenstaatliche Seiten keine sektoralen TLDs trugen. Insbesondere UN-Organisationen bevorzugten offensichtlich die TLD „.org“.

<i>Inlinkquellen</i>	<i>TLD</i>	<i>zwischenstaatl.</i>	
		<i>Bezug</i>	<i>EU-Bezug</i>
www.univ-brest.fr/IUEM/projects/ecco/ecco_uk.htm	fr	nein	ja
traccess.tubitak.gov.tr/fp6_yeni/FrameSetLink.aspx?grpId=104&am p;tabIndex=5&tabid=458&ald=461	tr	nein	ja
www.elme-eu.org/public/Consortium.aspx	org	nein	ja
www.eloisegroup.org/themes/habitats/introduction.htm	org	nein	ja
www.pumpsea.icat.fc.ul.pt/main.php?p=3&o=87	pt	nein	ja
www.uncover.eu/index.php?id=86	eu	nein	ja
www.novac-project.eu/partners-es.htm	eu	nein	ja
ec.europa.eu/fisheries/tools/links_fr.htm	eu	ja	ja
www.europa.eu.int/comm/research/endocrine/projects_completed_ en.html	int	ja	ja
www.fao.org/fi/website/FIRetrieveAction.do?dom=org&xml=FI DI_STAT_org.xml&xp_nav=5	org	ja	nein
www.oceanexpert.net/viewMemberRecordPrintable.php?memberI D=11314	net	ja	nein
ioc.unesco.org/odinafrica3/contents.php?id=45&action=send	org	ja	nein
oceanexpert.org/viewInstitutionRecord.php?&institutionID=107 39	org	ja	nein
oceanexpert.org/viewMemberRecord.php?&memberID=13651	org	ja	nein
www.iode.org/oceanexpert/viewInstitutionRecord.php?institutionID =7503	org	ja	nein

Tabelle 37: **Inlinkquellen mit inhaltlichem Bezug zur Europäischen Union und zu zwischenstaatlichen Einrichtungen und Programmen**

5.1.3 Zusammenfassung und Diskussion

Wie die Zusammenstellung der Daten (vgl. Abschnitt 5.1.1) gezeigt hat, ist es möglich, TLD-spezifische Daten von In- und Outlinks zu sammeln. Aus Crawlerdaten lassen sich die Top-Level-Domains der Outlinks automatisiert extrahieren, und Yahoo bietet die Option, die Linksuche auf einzelne TLDs zu beschränken. Unter zehntausenden Treffern fand sich hier ein einzelner Fehler, für den es keine naheliegende Erklärung gibt. Insgesamt hat sich Yahoos TLD-spezifische Suche nach Inlinks als äußerst präzise erwiesen.

Es stellt sich somit die Frage, ob die erhobenen Top-Level-Domain-Daten Aussagekraft im Hinblick auf die nationale Verortung von Webseiten besitzen. Es hat sich gezeigt, dass in diesem Sample nur die Hälfte der Outlinkziele und ein starkes Drittel der Inlinkquellen eine Top-Level-Domain mit Länderhinweis besaß (vgl. Abschnitt 5.1.2.1), so dass sich eine länderspezifische Analyse der Top-Level-Domains nur auf einen Bruchteil der Links stützen kann. Die Inhaltsanalyse ergab, dass fast drei Viertel der OLZ, aber deutlich weniger als die Hälfte der ILQ einen inhaltlichen Bezug zu einem Land aufwiesen (vgl. Abschnitt 5.1.2.2). Ein

Grund für dieses Gefälle liegt vermutlich darin, dass Startseiten, die oft grundlegende Informationen enthalten, unter den OLZ häufiger vertreten sind als unter den ILQ. Die Inhaltsanalyse zeigte, dass die Präsenz von Spamseiten unter den Inlinkquellen (vgl. Abschnitt 5.1.2.3) einen Beitrag zu dieser Diskrepanz liefert: Während Spam unter den Outlinkzielen eine seltene Ausnahme ist, spielen insbesondere geklonte Wikipedia- und Directory-Einträge sowie sinnlose Linklisten unter den Inlinkquellen eine substantielle Rolle. Spamseiten unterscheiden sich von den übrigen ILQ darin, dass sie signifikant seltener TLDs mit Länderhinweis aufweisen und vor allem seltener einen inhaltlichen Bezug zu einem Land enthalten. Dies verwundert kaum, da diese Seiten nur vorgeblich der Informationsvermittlung, faktisch aber – nicht immer legalen – kommerziellen Interessen dienen.

Diese Befunde zeigen, dass Outlinks auf den Websites von Forschungseinrichtungen eine belastbarere Datengrundlage liefern als deren Inlinks. Dies gilt aufgrund der beschriebenen Verzerrungen durch Spam speziell in Bezug auf die Analyse von Top-Level-Domains, aber auch im Allgemeinen. Bei der Analyse der Sichtbarkeit von Forschungseinrichtungen anhand von Inlinks muss hinterfragt werden, um welche Art der Sichtbarkeit es hier geht: Die geklonten Links existieren genauso wie die originalen; ihr Einfluss auf die Rezeption der Inlinkziele dürfte jedoch marginal sein. Gleichzeitig ist zu vermuten, dass sie einen Matthäuseffekt verursachen, da vermutlich vorwiegend bekannte Institute in den kopierten Verzeichnissen enthalten sind, und damit quantitative Linkanalysen verzerren. Bei der Untersuchung der Vernetzung von Forschungseinrichtungen durch Outlinks stellt sich dies spezielle Problem nicht.

Eine essentielle Voraussetzung für die Verwendung von Top-Level-Domains als Indikator für die nationale Anbindung ist, dass der Inhalt von Webseiten mit ihrer landesspezifischen Top-Level-Domain – sofern vorhanden – übereinstimmt. Hier zeigte die Inhaltsanalyse, dass sowohl Outlinkziele als auch Inlinkquellen diese Bedingung erfüllen (vgl. Abschnitt 5.1.2.4). Wenngleich nur für einen geringen Teil der Daten inhaltlicher und äußerlicher Länderbezug verglichen werden konnte, sprechen Übereinstimmungsquoten von 99 bzw. 95 Prozent für OLZ und ILQ für eine hohe Zuverlässigkeit. Ein Teil der abweichenden Fälle könnte zudem mit geringem Aufwand identifiziert und überprüft werden.

Als kritisch hat sich dagegen eine zweite Vorbedingung erwiesen: Seiten, die sich inhaltlich einzelnen Nationen zuordnen ließen, verteilten sich nicht gleichmäßig auf Top-Level-Domains mit und ohne Länderhinweis. So waren Seiten mit deutscher, südafrikanischer und britischer Anbindung meist unter einer nationalen TLD abgelegt, während dies bei indischen und US-amerikanischen Seiten deutlich seltener der Fall war (vgl. Abschnitt 5.1.2.5).

Da insbesondere bei den Inlinkquellen nur ein Teil aller Seiten national verortet werden konnte, sind die Ergebnisse eher als empirischer Hinweis denn als endgültiger Befund zu

verstehen. Sie sind jedoch konsistent mit historischen Betrachtungen nationaler Unterschiede in der Wahl der Top-Level-Domains (vgl. Gallup 2003, Gandhi 2003, Lindén 2003, StClair 2003, Wass 2003c).

Da die geringe Verbreitung einiger Top-Level-Domains wie „.in“ und „.us“ auf die ehemals restriktive Vergabep Praxis zurückzuführen ist, die sich inzwischen weitestgehend gelockert hat (vgl. Gandhi 2003, Wass 2003c), ließe sich über eine zukünftige Angleichung der Verbreitung nationaler TLDs spekulieren. Aus folgenden Gründen erscheint dies jedoch nicht sehr wahrscheinlich:

- die Persistenz nationaler Gewohnheiten,
- die fortbestehenden älteren Websites, für die keine nationalen TLDs gewählt worden sind,
- die weitgehende Öffnung der Vergabe neuer sektoraler Top-Level-Domains durch ICANN ab dem Jahr 2010 (vgl. ICANN 2008 und 2009),
- das anhaltende Wachstum des Webs und die Knappheit einprägsamer Domains, die eine Nutzung dieser neuen Top-Level-Domains wahrscheinlich erscheinen lassen.

Wenngleich also zuverlässige Möglichkeiten zur Erhebung TLD-spezifischer Daten existieren und die inhaltlichen Bezüge stark mit den genutzten landesspezifischen TLDs übereinstimmen, spricht dieser Punkt eindeutig gegen die Verwendung von Top-Level-Domains als Indikator von Nationalität: Die Übertragung der Ergebnisse von Links mit landesspezifischen TLDs auf solche mit länderneutralen TLDs ist nicht möglich.

Bisher nutzen nur drei internationale webometrische Studien nationale Top-Level-Domains als Indikatoren für die nationale Anbindung von Inlinkquellen oder Outlinkzielen (Thelwall 2004b; Li et al. 2005b; Thelwall/Zuccala 2008). Während Thelwall (2004b) zumindest Stellung dazu bezieht, dass Top-Level-Domains eine unvollkommene, aber einfach zugängliche Informationsquelle sind, gehen die AutorInnen der anderen beiden Artikel gänzlich unkritisch mit deren Verwendung um. Die Ergebnisse dieser Studien können allenfalls vage Hinweise auf die tatsächliche internationale Einbettung der untersuchten Universitäten und Institute geben. Größere Aussagekraft ließe sich möglicherweise durch detaillierte Kenntnisse der aktuellen Vergabep Praxis nationaler Top-Level-Domains erreichen, die jedoch nicht systematisch dokumentiert wird.

Eine Alternative besteht darin, eine Zufallsstichprobe aller Inlinkquellen bzw. Outlinkziele inhaltsanalytisch zu klassifizieren und die internationalen Kontakte der untersuchten Website auf diese Weise zu erschließen. Allerdings ist der Aufwand bei dieser Vorgehensweise erheblich.

Einen Nebenaspekt dieses Kapitels bildete die Auswertung überstaatlicher Seiten. Auch wenn *global governance* bei der Verwaltung des weltweiten Domain Name Systems durch die private NGO ICANN offensichtlich keine Priorität hat (Hofmann 2007), bietet es für die Analyse von Verlinkungen mit supranationalen Websites zwei Ansatzpunkte: die Top-Level-Domains „eu“ und „int“. Da zwischenstaatliche Einrichtungen der TLD „int“ häufig „org“ und andere vorziehen, erscheint sie als Indikator für nicht geeignet (vgl. Abschnitt 5.1.2.6. und 5.1.2.7). Aus welchem Grund dies geschieht, konnte die Inhaltsanalyse u.a. aufgrund geringer Fallzahlen nicht klären. Ebenso wenig kann die TLD „eu“ derzeit als Indikator für Webseiten mit Bezug zum EU-Raum dienen. Da sie noch sehr jung ist, könnte ihr Stellenwert in Zukunft steigen. Dagegen spricht allerdings, dass sie frei vergeben wird und beispielsweise von lokalen Organisationen in der EU, aber ohne Bezug zu ihrem Raum oder ihrer Steuerungsebene, als Ersatz für bereits vergebene nationale TLDs genutzt werden kann. Somit bietet das World Wide Web mit seinen primär national und sektoral ausgerichteten Top-Level-Domains zur Zeit keinen geeigneten Ansatzpunkt für die Untersuchung supranationaler Referenzen im World Wide Web.

Die ausgeführten Ergebnisse zur ersten Leitfrage betreffen die beiden folgenden Leitfragen (5.2 und 5.3): Die im Folgenden verwendeten Indikatoren der internationalen Orientierung und Sichtbarkeit beruhen auf der Auswertung nationaler Top-Level-Domains. Nur für einen Teil der Länder, die in den In- und Outlinkmustern präsent sind, liegen Informationen zur Vergabepaxis der nationalen Top-Level-Domains vor. Insofern besitzen die empirischen Ergebnisse nur Gültigkeit für den Teil der In- und Outlinks, die eine nationale Top-Level-Domain besitzen, nicht jedoch für die Links mit sektoralen TLDs.

5.2 Inwieweit ähneln sich die Bilder, die webometrische und bibliometrische Indikatoren von der internationalen Einbettung von Forschungseinrichtungen zeigen?

Diverse Studien (Thelwall 2001b; Thelwall 2002a; Smith/Thelwall 2002; Thelwall/Harries 2003; Thelwall/Tang 2003; Qiu/Chen/Wang 2004; Li et al. 2005a; Vaughan/Thelwall 2005, Aguillo et al. 2006) haben einen Zusammenhang zwischen Maßen von Webpräsenz bzw. -sichtbarkeit und Indikatoren der Forschungsproduktivität gezeigt. In diesem Abschnitt geht es darum zu prüfen, ob dieser Zusammenhang auch für die vorliegende, weltumspannende Auswahl von Meeresforschungseinrichtungen zu finden ist oder ob es Abweichungen gibt. Besondere Aufmerksamkeit wird der Frage gewidmet, wie sich webometrische und bibliometrische Maße der internationalen Einbettung zueinander verhalten.

5.2.1 Verwendete Daten

Verschiedene Web- und Publikationsdaten werden genutzt, um diese Leitfrage zu beantworten.

Webdaten

Die Präsenz der Meeresforschungsinstitute im World Wide Web wird durch die Anzahl der Webseiten und der übrigen Dateien bestimmt (vgl. Abschnitt 4.3.1). Ferner werden Outlink- und Inlinkdaten herangezogen, um die internationale Orientierung und Sichtbarkeit der Websites zu charakterisieren (vgl. Abschnitt 4.3.2 und 4.3.3). Aus zwei Gründen wurde für die folgende Auswertung der Linkdaten ausschließlich die Ebene von Domains gewählt, wenngleich dies zu umständlicheren Formulierungen führt: Erstens gibt Yahoo pro Inlinkdomain maximal zwei Inlinks aus, so dass die gesamte Inlinkzahl nicht bekannt ist (vgl. Abschnitt 4.3.3). Zweitens wird so der Einfluss von gehäuften Wiederholungen einzelner Outlinks minimiert. Solche Ausreißer treten beispielsweise auf, wenn jede Seite einen Link auf die Website der genutzten Websoftware enthält. Um die Unterscheidung zwischen Outlinks und Outlinkdomains bzw. Inlinks und Inlinkdomains zu erleichtern, werden alle vier Begriffe in Abbildung 8 an einem fiktiven Beispiel graphisch dargestellt.

Um die internationale Ausrichtung der Institute im World Wide Web und in ihren Publikationen zu charakterisieren, wurde eine Reihe der in Abschnitt 3.3 beschriebenen Indikatoren eingesetzt, die auf der Auswertung länderspezifischer Top-Level-Domains fußen. Hier handelt es sich z.B. um die Anzahl der In- und Outlinkdomains mit TLDs des eigenen Landes, anderer Länder, von Entwicklungs-, Industrie- und Übergangsländern und der einzelnen Kontinente. Bei der Betrachtung von Kopublikations- oder Linkkontakten zu Ländergruppen wird für jedes Institut das Sitzland ausgeschlossen, um jeweils nur die internationale Vernetzung analysieren zu können.

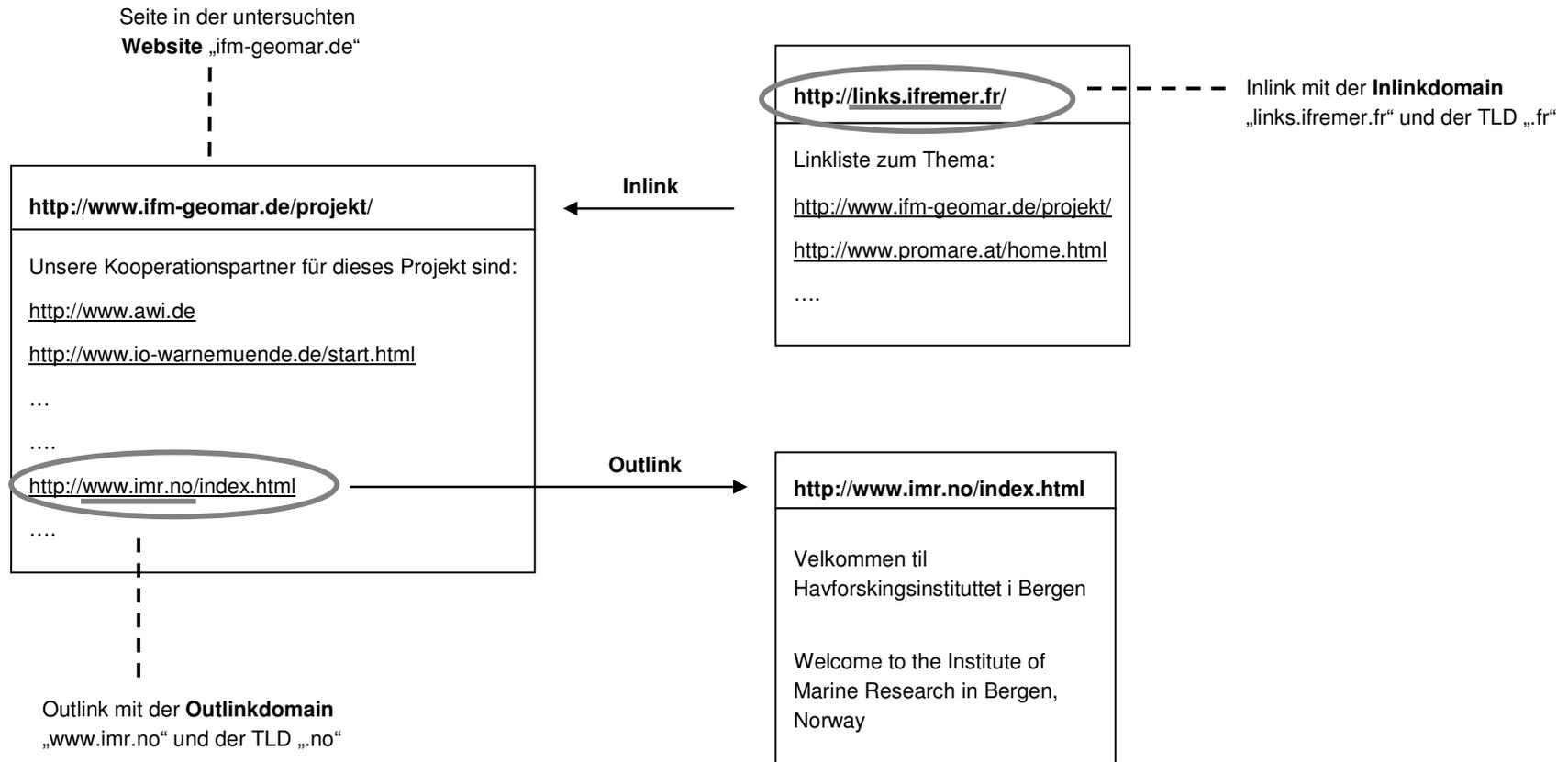


Abbildung 8: Fiktives Beispiel zur Unterscheidung von Inlinks, Inlinkdomains, Outlinks und Outlinkdomains, dargestellt aus Perspektive des IfM-Geomar (Graphik: Stephanie Zuber / Tina Ruschenburg).

Publikationsdaten

Für jedes Meeresforschungsinstitut werden die Einträge im Web of Science der Jahrgänge 2005 bis 2007 verwendet, um die rezente Präsenz im Mainstream der wissenschaftlichen Publikationen zu erfassen (vgl. Abschnitt 4.3.4). Es werden sämtliche Datensätze im Web of Science berücksichtigt, also nicht nur Fachartikel, sondern auch Editorials, Rezensionen etc.

Um Aufschluss über die internationalen Profile der Einrichtungen zu gewinnen, werden die Webindikatoren nachgebildet. Während sich die internationale Einbettung wissenschaftlicher Einrichtungen im Web anhand der Top-Level-Domains ihrer In- und Outlinks bestimmen lässt, werden bei Publikationen üblicherweise die Adressen herangezogen. Durch sie lassen sich beispielsweise Artikel mit nationalen und internationalen KoautorInnen unterscheiden.

Der zentrale Unterschied zu Links besteht darin, dass letztere einen gerichteten, einseitigen Bezug darstellen, Koautorenschaftsbeziehungen dagegen einen ungerichteten, wechselseitigen. Um dieser Asymmetrie Rechnung zu tragen und ein vollständiges Bild zu erhalten, werden den kopublikationsbasierten Indikatoren im Folgenden sowohl Outlink- als auch Inlinkmaße gegenübergestellt (5.2.2.3 und 5.2.2.4).

Die Zahl der AutorInnen wird dabei nicht berücksichtigt: Beiträge einzelner Personen mit mehreren Adressen werden als nationale bzw. internationale Kopublikationen betrachtet. Als internationale Kopublikationen werden alle Einträge im Web of Science gewertet, die mindestens eine Adresse in einem anderen Land aufweisen. Als nationale Kopublikationen zählen alle Einträge, die ausschließlich externe Adressen innerhalb des eigenen Landes enthalten. Immer wenn der Stellenwert von Ländergruppen in Kopublikationsmustern betrachtet wird – z.B. Industrie-, Entwicklungs- und Übergangsländer oder die Nationen einzelner Kontinente bzw. des eigenen Kontinents – wird das Sitzland der einzelnen Institute ausgeklammert. Dies gilt auch für die Linkanalyse. Es werden ausschließlich *whole counts* angegeben.

Neben der Zahl der Publikationen in Fachzeitschriften wird auch die Summe der zeitlich standardisierten Zitationen verwendet (vgl. Tabelle 16), die eine Einschätzung der allgemeinen Sichtbarkeit der gesamten Publikationsaktivität geben.

Weitere Daten

Um Effekte der Institutsgröße auszugleichen, werden die Daten standardisiert. Die Größe wird dabei auf zwei Arten bestimmt: zum einen an der Publikations- bzw. Webproduktivität (Anzahl der Publikationen bzw. der Webseiten), zum anderen am Personalumfang (Anzahl der wissenschaftlichen MitarbeiterInnen; vgl. Tabelle 17). Die Zugehörigkeit der Länder zu

Kontinenten und Weltregionen sowie der Entwicklungsstatus werden nach der Klassifikation der Vereinten Nationen bestimmt (vgl. Abschnitt 4.3.5).

An dieser Stelle sei nochmals auf das Verzeichnis „Abkürzungen und Lage der untersuchten Meeresforschungseinrichtungen“ hingewiesen (S. XI). Es enthält neben den Akronymen der Institute auch eine Zusammenstellung der Länder, Kontinente und Weltregionen, in denen sie angesiedelt sind. Dadurch kann es das Verständnis der folgenden geographischen Auswertungen erleichtern.

Auswahl des Korrelationskoeffizienten

Wenngleich alle Variablen ein metrisches Skalenniveau aufweisen, kann Pearsons Korrelationskoeffizient r nicht eingesetzt werden: Er setzt einen linearen Zusammenhang zwischen normalverteilten metrischen Variablen voraus (vgl. Degen/Lorscheid 2002: 71ff), der hier nur selten gegeben ist. Stattdessen wird Spearmans Rangkorrelationskoeffizient ρ verwendet, ein rangbasierter Korrelationskoeffizient für ordinalskalierte Variablen, der lediglich einen monotonen Zusammenhang erfordert (vgl. Degen/Lorscheid 2002: 75ff). Die Fragestellung fokussiert darauf, ob positive Zusammenhänge vorliegen oder nicht. Da die Fälle nicht auf einer Zufallsauswahl beruhen, wird auf Signifikanztests verzichtet.

5.2.2 Ergebnisse

Um die Leitfrage in angemessenem Umfang klären zu können, wird sie in ihre Teilaspekte zerlegt. Zunächst wird die Präsenz der ausgewählten Meeresforschungsinstitute im World Wide Web und im Web of Science verglichen (5.2.2.1). Gleiches folgt für ihre allgemeine Sichtbarkeit in beiden Medien (5.2.2.2). In den folgenden drei Abschnitten wird die internationale Orientierung der Institute im Web und im WoS verglichen (5.2.2.3). Beide Aspekte werden anschließend der internationalen Sichtbarkeit im Web gegenübergestellt (5.2.2.4 und 5.2.2.5). Abschließend wird für einzelne Meeresforschungseinrichtungen geprüft, ob die Kontinente und Regionen der Welt in den Publikations- und Linkprofilen einen ähnlichen Stellenwert einnehmen (5.2.2.6).

5.2.2.1 Gibt es einen Zusammenhang zwischen der Präsenz der Institute im World Wide Web und der Präsenz im Web of Science?

Um zu prüfen, ob die allgemeine Präsenz der Meeresforschungseinrichtungen im Web mit der Anwesenheit im Web of Science zusammenhängt, wurden die Einträge der Institute in letzterem mit elementaren Indikatoren der Webpräsenz – nämlich der Anzahl der Webseiten sowie der in dem Internetauftritt enthaltenen sonstigen Dateien und Outlinkdomains – korreliert. Um Effekte der Institutsgröße zu kontrollieren, wurden sämtliche Daten anhand der

MitarbeiterInnenzahl standardisiert (vgl. Tabelle 38). Da hier Publikationen und Webseiten korreliert werden, entfällt hier die Standardisierung am Publikations- und Web-Output.

<i>Publikationen im WoS vs.</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite/Publikation</i>
Webseiten	0,816	0,794	---
andere Dateien	0,794	0,745	---
Outlinkdomains	0,782	0,721	---

Tabelle 38: Spearmans ρ für Maße der Präsenz im World Wide Web und im Web of Science

Es zeigt sich, dass alle Maße der Webpräsenz stark mit der Anzahl der Publikationen im Web of Science zusammenhängen. Dies gilt, in etwas geringerem Umfang, auch für die am Personalumfang standardisierten Werte. Absolut gesehen verfügen IFM, IOW, NIO und MBARI sowohl über die größte Zahl von Publikationen als auch über die größten Websites. Dagegen besitzen MOI, NIOPK und ORI, die nur wenige Einträge im WoS haben, die kleinsten Websites (vgl. Tabelle 8 und Tabelle 15). Im Verhältnis zur Personalstärke ergibt sich eine leichte Verschiebung dieser Reihenfolge daraus, dass IFM und NIO nur eine mittlere Anzahl von Webseiten pro MitarbeiterIn produziert haben. Die Anzahl der anderen Dateien und der Outlinkdomains korrespondiert deutlich mit der Anzahl der Webseiten. Entsprechend korrelieren diese Werte ebenfalls stark mit der Publikationszahl.

5.2.2.2 Gibt es einen Zusammenhang zwischen der Sichtbarkeit der Institute im World Wide Web und im Web of Science?

Die Sichtbarkeit im Web wird durch die Zahl der Inlinkdomains operationalisiert. Durch die Anzahl der institutseigenen Webseiten geteilt, bildet dieser Wert den klassischen Web Impact Factor (WIF). Bei einer seiner Varianten steht die MitarbeiterInnenzahl im Nenner (vgl. Thelwall 2001b).

Ein exaktes Äquivalent dazu gibt es für Publikationsdaten nicht. Hier wird die Summe der zeitlich standardisierten Zitationen herangezogen (vgl. Tabelle 16). Auch sie wird jeweils durch die personelle Institutsgröße bzw. die Publikationszahl dividiert. Die Zitations- und Inlinkmaße entsprechen einander nicht direkt, sind jedoch die für beide Medien grundlegenden Indikatoren für Sichtbarkeit. Wie die Zahl der Inlinkdomains wird auch die Zahl der Zitationen in Beziehung zur Größe der Einrichtungen in Form von Personal- bzw. Publikationsstärke gesetzt. Auch hier ergeben sich für die rohen ebenso wie für die am Personalumfang standardisierten Werte hohe Korrelationen (vgl. Tabelle 39).

<i>Inlinkdomains vs.</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite/Publikation</i>
standardisierte Zitationen	0,794	0,818	-0,358

Tabelle 39: Spearmans ρ für Maße der Sichtbarkeit im World Wide Web und im Web of Science

Institute, die besonders viele Zitationen pro MitarbeiterIn erhalten, erhalten auch viele Inlinks aus verschiedenen Domains. In beiden Kategorien liegen sämtliche deutschen und US-amerikanischen Forschungseinrichtungen deutlich vor denen in Entwicklungsländern.

Der klassische WIF, der die Zahl der Webseiten im Nenner trägt, korreliert jedoch leicht negativ mit der Zahl der standardisierten Zitationen pro Veröffentlichung. In den Rohdaten zeigt sich, dass BLOS, MBARI und IFM besonders zahlreiche standardisierte Zitationen pro Publikation erhalten. Gleichzeitig erhalten MBARI und IFM eine deutlich unterdurchschnittliche Zahl an Inlinkdomains pro Webseite (berechnet aus Tabelle 8, Tabelle 12, Tabelle 15 und Tabelle 16). Ein Grund liegt darin, dass sie zwei der drei größten Websites im Sample besitzen. Die meisten Inlinks pro Seite ziehen dagegen ORI und NIOPK auf sich, die sehr kleine Internetauftritte aufweisen und die niedrigste Anzahl von Standardzitationen pro Publikation erhalten haben. Die negative Korrelation beruht somit auf Skaleneffekten.

5.2.2.3 Gibt es einen Zusammenhang zwischen der internationalen Orientierung im World Wide Web und im Web of Science?

Nachdem ein allgemeiner Zusammenhang zwischen der Präsenz und Sichtbarkeit der Meeresforschungseinrichtungen im World Wide Web und im Web of Science gefunden wurde (vgl. 5.2.2.1 und 5.2.2.2), wird in diesem Abschnitt geprüft, ob die internationale Orientierung der Meeresforschungsinstitute im World Wide Web und im Web of Science korrespondieren. Wie im vorangehenden Abschnitt 5.2.2.2 werden hier Indikatoren verwendet, die keine direkten, sondern funktionale Äquivalente sind.

Für die ersten Indikatoren wurde bestimmt, wieviele Outlinks in die nationalen TLDs des eigenen Landes, anderer Länder, des eigenen Kontinents und anderer Kontinente verweisen. Analog wurde die Anzahl der Publikationen mit externen Anschriften im eigenen Land und in anderen Ländern sowie im eigenen Kontinent (mit Ausnahme des eigenen Landes) und in anderen Kontinenten gezählt. Tabelle 40 ist zu entnehmen, dass die entsprechenden Web- und Publikationsmaße jeweils stark korrelieren.

Die Standardisierung an der Zahl der wissenschaftlichen MitarbeiterInnen schwächt die Zusammenhänge für Bezüge zum eigenen Land deutlich ab, während sie sich für andere Länder und Kontinente sowie für den eigenen Kontinent noch verstärken. Mit MBARI und HBOI haben zwei US-amerikanische Einrichtungen die meisten nationalen Kopublikationen

pro MitarbeiterIn, die Institute in Kenia, Mauritius, Pakistan und Südafrika dagegen die wenigsten. Mit Ausnahme des ORI haben sie auch nur sehr geringe nationale Outlinkwerte (berechnet aus Tabelle 17, Tabelle 65 und Tabelle 71). Bei den Outlinks innerhalb des eigenen Landes erreichen die beiden deutschen Institute sowie das BLOS in den USA besonders hohe Werte. Folglich findet sich nur ein mittlerer Zusammenhang (vgl. Tabelle 40).

Für internationale Publikationen sowie für entsprechende Outlinkdomains zeigt sich ein ähnliches Gefälle zwischen den beiden deutschen und allen übrigen Einrichtungen (berechnet aus Tabelle 17, Tabelle 65, Tabelle 66, Tabelle 71 und Tabelle 72). Bei den interkontinentalen Maßen mischen sich die Institute verschiedenen Entwicklungsstands dagegen stärker: Neben den beiden deutschen und zwei der US-amerikanischen Forschungseinrichtungen hat das mauritische MOI pro MitarbeiterIn hohe Werte an interkontinentalen Kopublikationen und Outlinkdomains. Die niedrigsten Werte weisen hier das NIOPK in Pakistan und das HBOI in den USA auf (berechnet aus Tabelle 17, Tabelle 65, Tabelle 66, Tabelle 71 und Tabelle 72).

<i>Kopublikationen im WoS vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite/Publikation</i>
eigenes Land	0,721	0,564	0,030
andere Länder	0,794	0,903	0,612
eigener Kontinent	0,811	0,878	0,768
andere Kontinente	0,799	0,915	0,636

Tabelle 40: Spearman's ρ für grundlegende Maße der internationalen Orientierung im World Wide Web und im Web of Science

Die Standardisierung am Web- bzw. Publikationsoutput der Institute zeigt einen entgegengesetzten Effekt: Sie verringert die Intensität der Korrelationen und nivelliert sie im Fall des eigenen Landes sogar vollständig ($\rho = 0,030$). Bei den drei US-amerikanischen Instituten machen nationale Kopublikationen jeweils einem bis zwei Drittel aller Veröffentlichungen aus, bei den übrigen – mit Ausnahme des NIOPK – dagegen weniger als ein Drittel, während die übrigen Institute sich im Mittelfeld bewegen (berechnet aus Tabelle 15 und Tabelle 71). Für die Anzahl der Outlinks in die eigene nationale TLD pro Webseite weisen die Meeresforschungseinrichtungen dagegen sehr homogene Werte auf. Ein deutlicher Ausreißer im oberen Bereich ist das ORI, bei dem nationale Kopublikationen jedoch einen unterdurchschnittlichen Stellenwert einnehmen (berechnet aus Tabelle 8, Tabelle 9 und Tabelle 65).

Als nächstes wurden Outlink- und Koautorenschaftsbeziehungen zu Übergangsländern, Entwicklungsländern und Industrieländern untersucht (vgl. Tabelle 41). Auch hier zeigen sich

bei den unstandardisierten Daten mittlere bzw. für Industrie- und Übergangsländer sogar hohe Korrelationen zwischen Web- und Publikationsdaten, die auch nach der Standardisierung anhand der Personalstärke erhalten bleiben. Institute, die besonders häufig mit anderen Einrichtungen in Industrie- und Übergangsländern kopublizieren, verlinken sich auch eher mit deren Top-Level-Domains als andere Meeresforschungseinrichtungen. Die Korrelationen für die Werte der Übergangsländer resultieren allerdings zum Teil daraus, dass es insgesamt nur wenige Outlinks in die Länder des ehemaligen Ostblocks gibt und dass KMFRI, MOI, NIOPK und ORI weder Outlink- noch Koautorenschaftsbeziehungen mit ihnen pflegen (vgl. Tabelle 66 und Tabelle 72). Diese Ergebnisse sind somit wenig belastbar.

<i>Kopublikationen im WoS vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite/Publikation</i>
Übergangsländer	0,803	0,777	0,700
Entwicklungsländer	0,568	0,578	0,862
Industrieländer	0,884	0,915	0,624

Tabelle 41: Spearman's ρ für Maße der Orientierung an Ländergruppen nach Entwicklungsstand im World Wide Web und im Web of Science

Die Standardisierung am Output schwächt die Zusammenhänge für Industrie- und Übergangsländer ab, verstärkt sie jedoch für Entwicklungsländer: Institute, bei denen Publikationen mit Entwicklungsländern einen hohen relativen Stellenwert haben, richten im Verhältnis zu ihrer Websitegröße auch besonders viele Outlinks in deren TLDs ein. Hierbei liegen die afrikanischen Institute ORI, MOI und KMFRI klar vorne, während NIO und NIOPK deutlich hinter den übrigen zurückbleiben (berechnet aus Tabelle 8, Tabelle 15, Tabelle 66 und Tabelle 72). Bei der Standardisierung am Personalumfang ist diese Divergenz weniger ausgeprägt (berechnet aus Tabelle 17, Tabelle 66 und Tabelle 72).

Betrachtet man die Kopublikationen mit den einzelnen Kontinenten und die entsprechenden Outlinkdomains, zeigt sich ein gemischtes Bild (vgl. Tabelle 42). Für Asien und Europa weisen die Rohdaten mittlere, für Ozeanien, Nord- und Südamerika hohe Spearman-Werte auf. Die am Personalumfang standardisierten Werte zeigen einen starken Zusammenhang für Europa, Nord- und Südamerika und Ozeanien, jedoch nicht für Afrika und Asien.

<i>Kopublikationen im WoS vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite/Publikation</i>
Afrika	0,494	0,297	0,452
Asien	0,584	0,400	0,244
Europa	0,662	0,721	0,127
Nordamerika	0,902	0,948	0,310
Ozeanien	0,732	0,607	0,512
Südamerika	0,848	0,848	0,889

Tabelle 42: Spearmans ρ für Maße der Orientierung an einzelnen Kontinenten im World Wide Web und im Web of Science

Die afrikanischen Institute KMFRI, MOI und ORI weisen die höchsten Werte von Kopublikationen mit afrikanischen Ländern pro MitarbeiterIn auf. Outlinkdomains in die Top-Level-Domains Afrikas sind bei ihnen jedoch selten (berechnet aus Tabelle 17, Tabelle 67 und Tabelle 73). Nur bei IFM und IOW korrespondieren Kopublikationen mit afrikanischen Einrichtungen mit entsprechenden Outlinks. Zu berücksichtigen ist, dass es sich um geringe absolute Zahlen handelt. Eine ähnliche Situation zeigt sich für die in Asien gelegenen Institute: Während das NIO die höchste Pro-Kopf-Zahl an Publikationen mit asiatischen Ländern aufweist, ist der entsprechende Linkwert äußerst gering. Das NIOPK weist weder Link- noch Publikationsbeziehungen innerhalb Asiens auf. Die höchste Zahl an asiatischen Outlinkdomains weist das BLOS auf, das aber nur wenige Kopublikationen mit diesem Kontinent veröffentlicht hat. Nur das IFM weist hohe Werte in beiden Dimensionen auf. Sowohl für Afrika als auch für Asien zeigt sich somit, dass sich Link- und Publikationsstrukturen nicht entsprechen: Die dortigen Institute verfügen mit Ausnahme des NIOPK über intrakontinentale Koautorenschaftsbeziehungen, die sich jedoch nicht in Outlinks niederschlagen.

Die deutschen Institute IFM und IOW verfügen über eine hohe Anzahl von europäischen Kopublikationen und Outlinkdomains pro MitarbeiterIn, ebenso wie BLOS und MBARI. Für Europa entsprechen sich auch die Werte der übrigen Institute gut. Dies gilt noch stärker für Nordamerika. Eine Besonderheit ist hier jedoch, dass die europäischen Institute für den nordamerikanischen Kontinent zum Teil höhere Publikations- und Outlinkwerte pro MitarbeiterIn aufweisen als die US-amerikanischen. Der Grund ist, dass rein nationale Kopublikationen und Outlinks nicht mitgezählt wurden, was bei Nordamerika stark ins Gewicht fällt, da die USA einen großen Teil des Kontinents ausmachen.

Auf zwei Kontinenten liegt keines der ausgewählten Institute: Die an der MitarbeiterInnenzahl standardisierten Werte für Ozeanien korrelieren insgesamt gut, jedoch gibt es drei Ausreißer: Das BLOS hat zahlreiche Outlinks nach Ozeanien, jedoch nur wenige Kopublikationen, während NIO und ORI australische Kopublikationen, aber keinerlei Outlinks in diese Welt-

region besitzen. Die sehr starke Korrelation für Südamerika geht primär darauf zurück, dass die Mehrzahl aller Institute keinerlei Kontakte hierhin besitzt: Nur die deutschen und US-amerikanischen Einrichtungen BLOS, MBARI, IFM und IOW besitzen Link- und Publikationsbeziehungen zu diesem Kontinent.

Die Standardisierung an der Anzahl der Webseiten bzw. Publikationen löscht fast alle starken Korrelationen aus. Die Ausnahme Südamerika ist auf die soeben beschriebene Begrenztheit der Kontakte zurückzuführen. Der relative Stellenwert der einzelnen Kontinente im Verhältnis zum Publikations- bzw. Web-Output ist somit über alle Institute hinweg ähnlich (berechnet aus Tabelle 8, Tabelle 15, Tabelle 67 und Tabelle 73).

5.2.2.4 Gibt es einen Zusammenhang zwischen der internationalen Sichtbarkeit im World Wide Web und der internationalen Orientierung im Web of Science?

In diesem Abschnitt wird die internationale Ausrichtung der Koautorenschaftsbeziehungen der internationalen Sichtbarkeit der Forschungseinrichtungen im World Wide Web, wie sie sich in Inlinks manifestiert, gegenübergestellt.

Für die nationalen, internationalen, intra- und interkontinentalen Kopublikationen und Inlinkdomains zeigen sich durchgängig hohe und sehr hohe Korrelationen (Tabelle 43). Die Standardisierung durch die personelle Institutsgröße mildert die Zusammenhänge insgesamt ab, so dass zwischen Publikations- und Inlinkbezügen zu anderen Ländern und Kontinenten nur noch ein mittlerer Zusammenhang bleibt. Die Standardisierung am Output nivelliert die Korrelationen deutlich stärker.

<i>Kopublikationen im WoS vs. Inlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite/Publikation</i>
eigenes Land	0,815	0,709	0,648
andere Länder	0,685	0,564	0,285
eigener Kontinent	0,890	0,805	0,110
andere Kontinente	0,644	0,479	0,492

Tabelle 43: Spearman's ρ für grundlegende Maße der internationalen Sichtbarkeit im World Wide Web und der internationalen Orientierung im Web of Science

Kopublikationen mit Einrichtungen im eigenen Land sind, ebenso wie Inlinks aus der eigenen TLD, im Verhältnis zur personellen Institutsgröße besonders häufig bei den US-amerikanischen und deutschen Meeresforschungsinstituten zu finden. Intrakontinentale Kopublikationen und Inlinkdomains spielen insbesondere bei den deutschen Instituten eine große Rolle, bei den afrikanischen und asiatischen Einrichtungen dagegen nur eine marginale (berechnet aus Tabelle 17, Tabelle 68 und Tabelle 71). Die internationalen Kopublikationen

und Inlinkdomains, standardisiert an der Zahl der MitarbeiterInnen, korrespondieren insgesamt gut. Allerdings erhalten die drei US-amerikanischen Institute wesentlich mehr internationale Inlinks als die übrigen, während die deutschen Einrichtungen besonders viele grenzüberschreitende Kopublikationen besitzen (berechnet aus Tabelle 17, Tabelle 68 und Tabelle 71). Diese Tendenz zeigt sich in abgeschwächter Form auch bei den interkontinentalen Daten.

Gruppiert man Kopublikationen und Inlinkdomains nach dem Entwicklungsstand der unter den Anschriften bzw. TLDs vertretenen Ländern, zeigen sich starke Zusammenhänge für die unstandardisierten Bezüge zu Industrieländern sowie, unabhängig von der Standardisierung, für die Bezüge zu Übergangsländern (Tabelle 44). Letztere gehen primär darauf zurück, dass KMFRI, MOI, NIOPK und ORI keinerlei Kopublikationen mit oder Inlinks aus Übergangsländern aufweisen, die übrigen Institute dagegen schon (vgl. Tabelle 69 und Tabelle 72). Die Kopublikations- und Inlinkbeziehungen zu Industrieländern ähneln denen zu anderen Ländern und Kontinenten (Tabelle 43): Im Verhältnis zur Anzahl der MitarbeiterInnen erhalten die drei Forschungsinstitute in den USA deutlich mehr Inlinks aus den TLDs anderer Industrieländer, als es deren Gewicht bei den Kopublikationen entspricht (berechnet aus Tabelle 17, Tabelle 69 und Tabelle 72).

Für die Kontakte zu Entwicklungsländern finden sich sehr geringe, nach der Standardisierung negative Korrelationen. Bezogen auf die personelle Institutsgröße resultieren sie daraus, dass MOI und ORI am häufigsten mit anderen Entwicklungsländern kopublizieren, während sie nur einen Bruchteil der Inlinks aus diesen Nationen erhalten, die die deutschen und US-amerikanischen Meeresforschungsinstitute auf sich ziehen.

<i>Kopublikationen im WoS vs. Inlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite/Publikation</i>
Entwicklungsländer	0,140	-0,261	-0,091
Industrieländer	0,833	0,503	0,236
Übergangsländer	0,913	0,871	0,742

Tabelle 44: Spearman's ρ für Maße der Sichtbarkeit im World Wide Web und der Orientierung im Web of Science, bezogen auf Ländergruppen nach Entwicklungsstand

Der Stellenwert der einzelnen Kontinente unter den rohen Inlinkdomains und Kopublikationen der Institute ähnelt sich mit Ausnahme von Afrika und Asien stark (Tabelle 45). Standardisiert man die Daten an der MitarbeiterInnenzahl, finden sich mittlere und hohe Spearman-Werte für Europa, Nord- und Südamerika. Bei Standardisierung an der Seiten- bzw.

Publikationszahl besteht ein ausgeprägter statistischer Zusammenhang nur zwischen nord-amerikanischen Inlinkdomains und Kopublikationen.

Die leichten negativen Korrelationen bei den Bezügen zu Entwicklungsländern aus Tabelle 44 zeigen sich in Tabelle 45 für Afrika. Sie resultieren daraus, dass ORI, KMFRI und IOW afrikanische Kopublikationen, aber keine Inlinks aufweisen, während BLOS, HBOI und MBARI Inlinks aus afrikanischen TLDs erhalten, ohne entsprechende Veröffentlichungen zu besitzen (vgl. Tabelle 70 und Tabelle 73). Auch für Asien und Südamerika, die ebenfalls zahlreiche Entwicklungsländer umfassen, gilt, dass die hohen Inlinkwerte der US-amerikanischen Institute sich nicht in ihren Koautorenschaftsprofilen widerspiegeln. Die in Afrika und Asien gelegenen Institute erhalten dagegen unabhängig von ihrer Kopublikationsaktivität geringe Inlinkzahlen pro MitarbeiterIn aus Afrika, Asien und Südamerika (berechnet aus Tabelle 17, Tabelle 70 und Tabelle 73).

<i>Kopublikationen im WoS vs. Inlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite/Publikation</i>
Afrika	-0,293	-0,294	-0,248
Asien	0,351	0,030	0,055
Europa	0,796	0,576	0,018
Nordamerika	0,869	0,745	0,559
Ozeanien	0,779	0,423	0,160
Südamerika	0,742	0,755	0,082

Tabelle 45: Spearmans ρ für Maße der Sichtbarkeit im World Wide Web und der Orientierung im Web of Science, bezogenen auf einzelne Kontinente

Sowohl aus Europa als auch aus Nordamerika erhalten BLOS, MBARI, IOW und IFM die meisten Inlinks pro Kopf. Gleichzeitig verfügen sie über die intensivsten Koautorenschaftsbeziehungen mit diesen Kontinenten. Aus Ozeanien erhalten die MitarbeiterInnen von MBARI, BLOS und HBOI besonders viele Inlinks, die von MOI, NIOPK und ORI dagegen gar keine. Die meisten Kopublikationen pro Kopf mit diesem Kontinent veröffentlichten dagegen IFM, MOI und MBARI, die wenigsten NIO, KMFRI und NIOPK, so dass sich kein systematischer Zusammenhang ergibt (berechnet aus Tabelle 17, Tabelle 70 und Tabelle 73).

5.2.2.5 Gibt es einen Zusammenhang zwischen der internationalen Sichtbarkeit und Orientierung im World Wide Web?

Aus Gründen der Vollständigkeit wird hier auch geprüft, inwiefern die internationale Sichtbarkeit und die Außenorientierung der Institute im World Wide Web korrespondieren.

Die rohen Werte der In- und Outlinkdomains des eigenen Landes, des eigenen Kontinents, anderer Länder und anderer Kontinente korrelieren stark (vgl. Tabelle 46). Bei einer Standardisierung an der Anzahl der MitarbeiterInnen schwächen sich die Zusammenhänge leicht ab. Es liegen generell die deutschen und US-amerikanischen Einrichtungen vor den übrigen. Zu den wenigen Abweichungen zählen die zahlreichen nationalen Outlinkdomains des ORI, die vielen internationalen Outlinkdomains des MOI und die außerhalb der eigenen nationalen TLDs generell niedrigen Werte des HBOI (berechnet aus Tabelle 17, Tabelle 65 und Tabelle 68).

<i>Inlinkdomains vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
eigenes Land	0,936	0,903	0,588
andere Länder	0,806	0,685	0,406
eigener Kontinent	0,694	0,756	-0,378
andere Kontinente	0,768	0,600	0,316

Tabelle 46: Spearman's ρ für grundlegende Maße der internationalen Orientierung und Sichtbarkeit im World Wide Web

Die Standardisierung an der Zahl der Webseiten bzw. Publikationen reduziert die Zusammenhänge stärker. Die mittlere Korrelation für die Verlinkungen mit der eigenen nationalen TLD ergibt sich u.a. daraus, dass das ORI über besonders viele In- und Outlinks mit südafrikanischer TLD verfügt, während NIO und KMFRI weder In- noch Outlinks mit ihrer eigenen nationalen TLD besitzen (vgl. Tabelle 65 und Tabelle 68). Eine Ursache für die leichte, negative Korrelation der Verlinkungen mit den TLDs des eigenen Kontinents resultiert daraus, dass das KMFRI und besonders das ORI im Verhältnis zu ihren sehr kleinen Websites viele Outlinks in afrikanische TLDs eingerichtet haben, aber von dort keine Inlinks erhalten (berechnet aus Tabelle 8, Tabelle 67 und Tabelle 70). Insbesondere für die Outlinks geht es auch hier um kleine absolute Werte.

Eine Aufschlüsselung der Links nach Ländergruppen ergibt ein gemischtes Bild (vgl. Tabelle 47). Die Korrelationen, die für Links von Übergangsländern gefunden wurden, sind auch hier nicht besonders aussagekräftig, da nur drei Institute über entsprechende Outlinks verfügen (vgl. Tabelle 66). Für Links aus und in Entwicklungsländer ergibt sich, unabhängig von der Standardisierung, kein statistischer Zusammenhang. Bezogen auf ihre Personalstärke weisen sämtliche Institute recht homogene Outlinkwerte auf niedrigem Niveau auf, während vor allem die US-amerikanischen, aber auch die deutschen Meeresforschungseinrichtungen deutlich mehr Inlinks aus Entwicklungsländern erhalten, als die Institute, die selbst in ebensolchen liegen (berechnet aus Tabelle 17, Tabelle 66 und Tabelle 69). Dagegen

bewegen sich die Institute aus den beiden Industrieländern überwiegend im Mittelfeld, wenn man die In- und Outlinkdomains aus Entwicklungsländern in Relation zur Websitegröße betrachtet. Hier gibt es mit NIO und HBOI allerdings zwei Institute, die keine Outlinks in Entwicklungsländer aufweisen, aber relativ viele Inlinks pro Webseite aus deren TLDs erhalten. Bei KMFRI tritt der umgekehrte Fall auf (berechnet aus Tabelle 8, Tabelle 66 und Tabelle 69).

<i>Inlinkdomains vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
Übergangsländer	0,742	0,623	0,439
Entwicklungsländer	0,448	0,301	0,135
Industrieländer	0,890	0,661	0,394

Tabelle 47: Spearmans ρ für Maße der Orientierung und Sichtbarkeit im World Wide Web, bezogen auf Ländergruppen nach Entwicklungsstand

Der mittlere Zusammenhang von personalbezogenen Linkwerten für Industrieländer resultiert aus dem starken Abschneiden der Institute aus den USA und Deutschland im Hinblick auf In- und Outlinks (berechnet aus Tabelle 17, Tabelle 66 und Tabelle 69). Die Standardisierung an der Größe der Website nivelliert die Outlinkwerte und lässt bei den Inlinkwerten die kleinen Websites aus den Entwicklungsländern wesentlich besser abschneiden (berechnet aus Tabelle 8, Tabelle 66 und Tabelle 69).

Tabelle 48 zeigt die Ergebnisse für die Korrelation von Inlink- und Outlinkdomains für die einzelnen Kontinente. Afrikanische In- und Outlinks sind selten und treten nur bei BLOS, IFM und MBARI gemeinsam auf, so dass keinerlei Zusammenhang festgestellt werden kann (vgl. Tabelle 67 und Tabelle 70). Outlinks nach Südamerika haben nur drei Institute eingerichtet, die gleichzeitig auch Inlinks erhalten. Die gefundene Korrelation ist daher wenig belastbar.

<i>Inlinkdomains vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
Afrika	0,059	-0,248	-0,352
Asien	0,667	0,668	-0,219
Europa	0,893	0,818	0,358
Nordamerika	0,817	0,796	0,180
Ozeanien	0,660	0,602	-0,255
Südamerika	0,803	0,755	0,232

Tabelle 48: Spearmans ρ für Maße der Orientierung und Sichtbarkeit im World Wide Web, bezogen auf einzelne Kontinente

Anderes gilt für die übrigen Kontinente: Generell erhalten die fünf Institute in Industrieländern mehr Inlinks pro MitarbeiterIn als die in Entwicklungsländern. Bei den personalbezogenen Outlinks dagegen schneidet das MOI wesentlich besser, MBARI und HBOI deutlich schlechter ab. Ähnlich ist die Situation bei den europäischen und ozeanischen In- und Outlinks. In Bezug auf Nordamerika ist die stärkste Auffälligkeit die hohe Zahl der Outlinks des IFM im Verhältnis zu den erhaltenen Inlinks (berechnet aus Tabelle 17, Tabelle 67 und Tabelle 70).

Nach der Standardisierung an der Seitenzahl zeigen sich für keinen der Kontinente deutliche Korrelationen von Inlink- und Outlinkdomains.

5.2.2.6 Nehmen die Kontinente und Regionen der Welt in den Koautorenschafts-, Outlink- und Inlinkprofilen der Institute einen ähnlichen Stellenwert ein?

Aus den Rohdaten (vgl. Tabelle 67, Tabelle 70 und Tabelle 73) lässt sich für jedes Institut ermitteln, welcher Kontinent unter den Inlink- und Outlinkdomains die größte Rolle spielt (Tabelle 49). Es wurden nur Institute berücksichtigt, die mindestens zehn internationale Publikationen, Inlink- oder Outlinkdomains aufweisen. Aus diesem Grund sind für KMFRI, MOI, NIOPK und NIO nicht alle Angaben vorhanden.

Während für jedes der Institute die meisten Inlinkdomains aus Europa kommen, gibt es bei den Outlinks und Publikationen etwas mehr Variation: Bei IFM und NIO richten sich die meisten Outlinks auf Nordamerika, und das HBOI hat ebensoviel mit nordamerikanischen wie mit europäischen Einrichtungen veröffentlicht. Diese Abweichungen erstrecken sich jedoch immer nur auf eine der Variablen, so dass sie nicht für eine generelle Ausrichtung der Institute auf Europa resp. Nordamerika sprechen.

<i>Institut</i>	<i>eigener Kontinent</i>	<i>Outlinkdomains</i>	<i>Inlinkdomains</i>	<i>Publikationen</i>
BLOS	Nordamerika	Europa	Europa	Europa
HBOI	Nordamerika	Europa	Europa	Europa und Nordamerika
IFM	Europa	Nordamerika	Europa	Europa
IOW	Europa	Europa	Europa	Europa
KMFRI	Afrika	---	Europa	Europa
MBARI	Nordamerika	Europa	Europa	Europa
MOI	Afrika	---	Europa	---
NIO	Asien	Nordamerika	Europa	Europa
NIOPK	Asien	---	Europa	---
ORI	Afrika	---	Europa	---

Tabelle 49: Am häufigsten auftretende Kontinente unter den Publikationen, Outlink- und Inlinkdomains der Institute

Um im Detail zu prüfen, inwieweit sich der Stellenwert der Weltregionen in den Publikations- und Linkprofilen der Meeresforschungsinstitute ähnelt, wurden für jedes Institut die Anzahl der Kopublikationen mit Einrichtungen in den Weltregionen bestimmt. Ebenso wurden die Inlink- und Outlinkdomains mit den länderspezifischen TLDs anderer Nationen gezählt. In diesen Vergleich wurden nur diejenigen Institute einbezogen, die über eine Mindestanzahl an internationalen Inlinks, Outlinks und Kopublikationen verfügten, um überhaupt ausgeprägte regionale Profile finden zu können und Zufallseffekte zu minimieren. Als Schwelle wurden ein Minimum von zehn In- und Outlinkdomains in TLDs anderer Länder sowie zehn internationale Kopublikationen definiert. Diese Werte erreichten BLOS, IFM, IOW, MBARI und NIO.

Für jedes Institut wurde der Zusammenhang der regionalen Verteilung von Kopublikationen und Outlinkdomains mit Spearmans Rangskorrelationskoeffizient bestimmt. Gleiches folgte für Kopublikationen und Inlinkdomains sowie Inlink- und Outlinkdomains. Alle drei Korrelationen ergaben für jedes Institut ρ -Werte, die größtenteils zwischen 0,7 und 0,8 liegen (Tabelle 50). Die Institute unterscheiden sich in der Stärke der Zusammenhänge:

<i>Institute</i>	<i>Publikationen vs. Outlinkdomains</i>	<i>Publikationen vs. Inlinkdomains</i>	<i>Outlinkdomains vs. Inlinkdomains</i>
BLOS	0,765	0,811	0,783
IFM	0,847	0,846	0,816
IOW	0,746	0,694	0,757
MBARI	0,905	0,850	0,806
NIO	0,665	0,743	0,675

Tabelle 50: Spearmans ρ für den Stellenwert der Weltregionen in den Publikations-, Inlink- und Outlinkprofilen

Die höchsten Werte erreichen IFM und MBARI, die auch insgesamt die größten Publikations- und Linkzahlen aufweisen. Keiner der drei berechneten Zusammenhänge erweist sich über die fünf Institute hinweg als besonders ausgeprägt: Bei IFM und MBARI hängt die Verteilung von Publikationen und Outlinkdomains auf die Weltregionen besonders stark zusammen ($\rho = 0,847$ resp. $\rho = 0,905$), bei BLOS und NIO sind Publikationen und Inlinkdomains ähnlich verteilt ($\rho = 0,811$ resp. $\rho = 0,743$) und bei IOW verteilen sich In- und Outlinkdomains am ähnlichsten auf die Regionen der Welt ($\rho = 0,757$).

5.2.3 Zusammenfassung und Diskussion

Die Ergebnisse des Vergleichs der web- und publikationsbasierten Institutsprofile werden in diesem Abschnitt thematisch zusammengefasst und diskutiert.

Präsenz im World Wide Web und im Web of Science

Die Meeresforschungseinrichtungen, die eine starke Präsenz im Web of Science aufweisen, sind auch besonders umfangreich im World Wide Web vertreten (vgl. 5.2.2.1). Dies gilt in absoluten Werten ebenso wie im Verhältnis zur Personalstärke der Institute. Dieser Befund ist konsistent mit anderen, zu Beginn des Teilkapitels 5.2 erwähnten Studien, die einen Zusammenhang von Forschungs- und Webproduktivität dokumentieren.

Anders als das erfolgreiche Publizieren in den Peer-Review-geprüften Fachzeitschriften des wissenschaftlichen Mainstreams, ist Produktivität im Web keine Frage von Exzellenz oder Originalität. Derzeit scheint es jedoch nicht üblich zu sein, geringeren wissenschaftlichen Erfolg mit der Ausweitung der Webpräsenz zu kompensieren. Da die größeren Internetauftritte ihren Umfang u.a. üppigen Datenbanken verdanken, die ein unmittelbares Ergebnis von Forschungstätigkeit sind, ist eine solche Kompensation vermutlich nicht ohne Weiteres möglich. Insofern kann eine gewisse Kausalität zwischen wissenschaftlicher und virtueller Produktivität angenommen werden. Zum anderen liegt auch ein Zusammenhang mit dem sozio-ökonomischen Arbeitsumfeld nahe: Forschungsinstitute, die mit stark begrenzten Ressourcen auskommen müssen und deren wissenschaftliche Leistung aufgrund dessen eingeschränkt ist, werden ihrem Internetauftritt nicht die höchste Priorität einräumen können. In Abschnitt 5.3 wird dieser Vermutung weiter nachgegangen.

Sichtbarkeit im World Wide Web und im Web of Science

Die rohen sowie die an der Institutgröße standardisierten Inlinkwerte hängen stark mit den entsprechenden Zitationswerten zusammen. Die leicht negative Korrelation zwischen den am wissenschaftlichen und virtuellen Output standardisierten Sichtbarkeitsindikatoren lässt sich auf Skaleneffekte beim WIF zurückführen, die bereits früher gefunden wurden (vgl. Thelwall/Smith 2002): Jeder Internetauftritt zieht ein Minimum an Inlinks auf sich, weil er grundlegende Funktionen erfüllt und zumindest einen Hinweis auf die Existenz der Forschungseinrichtung gibt. Jedoch steigt die Zahl der Inlinks meist nicht proportional zum Umfang der Website an, so dass kleinere Internetauftritte oft mehr Inlinks pro Seite erhalten als größere.

Institute, die im Verhältnis zu ihrer Größe viele Zitationen auf sich ziehen, erhalten auch viele Inlinks. Dieser Zusammenhang von wissenschaftlicher und virtueller Sichtbarkeit könnte eine

Konsequenz aus der Reputation sein, die sich nicht nur in den wissenschaftlichen Fachdiskursen abbildet, sondern auch auf das World Wide Web wirkt. Da sich die untersuchten Meeresforschungseinrichtungen auf die ganze Welt verteilen, könnten auch globale Disparitäten die Ursache sein. Auch hier wird der folgende Abschnitt 5.3 Aufschluss geben.

Nationale und intrakontinentale Orientierung und Sichtbarkeit

Die lokalen Publikations- und Linkprofile der Meeresforschungseinrichtungen zeigen starke Zusammenhänge: Das eigene Land und der eigene Kontinent haben in den Publikations-, Inlink- und Outlinkmustern der deutschen und US-amerikanischen Institute einen höheren Stellenwert als bei den Instituten anderer Länder. Dies gilt sowohl absolut gesehen als auch im Verhältnis zu ihrer Personalstärke; im Verhältnis zur Websitegröße und zur Gesamtpublikationszahl nur zum Teil. Angesichts der umfangreichen Forschungssysteme in Deutschland und den USA sowie deren Nachbarländern verwundert es nicht, dass die dortigen Institute sich im Hinblick auf Koautorenschaft und Outlinks eher an der eigenen Nation und deren Anrainern orientieren als die Einrichtungen in Afrika und Asien. Interessant ist, dass die nationalen und intrakontinentalen Inlinks mit den entsprechenden Outlink- und Koautorenschaftswerten korrelieren. Die Orientierung an der eigenen Nation und am eigenen Kontinent in den wissenschaftlichen Publikationen findet sich somit nicht nur in den Outlinkmustern wieder, sondern auch unter den Inlinks, auf die die Forschungsinstitute keinen Einfluss haben. Dieses Ergebnis deutet zum einen auf Ungleichheitseffekte hin, die in Abschnitt 5.3 näher untersucht und diskutiert werden. Zum anderen ist die besonders starke Korrelation nationaler In- und Outlinkdomains ein Hinweis auf mögliche Unterschiede in der Nutzung nationaler Top-Level-Domains: Vermutlich ist die Divergenz zwischen dem südafrikanischen ORI und dem indischen NIO in Teilen ein Artefakt, da die TLD „za“ wesentlich verbreiteter ist als die TLD „in“, die nicht einmal das NIO selbst nutzt. Da sich alle anderen Inlink- und Outlinkmaße aus den Links aus mehreren Ländern zusammensetzen, kommt dieser Effekt bei anderen Indikatoren nicht so deutlich zum Tragen. In der Konsequenz ist dieser Indikator für Vergleiche von Websites innerhalb eines Landes geeignet, nicht jedoch für internationale Vergleiche, die sich auf Länder mit heterogener Vergabepaxis erstrecken.

Internationale und interkontinentale Orientierung und Sichtbarkeit

Im Verhältnis zur Personalstärke hängen internationale Kopublikationen und Outlinks klar zusammen. Dies gilt besonders für die am Personalumfang standardisierten Daten. Dabei erzielen die beiden deutschen Institute besonders hohe Werte. Ein Grund dafür dürfte an ihrer Einbettung in die EU liegen, die inzwischen einen eng vernetzten, quasi-nationalen Raum in der Forschungslandschaft bildet (Leydesdorff 2000). Auch die interkontinentalen Outlink- und Publikationsdaten zeigen einen statistischen Zusammenhang.

Dies gilt jedoch nicht für die internationalen und interkontinentalen Inlinkdaten: Die an der Anzahl der MitarbeiterInnen standardisierten Inlinkwerte korrelieren nur schwach mit den internationalen Publikationen bzw. Outlinkdomains. Im Verhältnis zur Publikations- bzw. Webproduktivität zeigen sich keine entsprechenden Zusammenhänge. Der wesentliche Grund dafür liegt in der Dominanz der drei US-amerikanischen Institute bei den Inlinkdomains: Sie weisen in diesem Sample eine mittlere oder geringe internationale Orientierung in ihren Publikationen und Outlinks auf, besitzen aber mit Abstand die größte weltweite Sichtbarkeit im Web.

Der Stellenwert einzelner Ländergruppen

Unabhängig von der Standardisierung fanden sich deutliche Korrelationen zwischen den Kopublikationen mit Entwicklungsländern und den analog gruppierten Outlinkdomains. Hier weisen die drei afrikanischen Institute hohe Publikations- und Outlinkwerte pro MitarbeiterIn auf, die beiden asiatischen besonders geringe. Auch Publikations- und Outlinkkontakte zu Industrieländern hängen zusammen. Die zum Teil gefundenen Korrelationen von Koautorenschaft, Inlink- und Outlinkdomains für die Übergangsländer sind allerdings aufgrund geringer Häufigkeiten nicht belastbar und werden daher an dieser Stelle nicht weiter berücksichtigt.

Die an der Größe der Webseite standardisierten Inlinkwerte zeigen weder für Entwicklungs- noch für Industrieländer einen statistischen Zusammenhang mit Publikationen oder Outlinkdaten. Allerdings korrelieren im Verhältnis zum Personalumfang der untersuchten Meeresforschungseinrichtungen die Inlink- und Outlinkdomains, die Industrieländern zuzuordnen sind. Hier weisen die US-amerikanischen und deutschen Institute mehr Outlink- und Inlinkdomains pro MitarbeiterIn auf als die übrigen Einrichtungen. Die Standardisierung an der Websitegröße nivelliert diese Unterschiede.

Bei den untersuchten Meeresforschungseinrichtungen nehmen Kontakte zu Industrieländern (ebenso wie zu Entwicklungsländern) einen vergleichbaren Stellenwert in den Kopublikationen und in den Zielen der Outlinkdomains ein. Da die US-amerikanischen und zum Teil auch die deutschen Institute besonders viele Inlinks aus Industrie- wie Entwicklungsländern erhalten, ähneln die Inlinkprofile im gesamten Sample weder den Outlink- noch den Publikationsprofilen.

Der Stellenwert einzelner Kontinente

Für Afrika zeigt sich, dass die drei dort gelegenen Institute im Verhältnis zur MitarbeiterInnenzahl die intensivsten Kopublikationsbeziehungen auf dem eigenen Kontinent pflegen, jedoch nur in sehr geringem Umfang Outlinks in andere afrikanische Länder einrichten. Nur eins von ihnen ist auch das Ziel von Inlinks mit afrikanischen TLDs – hier sind die Institute in

den USA sowie das IFM die populärsten Ziele. Dass die Linkstrukturen den Publikationsstrukturen nicht entsprechen kann zum Teil daraus resultieren, dass die drei untersuchten afrikanischen Websites sehr klein sind und insgesamt nur wenige Outlinks enthalten. Andere afrikanische Websites wählen möglicherweise eher die besonders umfang- und ressourcenreichen Internetauftritte nordamerikanischer oder europäischer Forschungseinrichtungen als Ziel ihrer Links.

Für Asien zeigt sich ein ähnliches Bild: Das indische NIO unterhält pro Kopf besonders viele Kopublikationskontakte innerhalb Asiens, besitzt jedoch kaum asiatische Out- und Inlinks. Das in Pakistan gelegene NIOPK besitzt weder asiatische Kopublikationen noch Outlinks, jedoch in geringem Umfang Inlinks aus Asien. Dagegen ist das deutsche IFM in allen drei Kategorien stark. Ansonsten entsprechen sich Publikations- und Linkmuster in Asien kaum. Ein Grund für die geringe intrakontinentale Vernetzung der beiden asiatischen Institute könnte in der sprachlichen, politischen und kulturellen Heterogenität Asiens bestehen, die intensivere Kooperationen zwischen NIO und NIOPK oder mit wissenschaftlich aktiven Nachbarländern wie China oder Japan erschwert. Auch die Vielzahl der verwendeten Schriften behindert möglicherweise Verlinkungen innerhalb Asiens.

Für Europa und Nordamerika zeigen sich starke Korrelationen zwischen Publikations-, Inlink- und Outlinkmustern, die daraus resultieren, dass in jeder Kategorie die deutschen und US-amerikanischen Meeresforschungseinrichtungen besonders präsent sind.

In Südamerika ist keins der untersuchten Institute gelegen. Die hohen Korrelationen zwischen den Publikations- und Linkwerten Südamerikas resultieren daraus, dass lediglich vier US-amerikanische und deutsche Institute Publikationen mit südamerikanischen KoautorInnen im Erhebungszeitraum veröffentlicht haben, die gleichzeitig auch durch Out- und Inlinks mit Südamerika verbunden sind.

Für Ozeanien zeigen sich dagegen kaum systematische Übereinstimmungen zwischen Kopublikationen, Inlink- und Outlinkdomains, wenngleich sich auch hier eine gewisse Prävalenz von Instituten aus den USA und Deutschland zeigt. Das MOI erhält keine Inlinks von diesem Kontinent, obwohl es den zweithöchsten Pro-Kopf-Wert an Kopublikationen mit Ozeanien aufweist. Andererseits besitzt das BLOS hohe Inlink- und Outlinkwerte für Ozeanien, wenngleich es keine Publikationen mit dortigen KoautorInnen herausgebracht hat.

Wenngleich keines der untersuchten Institute in Ozeanien liegt, findet sich hier, anders als in Südamerika, keine durchgängige Dominanz der deutschen und US-amerikanischen Meeresforschungsinstitute: Mit Ausnahme des NIOPK haben alle Institute Kopublikationen mit ozeanischen Einrichtungen sowie In- oder Outlinkdomains. Sowohl Ozeanien als auch Südamerika verfügen über mehrere Meeresforschungseinrichtungen. Dass Ozeanien offensichtlich für die afrikanischen und asiatischen Institute etwas mehr Relevanz als Südamerika

besitzt, kann daran liegen, dass dort mit Australien und Neuseeland zwei englischsprachige Nationen mit gut ausgebauten Forschungssystemen zu finden sind.

Insgesamt zeigt sich, dass die asiatischen und afrikanischen Institute mit Ausnahme des NIOPK ihre Kopublikationskontakte auf den eigenen Kontinent sowie Europa und Nordamerika konzentrieren. Der Stellenwert dieser Kontinente spiegelt sich jedoch nicht in den Linkwerten wider. Dagegen besitzen die meisten Institute in Deutschland und den USA Kopublikationsbeziehungen auf vielen Kontinenten, die auch in ihren Outlink- und Inlinkprofilen präsent sind. Insbesondere die starken Pro-Kopf-Inlinkwerte für die Institute in den USA und Deutschland weisen auf eine globale Schiefelage hin, was die virtuelle Sichtbarkeit der Institute auf den einzelnen Kontinenten betrifft. Dies wird Gegenstand des folgenden Teilkapitels 5.3 sein.

Schlussfolgerungen aus den Standardisierungsweisen

Es hat sich gezeigt, dass die Standardisierung am Publikationsumfang bzw. an der Websitegröße fast alle sinnvollen Beziehungen zwischen Publikations- und Linkdaten nivelliert. Die wenigen Ausnahmen beruhen größtenteils auf geringen Fallzahlen (wie bei den gefundenen Korrelationen für die Übergangsländer oder Südamerika). Eine weitere Ausnahme stellt der Zusammenhang zwischen nationalen Inlink- und Outlinkdomains dar, der vermutlich auch auf Unterschiede in der Handhabung der nationalen Top-Level-Domains zurückgeht. Diese allgemeine Nivellierung ist ein Hinweis darauf, dass die relative Bedeutung einzelner Ländergruppen oder Kontinente im Verhältnis zum Publikations- oder Website-Output über alle Institute hinweg recht ähnlich ist. Räumliche oder strukturelle Schwerpunkte lassen sich besser durch die Standardisierung an der personellen Institutsgröße identifizieren.

Eine Alternative, die bereits in anderen Studien genutzt wurde (z.B. Li et al. 2005b), besteht darin, bestimmte Linkgruppen als Anteil an allen In- bzw. Outlinks auszudrücken. Dies weicht zwar von dem webometrischen Standard des Web Impact Factors ab, entspricht aber unmittelbar den hier verwendeten bibliometrischen Maßen. Ob diese Vorgehensweise ein klareres Bild von dem relativen Stellenwert der Links liefern kann, kann im Rahmen dieser Dissertation nicht geprüft werden.

Regionale Profile

Um eine Einschätzung der weltweiten Streuung von Inlinks, Outlinks und Kopublikationen zu bekommen, wurde ihre Verteilung auf die Weltregionen untersucht. Die Institute BLOS, IFM, IOW, MBARI und NIO wiesen jeweils mindestens zehn internationale Kopublikationen, Inlink- und Outlinkdomains auf. Für sie wurde geprüft, inwiefern sich jeweils zwei dieser Variablen in ähnlicher Weise auf die Regionen der Welt verteilen. Es wurden durchweg deutliche

Korrelationen gefunden. Über die Institute hinweg ließen sich keine systematischen Unterschiede in der Stärke des Zusammenhangs von Publikationen und Outlinkdomains, Publikationen und Inlinkdomains oder Outlink- und Inlinkdomains feststellen.

Insgesamt erhalten Institute, die mit AutorInnen aus bestimmten Regionen veröffentlichen, auch Inlinks aus diesen Regionen und richten selbst Outlinks dorthin ein. Ein Grund für diesen ausgeprägten Zusammenhang liegt in der Dominanz Europas und Nordamerikas, die bei Kopublikationen wie bei In- und Outlinks zu diagnostizieren ist. Hinzu kommt, dass diese Regionen auch die Heimatregionen von vier der fünf untersuchten Institute sind. Einzige Ausnahme ist hier das indische NIO. Es weist durchgehend hohe Korrelationswerte auf, die jedoch unter denen der übrigen Institute liegen. Zwei mögliche Ursachen liegen nahe:

- Es könnte sich um einen Größeneffekt handeln, der aus der geringeren Gesamtzahl an Publikationen und Links resultiert.
- Es könnte sich um einen geographischen Effekt handeln, der denjenigen Instituten besonders hohe Korrelationswerte beschert, deren Heimatregion von hochentwickelten, wissenschaftlich wie virtuell aktiven Regionen umgeben ist. Bei ihnen resultieren Kontakte zu Nachbarregionen vermutlich gleichzeitig in hohen Link- und Kopublikationswerten, während letztere in einem heterogeneren Umfeld stärker streuen.

Es stellt sich somit erneut die Frage, ob sich hier Unterschiede zwischen Instituten in Industrie- und Entwicklungsländern zeigen. Mit KMFRI, MOI, NIOPK und ORI erreichen vier der fünf Institute in Entwicklungsländern den Schwellenwert von mindestens zehn internationalen Kopublikationen, Inlink- und Outlinkdomains nicht. Entsprechend kann diese Frage hier nicht geklärt werden und weitere empirische Prüfung ist nötig.

5.3 Hängen Präsenz und Vernetzung wissenschaftlicher Einrichtungen im World Wide Web bzw. im Web of Science mit dem Entwicklungsstand ihrer Sitzländer zusammen?

Ziel dieses Abschnitts ist es zu prüfen, inwiefern der Umfang und die Vernetzung der untersuchten Internetauftritte mit dem Entwicklungsstand der Staaten zusammenhängt, in denen sie angesiedelt sind. Gleiches wird auch für die Publikationen der Meeresforschungsinstitute aus dem Web of Science untersucht, um herauszufinden, ob webometrische oder bibliometrische Daten den sozio-ökonomischen Forschungskontext in unterschiedlichem Ausmaß widerspiegeln.

5.3.1 Verwendete Daten

Es werden dieselben Web- und Publikationsdaten aus den Abschnitten 4.3.1 bis 4.3.4 herangezogen, die bereits im vorausgehenden Teilkapitel 5.2.1 spezifiziert wurden. Ebenso wird die Anzahl der wissenschaftlichen MitarbeiterInnen zur Standardisierung (Tabelle 17) und die UN-Länderklassifikation zur Bestimmung der geographischen und entwicklungspolitischen Gruppierung der vertretenen Staaten genutzt (4.3.5).

Als Maß für den Entwicklungsstand der Sitzländer kommt der Human Development Index (HDI) hinzu (Tabelle 18).

Zusätzlich werden auch die Literaturlisten der Institute aus dem Jahr 2006 herangezogen (vgl. 4.3.4), um die Präsenz der Institute im Web of Science mit ihrem gesamten Publikationsoutput abzugleichen.

Standardisierung

Im Hinblick auf die zwei Standardisierungsoptionen ist hervorzuheben, dass der HDI für das vorliegende Sample mit der Seitenzahl der Institute korreliert (Tabelle 51). Dies gilt jedoch nicht für den HDI und die Anzahl der Publikationen (vgl. Tabelle 61) oder die Anzahl der MitarbeiterInnen ($\rho = 0,092$).

Auswahl des Korrelationskoeffizienten

Aus den in Abschnitt 5.2.1 genannten Gründen wird auch in diesem Teil der Auswertung Spearmans Rangkorrelationskoeffizient verwendet und auf Signifikanztests verzichtet.

5.3.2 Ergebnisse

Dieses Teilkapitel ist ähnlich zu 5.2 aufgebaut. Auch hier wird die Leitfrage unterteilt, um unterschiedlichen Aspekten Rechnung zu tragen. Zunächst wird untersucht, inwiefern die Präsenz der Meeresforschungsinstitute im World Wide Web mit dem Human Development Index zusammenhängt (5.3.2.1). Anschließend wird der Zusammenhang des HDI mit der internationalen Orientierung (5.3.2.2) und der internationalen Sichtbarkeit (5.3.2.3) der Institutswebsites untersucht. Gleiches folgt für die Präsenz und Sichtbarkeit (5.3.2.4) sowie die internationale Orientierung (5.3.2.5) der Institute im Web of Science.

5.3.2.1 Gibt es einen Zusammenhang zwischen dem Human Development Index und der Präsenz der Meeresforschungsinstitute im World Wide Web?

Die Größe der Webpräsenzen der untersuchten Meeresforschungseinrichtungen korreliert, unabhängig von Standardisierung deutlich mit dem Human Development Index (Tabelle 51). Die USA und Deutschland weisen die höchsten HDI-Werte unter den im Sample vertretenen Nationen auf. Die dort angesiedelten Institute besitzen größere Websites als die meisten Einrichtungen in den übrigen Nationen. Sie stellen außerdem mehr andere Dateien (wie pdf, Word, Excel etc.) ins Netz, und zwar sowohl im Hinblick auf die Rohdaten als auch bei Standardisierung an der Anzahl der Webseiten oder MitarbeiterInnen.

<i>HDI vs.</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
Anzahl der Webseiten	0,573	0,714	---
Anzahl anderer Dateien	0,622	0,788	0,585

Tabelle 51: Spearman's ρ für HDI und Maße der Websitegröße

Bezogen auf die Personalstärke besitzen MBARI, IOW und BLOS (53,6, 44,5 und 26,3 Seiten pro MitarbeiterIn) die meisten Webseiten, KMFRI, ORI und NIOPK die wenigsten (0,9, 0,7 und 0,3 Seiten pro MitarbeiterIn). Lediglich das NIO hat wesentlich mehr Webseiten und andere Dateien, als angesichts des indischen HDI-Werts zu erwarten wäre (16,5 Seiten pro MitarbeiterIn) (alle Werte berechnet aus Tabelle 8 und Tabelle 17).

Bei der Standardisierung an der Seitenzahl schwächt sich der Zusammenhang leicht ab: Die höchste Anzahl von Dateien pro Seite enthalten die kleinste und die größte Website (ORI und MBARI), die geringste ebenfalls eine der kleinsten und eine der größten (KMFRI und IFM) (berechnet aus Tabelle 8 und Tabelle 18). Somit mischen sich hier Institute in Ländern mit eher hohem und eher niedrigem HDI-Wert.

5.3.2.2 Gibt es einen Zusammenhang zwischen dem Human Development Index und der internationalen Orientierung der Institute im World Wide Web?

Auch die Anzahl der Outlinkdomains hängt mit dem HDI zusammen (vgl. Tabelle 52). Dies gilt in noch stärkerem Umfang für die an der personellen Institutsgröße standardisierten Werte, bei denen die US-amerikanischen und deutschen Einrichtungen die höchsten Werte aufweisen (berechnet aus Tabelle 17 und Tabelle 65). Je höher der HDI eines Landes desto größer ist – absolut und pro MitarbeiterIn – die Anzahl der unterschiedlichen Domains, auf die die dort gelegenen Forschungseinrichtungen in ihren Websites verweisen. Bezogen auf die Websitegröße zeigen sich jedoch keine Korrelationen zwischen HDI und Outlinkwerten. Forschungseinrichtungen in Entwicklungsländern weisen sowohl die niedrigsten als auch die höchsten Outlinkwerte pro Webseite auf (berechnet aus Tabelle 8 und Tabelle 65).

<i>HDI vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
alle Outlinkdomains	0,800	0,899	-0,214
ohne Länderhinweis	0,775	0,960	0,025
mit Länderhinweis	0,769	0,800	0,135

Tabelle 52: Spearman's ρ für HDI und Maße der allgemeinen Außenorientierung im World Wide Web

Outlinkdomains, deren TLD für ein bestimmtes Land steht, können differenzierter betrachtet werden. Für jede Institutswebsite können speziell die Verweise auf andere Seiten mit der TLD des eigenen Landes oder anderer Staaten gezählt werden. Ebenso lassen sich auch Outlinkdomains mit den TLDs zusammenfassen, die die Länder des eigenen Kontinents sowie anderer Kontinente repräsentieren.

Hier korrelieren alle Werte bei Standardisierung am Personalumfang mit dem HDI, nicht jedoch bei Standardisierung am Umfang der Website (Tabelle 53). Mit Ausnahme des HBOI weisen die Institute in den USA und Deutschland pro MitarbeiterIn die höchste Anzahl von Outlinkdomains mit TLDs des eigenen Kontinents sowie anderer Länder und Kontinente auf. Schlusslicht ist jeweils das NIOPK, dessen Website nur eine einzige Outlinkdomain enthält, die in die eigene nationale TLD „.pk“ verweist (berechnet aus Tabelle 17 und Tabelle 65).

Bezogen auf die Anzahl der MitarbeiterInnen, weisen die Outlinkdomains in die eigenen nationalen TLDs der Institute einen starken Zusammenhang mit dem HDI auf. Bei der Interpretation ist zu berücksichtigen, dass sich nationale Unterschiede in der Praxis der Domainvergabe hier besonders stark auswirken und der Zusammenhang zumindest teilweise darauf zurückzuführen ist. Dies ist vermutlich auch der Grund dafür, dass dieser

Wert als einziger eine mittlere Korrelation mit dem HDI zeigt, wenn er an der Zahl der Webseiten standardisiert wird.

<i>HDI vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
in die eigenen nationalen TLDs	0,855	0,819	0,517
in andere nationale TLDs	0,628	0,665	-0,080
in TLDs des eigenen Kontinents	0,531	0,599	-0,019
in TLDs anderer Kontinente	0,659	0,628	-0,111

Tabelle 53: Spearmans ρ für HDI und Maße der internationalen Orientierung im World Wide Web

Die Anzahl der Outlinkdomains mit TLDs von Entwicklungsländern korreliert, unabhängig von der Standardisierung, nicht mit dem HDI (Tabelle 54). Dies gilt auch für die Übergangsländer, deren Top-Level-Domains allerdings nur bei drei Instituten (IFM, IOW und MBARI; vgl. Tabelle 66) das Ziel von Outlinks sind, so dass keine Schlussfolgerungen möglich sind.

<i>HDI vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
in TLDs von Entwicklungsländern	0,500	0,442	-0,131
in TLDs von Industrieländern	0,638	0,622	0,000
in TLDs von Übergangsländern	0,399	0,379	0,379

Tabelle 54: Spearmans ρ für HDI und Maße der Orientierung im World Wide Web an Ländergruppen nach Entwicklungsstand

Für die Industrieländer ähneln die Werte denen aus Tabelle 53: Im Verhältnis zur Personalstärke der Institute korrelieren diese Outlinkdomains deutlich mit dem HDI, nicht jedoch bei im Verhältnis zur Websitegröße. Erneut besitzen alle deutschen und US-amerikanischen Einrichtungen mit Ausnahme des HBOI pro MitarbeiterIn mehr Outlinkdomains mit TLDs von Industrieländern als die Meeresforschungsinstitute in Entwicklungsländern (berechnet aus Tabelle 17 und Tabelle 66).

Gruppiert man die TLDs der Outlinkdomains nach einzelnen Kontinenten, korreliert die Anzahl der Outlinkdomains bei Standardisierung an der Seitenzahl nur sehr schwach mit dem HDI (Tabelle 55). Eine Ausnahme ist Südamerika, in dessen TLDs allerdings nur drei Institute Outlinks gesetzt haben (BLOS, IFM, MBARI; vgl. Tabelle 67), so dass dieses Ergebnis nicht belastbar ist. Bei Standardisierung am Personalumfang ergeben sich mittlere bis hohe Werte für Asien, Europa, Nordamerika, Ozeanien und Südamerika, nicht jedoch für Afrika.

<i>HDI vs. Outlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
in die TLDs Afrikas	0,236	0,152	-0,089
in die TLDs Asiens	0,573	0,546	0,476
in die TLDs Europas	0,759	0,769	-0,068
in die TLDs Nordamerikas	0,533	0,599	0,235
in die TLDs Ozeaniens	0,681	0,748	0,188
in die TLDs Südamerikas	0,568	0,644	0,606

Tabelle 55: Spearman's ρ für HDI und Maße der Orientierung im World Wide Web nach einzelnen Kontinenten

Im Verhältnis zur Personalstärke besitzen vier deutsche und US-amerikanische Institute die meisten Outlinkdomains mit europäischen TLDs. Das HBOI weist geringere Werte auf als das fünftplatzierte MOI. Ähnlich stellt sich die Situation bei den Outlinks in nordamerikanische TLDs dar: Während die beiden deutschen Institute sowie BLOS und MBARI pro MitarbeiterIn die meisten Outlinkdomains mit diesen TLDs eingerichtet haben, folgt das HBOI erst nach MOI, NIO und ORI (berechnet aus Tabelle 17 und Tabelle 67).

Auch Ozeanien ist vorwiegend das Ziel von Outlinks aus den Websites von IFM, MBARI, BLOS und IOW. Die Internetauftritte aller übrigen Einrichtungen enthalten nur eine oder keine entsprechende Outlinkdomain (vgl. Tabelle 67).

HBOI, MOI, NIO und NIOPK besitzen keinerlei Outlinks mit afrikanischen TLDs, während das südafrikanische ORI die meisten Outlinks pro MitarbeiterIn in andere afrikanische Länder besitzt. Die Websites von HBOI, KMFRI, NIOPK und ORI enthalten keine Outlinks in die TLDs Asiens. Von allen übrigen Instituten weist das indische NIO die niedrigsten Outlinkwerte in asiatische TLDs auf, das afrikanische MOI die dritthöchsten (berechnet aus Tabelle 17 und Tabelle 67). Während sich für Afrika eine gewisse regionale Vernetzung bei ORI und KMFRI, nicht aber MOI zeigt, verlinken sich die beiden asiatischen Institute NIOPK und NIO nicht oder nur schwach mit anderen Ländern des eigenen Kontinents.

Insgesamt unterscheiden sich die ausgewählten Meeresforschungsinstitute in Ländern von unterschiedlichem Entwicklungsstand kaum, wenn man ihre Outlinkwerte in Relation zur Größe ihrer Websites setzt. Im Verhältnis zur personellen Größe der Einrichtungen divergieren die Outlinkmuster dagegen: Nur für Outlinkdomains mit den Top-Level-Domains von Entwicklungsländern und Übergangsländern sowie von Afrika und Asien finden sich keine Korrelationen mit dem HDI.

5.3.2.3 Gibt es einen Zusammenhang zwischen dem Human Development Index und der internationalen Sichtbarkeit der Institute im World Wide Web?

Ähnlich wie die Outlinkdomains korrelieren auch die Inlinkdomains der untersuchten Institute – mit und ohne Länderhinweis – stark mit dem HDI, wenn sie auf den Personalumfang bezogen werden (Tabelle 56). Bei den an der Seitenzahl standardisierten Inlinkdomains tritt dieser Zusammenhang nicht auf.

Pro MitarbeiterIn erhalten die drei US-amerikanischen Institute die meisten Inlinkdomains vor den beiden deutschen Einrichtungen. Es folgen mit deutlich niedrigeren Werten ORI, NIO, MOI, NIOPK und KMFRI (berechnet aus Tabelle 17 und Tabelle 68). Für Inlinkdomains ohne bzw. mit Länderhinweis stellt sich die Situation gleich bzw. ähnlich dar. Es ist zu berücksichtigen, dass Yahoo jeweils nur tausend Treffer pro Abfrage angezeigt hat und dass das IFM diese Schwelle mehrfach deutlich überschritten hat (vgl. Tabelle 11), so dass es de facto vermutlich höher platziert ist. Der Einfluss auf die gefundenen Zusammenhänge ist jedoch als gering einzuschätzen.

<i>HDI vs. Inlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
alle Inlinkdomains	0,837	0,948	-0,068
ohne Länderhinweis	0,769	0,948	-0,160
mit Länderhinweis	0,695	0,880	-0,117

Tabelle 56: Spearman's ρ für HDI und Maße der allgemeinen Sichtbarkeit im World Wide Web

Auch die Anzahl der Inlinkdomains mit Top-Level-Domains des eigenen Landes und des eigenen Kontinents korreliert – bei Standardisierung am Personalbestand – sehr stark mit dem HDI (Tabelle 57). Wie bei den Outlinkdomains weisen die nationalen Inlinkdomains einen hohen Zusammenhang auf. Dies stützt die Vermutung, dass die unterschiedliche Verbreitung der nationalen TLDs diesen Wert beeinflusst. Erneut liegen die Institute in den USA und Deutschland deutlich vorne. Von den übrigen Ländern weist das südafrikanische ORI pro MitarbeiterIn die höchste Zahl von Inlinkdomains auf (berechnet aus Tabelle 17 und Tabelle 68). Das KMFRI in Kenia erhält keinerlei Inlinks aus der TLD „.ke“, das MOI nur einen einzigen aus der nationalen mauritischen TLD „.mu“ (vgl. Tabelle 68).

Für den HDI und die Inlinkdomains aus dem eigenen Kontinent ergeben sich nur geringe Spearman-Werte (Tabelle 57). Sie resultieren daraus, dass die deutschen Institute hier deutlich vor den US-amerikanischen liegen und dann die beiden asiatischen Einrichtungen den drei afrikanischen vorausgehen (berechnet aus Tabelle 18 und Tabelle 69), obwohl die USA einen höheren HDI-Wert als Deutschland aufweisen und Mauritius und Südafrika einen

höheren HDI-Wert als Indien und Pakistan besitzen (vgl. Tabelle 17). Hier wirken sich erneut die starke europäische Vernetzung, die geringe Anzahl der nordamerikanischen Länder, sowie die allgemeine Webnutzung aus, die in Asien weiter vorangeschritten ist als in Afrika. Die drei afrikanischen Meeresforschungseinrichtungen erhalten keine Inlinks (KMFRI, ORI) oder nur einen einzigen Inlink (MOI) aus den Top-Level-Domains Afrikas.

<i>HDI vs. Inlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
aus den eigenen nationalen TLDs	0,735	0,868	0,277
aus anderen nationalen TLDs	0,726	0,935	-0,455
aus TLDs des eigenen Kontinents	0,502	0,685	0,161
aus TLDs anderer Kontinente	0,741	0,874	-0,432

Tabelle 57: Spearman's ρ für HDI und Maße der internationalen Sichtbarkeit im World Wide Web

Ein besonders ausgeprägter Zusammenhang mit dem HDI zeigt sich für die Inlinkdomains aus den nationalen TLDs anderer Länder und Kontinente (Tabelle 57). Hier unterscheiden sich die Inlink- und Outlinkmuster deutlich. Im Verhältnis zum wissenschaftlichen Personal erhalten die drei US-amerikanischen Institute bei Weitem die meisten Inlinks aus anderen nationalen Top-Level-Domains, gefolgt von den deutschen Einrichtungen sowie dem indischen NIO und dem südafrikanischen ORI, während MOI, NIOPK und KMFRI nur geringe Werte aufweisen (berechnet aus Tabelle 17 und Tabelle 68). Für die Inlinkdomains aus anderen Kontinenten ist die Situation ähnlich, mit einer deutlichen Ausnahme: Das deutsche IOW erhält deutlich weniger Inlinkdomains pro MitarbeiterIn aus anderen Kontinenten als NIO und ORI, die beide in Ländern mit geringerem HDI liegen (berechnet aus Tabelle 17 und Tabelle 69). Dies korrespondiert mit der regionalen Ausrichtung des IOW auf den Ostseeraum.

Die Standardisierung an der Websitegröße nivelliert allerdings auch hier sämtliche gefundenen Zusammenhänge. Die mittleren negativen Korrelationskoeffizienten für die internationalen und interkontinentalen Inlinkdomains resultieren aus Skaleneffekten: Hier erhält jeweils das ORI die höchste Anzahl der Inlinkdomains im Verhältnis zu seiner sehr kleinen Website, gefolgt von den übrigen Instituten mit geringer Webpräsenz (NIOPK, MOI, KMFRI) (berechnet aus Tabelle 8, Tabelle 68 und Tabelle 69).

Als Nächstes werden – jeweils separat – die Inlinkdomains mit TLDs von Entwicklungs-, Industrie- und Übergangsländern ausgewertet, die auf die Meeresforschungseinrichtungen verweisen. Inlinks aus der TLD des eigenen Landes werden hier nicht mitgezählt. Für alle drei Ländergruppen weist die Anzahl der Inlinkdomains bei Standardisierung an der Personalstärke sehr hohe Korrelationen mit dem HDI auf (Tabelle 58). Wie bei allen internationa-

len Inlinkdomains (s.o.) weisen die Institute in den USA sowie Deutschland pro MitarbeiterIn deutlich mehr Inlinkdomains aus Industrie-, Entwicklungs- und Übergangsländern auf als die Institute in den übrigen Ländern.

Bei Standardisierung an der Websitegröße zeigt sich ein heterogeneres Bild: Für die Inlinkdomains aus Entwicklungsländern ergibt sich kein Zusammenhang mit dem HDI. Hier nehmen die US-amerikanischen und deutschen Institute eine mittlere Position ein, während die in Schwellen- und Entwicklungsländern gelegenen Institute besonders hohe (ORI, NIOPK, MOI) und niedrige Werte (NIO, KMFRI) aufweisen (berechnet aus Tabelle 17 und Tabelle 69). Die seitenbezogene Anzahl der Inlinkdomains aus Industrieländern korreliert leicht negativ mit dem HDI. Dies erklärt sich erneut durch Skaleneffekte, da auch hier die vier Institute mit den kleinsten Websites, die alle in Ländern mit geringem HDI angesiedelt sind, die meisten Inlinks pro Seite erhalten. Für die Übergangsländer findet sich ein ausgeprägter positiver Zusammenhang, der daraus resultiert, dass die drei afrikanischen Institute sowie das NIOPK in Pakistan keinerlei Inlinks aus den Staaten des ehemaligen Ostblocks erhalten (berechnet aus Tabelle 8 und Tabelle 69).

<i>HDI vs. Inlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
aus TLDs von Entwicklungsländern	0,747	0,923	0,018
aus TLDs von Industrieländern	0,609	0,948	-0,468
aus TLDs von Übergangsländern	0,565	0,826	0,705

Tabelle 58: Spearman's ρ für HDI und Maße der Sichtbarkeit im World Wide Web bei Ländergruppen nach Entwicklungsstand

In einem weiteren Schritt lassen sich die Inlinkdomains, die auf die Meeresforschungseinrichtungen verweisen, anhand ihrer TLDs nach einzelnen Kontinenten gruppieren (Tabelle 59). Inlinks aus der TLD des eigenen Landes werden dabei ausgeschlossen. Es zeigen sich unterschiedlich starke Korrelationen mit dem HDI, die zum Teil auf Skaleneffekte und geringe Fallzahlen zurückgehen. Aus afrikanischen TLDs – ohne die des eigenen Landes – erhalten vier Institute keinerlei Inlinks (IOW, KMFRI, NIOPK, ORI), zwei erhalten Inlinks aus einer einzigen Domain (MOI und NIO). Auch bei den übrigen Instituten bewegen sich die rohen Zahlen der afrikanischen Inlinkdomains im einstelligen Bereich (Tabelle 70). Die gefundenen Korrelationen für Afrika beruhen somit – unabhängig von der Standardisierung – auf geringen Inlinkwerten.

Ähnlich ist die Situation für Inlinkdomains aus Ozeanien. Aus dessen TLDs erhalten MOI, NIOPK und KMFRI keinerlei Inlinks, das ORI einen einzigen (Tabelle 70). Bezogen auf die personelle Stärke der Meeresforschungseinrichtungen erhalten auch hier die deutschen und

US-amerikanischen Institute, gefolgt von dem indischen NIO, die meisten Inlinkdomains. Dies führt zu einer starken Übereinstimmung mit dem HDI und entsprechend auch zu einem hohen Korrelationskoeffizienten. Bei der Standardisierung fällt er geringer aus, da hier das ORI durch den singulären Inlink aus Ozeanien bezogen auf eine minimale Websitegröße den höchsten Wert erreicht.

Bei den Inlinkdomains aus südamerikanischen TLDs zeigen sich ebenfalls Effekte geringer Fallzahlen: MOI und KMFRI erhalten keine Inlinks von diesem Kontinent, ORI und NIOPK jedoch jeweils einen (Tabelle 70). Dies führt dazu, dass sie bei der Standardisierung an der Seitenzahl die niedrigsten bzw. höchsten Werte aufweisen, so dass sich keine Korrelation findet. Bei Standardisierung am Personalumfang liegen erneut MOI und KMFRI vorn, während ORI und NIOPK eine mittlere Position einnehmen, so dass die Korrelation mit dem HDI für Südamerika geringer ausfällt als für die restlichen Kontinente.

Dagegen erhalten alle untersuchten Institute Inlinks aus den Top-Level-Domains Asiens, Europas und Nordamerikas (Tabelle 70). Hier korrelieren die an der Anzahl der MitarbeiterInnen standardisierten Inlinkwerte sehr stark positiv mit dem HDI, während sich für die an der Seitenzahl standardisierten Werte leichte bis mittlere negative Zusammenhänge zeigen (vgl. Tabelle 59). Letztere fallen für Nordamerika am deutlichsten aus, da hier die vier Institute mit den kleinsten Websites (KMFRI, MOI, NIOPK und ORI) besonders hohe Werte erzielen (berechnet aus Tabelle 8 und Tabelle 70).

<i>HDI vs. Inlinkdomains</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Seite</i>
aus den TLDs Afrikas	0,767	0,702	0,619
aus den TLDs Asiens	0,747	0,917	-0,172
aus den TLDs Europas	0,683	0,917	-0,326
aus den TLDs Nordamerikas	0,407	0,788	-0,568
aus den TLDs Ozeaniens	0,791	0,910	0,586
aus den TLDs Südamerikas	0,762	0,747	0,185

Tabelle 59: Spearman's ρ für HDI und Maße der Sichtbarkeit im World Wide Web nach einzelnen Kontinenten

Somit korrelieren sämtliche an der Personalstärke standardisierten Inlinkwerte der Meereresforschungseinrichtungen stark mit dem HDI. Etwas geringere Werte zeigten sich lediglich für die Inlinks aus den TLDs des eigenen Kontinents sowie Afrikas. Jedoch fanden sich zum Teil auch bei Standardisierung an der Websitegröße deutliche Korrelationen mit dem HDI für die Inlinkdomains einiger Kontinente sowie der Übergangsländer, die jedoch überwiegend aus geringen Fallzahlen sowie Skaleneffekten resultierten.

5.3.2.4 Gibt es einen Zusammenhang zwischen dem Human Development Index und der Präsenz bzw. der Sichtbarkeit der Meeresforschungsinstitute im Web of Science?

Von allen Instituten mit Ausnahme des NIOPK liegen Publikationslisten für das Jahr 2006 vor. Die Liste enthält zwar keine Informationen über internationale Koautorenschaft o.ä., ihre Abdeckung durch das Web of Science gibt jedoch Aufschluss über die internationale Präsenz: Das WoS umfasst vorwiegend internationale Zeitschriften, die zum weltweiten Kern des wissenschaftlichen Mainstreams gehören. Zeitschriften mit eher lokalem oder regionalem Publikum werden nicht im gleichen Maß von dieser Datenbank erfasst. Eine geringe Abdeckung des Publikationsoutputs ist somit ein indirekter Hinweis auf begrenzte internationale Reichweite.

Die institutseigenen Publikationsverzeichnisse umfassen eine stark variierende Bandbreite von Veröffentlichungen, die von Fachartikeln in wissenschaftlichen Zeitschriften bis zu Berichten, Examensarbeiten, Newslettern, Softwaredokumentationen oder populärwissenschaftlichen Artikeln reicht. An dieser Stelle sind nur Beiträge in Fachzeitschriften relevant, da das Web of Science andere Veröffentlichungen nicht erfasst. Daher weist Tabelle 60 in Spalte A die Anzahl der darin enthaltenen Fachartikel aus.

Ein Abgleich mit dem Web of Science zeigt, dass dort ein Teil der aufgelisteten Artikel unter der Anschrift der Meeresforschungseinrichtungen zu finden ist (Tabelle 60, Spalte B), während andere Artikel im Web of Science unter anderen Adressen geführt werden (Spalte C). Eine mögliche Ursache ist, dass Mitglieder der untersuchten Institute Publikationen an früheren oder temporären Arbeitsstätten auf die Institutsliste gesetzt haben, obwohl sie die Veröffentlichungen unter abweichenden Adressen eingereicht haben. Daneben sind einige Publikationen in Zeitschriften erschienen, die nicht zum Corpus des Web of Science gehören (Spalte D). Schließlich sind im Web of Science Artikel mit der Adresse der Institute zu finden, die jedoch nicht in der institutseigenen Publikationsliste enthalten waren (Spalte E).

Aus diesen Daten lässt sich die Summe aller von den Instituten selbst benannten oder im Web of Science gefundenen Veröffentlichungen in Fachzeitschriften bilden (Spalte F). An dieser Stelle sind jedoch insbesondere diejenigen Publikationen wichtig, die der Einrichtung – und nicht den dort zu irgendeinem Zeitpunkt beschäftigten Personen – zuzuschreiben sind. Daher werden die Publikationen, die im Web of Science unter den Adressen anderer Institute geführt werden (Spalte C) nicht weiter berücksichtigt. Die Summe der bekannten Publikationen mit Institutsadresse (Spalte G) besteht somit aus den Publikationen auf der Liste – sofern sie nicht unter den Adressen anderer Einrichtungen gelistet sind (also Spalte B und D, nicht jedoch Spalte C) – sowie den Beiträgen, die zusätzlich zur Publikationsliste unter der Institutsanschrift im Web of Science enthalten sind (Spalte E).

<i>Institut</i>	<i>A) aufgelistete Publikationen in Fachzeitschriften</i>	<i>B) davon im WoS enthalten mit Institutsanschrift</i>	<i>C) davon im WoS enthalten ohne die Institutsanschrift</i>	<i>D) davon nicht im WoS enthalten (Zeitschrift nicht erfasst)</i>	<i>E) zusätzliche Artikel, die unter der Institutsanschrift im WoS zu finden sind</i>	<i>F) Summe aller bekannten Publikationen = A+E</i>	<i>G) Summe aller bekannten Publikationen mit Institutsadresse = B+D+E</i>	<i>H) Abdeckung aller bekannten Publikationen mit Institutsadresse durch das WoS = (B+E)/G</i>
BLOS	33	14	17	2	1	34	17	88,2%
HBOI	27	24	0	3	17	44	44	93,2%
IFM	175	155	16	4	73	248	232	98,3%
IOW	79	67	6	6	13	92	86	93,0%
KMFRI	4	4	0	0	9	13	13	100,0%
MBARI	68	55	11	2	29	97	86	97,7%
MOI	4	2	2	0	0	4	2	100,0%
NIO	163	154	0	9	2	165	165	94,5%
NIOPK	---	---	---	---	---	---	---	---
ORI	6	5	0	1	2	8	8	87,5%

Tabelle 60: Publikationen der Institute aus dem Jahr 2006 und ihre Abdeckung durch das Web of Science

Daraus lässt sich ermitteln, wie hoch die Abdeckung der bekannten, den Instituten zuzuschreibenden Fachartikel durch das Web of Science ist (Spalte H). Im Durchschnitt der zehn Meeresforschungseinrichtungen sind 94,7 Prozent ihrer bekannten Publikationen in Fachzeitschriften aus dem Jahr 2006 mit Institutsanschrift im Web of Science enthalten. Die einzelnen Abdeckungswerte liegen zwischen 88,2 und 100 Prozent. Somit wurden die Beiträge aller Institute in Fachzeitschriften annähernd vollständig oder sogar komplett durch das WoS abgedeckt. Es sind keine systematischen Unterschiede zwischen Instituten in Industrie- und Entwicklungsländern erkennbar, so dass die internationale Präsenz für das gesamte Sample als hoch bewertet werden kann.

Eine detailliertere Analyse zeigt, dass die nicht abgedeckten Zeitschriftenbeiträge (Spalte D) nicht notwendigerweise regional ausgerichtet, sondern zum Teil in neu gegründeten oder sozial- bzw. geisteswissenschaftlichen Zeitschriften erschienen sind, die das WoS in geringerem Umfang erfasst. Diese Auswertung erlaubt außerdem Rückschlüsse auf die Qualität der Publikationslisten: Fünf von neun Zusammenstellungen (BLOS, IFM, IOW, MBARI, MOI) enthalten Beiträge, die nicht unter der Institutsadresse erschienen sind. Für acht der neun

Institute (alle außer MOI) weist das WoS Veröffentlichungen aus, die nicht in der Liste enthalten sind. Beide Abweichungen haben zum Teil erheblichen Umfang und treten bei Instituten mit kleiner wie großer Publikationszahl auf. Eine umfassende und präzise Übersicht über die hauseigenen Publikationen scheinen die meisten untersuchten Forschungseinrichtungen nicht zu besitzen. Auch eine einheitliche Regelung, unter welcher Adresse die WissenschaftlerInnen z.B. bei Gastaufenthalten publizieren, scheint es nicht zu geben, so dass sich die Publikationslisten der Institute nur in Teilen mit dem WoS überschneiden, das von den faktischen Adressangaben in den Artikeln ausgeht.

Die Anzahl der Einträge im Web of Science korreliert nicht mit dem Human Development Index. Im Verhältnis zur personellen Institutsgröße zeigt sich jedoch ein stark ausgeprägter Zusammenhang (vgl. Tabelle 61): Die deutschen Institute haben pro MitarbeiterIn die meisten Einträge im WoS, KMFRI und NIOPK die wenigsten. Von den übrigen Einrichtungen haben das indische NIO, das südafrikanische ORI und das kenianische KMFRI viele Publikationen im Verhältnis zu den HDI-Werten ihres Landes. Die MitarbeiterInnen des höchstplatzierten IFM haben für die untersuchten Jahre sechsmal mehr Einträge im WoS als das KMFRI und 17mal mehr als das NIOPK (berechnet aus Tabelle 15, Tabelle 17 und Tabelle 18).

<i>HDI vs.</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Publikation</i>
Anzahl der Publikationen	0,388	0,591	---
standardisierte Zitationen	0,535	0,812	0,837

Tabelle 61: Spearmans ρ für HDI und die Publikationen resp. Zitationen der Institute im Web of Science

Bei den zeitlich standardisierten Zitationen, die die Institute – pro MitarbeiterIn oder pro Publikation – erhalten, zeigt sich ein stärkeres Gefälle: Sie korrelieren stark mit dem HDI (Tabelle 61). Hier weisen die deutschen und US-amerikanischen Institute jeweils deutlich höhere Werte auf als alle übrigen. KMFRI und NIOPK erhalten pro MitarbeiterIn nur ein Elftel bzw. ein Prozent der Zitationen, die das IFM verzeichnet (berechnet aus Tabelle 16, Tabelle 17 und Tabelle 18).

5.3.2.5 Gibt es einen Zusammenhang zwischen dem Human Development Index und der internationalen Orientierung der Institute im Web of Science?

Eine nach Koautorenschaftsstatus differenzierende Betrachtung der Publikationen zeigt Unterschiede zwischen verschiedenen Publikationstypen (Tabelle 62): Im Rohzustand korrelieren weder interne, nationale und internationale noch intra- oder interkontinentale Kopublikationen mit dem HDI. Bezogen auf die Personalstärke der Institute zeigt sich ein hoher

Zusammenhang mit dem HDI ausschließlich für die Anzahl der nationalen Kopublikationen. Hierzu werden alle Veröffentlichungen gezählt, die mindestens eine externe Koautorendresse innerhalb des eigenen Landes, aber keine weitere Adresse in anderen Ländern aufweisen.

Die fünf Meeresforschungsinstitute in den USA und Deutschland weisen pro MitarbeiterIn die meisten Publikationen mit einheimischen KoautorInnen auf, gefolgt von NIO in Indien und ORI in Südafrika und den übrigen drei Instituten (berechnet aus Tabelle 17 und Tabelle 71). Rein nationale Kopublikationen stellen bei den US-amerikanischen Instituten etwa ein bis zwei Drittel aller Einträge der Institute im Web of Science (BLOS: 32,3%, MBARI: 49,6%, HBOI: 60,1%). Auch das pakistanische NIOPK weist mit 60 Prozent einen hohen Wert bei einer sehr geringen Gesamtzahl von fünf Publikationen. Alle übrigen Einrichtungen veröffentlichen nicht mehr als 28 Prozent ihrer Publikationen exklusiv mit KollegInnen aus demselben Land.

Diese Ergebnisse korrespondieren grob mit der Größe der nationalen Forschungssysteme. So gibt es in den USA deutlich mehr Einrichtungen im Bereich der Meeresforschung als in den übrigen Ländern. Da KMFRI und MOI die einzigen Meeresforschungseinrichtungen in Kenia bzw. Mauritius sind, überrascht es nicht, dass sie mit 14,0 bzw. 12,5 Prozent den geringsten Anteil nationaler Kopublikationen aufweisen (berechnet aus Tabelle 15 und Tabelle 71). Dagegen gibt es in Südafrika noch ozeanographische Universitätsinstitute.

<i>HDI vs.</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Publikation</i>
interne Publikationen	0,382	0,339	0,246
nationale Kopublikationen	0,486	0,720	0,394
internationale Kopublikationen	0,234	0,468	-0,314
Kopublikationen mit eigenem Kontinent	0,415	0,463	0,265
Kopublikationen mit anderen Kontinenten	0,234	0,499	-0,345

Tabelle 62: Spearman's ρ für HDI und verschiedene Maße der internationalen Orientierung der Institute im Web of Science

Differenziert man die internationalen Kopublikationen nach Ländergruppen, zeigen sich für Entwicklungs- und Industrieländer unabhängig von der Standardisierung nur geringe und mittlere Korrelationen mit dem HDI (Tabelle 63). MOI und KMFRI haben von allen untersuchten Meeresforschungseinrichtungen den höchsten Anteil von Publikationen mit Adressen in anderen Entwicklungsländern sowie Industrieländern, NIO und HBOI den geringsten (berechnet aus Tabelle 15 und Tabelle 72). Dagegen haben im Verhältnis zur personellen Institutsgröße MOI und ORI die meisten Kopublikationen mit Entwicklungsländern, NIOPK und HBOI die wenigsten. Besonders viele Kopublikationen pro

MitarbeiterIn mit Industrieländern weisen IFM und IOW auf, besonders wenige ORI und NIOPK (berechnet aus Tabelle 17 und Tabelle 72). Insgesamt ist die Situation hier sehr heterogen, was sich in den geringen Werten der Korrelationskoeffizienten widerspiegelt (Tabelle 63).

Kopublikationen mit Übergangsländern sind für die drei afrikanischen Institute ebenso wie für das NIOPK im Untersuchungszeitraum nicht im WoS enthalten. Auch bei den übrigen – mit Ausnahme des IFM – sind die Werte gering (vgl. Tabelle 72). Die gefundenen Korrelationen resultieren somit erneut zum Teil aus geringen Fallzahlen (Tabelle 63).

<i>HDI vs.</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Publikation</i>
Kopublikationen mit Entwicklungsländern	-0,074	-0,043	-0,209
Kopublikationen mit Industrieländern	0,296	0,517	-0,203
Kopublikationen mit Übergangsländern	0,510	0,622	0,730

Tabelle 63: Spearmans ρ für HDI und Maße der Orientierung der Institute im Web of Science an Ländergruppen nach Entwicklungsstand

Die Publikationsaktivitäten mit KoautorInnen auf anderen Kontinenten zeigen nur in einigen Fällen einen Zusammenhang mit dem HDI (Tabelle 64). Bezogen auf die Anzahl der MitarbeiterInnen korreliert die Anzahl der Kopublikationen mit Adressen in Nordamerika, Ozeanien und Südamerika deutlich mit dem HDI, nicht jedoch die Anzahl der Kopublikationen mit Adressen in Afrika, Asien und Europa.

<i>HDI vs.</i>	<i>roh</i>	<i>pro MitarbeiterIn</i>	<i>pro Publikation</i>
Kopublikationen mit Afrika	-0,427	-0,464	-0,552
Kopublikationen mit Asien	-0,006	-0,135	-0,382
Kopublikationen mit Europa	0,185	0,394	-0,283
Kopublikationen mit Nordamerika	0,512	0,708	-0,166
Kopublikationen mit Ozeanien	0,544	0,591	0,383
Kopublikationen mit Südamerika	0,589	0,624	0,624

Tabelle 64: Spearmans ρ für HDI und Maße der Orientierung der Institute im Web of Science an einzelnen Kontinenten

Lediglich für Südamerika findet sich unabhängig von der Standardisierung ein höherer Korrelationskoeffizient, der sich daraus ergibt, dass nur vier Institute in den USA und Deutschland überhaupt und in eher geringem Umfang Kopublikationen mit Südamerika veröffentlicht haben (vgl. Tabelle 73).

Sowohl im Verhältnis zum Personalumfang als auch zur Gesamtzahl der Publikationen haben die drei afrikanischen Institute die meisten Fachartikel mit anderen afrikanischen Ländern veröffentlicht. Es folgen die beiden deutschen Einrichtungen und das indische Meeresforschungsinstitut, während die Institute in den USA und Pakistan im Untersuchungszeitraum keine entsprechenden Einträge im Web of Science haben (berechnet aus Tabelle 15, Tabelle 17 und Tabelle 73).

Bezogen auf den Personalumfang gibt es keinen statistischen Zusammenhang zwischen HDI und der Anzahl der Kopublikationen mit Asien. Hier mischen sich Institute in Ländern mit unterschiedlichem Entwicklungsstand stark. Interessant ist, dass das pakistanische NIOPK keinerlei Kopublikationsbeziehungen auf dem eigenen Kontinent besitzt, während das indische NIO den zweithöchsten Wert an Kopublikationen mit Asien pro MitarbeiterIn aufweist (berechnet aus Tabelle 17 und Tabelle 73). Der relative Stellenwert, den asiatische Kopublikationen unter den Publikationen der einzelnen Institute einnehmen, ähnelt – bei etwas höheren absoluten Werten – denen der afrikanischen: Die Meeresforschungseinrichtungen in Mauritius, Kenia und Südafrika veröffentlichen den größten Anteil von Publikationen mit asiatischen KoautorInnen, gefolgt vom NIO. Die deutschen und US-amerikanischen Institute weisen deutlich geringere Anteile auf (berechnet aus Tabelle 15 und Tabelle 73).

Die beiden deutschen Institute haben pro MitarbeiterIn mit Abstand die meisten europäischen Kopublikationen im Web of Science, gefolgt von den US-amerikanischen BLOS und MBARI (berechnet aus Tabelle 17 und Tabelle 73). Besonders niedrige Werte weisen hier NIOPK, MOI und HBOI auf, so dass sich kein statistischer Zusammenhang zwischen den Publikationen mit KoautorInnen aus europäischen Ländern und dem Entwicklungsstand findet (vgl. Tabelle 64). Der Anteil europäischer Kopublikationen an allen Publikationen im Web of Science ist beim KMFRI besonders hoch, gefolgt von den beiden deutschen Instituten. Besonders niedrig ist er bei HBOI, ORI und NIO (berechnet aus Tabelle 15 und Tabelle 73). Auch hier mischen sich die Institute in Ländern mit unterschiedlichem Entwicklungsstand so stark, dass es keinerlei Korrelation gibt.

Gleiches gilt für die Kopublikationen mit Nordamerika, die bei IFM, MOI und NIOPK den höchsten, bei NIO, HBOI und MBARI den geringsten Stellenwert einnehmen (berechnet aus Tabelle 15 und Tabelle 73). Pro MitarbeiterIn haben dagegen die beiden deutschen Einrichtungen sowie BLOS und MBARI die höchsten Werte, während NIOPK, KMFRI und NIO besonders niedrige Werte aufweisen (berechnet aus Tabelle 17 und Tabelle 73), woraus eine deutliche Korrelation resultiert (vgl. Tabelle 64).

Für Ozeanien findet sich eine mittlere Korrelation zwischen dem HDI und der Anzahl der Kopublikationen pro MitarbeiterIn (vgl. Tabelle 64), die sich daraus ergibt, dass die deutschen und US-amerikanischen Institute überwiegend höhere Werte haben als die

übrigen. Hier weichen jedoch das MOI und das BLOS deutlich ab, die im Verhältnis zu ihrem Personalumfang die zweithöchste resp. viertniedrigste Anzahl an Kopublikationen mit Ozeanien besitzen (berechnet aus Tabelle 17 und Tabelle 73).

Interessant ist, dass sich hier Unterschiede im Publikationsverhalten der einzelnen Institute zeigen: Während bei einigen Instituten wie dem deutschen IFM oder den US-amerikanischen BLOS und MBARI viele Kontinente und Länder in den Adressen der KoautorInnen vertreten sind, die im Hinblick auf geographische Lage und sozioökonomischen Status sehr heterogen sind, konzentrieren sich andere wie das HBOI oder das deutsche IOW eher auf das eigene Land bzw. die eigene Region. Ähnliche Divergenzen finden sich auch bei den Entwicklungsländern: Die drei afrikanischen Einrichtungen räumen bei geringer absoluter Anzahl von Publikationen gemeinschaftlichen Veröffentlichungen mit anderen Instituten auf dem eigenen Kontinent, aber auch in Asien, Europa und Nordamerika einen hohen Stellenwert ein. Dagegen konzentrieren sich die beiden asiatischen Institute stark auf Europa und – dies gilt nur für das NIO – auf den eigenen Kontinent.

5.3.3 Zusammenfassung und Diskussion

Präsenz und Sichtbarkeit im World Wide Web und im Web of Science

Im Hinblick auf die allgemeine Präsenz der Meeresforschungsinstitute hat sich gezeigt, dass die Websitegröße mit dem Entwicklungsstand zusammenhängt (standardisiert an der personellen Institutgröße, vgl. Tabelle 51). Hier ist zu berücksichtigen, dass die Websitegröße ein zentrales Kriterium bei der Auswahl der Fälle war (vgl. 4.2.3): Mehrere Meeresforschungseinrichtungen in den Deutschland und den USA wurden aufgrund ihrer umfangreichen Websites ausgeschlossen, um die Vergleichbarkeit der Fälle zu gewährleisten. Ebenso wurden zwei Institute in Entwicklungsländern nicht berücksichtigt, weil sie keine eigene Internetpräsenz besaßen (Tabelle 4). Somit ist anzunehmen, dass die Unterschiede in der Webpräsenz zwischen allen existierenden Meeresforschungseinrichtungen in den USA, Deutschland und englischsprachigen Entwicklungsländern noch deutlich größer ausfallen.

Bemerkenswert ist, dass die Webpräsenz der untersuchten Einrichtungen stärker mit dem Entwicklungsstand ihrer Sitzländer korreliert als die Zahl der Einträge im Web of Science (standardisiert an der personellen Institutgröße, vgl. Tabelle 51 und Tabelle 61). Noch deutlicher als die grundlegenden Maße der Präsenz im World Wide Web und im Web of Science korrelieren Maße der Sichtbarkeit, wie die Anzahl der Inlinkdomains oder der standardisierten Zitationen pro Mitarbeiterin, mit dem Human Development Index (vgl. Tabelle 56 und Tabelle 61).

Sowohl im Hinblick auf die Präsenz als auch auf die Sichtbarkeit der Institute weisen die Webmaße (Anzahl der Webseiten und anderer Dateien, Inlink- und Outlinkdomains) durchweg einen stärkeren Zusammenhang mit dem Entwicklungsstand der Sitzländer auf als die Publikationsmaße (Anzahl der Publikationen im Web of Science und ihrer zeitlich standardisierten Zitationen). Dies ist ein erster Hinweis darauf, dass das Bild vom *global research village*, in der das Internet die ForscherInnen zu virtuellen Nachbarn werden lässt, für die untersuchten Meeresforschungseinrichtungen unzutreffend ist. Das World Wide Web scheint die globalen Disparitäten, die sich in der Präsenz und Sichtbarkeit der Institute im Web of Science zeigen, eher zu verstärken. Alle untersuchten Einrichtungen in Entwicklungsländern besaßen englischsprachige Websites und boten somit vergleichsweise günstige Voraussetzungen. Es ist zu erwarten, dass Websites, die überwiegend in anderen Sprachen bzw. nicht-lateinischer Schrift verfasst sind, noch seltener das Ziel von internationalen Links sind. Während der weltweiten Recherche nach Meeresforschungsinstituten für diese Fallstudie wurden mehrere Internetauftritte dieser Art, insbesondere in Asien und Südamerika, gefunden.

Eine mögliche Erklärung ist, dass die Produktivität der Institute in Form von wissenschaftlichen Publikationen und von Websites vom sozio-ökonomischen Umfeld abhängt, aus dem in der Regel die Ressourcen stammen, die den Instituten zur Verfügung stehen. Sind die Ressourcen knapp bemessen, konzentrieren sich die Forschungseinrichtungen offensichtlich primär auf die Kernaktivitäten wissenschaftlichen Arbeitens. Der Auf- und Ausbau der eigenen Webpräsenz steht nicht an erster Stelle.

Nationale und internationale Orientierung und Sichtbarkeit

Die web- und publikationsbasierten Maße liefern auch Informationen über die nationale und internationale Einbettung der Meeresforschungsinstitute. Die Anzahl der nationalen Kopublikationen pro MitarbeiterIn hängt in ähnlich starkem Maße mit dem Entwicklungsstand der Sitzländer zusammen wie die Anzahl der Inlink- und Outlinkdomains der eigenen nationalen Top-Level-Domain (vgl. Tabelle 53, Tabelle 57 und Tabelle 62): Die Forschungseinrichtungen in den USA und Deutschland weisen eine bessere nationale Vernetzung ihrer Websites auf als die übrigen Institute. Die Institute in den USA haben außerdem die meisten Publikationen mit KoautorInnen aus dem eigenen Land veröffentlicht.

Die internationale und interkontinentale Orientierung und Sichtbarkeit der Institute im World Wide Web (jeweils gemessen an In- und Outlinkdomains mit den entsprechenden Top-Level-Domains pro MitarbeiterIn) hängt ebenfalls deutlich mit dem Human Development Index zusammen. Anders ist die Situation jedoch bei den Publikationen: Die Anzahl der Kopublika-

tionen mit Einrichtungen in anderen Ländern korreliert weniger stark mit dem HDI. Gleiches gilt für Veröffentlichungen mit KoautorInnen von dem eigenen oder von anderen Kontinenten.

Dass MeeresforscherInnen in Entwicklungsländern weniger nationale Kopublikationen veröffentlichen als die in Industrieländern, ist nachvollziehbar: Sowohl die USA als auch Deutschland besitzen insgesamt gut ausgebaute Forschungssysteme mit jeweils mehreren Meeresforschungseinrichtungen. Mit Ausnahme von Indien gilt dies nicht für die Sitzländer der übrigen Institute. Mit den EU-Staaten bzw. Kanada besitzen Deutschland und die USA außerdem Nachbarländer, die ebenfalls umfangreiche und stark ausdifferenzierte Forschungslandschaften aufweisen. Dennoch weist die geringe Korrelation intrakontinentaler Publikationen pro MitarbeiterIn mit dem HDI darauf hin, dass auch in Afrika und Asien regionale Forschungsk Kooperationen bestehen, die sich jedoch nicht in den Outlink- und Inlinkwerten niederschlagen. Die vergleichsweise gute Einbindung der afrikanischen und asiatischen Institute in internationale und interkontinentale Publikationsaktivitäten, die die schwache Korrelation von HDI und Anzahl der internationalen Kopublikationen pro MitarbeiterIn impliziert, findet sich nicht in gleichem Maße bei den Outlinkdomains wieder. Besonders ungleich verteilt ist jedoch die internationale und interkontinentale Sichtbarkeit der untersuchten Institute, gemessen an den Inlinkdomains pro ForscherIn.

Der Stellenwert einzelner Ländergruppen

Betrachtet man die Kopublikationen nach Ländergruppen, zeigt sich lediglich für die Übergangsländer ein deutlicher Zusammenhang mit dem HDI, der u.a. darauf zurückgeht, dass von den in Entwicklungsländern gelegenen Meeresforschungseinrichtungen lediglich das NIO entsprechende Veröffentlichungen aufweisen kann (vgl. Tabelle 63). Bei den Outlinkdomains fand sich eine stärkere Korrelation nur für Verweise auf Seiten in Industrienationen, nicht aber in Entwicklungs- oder Übergangsländern (vgl. Tabelle 54). Dagegen erhalten die Forschungseinrichtungen in den USA und Deutschland pro MitarbeiterIn wesentlich mehr Inlinks aus Industrie-, Entwicklungs- und Übergangsländern (vgl. Tabelle 58). Die ungleiche internationale Sichtbarkeit im World Wide Web besteht somit unabhängig von der Länderkategorie. Sie ist nicht allein durch das starke Gewicht von Inlinks aus hoch entwickelten Ländern mit starker Webaktivität zu erklären. Auch die Inlinks, die die Top-Level-Domain von Entwicklungsländern tragen, richten sich vorwiegend auf die Websites der US-amerikanischen und deutschen Forschungseinrichtungen. Vermutlich spielen hier mehrere Gründe eine Rolle: Viele Meeresforschungseinrichtungen in den USA und in Deutschland besitzen eine starke Reputation innerhalb der *scientific community*, so dass sich ForscherInnen in anderen Ländern an ihnen orientieren, ihre Website auch ohne konkrete Kooperation zur Kenntnis nehmen und besonders häufig auf sie verweisen. Darüber hinaus besitzen diese Einrichtungen eine gewisse Bekanntheit in der außerwissenschaftlichen Öffentlichkeit. Ihre

umfangreichen und vielfältigen Websites erhalten auch Aufmerksamkeit von Fachleuten aus Medien, Politik und Wirtschaft sowie von interessierten Laien, die sich mit Ozeanen, Meeres-tieren oder Klimawandel befassen. Nicht wenige der inhaltsanalytisch klassifizierten Seiten stammten beispielsweise aus Aquaristikforen oder –themenseiten, die auf Fischfotografien auf den Websites deutscher und US-amerikanischer Meeresforschungseinrichtungen verwiesen. Die untersuchten Institute in Kenia, Mauritius, Pakistan und Südafrika besitzen diese Popularität nicht in gleichem Maße. Gleichzeitig umfassen ihre minimalistischen Internetauftritte keinen großen Fundus an Informationen und Ressourcen. Es ist ferner nicht auszuschließen, dass nationale Top-Level-Domains in Industrieländern insgesamt häufiger genutzt werden als in Entwicklungsländern. Systematische Informationen zu deren Verbreitung und zu Unterschieden in den Zugangsmodalitäten liegen bisher nicht vor.

Der Stellenwert einzelner Kontinente

Auch hier korrelieren die an dem wissenschaftlichen Personal standardisierten Inlinkdomains durchgängig stark mit dem Entwicklungsstand der Sitzländer der untersuchten Institute (Tabelle 59). Je höher entwickelt ein Land ist, desto sichtbarer sind seine Meeresforschungseinrichtungen in den Top-Level-Domains der einzelnen Kontinente. Geographische Nähe scheint hier eine geringere Rolle zu spielen als die Zentralität im globalen Wissenschaftssystem, wie die fehlenden oder niedrigen Inlinkwerte der afrikanischen und asiatischen Institute aus afrikanischen, asiatischen, aber auch südamerikanischen und ozeanischen TLDs zeigen.

Bei den entsprechenden Outlinkdomains zeigen sich statistische Zusammenhänge für Europa, Ozeanien und Nord- und Südamerika, nicht jedoch für Afrika und Asien, da sich die dort gelegenen Institute mit ihren Nachbarstaaten verlinken (Tabelle 54). Allerdings sind europäische und nordamerikanische Top-Level-Domains für die meisten Institute am häufigsten das Ziel von Outlinks (Tabelle 67). Dies ist konsistent mit den Ergebnissen von Park und Thelwall (2006), die die Linkstrukturen europäischer und asiatischer Universitäten untersuchten und fanden, dass beide Gruppen in erster Linie europäische Outlinkziele aufwiesen.

Regionale Verbindungen zeigen sich auch bei den Publikationen (Tabelle 64): Die Anzahl der Kopublikationen mit asiatischen bzw. afrikanischen Partnern korreliert unabhängig von der Standardisierung negativ mit dem HDI, weil die dort gelegenen Institute mit Ausnahme des NIOPK regionale Kopublikationskontakte pflegen. Pro MitarbeiterIn besitzen die US-amerikanischen und deutschen Institute die meisten Kopublikationen mit Ozeanien, Nord- und Südamerika, so dass sich statistische Zusammenhänge ergeben. Dies gilt jedoch nicht für Europa, das bei fast allen Instituten den höchsten Stellenwert als Publikationspartner einnimmt (Tabelle 73).

Insgesamt zeigen sich im Hinblick auf Verbindungen mit einzelnen Kontinenten ähnliche Muster für Kopublikationen, Outlink- und Inlinkdomains. Bei den Inlinkdomains ist die regionale Vernetzung am schwächsten ausgeprägt. Entsprechend hängt für alle Kontinente die Anzahl der Inlinkdomains, die auf die Institute verweisen, besonders stark mit dem Entwicklungsstand ihrer Sitzländer zusammen.

Schlussfolgerungen aus der Standardisierungsweise

In den Inlink- und Outlinkmustern zeigten sich fast durchgehend Korrelationen mit dem HDI, wenn die Werte an der Personalstärke der Meeresforschungsinstitute standardisiert waren. Wenn sie auf die Websitegröße bezogen waren, traf dies jedoch nur in Ausnahmefällen zu. Die untersuchten Meeresforschungseinrichtungen in Entwicklungsländern sind im World Wide Web in ähnlichem Maß durch Inlinks sichtbar und durch Outlinks vernetzt wie die Institute in Industrieländern, wenn ihre Websitegröße berücksichtigt wird. Im Verhältnis zu ihrer personellen Institutsgröße erhalten sie jedoch weniger Inlinks als die Institute in den USA oder Deutschland. Somit entspricht ihre Sichtbarkeit im Web (in Form externer Inlinkdomains) ihrer virtuellen Größe (in Form von Webseiten), jedoch nicht ihrer realen Größe (im Hinblick auf den Personalumfang).

Dieser Befund enthält aus der Perspektive der Entwicklungsländer zwiespältige Implikationen: Die kleinen, grundlegenden Internetauftritte, die die untersuchten Meeresforschungseinrichtungen in Südafrika, Mauritius, Kenia und Pakistan aufgebaut haben, sind im Verhältnis zu ihrer Größe nicht weniger sichtbar als die größeren Websites der übrigen Institute. Somit resultiert die Ungleichheit auch aus der geringeren Größe der Webpräsenzen. Ihr Ausbau ließe sich, sofern politisch gewünscht, durch gezielte Programme im Rahmen des *capacity building* vorantreiben. Sofern die gefundenen Zusammenhänge stabil bleiben, besteht so die Möglichkeit, die internationale Vernetzung durch In- und Outlinks zu erhöhen.

Allerdings ist es noch ein weiter Weg, bis diese Einrichtungen die virtuelle Sichtbarkeit der Institute in den USA und Deutschland im Verhältnis zu ihrer Personalstärke erreichen. Dies ist weniger eine Frage des *capacity building* oder der zielgerichteten Strategien der Forschungseinrichtungen als vielmehr der Verteilung von Ressourcen und der Zentralität in der weltweiten *scientific community*. Die Tatsache, dass bei Weitem die meisten Inlinks mit nationalen Top-Level-Domains aus Europa und Nordamerika stammten, untermauert den Stellenwert globaler Disparitäten im Internetzugang und der Webpräsenz sowie in der Verwendung nationaler TLDs. Folglich kann das World Wide Web nicht die tatsächliche Bekanntheit und Sichtbarkeit der untersuchten Forschungseinrichtungen in Entwicklungsländern abbilden.

Im Hinblick auf Publikationen ist die Situation eine andere: Zwar weisen auch die Kopublikationsmaße fast durchgängig höhere Korrelationen auf, wenn sie an der Institutsgröße statt an der Publikationszahl standardisiert werden. Jedoch ist, bezogen auf die personelle Institutsgröße, die Ungleichheit in Präsenz und internationaler Orientierung durchweg geringer als in den Webdaten. Lediglich die Anzahl der nationalen Kopublikationen unterscheidet sich deutlich für Institute in weit und weniger weit entwickelten Ländern. Dies korrespondiert mit der Größe und Differenzierung der nationalen Forschungssysteme, die in Deutschland und vor allem in den USA eine besonders große Auswahl von landesinternen Kooperationsmöglichkeiten bieten. Diese Art von Größeneffekt in Kopublikationen wurde bereits von Frame und Carpenter (1979) beschrieben. Die globalen Web- und Publikationsdaten zeigen somit spezifische Disparitäten.

Anders als die Publikationswerte korrelieren die zeitlich standardisierten Zitationen unabhängig von der Standardisierung stark mit dem HDI. Die Veröffentlichungen der Forschungseinrichtungen in Deutschland und den USA sind zwar nicht wesentlich präsenter im Web of Science als die der übrigen untersuchten Institute, sie werden aber wesentlich häufiger zitiert. Während also der Stellenwert internationaler Kopublikationen – bezogen auf die Anzahl der MitarbeiterInnen – nicht mit dem Entwicklungsstand der Sitzländer der Institute zusammenhängt, korrespondiert die relative Sichtbarkeit – die Zahl der Zitationen pro MitarbeiterIn – stark damit (vgl. Tabelle 61). Hier zeigt sich ein wesentlicher Unterschied zur Sichtbarkeit im World Wide Web: Die Einrichtungen in Entwicklungsländern erhalten pro MitarbeiterIn, aber nicht pro Seite weniger Inlinkdomains als die in den USA und Deutschland. Ihre geringe Sichtbarkeit im World Wide Web steht im Verhältnis zu der geringen Größe ihrer Websites. Dagegen nivellieren sich die Disparitäten bei den Veröffentlichungen bei Standardisierung an der Publikationszahl nicht.

In methodologischer Hinsicht hat sich außerdem gezeigt, dass einige Indikatoren nur sinnvoll einsetzbar sind, wenn Institute innerhalb eines Landes verglichen werden: Dies gilt für In- und Outlinkkontakte zu den eigenen nationalen TLDs, da deren einheitliche Nutzung im internationalen Vergleich nicht vorausgesetzt werden kann. Eine Alternative wäre die Standardisierung der nationalen Linkkontakte im Hinblick auf die Verbreitung der entsprechenden TLDs, über die jedoch aussagekräftige quantitative Daten nicht erhältlich sind. Auch Web- und Publikationsindikatoren, die sich auf den eigenen Kontinent beziehen, sind nur bedingt aussagekräftig, wenn Einrichtungen auf verschiedenen Erdteilen verglichen werden, da sich die Anzahl der Nachbarländer und der Umfang ihrer Forschungssysteme stark unterscheidet. Dies gilt auch für Gebiete mit vergleichbarem Entwicklungsstand wie etwa Europa und Nordamerika.

6. Fazit

In den zwanzig Jahren seit seiner Entstehung hat das World Wide Web die Vernetzung und den Austausch von Informationen innerhalb der wissenschaftlichen Gemeinschaft enorm erleichtert. Zu seinen grundlegenden Eigenschaften gehört die Irrelevanz räumlicher Nähe, die die zentrale Innovation gegenüber älteren Hypertextkonzepten darstellt (vgl. Abschnitt 2.1.1). Die globale Dimension, auf die Berners-Lee das von ihm entwickelte Informationssystem ausgelegt hat, zeigt sich auch in der Wahl des Namens „World Wide Web“. Mit der Entwicklung der Webometrie hat sich die Wissenschaftsforschung das World Wide Web als Datenquelle erschlossen. In der vorliegenden Dissertation wurde diskutiert, inwiefern dieses Medium seinem globalen Anspruch gerecht wird und eine geeignete Grundlage für die internationale Untersuchung wissenschaftlicher Strukturen darstellt.

Zu diesem Zweck wurden auf der Basis von Literatur aus Webometrie und Wissenschaftsforschung zunächst die grundlegenden Eigenschaften der webometrischen Methode sowie verschiedene Aspekte hervorgehoben, die in internationalen Zusammenhängen besonders relevant sind (Kapitel 2). Dabei blieb festzuhalten, dass es zahlreiche essentielle Unterschiede zwischen der Webometrie und der Bibliometrie, ihrem Vorbild in den ersten Jahren der Entwicklung, gibt. Im Wesentlichen liegt dies an der Beschaffenheit des World Wide Webs, das äußerst heterogene Dokumente umfasst. Es kennt weder allgemeine formale Standards noch institutionalisierte Qualitätskontrolle und bietet kaum äußerliche Ansatzpunkte, um wissenschaftliche und nichtwissenschaftliche Dokumente zu unterscheiden. Es kann sich permanent wandeln und umfasst keine Archivfunktion. Daher erlaubt es weder retrospektive Analysen noch die Replikation von Untersuchungen, weshalb es keine Grundlage für die Nutzung zur Evaluation wissenschaftlicher Leistung bietet. Die spezifischen Stärken des Webs – der freie Zugang zu Informationen, seine Aktualität, die Vielfalt der Nutzungsmöglichkeiten für die verschiedensten gesellschaftlichen Gruppen – lassen es dennoch als wertvolle Datenquelle für die Wissenschaftsforschung erscheinen, wenn es um explorative Erhebungen oder Analysen des sozialen Systems Wissenschaft einschließlich seiner außerwissenschaftlichen Vernetzung geht. Im Hinblick auf internationale Fragestellungen besteht jedoch spezifischer methodologischer Klärungsbedarf (vgl. Abschnitt 2.6).

Außerdem wurde die besondere Bedeutung von Internationalität in der Wissenschaft herausgearbeitet und das Spektrum der möglichen webometrischen Indikatoren aufgezeigt (Kapitel 3). Kommunikation über Ländergrenzen hinweg gibt es seit Beginn der modernen Wissenschaft. Politische, wirtschaftliche und technologische Entwicklungen haben sie – insbesondere in den letzten Jahrzehnten – enorm erleichtert. Nicht zuletzt die Möglichkeiten des Internets haben dazu beigetragen. Die Metapher des *global research village* verbildlicht

den Eindruck, dass das Internet einen weltumspannenden Kommunikationsraum schafft, in dem sich alle ForscherInnen unabhängig von ihrem Aufenthaltsort nach Belieben austauschen können. An dieser Stelle ist jedoch zu berücksichtigen, dass Forschung trotz des globalen normativen Horizonts der *scientific community* von lokalen und nationalen Bedingungen abhängt. Forschungsförderung ist vorwiegend im nationalen Rahmen organisiert, das Gleiche gilt für die Bildungssysteme. Auch das Internet unterliegt lokalen Einflüssen. Anders als beispielsweise der Name World Wide Web suggeriert, ist der Zugang zu diesem Informations- und Kommunikationssystem keineswegs gleichmäßig auf dem gesamten Globus verteilt. Unter dem Schlagwort *digital divide* wurde die These aufgestellt, dass das Internet bestehende Ungleichheit eher verstärkt als reduziert hat. Empirische Studien im akademischen Bereich deuten darauf hin, dass ForscherInnen in verschiedenen Entwicklungsländern über einen qualitativ minderwertigen Internetzugang verfügen und sich ihre Internetaktivitäten vorwiegend auf Emailverkehr beschränken. Webometrische Untersuchungen zur Nutzung des World Wide Webs im internationalen Vergleich gibt es bisher kaum. Als klärensenswert wurde daher die Frage angesehen, inwiefern die Präsenz und Sichtbarkeit von Forschungseinrichtungen aus verschiedenen Ländern mit ihrem Entwicklungsstand variieren.

Auf Basis der beiden literaturbasierten Kapitel 2 und 3 wurden drei zentrale Leitfragen für eine Fallstudie zusammengestellt (vgl. Abschnitt 4.1):

1. Wie aussagekräftig sind Top-Level-Domains als Indikatoren der nationalen Verortung?
2. Inwieweit ähneln sich die Bilder, die webometrische und bibliometrische Indikatoren von der internationalen Einbettung von Forschungseinrichtungen zeigen?
3. Hängen Präsenz und Vernetzung wissenschaftlicher Einrichtungen im World Wide Web (bzw. im Web of Science) mit dem Entwicklungsstand ihrer Sitzländer zusammen?

Als geeignetes empirisches Feld wurde die Meeresforschung identifiziert, die im Vergleich zu anderen Fachgebieten außergewöhnlich international arbeitet und somit eine gute Grundlage für die Untersuchung grenzüberschreitender Bezüge bietet. Zehn Meeresforschungseinrichtungen in deutsch- und englischsprachigen Staaten in Afrika, Asien, Nordamerika und Europa wurden ausgewählt (vgl. Abschnitt 4.2). Für jedes der Institute wurden Personal-, Web-, Link- und Publikationsdaten erhoben (vgl. Abschnitt 4.3). Entlang der drei Leitfragen wurde in Kapitel 5 empirisch geprüft, inwiefern länderspezifische Top-Level-Domains zuverlässige Hinweise auf die nationale Anbindung einer Webseite geben, inwiefern sich die Bilder ähneln, die web- und publikationsbasierte Maße von der internationalen Einbettung der Meeresforschungseinrichtungen liefern, und inwiefern der Umfang und die Vernetzung ihrer Internetauftritte mit dem Entwicklungsstand ihrer Sitzländer zusammenhängen.

In diesem Gesamtfazit wird entlang der drei Leitfragen die in der Einleitung angeführte Forschungsfrage beantwortet, welches Potential die Webometrie zur Untersuchung der

internationalen Einbettung von Forschungseinrichtungen besitzt (6.1). Dabei wird herausgearbeitet, welche Möglichkeiten die Webometrie in dieser Hinsicht bereithält und welche Unklarheiten und Hindernisse bestehen. Die dritte Leitfrage hat sich als besonders aufschlussreich erwiesen und wird daher ausführlicher diskutiert als die ersten beiden. Ein wesentlicher Faktor ist außerdem der Zeitaufwand, mit dem Daten hinreichender Qualität gewonnen werden können (6.2). Abschließend werden in einem Ausblick Möglichkeiten zur Weiterentwicklung der Webometrie aufgezeigt (6.3).

6.1 Zum Potential der Webometrie zur Untersuchung der internationalen Einbettung von Forschungseinrichtungen

Mit Hilfe der Webometrie lässt sich die Präsenz von Forschungseinrichtungen im World Wide Web durch die Größe ihrer Internetauftritte erfassen und vergleichen. Anhand von Inlink- und Outlinkmustern lässt sich ihre virtuelle Sichtbarkeit bzw. ihre Außenorientierung untersuchen.

Leitfrage 1:

Wie aussagekräftig sind Top-Level-Domains als Indikatoren der nationalen Verortung?

Im Hinblick auf internationale Fragestellungen ist von besonderer Relevanz, ob es Möglichkeiten gibt herauszufinden, aus welchen Ländern Inlinks stammen bzw. in welche Staaten Outlinks verweisen (vgl. Kapitel 5.1). Der einzige Ansatzpunkt, der über die nationale Anbindung von Webseiten oder Links Aufschluss geben und automatisiert ausgewertet werden kann, ist die Top-Level-Domain. Eine Auswertung aller Outlinkziele der zehn untersuchten Meeresforschungseinrichtungen hat ergeben, dass nur die Hälfte eine länderspezifische Top-Level-Domain in der URL trug. Bei den Inlinkquellen war es ein gutes Drittel. Somit kann sich eine Untersuchung der internationalen Einbettung nur auf einen Bruchteil der Links stützen, die in den Websites enthalten sind oder auf sie verweisen.

Durch eine Inhaltsanalyse wurde für eine Stichprobe von Inlinkquellen und Outlinkseiten geprüft, ob der Inhalt der verlinkten Seiten einen Bezug zu einem bestimmten Land enthielt. Es zeigte sich, dass die Übereinstimmung der inhaltlichen Landesbezüge mit den landesspezifischen Top-Level-Domains überaus hoch war. In einem weiteren Analyseschritt stellte sich allerdings heraus, dass einzelne Nationen unter den inhaltlichen Landesbezügen der verlinkten Seiten mit nationalen und mit länderneutralen TLDs nicht gleichermaßen häufig zu finden sind. Dies ist konsistent mit Literatur, die auf Unterschiede in der Vergabep Praxis und Verbreitung der einzelnen nationalen Top-Level-Domains hinweist (vgl. Gallup 2003, Gandhi 2003, Lindén 2003, StClair 2003, Wass 2003c). Eine Auswertung und nationale Zuordnung von Seiten, die mit einem Internetauftritt verlinkt sind, kann sich also nicht ausschließlich auf die landesspezifischen TLDs stützen. Sie würde ein verzerrtes Bild von dem tatsächlichen

Stellenwert einzelner Länder unter den Inlinkquellen und Outlinkzielen liefern. Landesspezifische Top-Level-Domains können somit nicht als valide Indikatoren für die nationale Verortung von Webseiten dienen.

Leitfrage 2:

Inwieweit ähneln sich die Bilder, die webometrische und bibliometrische Indikatoren von der internationalen Einbettung von Forschungseinrichtungen zeigen?

Von grundsätzlicher Bedeutung ist die Frage, inwiefern sich die Informationen über Forschungseinrichtungen ähneln, die mit Hilfe bibliometrischer und webometrischer Verfahren gewonnen werden können (vgl. Kapitel 5.2). Für die zehn Meeresforschungseinrichtungen wurde die Anzahl der Publikationen in den Jahren 2005 bis 2007 im Web of Science sowie die Größe ihrer Websites im Juli 2007 gegenübergestellt, um die Publikationsaktivität mit der Präsenz im World Wide Web vergleichen zu können. Im Folgenden werden die Ergebnisse zusammengefasst, die sich nach der Standardisierung der rohen Publikations- und Linkwerte an der Anzahl der wissenschaftlichen MitarbeiterInnen ergaben. Es zeigte sich, dass Einrichtungen, die stark im Web of Science vertreten sind, auch besonders umfangreiche Websites besitzen.

Als Indikatoren für die Sichtbarkeit der Meeresforschungseinrichtungen wurde die Zahl der Zitationen mit der Anzahl der Inlinks kontrastiert. Wiederum standardisiert an der Anzahl der MitarbeiterInnen, zeigt sich eine starke Korrelation zwischen beiden Maßen. Möglicherweise wirkt sich die Reputation der Einrichtungen auf die virtuelle Sichtbarkeit ebenso aus wie auf die innerhalb der Forschung. Da die Inlinks jedoch nicht allein aus den Websites von Forschungseinrichtungen und anderen wissenschaftlichen Webdokumenten stammen, kommen auch allgemeine globale Disparitäten in der Webnutzung als Ursache in Frage (vgl. Leitfrage 3).

Diese Befunde zu dem internationalen Sample von Meeresforschungseinrichtungen stimmen mit früheren Studien überein, in denen für einzelne Länder ein Zusammenhang zwischen Maßen der Forschungsproduktivität und grundlegenden Webindikatoren festgestellt wurde (Thelwall 2001b; Smith/Thelwall 2002; Vaughan/Thelwall 2005; Thelwall/Tang 2003; Qiu/Chen/Wang 2004; Aguillo et al. 2006; Payne/Thelwall 2008). In gewisser Hinsicht wirft das Ergebnis jedoch neue Fragen auf: Während erfolgreiches wissenschaftliches Publizieren Exzellenz und Originalität voraussetzt, gilt dies nicht für das Gestalten von Websites. Trotzdem kompensieren Einrichtungen mit geringer Publikationsaktivität diese Schwäche nicht mit starker Webpräsenz. Vier Meeresforschungseinrichtungen begnügen sich mit Websites, die aus weniger als dreißig Seiten bestehen. Eine direkte kausale Abhängigkeit zwischen Webpräsenz und Publikationsaktivität ist nicht gegeben. Umso interessanter ist es, nach den

Gründen zu suchen, die die starke Übereinstimmung erklären können, die für die zehn Meeresforschungseinrichtungen gefunden wurde. Einen Ansatz bietet die dritte Leitfrage, die auf den sozio-ökonomischen Kontext abstellt, in dem die Institute arbeiten.

Um die Kontakte der Einrichtungen zu einzelnen Ländern oder Ländergruppen einschätzen zu können, wurden der Koautorenschaftsstatus in den Publikationen und die länderspezifischen Top-Level-Domains unter den Inlinks und Outlinks ausgewertet. Koautorenschafts- und Outlinkmuster entsprechen sich in großen Teilen. Dies gilt für den Stellenwert des eigenen Landes und anderer Länder sowie des eigenen Kontinents und anderer Kontinente. Auch zwischen Koautorenschafts- und Outlinkbezügen zu Entwicklungsländern, Industrieländern sowie zu den Ländern der einzelnen Kontinente finden sich starke Übereinstimmungen: Veröffentlicht ein Meeresforschungsinstitut häufig mit AutorInnen aus Asien, enthält seine Website auch in asiatische Outlinks Top-Level-Domains. Bei den Inlinkwerten ist die Situation eine andere: Unabhängig von der Anzahl der Kopublikationen oder der Outlinks erhalten in der Regel die US-amerikanischen Institute die meisten Inlinks pro MitarbeiterIn für eine Ländergruppe. Es liegt die Schlussfolgerung nahe, dass sich die globale Aufmerksamkeit im Web, wie sie sich in Inlinks abbildet, im Wesentlichen auf Nordamerika und Europa konzentriert. Dies wurde im Rahmen der dritten Leitfrage geprüft.

Leitfrage 3:

Hängen Präsenz und Vernetzung wissenschaftlicher Einrichtungen im World Wide Web (bzw. im Web of Science) mit dem Entwicklungsstand ihrer Sitzländer zusammen?

Die globale Ungleichheit ist ein besonders relevanter Aspekt des Verhältnisses von Bibliometrie und Webometrie im internationalen Kontext. Fachpublikationen aus dem Großteil der Entwicklungsländer nehmen in bibliometrischen Datenbanken nur einen marginalen Stellenwert ein, so dass ein internationaler Bias des Web of Science diskutiert wurde (vgl. Abschnitt 3.1.3). Zu klären war, ob Forschung aus Entwicklungsländern in dem weltweit und frei zugänglichen World Wide Web in stärkerem Maße präsent und sichtbar ist als in dieser Publikationsdatenbank. Daher wurde für die zehn auf verschiedenen Kontinenten gelegenen Meeresforschungseinrichtungen geprüft, inwiefern webometrische und bibliometrische Maße mit dem Human Development Index der Vereinten Nationen korrelieren (vgl. Kapitel 5.3).

Die Ergebnisse der Untersuchung zeigen einen klaren Zusammenhang zwischen den nationalen sozio-ökonomischen Bedingungen in den Sitzländern der Forschungseinrichtungen und ihrer *Präsenz* im World Wide Web. Die untersuchten Meeresforschungsinstitute in den USA und Deutschland haben – im Verhältnis zu ihrer personellen Größe – deutlich größere und besser verlinkte Internetauftritte als die in den untersuchten Entwicklungsländern. An dieser Stelle ist zu berücksichtigen, dass zwei ozeanographische Institute in Entwicklungs-

ländern zum Zeitpunkt der Erhebung im Jahr 2007 nicht über Internetauftritte verfügten, obwohl sie die nationale Kontaktstelle für die internationalen Programme der Vereinten Nationen darstellten. Sie konnten daher nicht in die Untersuchung einbezogen werden. Gleiches gilt für mehrere US-amerikanische Institute, deren Websites zu umfangreich waren, um sinnvoll mit den kleinen und mittleren Websites der afrikanischen und asiatischen Institute verglichen werden zu können. Somit ist anzunehmen, dass das Gefälle zwischen der Gesamtzahl der existierenden Meeresforschungseinrichtungen noch größer ausfällt. Dieser Befund stimmt mit der Studie von Thelwall et al. (2002) überein, in der ein Zusammenhang zwischen der Webpräsenz europäischer Universitäten und dem Pro-Kopf-Einkommen in ihren Sitzländern gefunden wurde. Vergleichbare Untersuchungen, die über den europäischen Raum hinausgehen, gab es bisher nicht.

Interessant ist, dass der Zusammenhang mit dem Entwicklungsstand der Sitzländer für die Publikationszahl der Institute deutlich geringer ausfällt als für die Websitegröße. Eine Differenzierung der Publikationen in solche ohne externe KoautorInnen, mit KoautorInnen ausschließlich im eigenen Land sowie mit KoautorInnen in anderen Ländern erlaubt eine detailliertere Betrachtung: Lediglich für die nationalen Kopublikationen zeigt sich ein statistischer Zusammenhang mit dem Entwicklungsstand der Länder, in denen die Forschungseinrichtungen angesiedelt sind. Er lässt sich als Größeneffekt interpretieren, da die Zahl der möglichen Forschungskontakte in Mauritius oder Pakistan deutlich kleiner ist als in Deutschland oder den USA (vgl. Frame und Carpenter 1979).

Wenngleich ein statistischer Zusammenhang nichts über Kausalitäten aussagt, erscheint im Hinblick auf die Präsenzindikatoren die folgende Interpretation plausibel: Die ungleiche globale Verteilung von Ressourcen wirkt sich in der Meeresforschung stärker auf die Internetauftritte als auf die Publikationen aus. Offensichtlich konzentrieren sich die Forschungseinrichtungen mit begrenzten Forschungsmitteln primär auf wissenschaftliche Kernaktivitäten, zu denen das Publizieren von Ergebnissen in Fachzeitschriften gehört. Dagegen wird der Gestaltung der Institutswebsite eine weitaus geringere Priorität eingeräumt, da sie keine wissenschaftliche Funktion im engeren Sinne erfüllt. Dies erscheint strategisch sinnvoll, da die Internetpräsenz – anders als die Publikationen – in der Regel nicht Gegenstand von Evaluationen ist. Doch das Präsenzgefälle zwischen den Meeresforschungseinrichtungen fügt der globalen Ungleichheit in der *scientific community* eine weitere Dimension hinzu, die sich indirekt auf den wissenschaftlichen Erfolg auswirken könnte. Dies wäre der Fall, wenn die Webpräsenz denjenigen Einrichtungen eine stärkere Rezeption der Forschungsergebnisse, mehr Kooperationsoptionen o.ä. verschafft, die mit entsprechendem finanziellem und technischem Einsatz die Möglichkeiten ihrer Website ausreizen, um ihre Projekte, Aktivitäten und Publikationen vorzustellen und umfangreiche Datensammlungen zu präsentieren.

Im Hinblick auf die *Sichtbarkeit* der Einrichtungen zeigen sich sowohl im World Wide Web als auch im Web of Science größere Disparitäten als bei den Präsenzmaßen. Die Anzahl der Domains, die Inlinks auf die untersuchten Websites enthalten, steigt mit dem Entwicklungsstand der Länder, in denen die Meeresforschungseinrichtungen liegen, stark an. Fast in ähnlichem Maß gilt dies auch für die Zitationen, die die Publikationen der Institute erhalten.

Nicht eindeutig zu beantworten ist die Frage, welchen Anteil die *Erhebungsinstrumente* an den gefundenen Divergenzen haben (vgl. Kapitel 1.1). In der Fallstudie wurde die Einbindung der zehn Meeresforschungseinrichtungen in globale Kommunikationsstrukturen mit Hilfe des Web of Science, des SocSciBot und der Suchmaschine Yahoo untersucht. Im Rahmen dieses Forschungsdesigns lässt sich die faktische strukturelle Integration der Institute nicht von dem Bild trennen, das die Erhebungsinstrumente liefern. Allerdings können einige begründete Annahmen formuliert werden: Bei dem Abgleich der institutseigenen Publikationslisten für das Jahr 2006 mit den Einträgen im Web of Science wurden weder starke noch systematische Unterschiede zwischen den Meeresforschungseinrichtungen gefunden (vgl. Tabelle 60). Ein Bias zugunsten der US-amerikanischen oder deutschen Institute wurde nicht festgestellt. Vielmehr gab es reale Unterschiede in der Publikationsaktivität der Einrichtungen.

Die gefundenen Korrelationen zwischen den Zitationen, die die Meeresforschungseinrichtungen erhalten, und dem Entwicklungsstand ihres Sitzlandes decken sich mit einem früheren Befund von King (2004, Tabelle 1). Ihm zufolge machten Artikel mit (Ko-)AutorInnen aus den USA in den Jahren 1997 bis 2001 etwa ein Drittel aller Publikationen im Web of Science aus; sie erhielten jedoch die Hälfte aller Zitationen bis 2002 (vgl. Abschnitt 3.1.2). Die Veröffentlichungen anderer Nationen, wie z.B. Südafrikas oder Indiens, wurden dagegen deutlich seltener zitiert, als angesichts ihrer Präsenz im Web of Science zu erwarten gewesen wäre (King 2004, Tabelle 1). Kenia, Pakistan und Mauritius waren aufgrund ihrer geringen Präsenz im SCI nicht Bestandteil der Analyse. Während die EU-Staaten im Vergleich zu dem Vierjahreszeitraum 1993 bis 1997 stärker Anschluss an die USA finden, hat sich das Gefälle gegenüber anderen Ländern erhöht:

„The nations with the most citations are pulling away from the rest of the world“ (King 2004: 311).

King (2004: 311) diagnostiziert einen exponentiellen Zusammenhang zwischen dem Verhältnis der Zitationen zu den Publikationen einer Nation und ihrem Pro-Kopf-Einkommen, bietet aber keine Erklärung an. Als mögliche Ursachen ist die inhaltliche Qualität der Veröffentlichungen ebenso zu diskutieren wie der quantitative Stellenwert von Zitationen innerhalb nationaler Forschungsräume (oder der EU). Hinzu kommt, dass es internationale Unterschiede in der Bekanntheit und der Reputation von Universitäten und Forschungseinrichtungen gibt. Hier verfügen die USA mit ihrem einzigartigen Forschungssystem über eine beson-

dere Position. Abschließend sei nochmals auf die Rolle des World Wide Webs hingewiesen, das gezielt genutzt werden kann, um den eigenen Publikationen über Institutswebsites, Preprintforen oder *repositories* Aufmerksamkeit zu verschaffen. Vor dem Hintergrund der oben diskutierten Ergebnisse ist zu vermuten, dass ForscherInnen in den USA oder in Deutschland in der Lage sind, diese Möglichkeiten besonders intensiv zu nutzen.

Für den SocSciBot kann angenommen werden, dass er keinen Bias zugunsten bestimmter Länder aufweist: Er hat alle Internetauftritte systematisch mit demselben Algorithmus erfasst, nachdem die Eignung der Websites manuell geprüft wurde. Somit kann man für die Befunde zur Präsenz – sowohl im Web of Science als auch im World Wide Web – von einem umfassenden, wirklichkeitsnahen Bild ausgehen.

Ungewiss ist dagegen die Situation bei den Inlinkdaten: Im Rahmen dieser Fallstudie ist nicht auszumachen, inwiefern die stark ungleichen Inlinkwerte die Struktur des World Wide Webs realistisch wiedergeben oder ob auch die ungleichmäßige Erfassung des Webs durch die verwendete Suchmaschine eine Rolle spielt. Für letzteres spricht der empirische Befund von Vaughan und Thelwall (2004), dass Suchmaschinen ältere Webbereiche besser abdecken als neue, so dass Länder mit einer langen Tradition der Webnutzung präsenter erscheinen (vgl. Vaughan/Thelwall 2004). Möglicherweise lassen sich die eingangs unterschiedenen Aspekte der faktischen strukturellen Einbettung und deren Abbildung durch Erhebungsinstrumente nicht vollständig trennen, da die Suchmaschinen für viele NutzerInnen den Eingang zum World Wide Web darstellen und somit eine starke Lenkungsfunktion besitzen.

Auch aus den unterschiedlichen *Standardisierungsweisen* lassen sich Schlussfolgerungen ziehen. Alle bisher beschriebenen Werte wurden an der personellen Institutsgröße standardisiert. Alternativ wurden in der Fallstudie alle Werte ins Verhältnis zur Websitegröße gesetzt, also durch die Anzahl der Webseiten dividiert. Diese Art der Standardisierung nivelliert die Zusammenhänge zwischen Inlinkmaßen und Human Development Index fast durchgängig. Im Verhältnis zu ihrer Websitegröße erhalten die untersuchten Forschungseinrichtungen ähnlich viele Inlinks. Im Verhältnis zu ihrer Institutsgröße ist die Sichtbarkeit der Institute in den USA, aber auch in Deutschland wesentlich höher als die der Einrichtungen in Südafrika, Mauritius, Indien, Kenia und Pakistan. Dieses Ergebnis hat ambivalente Implikationen für die Institute in Entwicklungsländern: Auch sehr kleine Internetauftritte werden wahrgenommen. Insofern lohnt sich die Mühe, zumindest eine minimale Webpräsenz einzurichten. In Relation zu ihrem Umfang sind sie nicht seltener das Ziel von Links als Websites in Deutschland oder den USA. Die Ungleichheit im Hinblick auf die Sichtbarkeit ist offensichtlich auch eine Folge der geringen Größe der Internetauftritte, so dass auch ein gezielter Ausbau lohnenswert scheint. Allein mit etwas mehr Mitteln und einer guten Strategie wird sich allerdings nicht eine ähnliche Sichtbarkeit erreichen lassen, wie sie die deutschen und US-

amerikanischen Meeresforschungseinrichtungen im Verhältnis zur Institutsgröße aufweisen. Hier scheinen globale Disparitäten eine Rolle zu spielen, die sich auf Quantität und Qualität des Internetzugangs erstrecken und die nicht durch das Handeln Einzelner zu beheben sind. Einen möglichen Ansatzpunkt für gezielte Maßnahmen stellt dagegen der Aufbau informationstechnologischer Kompetenz an Forschungseinrichtungen in Entwicklungsländern dar.

Es lässt sich festhalten, dass Webmaße die internationale Ungleichheit stärker abbilden als Publikationsmaße. Außerdem hängen web- ebenso wie publikationsbasierte Sichtbarkeitsmaße deutlich stärker mit dem Entwicklungsstand zusammen als die jeweiligen Präsenzmaße. Insofern hat sich die Vermutung, die bei der Diskussion von Leitfrage 2 formuliert wurde (s.o. in diesem Kapitel), bestätigt: Die globale, durch Inlinks abgebildete Aufmerksamkeit im Web konzentriert besonders auf Nordamerika und Europa. In der Einleitung wurde eine These von Aguillo et al. (2006) angeführt, derzufolge webometrische Indikatoren besser geeignet sind als bibliometrische, um Universitäten in Entwicklungsländern zu erfassen. Die Befunde dieser Fallstudie zeigen das genaue Gegenteil. Insgesamt sprechen die Ergebnisse dafür, dass das World Wide Web die internationale scientific community nicht in ein virtuelles *global research village* geführt hat. Vielmehr scheint das Einrichten und Pflegen einer Webpräsenz eine Anforderung zu sein, die Forschungseinrichtungen in Industrieländern mit Hilfe von spezialisiertem Personal besser bewältigen und für sich nutzen können als solche in Entwicklungsländern. Dort erschwert die *digital divide* im Hinblick auf Kosten und Qualität des Internetzugangs diese zusätzliche, wissenschaftsfremde Aufgabe, die die ohnehin knappen Forschungsressourcen belastet. Bisher verschärft das World Wide Web somit die bestehende Ungleichheit in der globalen Forschungslandschaft, eher als dass es sie mildert. Einer Forschungseinrichtung, die ihre eigene Existenz nicht oder nur in minimalem Umfang im World Wide Web anzeigen und Kontaktmöglichkeiten nennen kann, entgehen mit großer Wahrscheinlichkeit Kooperationsoptionen. Aus diesem Grund erscheint es ratsam, dieses Feld stärker in das *capacity building* einzubeziehen und gezielte Programme zu entwickeln.

Für die quantitative Wissenschaftsforschung ist diese Fallstudie ein weiterer Beleg dafür, dass die Webometrie keinen Ersatz für die Bibliometrie darstellt. Sicherlich wird die Bibliometrie der Entwicklung folgen, dass Publikationen zunehmend online verbreitet werden, und entsprechende webbasierte Erhebungsformen entwickeln. Webometrische Studien im Wissenschaftskontext besitzen mit dem World Wide Web einen eigenständigen Forschungszugang und bringen spezifische Ergebnisse hervor. Beide Methoden können sich nicht gegenseitig ersetzen. Ihre unterschiedlichen Perspektiven können sich jedoch ergänzen.

6.2 Zur Praktikabilität webometrischer Analysen

Die Praktikabilität eines Verfahrens wird auch durch den Aufwand bestimmt, der betrieben werden muss, um Daten von hinreichender Qualität zu gewinnen. Der logistische Aufwand zur Erhebung der Daten durch Suchmaschinen und Webcrawler war in dieser Studie gering, hoch dagegen der Aufwand für die Datenbereinigung, die maßgeschneiderte Software-Skripte sowie manuelle Prüfung erforderte.

Bei der Inlinksuche nannte Yahoo insgesamt gut 50.000 Treffer, gab jedoch nur 13.561 aus. Die angegebene Trefferzahl ging somit größtenteils auf Überschätzung und Dubletten zurück. Folglich sollte man sich nicht auf die bloßen Trefferangaben einer Suchmaschine verlassen, sondern Aufwand für die Säuberung der Daten einkalkulieren. Webcrawlerdaten erfordern ebenfalls eine Datenbereinigung sowie eine sorgfältige Vorbereitung der Erhebung, da die Größe von Websites schlecht im Voraus abzuschätzen ist, nicht jeder Internetauftritt geeignet ist und gegebenenfalls Teile ausgeschlossen werden müssen. Das Sammeln und Säubern bibliometrischer Daten aus dem Web of Science hat sich im Vergleich dazu als deutlich unaufwendiger erwiesen. Hinzu kommt, dass insbesondere bei der Nutzung von Suchmaschinen zur Datenerhebung weitere Einschränkungen und Unklarheiten bestehen, die nicht behoben werden können (vgl. Abschnitt 2.4). Durch die Kombination aus schnellem und freiem Zugang zu Webdaten und dem großen Bereinigungsbedarf besteht hier – stärker als bei der Bibliometrie – die Gefahr, dass Personen ohne hinreichende Methodenkenntnis Daten erheben und unsachgemäß verwenden.

Eine Analyse von Daten zur Rezeption der untersuchten Websites durch Internet-NutzerInnen hätte die Linkdaten auf interessante Weise ergänzen können. Gerade im internationalen Vergleich wäre es spannend gewesen zu prüfen, inwiefern sich die faktische Website-nutzung von inlinkbasierten Sichtbarkeitsmaßen unterscheidet, da das Besuchen einer Website im Hinblick auf Zeit, Ressourcen und Know-how weniger voraussetzungsvoll ist als das Einrichten von Links. Nutzungsdaten werden in der Regel über Logfiles gewonnen, deren Auswertung das Einverständnis der Website-BetreiberInnen sowie spezifische Fachkenntnisse voraussetzt und erheblichen Zeitaufwand erfordert. Eine einfachere Alternative stellt die Nutzung des Dienstes Alexa Internet dar, der Nutzungsinformationen über eine Suchmaschinenleiste bündelt. Er wurde in dieser Fallstudie jedoch aus methodologischen und forschungsethischen Gründen ausgeschlossen (vgl. Abschnitt 4.4.3). Stattdessen wurde ein dritter Weg erprobt, der zwar der Kooperation der Institute, aber nur geringen Forschungsaufwands bedarf: Die Institute wurden gebeten, Nutzungsstatistiken für ihre Websites zur Verfügung zu stellen, die häufig mit Hilfe entsprechender Software geführt werden, um die Rezeption und Funktionalität des eigenen Internetauftritts zu beobachten. Es stellte sich heraus, dass die Dokumentation dieser Statistiken weder bei den Meeres-

forschungseinrichtungen in Industriestaaten noch bei denen in Entwicklungsländern Standard war und dass die Vergleichbarkeit der Daten, die die verwendeten Programme lieferten, zu gering war. Aus diesem Grund wird hier das Fazit gezogen, dass es derzeit keinen gut zugänglichen Weg gibt, mit geringem Aufwand die Rezeption von Websites auszuwerten.

6.3 Ausblick

Das Interesse an der Webometrie als Instrument für die Wissenschaftsforschung scheint nach einem Boom zu Beginn des Jahrtausends inzwischen nachzulassen. Zwar erscheinen nach wie vor empirische Studien zu einzelnen Ländern oder Disziplinen; eine ausgeprägte Fokussierung der Webometrie auf die wissenschaftliche Nutzung des World Wide Webs findet sich jedoch nicht länger. Stattdessen verlagert sich der Schwerpunkt zunehmend auf die Erforschung des Web 2.0, z.B. durch die Analyse von Blogs und sozialen Netzwerken wie Facebook oder Myspace (vgl. Abschnitt 2.2.1).

Für die Wissenschaftsforschung bilden sich mit der zunehmenden Verbreitung von Open Access und webbasierten Volltextliteraturdatenbanken (*digital repositories*) neue Ansatzpunkte für die Untersuchung wissenschaftlicher Publikationen. Auch Google Scholar – eine kostenlos nutzbare Datenbank, die wissenschaftliche Publikationen sowie Zitationen umfasst und neben Fachartikeln auch Beiträge in Sammelbänden, Monographien, Qualifikationsarbeiten u.a. enthält – bietet ebenfalls neue Optionen für webbasierte bibliometrische Analysen. Allerdings setzen Datenqualität und Abfragemöglichkeiten der wissenschaftlichen Nutzung bisher enge Grenzen.

In dieser Dissertation wurde speziell untersucht, inwieweit webometrische Verfahren zur Untersuchung internationaler Strukturen in der Wissenschaft geeignet sind. Die Limitierungen, die dabei aufgezeigt wurden, sprechen dafür, dass ihre Aussagekraft in dieser Hinsicht begrenzt ist (vgl. Abschnitt 6.1). Es gibt jedoch zwei Einsatzbereiche, in denen sie weiterhin nützlich sein können: Erstens können webometrische Daten mit geringem Aufwand einen groben ersten Eindruck von wissenschaftlichen Strukturen geben. Zweitens stellt die Internetpräsenz von Forschungseinrichtungen einen eigenständigen Forschungsgegenstand dar, über den die Webometrie detaillierte Informationen bereitstellen kann. Nicht allen Hochschulen und anderen wissenschaftlichen Einrichtungen gelingt es gleichermaßen, im World Wide Web präsent und sichtbar zu sein. Die Unterschiede zu identifizieren und zu erklären ist eine vielversprechende, praxisrelevante Forschungsaufgabe. Einzelne wissenschaftliche Organisationen können ebenso wie die Forschungs- oder Entwicklungspolitik Schlussfolgerungen aus solchen Informationen ziehen. Selbstverständlich sind auch hier bei Datenerhebung und -analyse besondere Sorgfalt und genaue Methodenkenntnis erforderlich.

In der vorliegenden Dissertation wurde weiterer Forschungsbedarf aufgezeigt. Für die Verwendung der Webometrie zur Untersuchung internationaler Zusammenhänge in der Wissenschaft sind beispielsweise folgende Fragen relevant:

- Wie unterscheiden sich Websites von Forschungseinrichtungen in Industrie- und Entwicklungsländern inhaltlich?
- Sind nationale und internationale Links hinsichtlich ihrer Funktionen vergleichbar?
- Welchen Stellenwert nehmen nicht-wissenschaftsbezogene Links auf den Websites wissenschaftlicher Einrichtungen ein?
- Wie zuverlässig sind die sprachspezifischen Rechercheoptionen von Suchmaschinen?
- Können TLDs und SLDs als Indikatoren der sektoralen Verortung und speziell zur Abgrenzung von Wissenschaft eingesetzt werden?

Sämtliche Fragen sind mit hohem manuellem Aufwand verbunden und erfordern qualitative Analysen. Ihre Klärung ist notwendig, wenn webometrische Indikatoren zukünftig in internationalen Dimensionen eingesetzt werden sollen.

Dass dies wahrscheinlich ist, zeigt beispielsweise der Erfolg des Rankings der Internetauftritte von Universitäten auf der ganzen Welt des spanischen Cybermetrics Lab unter www.webometrics.info. Wenngleich webometrische Indikatoren vermutlich auf absehbare Zeit nicht Bestandteil von Evaluations-, Steuerungs- und Mittelvergabeprozessen sein werden, reihen sie sich offensichtlich doch in den Kanon der Wissenschaftsstatistik ein. So plant die Europäische Kommission, einen Abschnitt zu Webindikatoren in ihre Serie *European Report on Science and Technology Indicators* aufzunehmen (Thelwall 2009b).

Die weitere Prüfung der methodologischen Aussagekraft dieser interpretationsbedürftigen Kennzahlen sowie ihrer Fehleranfälligkeit ist unabdingbar, um zu einem kompetenten Umgang mit dieser Methode und zu einer angemessenen Abbildung und Interpretation wissenschaftlicher Webstrukturen zu gelangen.

7. Literaturverzeichnis

- Abraham, Ralph H. (1996): „*Webometry: measuring the complexity of the World Wide Web*“, <http://www.ralph-abraham.org/vita/redwood/vienna.html>, zuletzt abgerufen am 19.11.09.
- Adam, Lishan; Wood, Frances (1999): „An investigation of the impact of information and communication technologies in Sub-Saharan Africa“, *Journal of Information Science* 25 (4), 307-318.
- Aguillo, Isidro F. (2002): „Cybermetrics: definitions and methods for an emerging discipline“, *Seminaires de l'ADEST*, Paris, 14.02.2002, <http://www.upmf-grenoble.fr/adept/seminaires/ISIDRO/Cybermetrics.ppt>, zuletzt abgerufen am 19.11.09.
- Aguillo, Isidro F.; Granadino, Begoña; Ortega, Jose Luis; Prieto, José Antonio (2005): „What the internet says about science“, *The Scientist* 19 (14), 18.07.05, 10.
- Aguillo, Isidro F.; Granadino, Begoña; Ortega, Jose Luis; Prieto, José Antonio (2006): „Scientific research activity and communication measured with cybermetrics indicators“, *Journal of the American Society for Information Science and Technology* 57 (10), 1296-1302.
- Aguillo, Isidro F.; Ortega, José Luis; Fernández, Mario; Utrilla, Ana M. (2008): „Web mediators for cybermetric purposes: A comparative analysis“, in: Kretschmer, Hildrun; Frank Havemann: *Proceedings of the Fourth International Conference on Webometrics, Informetrics, and Scientometrics & Ninth COLLNET Meeting*, 28.07.-01.08.08, Berlin, <http://www.collnet.de/Berlin-2008/AguilloWIS2008wmc.pdf>, zuletzt abgerufen am 19.11.09.
- Al-Ansari, Husain (2005): „Internet use by the faculty members of Kuwait University“, *The Electronic Library* 24 (6), 791-803.
- Almind, Tomas C.; Ingwersen, Peter (1996): „Informetric analysis on the World Wide Web: A methodological approach to ‚internetometrics‘“, *CIS Report 2*, Kopenhagen: Centre for Informetric Studies, Royal School of Information Science.
- Almind, Tomas C.; Ingwersen, Peter (1997): „Informetric analyses on the World Wide Web: methodological approaches to ‚webometrics‘“, *Journal of Documentation* 53 (4), 404-426.
- Aminpour, Farzaneh; Kabiri, Payam; Otraj, Zahra; Keshtkar, Abbas Ali (2009): „Webometric analysis of Iranian universities of medical sciences“, *Scientometrics* 80 (1), 253-264.
- Angus, Emma; Stuart, David; Thelwall, Mike (2008): „General patterns of tag usage among university groups in Flickr“, *Online Information Review*, 32 (1), 89-101.
- Arakaki, Mónica; Willett, Peter (2009): „Webometric analysis of departments of librarianship and information science: a follow-up study“, *Journal of Information Science* 35 (2), 143-152.
- Arunachalam, Subbiah (1999): „Information and knowledge in the age of electronic communication: A developing country perspective“, *Journal of Information Science* 25 (6), 465-476.
- Azzi, Angelo (2005): „Scientific publishing in non industrialized countries: A pilot wireless Internet project for Africa“, *IUBMB Life* 57 (4/5), 259-261.
- Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier (1999): *Modern Information Retrieval*, New York: ACM Press.

- Bar-Ilan, Judit (1998/99): „Search engine results over time. A case study on search engine stability“, *Cybermetrics* 2/3 (1), Paper 1, <http://www.cindoc.csis.es/cybermetrics/articles/v2i1p1.html>, zuletzt abgerufen am 19.11.09.
- Bar-Ilan, Judit (2001): „Data collection methods on the Web for informetric purposes – A review and analysis“, *Scientometrics* 50 (1), 7-32.
- Bar-Ilan, Judit (2002a): „How much information do search engines disclose on the links to a web page? A longitudinal case study of the ‘Cybermetrics’ home page“, *Journal of Information Science* 28 (6), 455-466.
- Bar-Ilan, Judit (2002b): „Methods for measuring search engine performance over time“, *Journal of the American Society for Information Science and Technology* 53 (4), 308-319.
- Bar-Ilan, Judit (2004a): „A microscopic link analysis of academic institutions within a country – the case of Israel“, *Scientometrics* 59 (3), 391-403.
- Bar-Ilan, Judit (2004b): „The use of web search engines in information science research“, *Annual Review of Information Science and Technology* 38, 231-288.
- Bar-Ilan, Judit (2005a): „Comparing rankings of search results on the Web“, *Information Processing and Management* 41 (6), 1511-1519.
- Bar-Ilan, Judit (2005b): „Expectations versus reality – web search engines at the beginning of 2005“, in: Ingwersen/Larsen 2005, 87-96.
- Bar-Ilan, Judit (2005c): „What do we know about links and linking? A framework for studying links in academic environments“, *Information Processing and Management* 41 (4), 973-986.
- Bar-Ilan, Judit; Fink, Noa (2005): „Preference for electronic format of scientific journals. A case studies of the Science Library users at Hebrew University“, *Library and Information Science Research* 27 (3), 363-376.
- Bar-Ilan, Judit; Gutman, Tatyana (2005): „How do search engines respond to some non-English queries?“, *Journal of Information Science* 31 (1), 13-28.
- Bar-Ilan, Judit; Peritz, Bluma C. (2004): „Evolution, continuity, and disappearance of documents on a specific topic on the web: a longitudinal study of ‚Informetrics‘“, *Journal of the American Society for Information Science and Technology* 55 (11), 980-990.
- Bar-Ilan, Judit; Peritz, Bluma C.; Wolman, Yechezkel (2003): „A survey on the use of electronic databases and electronic journals accessed through the web by the academic staff of Israeli universities“, *Journal of Academic Librarianship* 29 (6), 346-361.
- Barjak, Franz; Thelwall, Mike (2008): „A statistical analysis of the web presences of European life sciences research teams“, *Journal of the American Society for Information Science and Technology* 59 (4): 628-643.
- Beaulieu, Anne (2005): „Sociable hyperlinks: An ethnographic approach to connectivity“, in: Christine Hine (Hrsg.) (2005): *Virtual methods: Issues in social research on the Internet*. Oxford und New York: Berg, 183-197.
- Beaver, Donald DeB. (2001): „Reflections on scientific collaboration (and its study): past, present, and future“, *Scientometrics* 52 (3), 365-377.
- Beaver, Donald DeB.; Rosen, Richard (1978): „Studies in scientific collaboration. Part I. Professional origins of scientific co-authorship“, *Scientometrics* 1 (1), 65-84.

- Beaver, Donald DeB.; Rosen, Richard (1979): „Studies in scientific collaboration. Part II. Scientific co-authorship, research productivity and visibility in the French scientific elite“, *Scientometrics* 1 (2), 133-149.
- Bergman, Michael K. (2001): „The Deep Web: Surfacing hidden value“, *Journal of Electronic Publishing* 7 (1), <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0007.104>, zuletzt abgerufen am 19.11.09.
- Berners-Lee, T.; Cailliau, R.; Groff, J.-F. (1992): „The world-wide web“, *Computer Networks and ISDN Systems* 25 (4/5), 454-459.
- Berners-Lee, Tim (1989): „Information management: a proposal“, <http://www.w3.org/History/1989/Proposal.html>, zuletzt abgerufen am 19.11.09.
- Berners-Lee, Tim; Cailliau, Robert (1990): „WorldWideWeb: proposal for a hypertext project“, <http://www.w3.org/Proposal.html>, zuletzt abgerufen am 19.11.09.
- Berners-Lee, Tim; Hendler, James (2001): „Publishing on the semantic web“, *Nature* 410, 26. April 2001, 1023-1024.
- Bernhofer, Martin (2001): „Cyberscience – Was macht die Wissenschaft im Internet?“, *Gegenworte. Zeitschrift für den Disput über Wissen* (8), 27-32.
- Bharat, Krishna; Chang, Bay-Wie; Henzinger, Monika; Ruhl, Matthias (2001): „Who links to whom: mining linkage between Web sites“, <http://people.csail.mit.edu/ruhl/papers/2001-icdm.pdf>, zuletzt abgerufen am 19.11.09.
- Björneborn, Lennart (2004): *Small-world link structures across an academic web space: a library and information science approach*. PhD thesis from the Department of Information Studies, Royal School of Library and Information Science, Denmark. Im Web zu finden unter: <http://vip.db.dk/lb/phd/phd-thesis.pdf>, zuletzt abgerufen am 19.11.09.
- Björneborn, Lennart; Ingwersen, Peter (2001): „Perspectives of webometrics“, *Scientometrics* 50 (1), 65-82.
- Björneborn, Lennart; Ingwersen, Peter (2004): „Towards a basic framework for webometrics“, *Journal of the American Society for Information Science and Technology* 55 (14), 1216-1227.
- Bordons, María; Gómez, Isabel (2000): „Collaboration networks in science“, in: Blaise Cronin; Helen Barsky Atkins (Hrsg.): *The web of knowledge: a festschrift in honour of Eugene Garfield*. Medford (New Jersey): Information Today, 197-213.
- Bordons, María; Gómez, Isabel; Fernández, M. Teresa; Zulueta, M. Angeles; Méndez, Aida (1996): „Local, domestic and international collaboration in biomedical research“, *Scientometrics* 37 (2), 279-295.
- Borgman, Christine L.; Furner, Jonathan (2002): „Scholarly communication and bibliometrics“, *Annual Review of Information Science and Technology* 36, 3-72.
- Bossy, Marcia J. (1995): „The last of the litter: ‚Netometrics‘“, <http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.html>, zuletzt abgerufen am 19.11.09.
- Brooks, Sam; Donovan, Paul; Rumble, Clint (2005): „Developing nations, the digital divide and research databases“, *Serials Review* 31 (4), 270-278.
- Brooks, Terrence A. (2003): „Web search: how the Web has changed information retrieval“, *Information Research* 8 (3), <http://InformationR.net/ir/8-3/paper154.html>, zuletzt abgerufen am 19.11.09.
- Brown, Cecilia (2004): „The Matthew effect of the *Annual Reviews* series and the flow of scientific communication through the World Wide Web“, *Scientometrics* 60 (1): 25-36.

- Brumfiel, Geoff (2004): "As one door closes...", *Nature* 427, 15. Januar 2004, 190-195.
- Bush, Vannevar (1945): „As we may think“, *The Atlantic Monthly* (176), 101-108.
- Carpenter, Mark P.; Narin, Francis (1981): „The adequacy of the Science Citation Index (SCI) as an indicator of international scientific activity“, *Journal of the American Society for Information Science* 32, 430-439.
- Castells, Manuel (2005): *Die Internet-Galaxy: Internet, Wirtschaft und Gesellschaft*, Wiesbaden: Verlag für Sozialwissenschaften.
- Chakrabarti, Soumen; Joshi., Mukul M.; Punera, Kunul; Pennock, David M. (2002): „The structure of broad topics on the Web“, *Proceedings of the WWW 2002 – the 11th World Wide Web Conference*, <http://www2002.org/CDROM/refereed/338/>, zuletzt abgerufen am 19.11.09.
- Chu, Heting (2005): „Taxonomy of inlinked Web entities: What does it imply for webometric research?“, *Library and Information Research* 27 (1): 8-27.
- Chung, Young Mee; Yu, So Young; Kim, Yong Kwang; Kim, Su Yeon (im Erscheinen): „Characteristics and link structure of a national scholarly Web space: The case of South Korea“, *Scientometrics*.
- Cothey, Viv (2004): „Web-crawling reliability“, *Journal of the American Society for Information Science and Technology* 55 (14), 1228-1238.
- Cothey, Viv (2005): „Some preliminary results from a link-crawl of the European Union Research Area Web“, in: Ingwersen/Larsen 2005, 212-220.
- Crane, Diana (1972): „Transnational networks in basic science“, in: Robert O. Keohane; Joseph S. Nye (Hrsg.) (1972): *Transnational relations and world politics*, Cambridge (Mass.): Harvard University Press, 235-251.
- Crawford, Elisabeth; Shinn, Terry; Sörlin, Sverker (1993): „The nationalization and denationalization of the sciences: An introductory essay“, in: Dies., *Denationalizing science. The contexts of international scientific practice*, Dordrecht et al: Kluwer Academic Publishers, 1-42.
- Cronin, Blaise (2001): „Bibliometrics and beyond: some thoughts on web-based citation analysis“, *Journal of Information Science* 27 (1), 1-7.
- Cronin, Blaise; Snyder, Herbert W.; Rosenbaum, Howard; Martinson, Anna; Callahan, Eva (1998): „Invoked on the Web“, *Journal of the American Society for Information Science* 49 (14), 1319-1328.
- Cyr, Dianne; Trevor-Smith, Haizley (2004): „Localization of web design: An empirical comparison of German, Japanese, and United States web site characteristics“, *Journal of the American Society for Information Science and Technology* 55 (13), 1199-1208.
- Dahdouh-Guebas, Farid; Ahimbisibwe, J.; van Moll, Rita; Koedam, Nico (2003): „Neo-colonial science by the most industrialised upon the least developed countries in peer-review publishing“, *Scientometrics* 56 (3), 329-343.
- Dastidar, Prabis G. (2004): „Ocean science & technology research across the countries: A global scenario“, *Scientometrics* 59 (1), 15-27.
- Davidson, Theresa; Sooryamoorthy, R.; Shrum, Wesley (2002): „Kerala connections: Will the Internet affect science in developing areas“, in: Barry Wellman; Caroline Haythornthwaite (Hrsg): *The Internet in everyday life*, Oxford: Blackwell Publishing, 496-519.
- de Roy, Olivier Coeur (1997): „The African challenge: Internet, networking and connectivity activities in a developing environment“, *Third World Quarterly* 18 (5), 883-898.

- Dedijer, Stevan (1963): „Underdeveloped science in underdeveloped countries“, *Minerva* 2 (1), 61-81.
- Degen, Horst; Lorscheid, Peter (2002): *Statistik-Lehrbuch mit Wirtschafts- und Bevölkerungsstatistik*, München und Wien: Oldenbourg.
- Der Landesbeauftragte für Datenschutz und Informationsfreiheit Bremen (2007): „Spyware – Spione sind unter uns“, http://www.datenschutz-bremen.de/sv_internet/spyware.php, zuletzt abgerufen am 19.11.09.
- DiMaggio, Paul; Hargattai, Eszter; Neuman, W. Russell; Robinson, John (2001): „Social Implications of the Internet“, *Annual Review of Sociology* 27 (1), 307-336.
- Dodge, Martin (1999): „The geography of cyberspace“, *CASA Working Paper 8*, Centre for Advanced Spatial Analysis, University College London, http://www.casa.ucl.ac.uk/working_papers/paper8.pdf, zuletzt abgerufen am 19.11.09.
- Dodge, Martin; Kitchin, Robert (2002): „New cartographies to chart cyberspace“, *Geoinformatics* 5 (April/May), 38-41.
- Drori, Gili S.; Jang, Yong Suk (2003): „The Global Digital Divide: A Sociological Assessment of Trends and Causes“, *Social Science Computer Review* 21 (2), 144-161.
- Drori, Gili; Meyer, John W.; Ramirez, Francisco O.; Schofer, Evan (2003): *Science in the modern world polity*, Stanford: Stanford University Press.
- Dumlao, Rebecca; Duke, Shearlean (2003): „The Web and e-mail in science communication“, *Science Communication* 24 (3), 283-308.
- Duque, Ricardo B.; Ynalvez, Marcus; Sooryamoorthy, R.; Mbatia, Paul; Dzorgbo, Dan-Bright S.; Shrum, Wesley (2005): „Collaboration paradox: Scientific productivity, the Internet, and problems of research in developing areas“, *Social Studies of Science* 35 (5), 755-785.
- EC (Hrsg.) (2003): *Third European Report on Science and Technology Indicators 2003. Towards a knowledge-based economy*, Luxembourg: Office for Official Publications for the European Communities.
- EC (Hrsg.) (2009): „The .eu Top Level Domain“, http://ec.europa.eu/information_society/policy/doteu/index_en.htm, zuletzt abgerufen am 19.11.09.
- Egghe, Leo (2000): „New informetric aspects of the Internet: some reflections – many problems“, *Journal of Information Science* 26 (5), 329-335.
- Ehikhamenor, Fabian A. (2003a): „Internet facilities: use and non-use by Nigerian university scientists“, *Journal of Information Science* 29 (1), 35-48.
- Ehikhamenor, Fabian A. (2003b): „Internet resources and productivity in scientific research in Nigerian universities“, *Journal of Information Science* 29 (2), 107-116.
- Eisend, Martin (2002): „The Internet as a new medium for the sciences? The effects of Internet use on traditional scientific communication media among social scientists in Germany“, *Online Information Review* 26 (5), 307-317.
- Elzinga, Aant (1993): „Antarctica: The construction of a continent by and for science“, in: Elisabeth Crawford, Terry Shinn, Sverker Sörlin (1993): *Denationalizing science. The contexts of international scientific practice*, Dordrecht et al: Kluwer Academic Publishers, 73-106.
- Elzinga, Aant (1996): „Introduction: Modes of internationalism“. in: Aant Elzinga, Catharina Landström (Hrsg.): *Internationalism and science*. London, Los Angeles: Taylor Graham, 3-20.

- Engels, Anita; Ruschenburg, Tina (2006): „Die Ausweitung kommunikativer Räume: Reichweite, Mechanismen und Theorien der Globalisierung der Wissenschaft“, *Soziale Welt* 57 (1), 5-29.
- Engels, Anita; Ruschenburg, Tina; Weingart, Peter (2005): „Recent internationalization in global environmental change research in Germany and the U.S.“, *Scientometrics* 62 (1), 67-85.
- Felt, Ulrike; Nowotny, Helga; Taschwer, Klaus (1995): *Wissenschaftsforschung. Eine Einführung*, Frankfurt a.M./New York: Campus Verlag.
- Finholt, Thomas A. (2002): „Collaboratories“, *Annual Review of Information Science and Technology* 36, 73-107.
- Frame, J. Davidson; Carpenter, Mark P. (1979): „International Research Collaboration“, *Social Studies of Science* 9 (4): 481-497.
- Früh, Werner (1998): *Inhaltsanalyse. Theorie und Praxis*, 4. überarbeitete Auflage, Konstanz: UVK.
- Gallup, Dana M. (2003): „Moldova's .md: The little domain that roared“, in: Wass 2003a, 87-99.
- Gandhi, Tushar A. (2003): „India's .in: Underused and underappreciated“, in: Wass 2003a, 43-54.
- Garfield, Eugene (2002): [ohne Titel], Beitrag zur Mailinglist *Sigmatrics* vom 21.06.02, <http://listserv.vide.net/cgi-bin/wa?A2=ind0206&L=sigmatrics&D=1&O=D&F=&S=&P=5617>, zuletzt abgerufen am 06.01.06.
- Glänzel, Wolfgang (2001): „National characteristics in international scientific co-authorship relations“, *Scientometrics* 51 (1), 69-115.
- Glänzel, Wolfgang; Schubert, András (2005): „Domesticity and internationality in co-authorship, references and citations“, *Scientometrics* 65 (3), 323-342.
- Gläser, Jochen (2003): „What Internet use does and does not change in scientific communities“, *Science Studies* 16 (1), 38-51.
- Gouveia, Fábio C.; Kurtenbach, Eleonora (2009): „Mapping the web relations of science centres and museums from Latin America“, *Scientometrics* 79 (3), 491-505.
- Greenberg, Daniel S. (2002): „New restrictions on foreign students“, *The Lancet* 359, 18.05.02, 1757.
- Guillén, Mauro F.; Suárez, Sandra L. (2005): „Explaining the global digital divide: Economic, political and sociological drivers of cross-national Internet use“, *Social Forces* 84 (2), 681-708.
- Haas, Stephanie W.; Grams, Erika S. (2000): „Readers, authors, and page structure: A discussion of four questions arising from a content analysis of Web pages“, *Journal of the American Society for Information Science* 51 (2): 181-192.
- Harries, Gareth; Wilkinson, David; Price, Liz; Fairclough, Ruth; Thelwall, Mike (2004): „Hyperlinks as a data source for science mapping“, *Journal of Information Science* 30 (5), 436-447.
- Heimeriks, Gaston; Hörlesberger, Marianne; Van den Besselaar, Peter (2003): „Mapping communication and collaboration in heterogeneous research networks“, *Scientometrics* 58 (2), 391-413.
- Herring, Susan Davis (2001): „Using the World Wide Web for research: Are faculty satisfied?“, *The Journal of Academic Librarianship* 27 (3), 213-219.

- Hofmann, Jeanette (2005): „Internet Governance: Zwischen staatlicher Autorität und privater Koordination“, *Internationale Politik und Gesellschaft* (3), 10-29.
- Hofmann, Jeanette (2007): „Internet Corporation for Assigned Names and Numbers (ICANN)“, in: *Global Information Society Watch* 2007, 39-47.
- Holmberg, Kim; Thelwall, Mike (2009): „Local government web site in Finland: A geographic and webometric analysis“, *Scientometrics* 79 (1), 157-169.
- Holmes, Alison; Oppenheim, Charles (2001): „Use of citation analysis to predict the outcome of the 2001 Research Assessment Exercise for Unit of Assessment (UoA) 61: Library and Information Management“. *Information Research* 6 (2), <http://InformationR.net/ir/6-2/paper103.html>, zuletzt abgerufen am 19.11.09.
- Hornbostel, Stefan (1997): *Wissenschaftsindikatoren: Bewertungen in der Wissenschaft*, Opladen: Westdeutscher Verlag.
- Hurd, Julie M. (2000): „The transformation of scientific communication: a model for 2020“, *Journal of the American Society for Information Science* 51 (14), 1279-1283.
- IANA (Hrsg.) (2005): „Index by TLD code“, <http://www.iana.org/cctld/cctld-whois.htm>, zuletzt abgerufen am 11.09.06.
- ICANN (Hrsg.) (2008): „Biggest Expansion in gTLDs Approved for Implementation“, <http://www.icann.org/en/announcements/announcement-4-26jun08-en.htm>, zuletzt abgerufen am 19.11.09.
- ICANN (Internet Corporation for Assigned Names and Numbers) (Hrsg.) (2009): „New TLD program“, <http://www.icann.org/en/topics/new-gtld-program.htm>, zuletzt abgerufen am 19.11.09.
- Ingwersen, Peter (1998): „The calculation of web impact factors“, *Journal of Documentation* 54 (2), 236-243.
- Ingwersen, Peter; Larsen, Birger (Hrsg.) (2005): *Proceedings of ISSI 2005 - the 10th International Conference of the International Society for Scientometrics and Informetrics*, Band 1, Stockholm: Karolinska University Press.
- InternetLab (Hrsg.) (2006a): „Universities: Distribution by country“, http://www.webometrics.info/Distribution_by_Country.asp.htm, zuletzt abgerufen am 11.08.06.
- InternetLab (Hrsg.) (2006b): „Webometrics Ranking of World Universities“, <http://www.webometrics.info/top3000.asp.htm>, zuletzt abgerufen am 11.08.06.
- IODE (Hrsg.) (2009): „IODE is...“, http://www.ioode.org/index.php?option=com_content&task=view&id=2&Itemid=34, zuletzt abgerufen am 19.11.09.
- ISC (Hrsg.) (2009): „Internet Domain Survey“, <https://www.isc.org/solutions/survey>, zuletzt abgerufen am 19.11.09.
- Jappe, Arlette (2007): „Explaining international collaboration in global environmental change research“, *Scientometrics* 71 (3), 367-390.
- Jepsen, Erik Thorlund; Seiden, Piet; Ingwersen, Peter; Björneborn, Lennart; Borlund, Pia (2004): „Characteristics of scientific web publications: preliminary data gathering and analysis“, *Journal of the American Society for Information Science and Technology* 55 (14), 1239-1249.
- Katz, J. Sylvan (1994): „Geographic proximity and scientific collaboration“, *Scientometrics* 31 (1), 31-43.
- Katz, J. Sylvan (2000): „Scale independent indicators and research evaluation“, *Science and Public Policy* 27 (1), 23-36.

- Katz, J. Sylvan; Martin, Ben R. (1997): „What is research collaboration?“, *Research Policy* 26 (1), 1-18.
- Kennedy, Don (2002): „Balancing terror and freedom“, *Science* 298, 13.12.02, 2091.
- Kim, Hak Joon (2000): „Motivations for hyperlinking in scholarly electronic articles: A qualitative study“, *Journal of the American Society for Information Science* 51 (10), 887-899.
- King, David A. (2004): „The scientific impact of nations. What different countries get for their research spending“, *Nature* 430, 15. Juli 2004, 311-316.
- Kling, Rob; McKim, Geoffrey (2000): „Not just a matter of time: field differences and the shaping of electronic media in supporting scientific communication“, *Journal of the American Society for Information Science* 51 (14): 1306-1320.
- Koehler, Wallace (2002): „Web page change and persistence – a four-year longitudinal study“, *Journal of the American Society for Information Science and Technology* 53 (2), 162-171.
- Koku, Emmanuel F.; Nazer, Nancy; Wellman, Barry (2001): „Netting scholars: Online and offline“, *American Behavioral Scientist* 44 (10), 1752-1774.
- Koll, Thomas R. (2004) (Hrsg.): *WikiReader Internet. Eine Artikelsammlung aus Wikipedia, der freien Enzyklopädie*, <http://de.wikipedia.org/wiki/Wikipedia:WikiReader/Internet.pdf>, zuletzt abgerufen am 15.11.05.
- Koster, Martijn (ohne Jahr): „The Robots Exclusion Protocol“, <http://www.robotstxt.org/wc/exclusion.html>, zuletzt abgerufen am 28.03.06.
- Kousha, Kayvan; Thelwall, Mike (2008): „Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines“, *Scientometrics* 74 (2), 273-294.
- Kousha, Kayvan; Thelwall, Mike (2009): „Google book search: Citation analysis for Social Science and the Humanities“, *Journal of the American Society of Information Science and Technology* 68 (8), 1537-1549.
- Kretschmer, Hildrun; Aguillo, Isidro F. (2004): „Visibility of collaboration on the Web“, *Scientometrics* 61 (3), 405-426.
- Larson, Ray R. (1996): „Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace“, <http://Sherlock.berkeley.edu/asis96/asis96.html>, zuletzt abgerufen am 19.11.09
- Laudel, Grit (2002): „What do we measure by co-authorships?“, *Research evaluation* 11 (1), 3-16.
- Lawrence, Steve (2001): „Free online availability substantially increases a paper's impact“, *Nature* 411, 31. Mai 2001, 521-521.
- Lawrence, Steve; Giles, C. Lee (1999): „Accessibility of information on the Web“, *Nature* 400, 8. Juli 1999, 107-109.
- Leclerc, M.; Gagné, J. (1994): „International scientific collaboration: The continentalization of science“, *Scientometrics* 31 (3), 261-292.
- Lewandowski, Dirk (2005): „Web information retrieval“, *Information: Wissenschaft und Praxis* 56 (1), 5-12.
- Lewandowski, Dirk; Wahlig, Henry; Meyer-Bautor, Gunnar (2006): „The freshness of Web search engines' databases“, *Journal of Information Science* 32 (2), 131-148.

- Leydesdorff, Loet (2006): „Re: The Metaphor Unchained by Roald Hoffmann, American Scientist Sep/Oct 2006“, Beitrag zur Mailingliste *SIGMETRICS* vom 18.08.2006; auch abrufbar unter <http://listserv.utk.edu/cgi-bin/wa?A2=ind0608&L=sigmetrics&D=1&O=D&P=3042>, zuletzt abgerufen am 19.11.09.
- Leydesdorff, Loet; Curran, Michael (2000): „Mapping university-industry-government relations on the Internet: The construction of indicators for a knowledge-based economy“, *Cybermetrics* 4 (1), Paper 2, <http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html>, zuletzt abgerufen am 19.11.09.
- Li, Xuemei (2003): „A review of the developments and application of the Web impact factor“, *Online Information Review* 27 (6), 407-417.
- Li, Xuemei; Thelwall, Mike; Wilkinson, David; Musgrove, Peter (2005a): „National and international university departmental Web site interlinking. Part 1: Validation of departmental link analysis“, *Scientometrics* 64 (2), 151-185.
- Li, Xuemei; Thelwall, Mike; Wilkinson, David; Musgrove, Peter (2005b): „National and international university departmental Web site interlinking. Part 2: Link patterns“, *Scientometrics* 64 (2), 187-208.
- Lindén, Patrik (2003): „Sweden's .se: Reestablishing itself as the best choice for all Swedes“, in: Wass 2003a, 67-76.
- Liu, Mengxiong (1993): „The complexity of citation practice: A review of citation studies“, *Journal of Documentation* 49 (4), 370-408.
- Liu, Ziming; Huang, Xiaobin (2005): „Evaluating the credibility of scholarly information on the web: A cross cultural study“, *The International Information and Library Review* 37 (2), 99-106.
- Luukkonen, Terttu; Tijssen, Robert J.W.; Persson, Olle; Sivertsen, Gunnar (1993): „The measurement of international scientific collaboration“, *Scientometrics* 28 (1), 15-36.
- Luukkonen, Terttu; Persson, Olle; Sivertsen, Gunnar (1992): „Understanding patterns of international scientific collaboration“, *Science, Technology and Human Values* 17 (1), 101-126.
- Mayr, Philipp; Walter, Anne-Kathrin (2007): „An exploratory study of Google Scholar“, *Online Information Review* 31 (6), 814-2007.
- McKiernan, Gerry (ohne Jahr): *CitedSites(sm): Citation Indexing of Web Resources*, <http://www.public.iastate.edu/~CYBERSTACKS/Cited.htm>, zuletzt abgerufen am 19.11.09.
- Merten, Klaus (1983): *Inhaltsanalyse. Einführung in Theorie, Methode und Praxis*. Opladen: Westdeutscher Verlag.
- Merton, Robert K. (1972): „Wissenschaft und demokratische Sozialstruktur“, in: Peter Weingart (Hrsg.): *Wissenschaftssoziologie 1. Wissenschaft als sozialer Prozeß*, Frankfurt: Athenäum; 45-59.
- Mettrop, Wouter; Nieuwenhuysen, Paul (2001): „Internet search engines – fluctuations in document accessibility“, *Journal of Documentation* 57 (5), 623-651.
- Meyer, John W.; Ramirez, Francisco O. (2005): „Die globale Institutionalisierung der Bildung“, in: John Meyer (herausgegeben von Georg Krücken): *Weltkultur. Wie die westlichen Prinzipien die Welt durchdringen*, Frankfurt a.M.: Suhrkamp 212-234.
- Middleton, Iain; McConnell, Mike; Davidson, Grant (1999): „Presenting a model for the structure and content of a university World Wide Web site“, *Journal of Information Science* 25 (3), 219-227.

- Miles, Edward (1972): „Transnationalism in space“, in: Robert O. Keohane; Joseph S. Nye (Hrsg.) (1972): *Transnational relations and world politics*, Cambridge (Mass.): Harvard University Press, 252-275.
- Nelson, Michael (2005): „Academic home pages and nobel laureates“, in: Ingwersen/Larsen 2005, 193-196.
- Nelson, Theodor H. (1967): „Getting it out of our system“, in: G. Schechter (Hrsg.): *Information retrieval: A critical review*, Washington (D.C.): Thomson Books, 191-210.
- Nielsen Netratings (2007): „Nielsen//Natratings announces February U.S. search share rankings“, http://www.netratings.com/pr/pr_070320.pdf, zuletzt abgerufen am 21.03.2007.
- NSB (Hrsg.) (2004): *Science and Engineering Indicators 2004*, Band 1, Arlington (VA): National Science Foundation.
- NSB (Hrsg.) (2006): *Science and Engineering Indicators 2006*, Band 1, Arlington (VA): National Science Foundation.
- NTIA (Hrsg.) (1998): *Falling through the net II: New data on the digital divide*, <http://www.ntia.doc.gov/ntiahome/net2/>, zuletzt abgerufen am 19.11.09.
- NTIA (Hrsg.) (1999): *Falling through the net: Defining the digital divide*, <http://www.ntia.doc.gov/ntiahome/ftn99/>, zuletzt abgerufen am 19.11.09.
- OECD (Hrsg.) (1998): *The global research village: How information and communication technologies affect the science system*, Paris: OECD.
- Onyancha, Omwoyo B.; Ocholla, Dennis N. (2007): „The Performance of South African and Kenyan universities on the World Wide Web: A Web link analysis“, *Cybermetrics* 11 (1). <http://www.cindoc.csic.es/cybermetrics/articles/v11i1p2.html>, zuletzt abgerufen am 06.04.09.
- Paasi, Anssi (2005): „Globalisation, academic capitalism, and the uneven geographies of international journal publishing spaces“, *Environment and Planning A* 37 (5), 769-789.
- Paisley, William (1990): „The future of bibliometrics“, in: Christine L. Borgman (ed.): *Scholarly communication and bibliometrics*, Newbury Park (Kalifornien): Sage, 281-299.
- Pandia Search Engine News (2004): „The death of AltaVista and AllTheWeb“, <http://www.pandia.com/sw-2004/08-yahoo.html>, zuletzt abgerufen am 19.11.09.
- Park, Han Woo; Thelwall, Mike (2005): „The network approach to web hyperlink research and its utility for science communication“, in: Christine Hine (Hrsg.) (2005): *Virtual methods: Issues in social research on the Internet*. Oxford und New York: Berg, 171-181.
- Park, Han Woo; Thelwall, Mike (2006): „Web science communication in the age of globalization: Links among universities' websites in Asia and Europe“, *New Media and Society* 8 (4), 631-652.
- Payne, Nigel; Thelwall, Mike (2007): „A longitudinal study of academic webs: Growth and stabilisation“, *Scientometrics* 71 (3), 523-539.
- Payne, Nigel; Thelwall, Mike (2008): „Longitudinal trends in academic web links“, *Journal of Information Science* 34 (1), 3-14.
- Perras, Arne (2009): „Neue Internetverbindungen: Ostafrika online“, *Süddeutsche Zeitung*, 16.06.09, <http://www.sueddeutsche.de/computer/647/472174/text/>, zuletzt abgerufen am 19.11.09.
- Persson, Olle (ohne Jahr): „Influential authors in library and information science 2000-2002“, <http://www.umu.se/inforsk/LIS/LIS2000-2002.htm>, zuletzt abgerufen am 19.11.09.

- Peters, Isabella; Stock, Wolfgang (2008): „Folksonomy and information retrieval“, *Proceedings of the American Society of Information Science and Technology* 44 (1), 1-28.
- Prabowo, Rudy; Thelwall, Mike, Alexandrov, Mikhail (2007): „Generating overview timelines for major events in an RSS corpus“, *Journal of Informetrics* 1 (2), 131-144.
- Prime, Camille; Bassecoulard, Elise; Zitt, Michel (2002): „Co-citations and co-sitations: a cautionary view on an analogy“, *Scientometrics* 54 (2), 291-308.
- Qiu, Junping; Chen, Jingquan; Wang, Zhi (2004): „An analysis of backlink counts and Web Impact Factors for Chinese university websites“, *Scientometrics* 60 (3), 463-473.
- Rodríguez i Gairín, Josep Manuel (1997): „Valoración del impacto de la información en Internet: AltaVista, el 'Citation Index' de la Red“, *Revista Espanola de Documentacion Científica* 20 (2), 175-181.
- Rössler, Patrick (2005): *Inhaltsanalyse*. Konstanz: UVK.
- Rousseau, Ronald (1997): „Sitations: an exploratory study“, *Cybermetrics* 1 (1), Paper 1, www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html, zuletzt abgerufen am 19.11.09.
- Rousseau, Ronald (1999): „Daily time series of common single word searches in AltaVista and NorthernLight“, *Cybermetrics* 2/3 (1), Paper 2, <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>, zuletzt abgerufen am 19.11.09.
- Ru, Yanbo; Horowitz, Ellis (2005): „Indexing the invisible web: a survey“, *Online Information Review* 29 (3), 249-265.
- Salam, Abdus (1966): „The isolation of the scientist in developing countries“, *Minerva* 4 (4), 461-465.
- Sangowusi, F.O. (2003): „Problems of accessing scholarly publications by Nigerian scientists: a study of the University of Ibadan“, *Journal of Information Science* 29 (2), 127-134.
- Scharnhorst, Andrea; Thelwall, Mike (2005): „Citation and hyperlink networks“, *Current Science* 89 (9), 1518-1523.
- Schott, Thomas (1991): „The world scientific community: globality and globalization“, *Minerva* 29 (4), 440-462.
- Schott, Thomas (1993): „World Science: Globalization of Institutions and Participation“, *Science, Technology, and Human Values* 18: 196-208.
- Schott, Thomas (1998): „Ties between center and periphery in the scientific world system: accumulation of rewards, dominance and self-reliance in the center“, *Journal of World Systems Research* 4 (2), 112-144.
- Sherman, Chris; Price, Gary (2003): „The invisible web: uncovering sources search engines can't see“, *Library Trends* 52 (2), 282-298.
- Shrum, Wesley (1997): „View from afar: 'visible' productivity of scientists in the developing world“, *Scientometrics* 40 (2): 215-235.
- Smith, Alastair G. (2003): „Think local, search global? Comparing search engines for searching geographically specific information“, *Online Information Review* 27 (2), 102-109.
- Smith, Alastair G. (2004): „Web links as analogues of citations“, *Information Research* 9 (4), <http://InformationR.net/ir/9-4/paper188.html>, zuletzt abgerufen am 19.11.09.
- Smith, Alastair G.; Thelwall, Mike (2002): „Web Impact Factors for Australasian universities“, *Scientometrics* 54 (3), 363-380.

- Snyder, Herbert; Rosenbaum, Howard (1999): „Can search engines be used as tools for web-link analysis? A critical view“, *Journal of Documentation* 55 (4), 375-384.
- Spinelli, Diomidis (2003): „The decay and failures of Web references“, *Communications of the ACM* 46 (1), 71-77.
- StClair, Richard (2003): „Niue's .nu: Providing a free Internet to an isolated nation“, in: Wass 2003a: 77-85.
- Stichweh, Rudolf (1996): „Science in the system of world society“, *Social Science Information* 35 (2), 327-340.
- Storer, Norman W. (1970): „The internationality of science and the nationality of scientists“, *International Social Science Journal* 22, 80-93.
- Stuart, David; Thelwall, Mike (2005): „What can university-to-government web links reveal about university-government collaboration?“, in: Ingwersen/Larsen 2005, 188-192.
- Stuart, David; Thelwall, Mike; Harries, Gareth (2007): „UK academic web links and collaboration – an exploratory study“, *Journal of Information Science* 33 (2), 231-246.
- Tang, Rong; Thelwall, Mike (2003): „US academic departmental web-site interlinking in the United States: disciplinary differences“, *Library and Information Science Research*, 25 (4), 437-458.
- Tang, Rong; Thelwall, Mike (2004): „Patterns of national and international Web inlinks to US academic departments: An analysis of variations“, *Scientometrics* 60 (3), 475-485.
- Thelwall, Mike (2000): „Commercial Web sites: lost in cyberspace?“ *Internet Research: Electronic Networking and Applications* 10 (2), 150-159.
- Thelwall, Mike (2001a): „A web crawler design for data mining“, *Journal of Information Science* 27 (5), 319-325.
- Thelwall, Mike (2001b): „Extracting macroscopic information from Web links“, *Journal of the American Society for Information Science and Technology* 52 (13), 1157-1168.
- Thelwall, Mike (2001c): „Results from a web impact factor crawler“, *Journal of Documentation* 57 (2), 177-191.
- Thelwall, Mike (2002a): „A comparison of sources of links for academic Web Impact Factor calculations“, *Journal of Documentation* 58 (1), 60-72.
- Thelwall, Mike (2002b): „Evidence for the existence of geographic trends in university Web site interlinking“, *Journal of Documentation* 58 (5), 563-574.
- Thelwall, Mike (2002c): „Research dissemination and invocation on the Web“, *Online Information Review* 26 (6), 413-420.
- Thelwall, Mike (2002d): „The top 100 linked-to pages on UK university web sites: high inlink counts are not usually associated with quality scholarly content“, *Journal of Information Science* 28 (6), 483-491.
- Thelwall, Mike (2003a): „A layered approach for investigating the topological structure of communities in the web“, *Journal of Documentation* 59 (4), 410-429.
- Thelwall, Mike (2003b): „What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation“, *Information Research* 8 (3), <http://informationr.net/ir/8-3/paper151.html>, zuletzt abgerufen am 19.11.09.
- Thelwall, Mike (2004a): *Link analysis. An information science approach*. Elsevier: San Diego.
- Thelwall, Mike (2004b): „Methods for reporting on the targets of links from national systems of university Web sites“, *Information Processing and Management* 20, 125-144.
- Thelwall, Mike (2005): „Text characteristics of English language university web sites“, *Journal of the American Society for Information Science and Technology* 56 (6), 609-619.

- Thelwall, Mike (2007): „Blog searching: The first general-purpose source of retrospective public opinion in the social sciences?“, *Online Information Review* 31(3), 277-289.
- Thelwall, Mike (2008a): „Bibliometrics to Webometrics“, *Journal of Information Science* 34 (4), 605-621
- Thelwall, Mike (2008b): „Extracting accurate and complete results from search engines: Case study of Windows Life“, *Journal of the American Society for Information Science and Technology* 59 (1), 38-50.
- Thelwall, Mike (2008c): „How are social network sites embedded in the web? An exploratory link analysis“, *Cybermetrics* 12 (1), <http://www.cindoc.csic.es/cybermetrics/articles/v12i1p1.html>, zuletzt abgerufen am 19.11.09.
- Thelwall, Mike (2008d): „Quantitative comparisons of search engine results“, *Journal of the American Society for Information Science and Technology* 59 (11), 1702-1710.
- Thelwall, Mike (2008e): „Social networks, gender and friending: An analysis of MySpace member profiles“, *Journal of the American Society for Information Science and Technology* 59 (8), 1321-1330.
- Thelwall, Mike (2009a): *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*. Synthesis Lectures on Information Concepts, Retrieval, and Services 1 (1), San Francisco: Morgan and Claypool.
- Thelwall, Mike (2009b): *Persönliche Kommunikation mit der Promovendin vom 23.01.2009*.
- Thelwall, Mike; Binns, Ray; Harries, Gareth; Page-Kennedy, Teresa; Price, Liz; Wilkinson, David (2002): „European Union associated university websites“, *Scientometrics* 53 (1), 95-111.
- Thelwall, Mike; Harries, Gareth (2003): „The connection between the research of a university and counts of links to its web pages: An investigation based upon a classification of the relationships of pages to the research of the host university“, *Journal of the American Society for Information Science and Technology* 54 (7), 594-602.
- Thelwall, Mike; Harries, Gareth (2004): „Can personal web pages that link to universities yield information about the wider dissemination of research?“, *Journal of Information Science* 30 (3), 240-253.
- Thelwall, Mike; Harries, Gareth; Wilkinson, David (2003): „Why do web sites from different academic subjects interlink?“, *Journal of Information Science* 29 (6), 453-472.
- Thelwall, Mike; Price, Liz (2003): „Disciplinary differences in academic web presence – a statistical study of the UK“, *Libri* 53 (4), 242-253.
- Thelwall, Mike; Ruschenburg, Tina (2006): „Grundlagen und Forschungsfelder der Webometrie“, *Information – Wissenschaft und Praxis* 57 (8), 401-406.
- Thelwall, Mike; Smith, Alastair G. (2002): „Interlinking between Asia-Pacific university Web sites“, *Scientometrics* 55 (3), 363-376.
- Thelwall, Mike; Tang, Rong (2003): „Disciplinary and linguistic considerations for academic Web linking: An exploratory hyperlink mediated study with Mainland China and Taiwan“, *Scientometrics* 58 (1), 155-181.
- Thelwall, Mike; Tang, Rong; Price, Liz (2003): „Linguistic patterns of academic Web use in Western Europe“, *Scientometrics* 56 (3), 417-432.
- Thelwall, Mike; Vaughan, Liwen (2004): „A fair history of the Web? Examining country balance in the Internet Archive“, *Library and Information Science Research* 26 (2), 162-176.

- Thelwall, Mike; Vaughan, Liwen; Björneborn, Lennart (2005): „Webometrics“, *Annual Review of Information Science and Technology* 39, 81-138.
- Thelwall, Mike; Vaughan, Liwen; Cothey, Viv; Li, Xuemei; Smith, Alastair G. (2003): „Which academic subjects have most online impact? A pilot study and a new classification process“, *Online Information Review* 27 (5), 333-343.
- Thelwall, Mike; Wilkinson, David (2003): „Three target document range metrics for university web sites“, *Journal of the American Society for Information Science and Technology* 54 (6), 490-497.
- Thomson Reuters (ohne Jahr): „The Thomson Reuters Impact Factor“, http://thomsonreuters.com/products_services/science/free/essays/impact_factor/, zuletzt abgerufen am 19.11.09.
- Uddin, Mohammad Nadir (2003): „Internet use by university academics: a bipartite study of information and communication needs“, *Online Information Review* 27 (4): 225-237.
- UNESCO (Hrsg.) (2006): „Expenditure on R&D“, <http://uis.unesco.org/TEMPLATE/html/Exceltables/science/R&DTables.xls>, zuletzt abgerufen am 04.01.07.
- UNDP (Hrsg.) (2007): *Human Development Report 2007/2008*, New York: UNDP, abrufbar unter <http://hdr.undp.org/en/reports/global/hdr2007-2008/>, zuletzt abgerufen am 19.11.2009.
- United Nations Statistics Division (Hrsg.) (2006): „Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings“, <http://unstats.un.org/unsd/methods/m49/m49regin.htm>, zuletzt abgerufen am 04.01.07.
- van Raan, Anthony F.J. (1997): „Science as an international enterprise“, *Science and Public Policy* 24 (5), 290-300.
- van Raan, Anthony F.J. (2001): „Bibliometrics and internet: some observations and expectations“, *Scientometrics* 50 (1), 59-63.
- van Raan, Anthony F.J. (2003): „The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments“, *Technikfolgenabschätzung – Theorie und Praxis* 12 (1), 20-29.
- Vaughan, Liwen (2004): „New measurements for search engine evaluation proposed and tested“, *Information Processing and Management* 40 (4), 677-691.
- Vaughan, Liwen (2008): „A new frontier of informetric and webometric research: Mining web usage data“, in: Hildrun Kretschmer; Frank Havemann *Proceedings of the Fourth International Conference on Webometrics, Informetrics, and Scientometrics & Ninth COLLNET Meeting*, 28.07.-01.08.08, Berlin, <http://www.collnet.de/Berlin-2008/VaughanWIS2008nfi.pdf>, zuletzt abgerufen am 19.11.09.
- Vaughan, Liwen; Shaw, Debora (2003): „Bibliographic and web citations: What’s the difference?“ *Journal of the American Society for Information Science and Technology* 54 (4), 1313-1322.
- Vaughan, Liwen; Shaw, Debora (2005): „Web citation data for impact assessment“, *Journal of the American Society for Information Science and Technology* 56 (10), 1075-1087.
- Vaughan, Liwen; Thelwall, Mike (2004): „Search engine coverage bias: evidence and possible causes“, *Information Processing and Management* 40 (4), 693-707.
- Vaughan, Liwen; Thelwall, Mike (2005): „A modelling approach to uncover hyperlink patterns: The case of Canadian universities“, *Information Processing and Management* 41 (2), 347-359.

- Voß, Jakob (2005): *Informetrische Untersuchungen an der Online-Enzyklopädie Wikipedia*, Magisterarbeit, Berlin: Humboldt-Universität.
- Wagner, Carolyn S.; Leydesdorff, Loet (2005): „Mapping the network of global science: Comparing international co-authorships from 1990 to 2000“, *International Journal of Technology and Globalisation* 1 (2), 185-208.
- Wagner-Döbler, Roland (2001): „Continuity and discontinuity of collaboration behaviour since 1800 – from a bibliometric point of view“, *Scientometrics* 52 (3), 503-517.
- Wallsten, Scott (2005): „Regulation and Internet use in developing countries“, *Economic Development and Cultural Change* 53 (2), 501-523.
- Wooster, Warren (1969): "The Ocean and Man", *Scientific American* 221 (3), 218-234.
- Wass, Erica Schlesinger (2003b): „Lots of dots“, in: Wass 2003a, ix-xix.
- Wass, Erica Schlesinger (2003c): „The United States‘ .us: Striving for the American dream“, in: Wass 2003a, 123-136.
- Wass, Erica Schlesinger (Hrsg.) (2003a): *Addressing the world. National identity and Internet country code domains*, Lanham (MD) et al.: Rowman and Littlefield.
- Weingart, Peter (2005): „Impact of bibliometrics on the science system: Inadvertant consequences?“, *Scientometrics* 62 (1), 117-131.
- Whitley, Richard (1984): *The intellectual and social organization of the sciences*, Oxford: Clarendon Press.
- Wilkinson, David; Harries, Gareth; Thelwall, Mike; Price, Liz (2003): „Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication“, *Journal of Information Science* 29 (1), 49-56.
- Wilkinson, David; Thelwall, Mike; Li, Xuemei (2003): „Exploiting hyperlinks to study academic web use“, *Social Science Computer Review* 21 (3), 340-351.
- World Bank (Hrsg.) (2005): *Financing information and communication infrastructure needs in the developing world: public and private roles*, World Bank Working Paper Nr. 65, Washington D.C.: World Bank, <http://info.worldbank.org/etools/docs/library/240325/Financing%20information%20and%20communication%20infrastructure%20needs%20in%20the%20developing%20world%20-%20public%20and%20private%20roles.pdf>, zuletzt abgerufen am 19.11.09.
- Wormell, Irene (2001): „Informetrics and webometrics for measuring impact, visibility, and connectivity in science, politics, and business“, *Competitive Intelligence Review* 12 (1), 12-23.
- Wouters, Paul; de Vries, Repke (2004): „Formally citing the web“, *Journal of the American Society for Information Science and Technology* 55 (14), 1250-1260.
- Ynalvez, Marcus; Duque, Ricardo B.; Mbatia, Paul; Sooryamoorthy, R.; Palackal, Anthony; Shrum, Wesley (2005): „When do scientists ‚adopt‘ the Internet? Dimensions of connectivity in developing areas“, *Scientometrics* 63 (1), 39-67.
- Zhang, Yin (2001): „Scholarly use of Internet-based electronic resources“, *Journal of the American Society for Information Science and Technology* 52 (8), 628-654.
- Zuccala, Alesia; Thelwall, Mike (2008): „A university-centred European Union link analysis“, *Scientometrics* 75 (3), 407-420.

8. Anhang

8.1 Abfrage der Inlinks bei Yahoo nach Top-Level-Domains

Wie bei den allgemeinen Abfragen wurde bei der Suche nach Links aus der eigenen TLD die Institutsdomain ausgeschlossen (z.B.: *linkdomain:bigelow.org -site:bigelow.org site:org*). Bei den übrigen Queries wurde darauf verzichtet (z.B: *linkdomain:bigelow.org site:de*).

Abfragen nach country-code Top-Level-Domains (ccTLDs)

linkdomain:[Institutsdomain] site:ac	linkdomain:[Institutsdomain] site:lc
linkdomain:[Institutsdomain] site:ae	linkdomain:[Institutsdomain] site:li
linkdomain:[Institutsdomain] site:af	linkdomain:[Institutsdomain] site:lk
linkdomain:[Institutsdomain] site:ag	linkdomain:[Institutsdomain] site:lr
linkdomain:[Institutsdomain] site:ai	linkdomain:[Institutsdomain] site:ls
linkdomain:[Institutsdomain] site:al	linkdomain:[Institutsdomain] site:lt
linkdomain:[Institutsdomain] site:am	linkdomain:[Institutsdomain] site:lu
linkdomain:[Institutsdomain] site:an	linkdomain:[Institutsdomain] site:lv
linkdomain:[Institutsdomain] site:ao	linkdomain:[Institutsdomain] site:ly
linkdomain:[Institutsdomain] site:aq	linkdomain:[Institutsdomain] site:ma
linkdomain:[Institutsdomain] site:ar	linkdomain:[Institutsdomain] site:mc
linkdomain:[Institutsdomain] site:as	linkdomain:[Institutsdomain] site:md
linkdomain:[Institutsdomain] site:at	linkdomain:[Institutsdomain] site:me
linkdomain:[Institutsdomain] site:au	linkdomain:[Institutsdomain] site:mg
linkdomain:[Institutsdomain] site:aw	linkdomain:[Institutsdomain] site:mh
linkdomain:[Institutsdomain] site:az	linkdomain:[Institutsdomain] site:mk
linkdomain:[Institutsdomain] site:ba	linkdomain:[Institutsdomain] site:ml
linkdomain:[Institutsdomain] site:bb	linkdomain:[Institutsdomain] site:mm
linkdomain:[Institutsdomain] site:bd	linkdomain:[Institutsdomain] site:mn
linkdomain:[Institutsdomain] site:be	linkdomain:[Institutsdomain] site:mo
linkdomain:[Institutsdomain] site:bf	linkdomain:[Institutsdomain] site:mp
linkdomain:[Institutsdomain] site:bg	linkdomain:[Institutsdomain] site:mq
linkdomain:[Institutsdomain] site:bh	linkdomain:[Institutsdomain] site:mr
linkdomain:[Institutsdomain] site:bi	linkdomain:[Institutsdomain] site:ms
linkdomain:[Institutsdomain] site:bj	linkdomain:[Institutsdomain] site:mt
linkdomain:[Institutsdomain] site:bm	linkdomain:[Institutsdomain] site:mu
linkdomain:[Institutsdomain] site:bn	linkdomain:[Institutsdomain] site:mv
linkdomain:[Institutsdomain] site:bo	linkdomain:[Institutsdomain] site:mw
linkdomain:[Institutsdomain] site:br	linkdomain:[Institutsdomain] site:mx
linkdomain:[Institutsdomain] site:bs	linkdomain:[Institutsdomain] site:my
linkdomain:[Institutsdomain] site:bt	linkdomain:[Institutsdomain] site:mz
linkdomain:[Institutsdomain] site:bv	linkdomain:[Institutsdomain] site:na
linkdomain:[Institutsdomain] site:bw	linkdomain:[Institutsdomain] site:nc
linkdomain:[Institutsdomain] site:by	linkdomain:[Institutsdomain] site:ne
linkdomain:[Institutsdomain] site:bz	linkdomain:[Institutsdomain] site:nf
linkdomain:[Institutsdomain] site:ca	linkdomain:[Institutsdomain] site:ng
linkdomain:[Institutsdomain] site:cc	linkdomain:[Institutsdomain] site:ni
linkdomain:[Institutsdomain] site:cd	linkdomain:[Institutsdomain] site:nl
linkdomain:[Institutsdomain] site:cf	linkdomain:[Institutsdomain] site:no
linkdomain:[Institutsdomain] site:cg	linkdomain:[Institutsdomain] site:np
linkdomain:[Institutsdomain] site:ch	linkdomain:[Institutsdomain] site:nr
linkdomain:[Institutsdomain] site:ci	linkdomain:[Institutsdomain] site:nu
linkdomain:[Institutsdomain] site:ck	linkdomain:[Institutsdomain] site:nz

linkdomain:[Institutsdomain]	site:il	linkdomain:[Institutsdomain]	site:tw
linkdomain:[Institutsdomain]	site:im	linkdomain:[Institutsdomain]	site:tz
linkdomain:[Institutsdomain]	site:in	linkdomain:[Institutsdomain]	site:ua
linkdomain:[Institutsdomain]	site:io	linkdomain:[Institutsdomain]	site:ug
linkdomain:[Institutsdomain]	site:iq	linkdomain:[Institutsdomain]	site:uk
linkdomain:[Institutsdomain]	site:ir	linkdomain:[Institutsdomain]	site:um
linkdomain:[Institutsdomain]	site:is	linkdomain:[Institutsdomain]	site:us
linkdomain:[Institutsdomain]	site:it	linkdomain:[Institutsdomain]	site:uy
linkdomain:[Institutsdomain]	site:je	linkdomain:[Institutsdomain]	site:uz
linkdomain:[Institutsdomain]	site:jm	linkdomain:[Institutsdomain]	site:va
linkdomain:[Institutsdomain]	site:jo	linkdomain:[Institutsdomain]	site:vc
linkdomain:[Institutsdomain]	site:jp	linkdomain:[Institutsdomain]	site:ve
linkdomain:[Institutsdomain]	site:ke	linkdomain:[Institutsdomain]	site:vg
linkdomain:[Institutsdomain]	site:kg	linkdomain:[Institutsdomain]	site:vi
linkdomain:[Institutsdomain]	site:kh	linkdomain:[Institutsdomain]	site:vn
linkdomain:[Institutsdomain]	site:ki	linkdomain:[Institutsdomain]	site:vu
linkdomain:[Institutsdomain]	site:km	linkdomain:[Institutsdomain]	site:wf
linkdomain:[Institutsdomain]	site:kn	linkdomain:[Institutsdomain]	site:ws
linkdomain:[Institutsdomain]	site:kp	linkdomain:[Institutsdomain]	site:ye
linkdomain:[Institutsdomain]	site:kr	linkdomain:[Institutsdomain]	site:yt
linkdomain:[Institutsdomain]	site:kw	linkdomain:[Institutsdomain]	site:yu
linkdomain:[Institutsdomain]	site:ky	linkdomain:[Institutsdomain]	site:za
linkdomain:[Institutsdomain]	site:kz	linkdomain:[Institutsdomain]	site:zm
linkdomain:[Institutsdomain]	site:la	linkdomain:[Institutsdomain]	site:zw
linkdomain:[Institutsdomain]	site:lb		

Abfragen nach generischen Top-Level-Domains (gTLDs)

linkdomain:[Institutsdomain]	site:aero	linkdomain:[Institutsdomain]	site:jobs
linkdomain:[Institutsdomain]	site:asia	linkdomain:[Institutsdomain]	site:mil
linkdomain:[Institutsdomain]	site:biz	linkdomain:[Institutsdomain]	site:mobi
linkdomain:[Institutsdomain]	site:com	linkdomain:[Institutsdomain]	site:museum
linkdomain:[Institutsdomain]	site:coop	linkdomain:[Institutsdomain]	site:net
linkdomain:[Institutsdomain]	site:edu	linkdomain:[Institutsdomain]	site:org
linkdomain:[Institutsdomain]	site:gov	linkdomain:[Institutsdomain]	site:pro
linkdomain:[Institutsdomain]	site:info	linkdomain:[Institutsdomain]	site:tel
linkdomain:[Institutsdomain]	site:int	linkdomain:[Institutsdomain]	site:travel

Nachträglich ausgeschlossene Abfragen

linkdomain:[Institutsdomain]	site:ad
linkdomain:[Institutsdomain]	site:ax
linkdomain:[Institutsdomain]	site:cat
linkdomain:[Institutsdomain]	site:do
linkdomain:[Institutsdomain]	site:name

8.2 Datentabellen

<i>Institut</i>	<i>Outlinkdomains mit sektoraler TLD</i>	<i>Outlinkdomains mit länderspezifischer TLD</i>	<i>Outlinkdomains mit der eigenen nationalen TLD</i>	<i>Outlinkdomains mit internationaler TLD</i>
BLOS	132	290	221	69
HBOI	148	62	54	8
IFM	216	681	353	328
IOW	118	358	203	155
KMFRI	12	3	0	3
MBARI	496	522	426	96
MOI	5	15	6	9
NIO	30	36	12	24
NIOPK	0	1	1	0
ORI	10	28	23	5

Tabelle 65: Outlinkdomains nach TLD-Typ sowie nach nationaler und internationaler TLD

<i>Institut</i>	<i>mit TLDs des eigenen Kontinents</i>	<i>mit TLDs anderer Kontinente</i>	<i>Übergangsländer</i>	<i>Entwicklungsländer</i>	<i>Industrieländer</i>
BLOS	15	54	0	10	59
HBOI	1	7	0	0	8
IFM	130	198	16	27	285
IOW	101	54	2	6	147
KMFRI	1	2	0	1	2
MBARI	30	66	2	12	82
MOI	0	9	0	1	8
NIO	1	23	0	0	24
NIOPK	0	0	0	0	0
ORI	3	2	0	3	2

Tabelle 66: Outlinkdomains mit internationaler TLD nach eigenem und anderen Kontinenten sowie nach Entwicklungsstatus

<i>Institut</i>	<i>Afrika</i>	<i>Asien</i>	<i>Europa</i>	<i>Nordamerika</i>	<i>Ozeanien</i>	<i>Südamerika</i>
BLOS	0	3	21	11	2	1
HBOI	0	1	8	8	6	0
IFM	12	53	307	194	43	18
IOW	8	6	101	23	8	2
KMFRI	11	8	29	6	2	0
MBARI	0	13	42	22	16	3
MOI	2	3	1	2	2	0
NIO	3	40	45	20	6	0
NIOPK	0	0	2	1	0	0
ORI	2	3	2	2	1	0

Tabelle 67: Outlinkdomains mit internationaler TLD nach Kontinenten

<i>Institut</i>	<i>Inlinkdomains mit sektoraler TLD</i>	<i>Inlinkdomains mit länderspezifischer TLD</i>	<i>Inlinkdomains mit der eigenen nationalen TLD</i>	<i>Inlinkdomains mit internationaler TLD</i>
BLOS	727	599	334	265
HBOI	921	511	267	244
IFM	1307	1456	709	747
IOW	339	711	454	257
KMFRI	18	12	0	12
MBARI	1575	1283	635	648
MOI	22	11	1	10
NIO	413	288	47	241
NIOPK	38	38	23	15
ORI	37	42	23	19

Tabelle 68: Inlinkdomains nach TLD-Typ sowie nach nationaler und internationaler TLD

<i>Institut</i>	<i>mit TLDs des eigenen Kontinents</i>	<i>mit TLDs anderer Kontinente</i>	<i>Übergangsländer</i>	<i>Entwicklungsländer</i>	<i>Industrieländer</i>
BLOS	31	234	7	50	208
HBOI	21	223	10	58	176
IFM	414	333	39	191	517
IOW	186	71	14	30	213
KMFRI	0	12	0	0	12
MBARI	40	608	44	136	468
MOI	1	9	0	2	8
NIO	40	201	23	35	192
NIOPK	3	12	0	4	11
ORI	0	19	0	2	17

Tabelle 69: Inlinkdomains mit internationaler TLD nach eigenem und anderen Kontinenten sowie nach Entwicklungsstatus

<i>Institut</i>	<i>Afrika</i>	<i>Asien</i>	<i>Europa</i>	<i>Nordamerika</i>	<i>Ozeanien</i>	<i>Südamerika</i>
BLOS	6	45	158	31	15	9
HBOI	3	51	143	21	18	8
IFM	6	183	414	89	37	16
IOW	0	26	186	35	7	3
KMFRI	0	1	8	3	0	0
MBARI	7	125	404	40	49	23
MOI	1	2	4	3	0	0
NIO	1	40	124	60	10	6
NIOPK	0	3	7	4	0	1
ORI	0	2	13	2	1	1

Tabelle 70: Inlinkdomains mit internationaler TLD nach Kontinenten

<i>Institut</i>	<i>Publikationen ohne externe Adressen</i>	<i>Kopublikationen mit nationalen Adressen</i>	<i>Kopublikationen mit internationalen Adressen</i>
BLOS	8	20	34
HBOI	20	65	23
IFM	99	130	466
IOW	55	53	126
KMFRI	2	6	35
MBARI	42	111	71
MOI	1	1	6
NIO	206	97	95
NIOPK	0	3	2
ORI	6	5	7

Tabelle 71: Publikationen nach Koautorenschaftsstatus

<i>Institut</i>	<i>mit eigenem Kontinent</i>	<i>mit anderen Kontinenten</i>	<i>Übergangsländer</i>	<i>Entwicklungsländer</i>	<i>Industrieländer</i>
BLOS	8	26	2	3	31
HBOI	8	15	1	1	22
IFM	197	269	52	70	406
IOW	86	40	5	16	113
KMFRI	4	31	0	15	31
MBARI	16	55	3	14	62
MOI	0	6	0	4	6
NIO	29	66	3	14	87
NIOPK	0	2	0	0	2
ORI	2	5	0	5	4

Tabelle 72: Internationale Kopublikationen mit dem eigenen und anderen Kontinenten, sowie mit Übergangsländern, Entwicklungs- und Industrieländern

<i>Institut</i>	<i>Afrika</i>	<i>Asien</i>	<i>Europa</i>	<i>Nordamerika</i>	<i>Ozeanien</i>	<i>Südamerika</i>
BLOS	0	3	21	11	2	1
HBOI	0	1	8	8	6	0
IFM	12	53	307	194	43	18
IOW	8	6	101	23	8	2
KMFRI	11	8	29	6	2	0
MBARI	0	13	42	22	16	3
MOI	2	3	1	2	2	0
NIO	3	40	45	20	6	0
NIOPK	0	0	2	1	0	0
ORI	2	3	2	2	1	0

Tabelle 73: Internationale Kopublikationen nach Kontinenten

Weltregionen	BLOS			IFM			IOW			MBARI			NIO		
	Kopublikationen	Outlinkdomains	Inlinkdomains												
Eastern Africa	0	0	0	0	2	0	0	0	0	0	0	0	1	0	1
Middle Africa	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
Northern Africa	0	0	0	3	1	0	0	0	0	0	0	0	1	0	0
Southern Africa	0	2	5	8	4	6	8	5	0	0	1	6	2	0	0
Western Africa	0	0	0	1	2	0	0	0	0	0	0	0	1	0	0
Caribbean	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0
Central America	0	2	4	7	0	5	1	0	1	8	6	5	1	0	1
Northern America	11	13	27	191	139	84	23	43	33	14	24	34	18	19	59
South America	1	2	9	18	7	16	2	0	3	3	3	23	0	0	6
Subantarctica	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
Antarctica	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Central Asia	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Eastern Asia	2	9	39	33	22	159	0	2	24	13	2	109	34	1	30
South-Eastern Asia	0	1	2	2	1	12	3	0	1	0	1	6	3	0	6
Southern Asia	1	1	0	16	0	1	1	0	0	0	0	4	0	0	0
Western Asia	0	0	2	6	3	11	2	0	1	0	0	6	4	0	2
Eastern Europe	2	0	27	59	19	105	18	10	36	5	1	89	1	0	27
Northern Europe	15	22	47	176	59	94	81	68	74	21	24	93	11	1	27
Southern Europe	3	3	23	40	8	76	12	6	26	11	4	65	5	0	11
Western Europe	7	6	61	127	44	139	24	17	50	16	17	157	32	3	59
Australia and New Zealand	2	7	15	41	17	32	8	3	3	16	11	39	6	0	8
Melanesia	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0
Micronesia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Polynesia	0	1	0	0	0	5	0	1	2	1	1	10	0	0	2

Tabelle 74: Kopublikationen, Outlinks und Inlinks von BLOS, IFM, IOW, MBARI und NIO nach Weltregionen (die eigene Region ist jeweils grau unterlegt)

8.3 Codebuch zur Codierung der nationalen Anbindung verlinkter Seiten

Die hier beschriebene Codierung ist Teil einer Untersuchung der internationalen Verlinkung von Internetauftritten. Ein möglicher Indikator für die nationale Anbindung einer Internetseite oder eines Links ist die in ihrer URL enthaltene Länder-Top-Level-Domains (TLDs) wie „.de“ oder „.uk“. Da die meisten TLDs frei wählbar sind, ist zunächst die Übereinstimmung der verwendeten TLDs und der tatsächlichen nationalen Verortung der Seiteninhalte zu prüfen. Aus diesem Grund wird hier ein Sample von Internetseiten einer Inhaltsanalyse unterzogen, um später das Ausmaß der Übereinstimmung von inhaltlicher und formalen Hinweisen (Inhalt der Seite vs. TLD) auf die nationale Anbindung zu bestimmen.

Ziel der Codierung ist es festzuhalten, ob und welche inhaltlichen Hinweise auf die nationale Verortung in den Webseiten enthalten sind.

Die aktuelle Seiten werden nicht zur Codierung herangezogen. Die **Internetverbindung** des Computers, auf dem gearbeitet wird, ist für die Zeit der Codierung zu **unterbrechen**, damit die Seiten nicht unbemerkt nachgeladen werden können.

Enthält die Datei den Namen einer unbekanntes Stadt oder Region, kann er in passenden Ressourcen (Atlas, Google Maps, Wikipedia) **nachgeschlagen** werden.

Auswahleinheit:

Es wird eine Auswahl von Webseiten kodiert, die durch Links mit den Websites von zehn ausgewählten Meeresforschungseinrichtungen verbunden sind.

Analyseeinheit:

Analysiert wird jeweils eine einzelne gespeicherte Webseite.

Kontexteinheit:

Über Analyseeinheit hinaus werden nur Informationen aus anderen gespeicherten Webseiten in die Auswertung einbezogen, die zur selben Domain gehören und ihren Inhalten zufolge mit der zu analysierenden Seite in Beziehung stehen und Aufschluss auf ihre nationale Anbindung geben.

Beispiel: Das Sample enthält die Seiten www.eisler-schule.de und www.eisler-schule.de/sv/. Die erste Seite enthält die Anschrift der Schule, die zweite nicht. Wenn angenommen werden kann, dass sich beide Seiten auf die gleiche Organisation beziehen, wird auch die zweite Datei mit dem Land der ersten codiert.

Kategorien:

Es werden drei Aspekte codiert:

- A) Für jede Seite wird codiert, in welchem Land das durch sie Repräsentierte (Person / Organisation / Projekt / Publikation / Ereignis o.ä.) angesiedelt ist.
- B) Handelt es sich um eine binationale, multinationale oder zwischenstaatliche Einrichtung wird dies separat codiert.
- C) Die Quelle der Information wird festgehalten.

	A – Nation		B – mehrstaatlich		C – Informationsquelle
A1	ein Land der Länderliste (wenn binational: zwei Länder)	B1	binational	C1	Anschrift
		B2	multinational	C2	Ort im Namen
		B3	zwischenstaatlich	C3	LLC. / § 501(c)(3)
		B4	EU	C4	andere Hinweise im Text
				C5	bekannte Organisation, Veranstaltung o.ä.
A7	Sprache unverständlich	B7	Sprache unverständlich		
A8	keine codierbaren Inhalte (Fehlermeldung / kaum Textinhalt / domain expired)	B8	keine codierbaren Inhalte (Fehlermeldung / kaum Textinhalt / domain expired)		
A9	kein Hinweis	B9	kein Hinweis	C9	kein Hinweis

A) Nationale Anbindung

- A1) Sofern die Webseiten einen Bezug zu einem **Land** aufweisen, wird es hier eingetragen. Möglich sind alle Staaten, die bei der UN geführt werden (siehe Datei UN_country_codes.htm).
- Für Webseiten, die eine Person oder Organisation vertreten, gilt es ihren Hauptsitz zu klassifizieren.
 - Bei Themenseiten oder Foren wird der Sitz des Urhebers klassifiziert.
 - Bei Veranstaltungen geht es um den Ort der Veranstaltungen.
- A7) Ist der Inhalt in einer Sprache verfasst, die der Codiererin nicht verständlich ist, ist der Code **Sprache unverständlich** zu wählen. Die Seite wird dann von einer weiteren, der Sprache mächtigen Person codiert.
- A8) Seiten mit wenig, nicht sinnvoll interpretierbarem Inhalt, werden mit **keine codierbaren Inhalte** codiert.

Beispiel: Eine Datei enthält nur eine Abbildung, nur einen html-Code ohne Textteile oder ähnliches.

In diese Kategorie fallen auch Dateien, die eine Fehlermeldung enthalten, die eine Suche nach einem Landesbezug unmöglich macht. Dateien, die nur partielle Fehlermeldungen enthalten, dürfen ausgewertet werden. *Beispiel:* Enthält eine Seite die Fehlermeldung 404 / page not found im Body, aber auch den Rahmen der Uni Bielefeld, darf sie mit Deutschland kodiert werden. Häufige Fehlermeldungen sind „Server not found“ oder „page not found“, oft auch mit Codes wie error 403, 404, 500 etc. bezeichnet.

- A9) Dateien, die zwar prinzipiell interpretierbare, sprachlich verständliche Inhalte aufweisen, aber keinen Hinweis auf die nationale Verortung der repräsentierten Person/Organisation etc. bieten, werden mit **kein Hinweis** codiert.

B) Mehrstaatliche Anbindung

- B1) Als **binational** werden Seiten klassifiziert, die Urheber aus zwei Ländern haben oder bei denen Trägerland und Sitzland auseinanderfallen:

Beispiel: Ein Büro des deutschen Ministeriums für Zusammenarbeit in Kenia wird unter B als binational gelabelt; unter A wird „Deutschland und Kenia“ eingetragen.

- B2) Als **multinational** werden Seiten klassifiziert, die Urheber aus mehr als zwei Ländern haben. Wird darüber hinaus ein Sitz genannt, wird dieser unter A eingetragen.

Beispiel: Für ein Forschungsprojekt mit Teilnehmern aus den USA, Frankreich, Italien und Japan bei dem eine Kontaktadressen in Japan angegeben ist, wird unter A „Japan“ und unter B „multinational“ eingetragen.

- B3) Als **zwischenstaatlich** werden Seiten klassifiziert, die auf internationale Abkommen zwischen Staaten zurückgehen. Hierzu zählen in diesem Kontext vorwiegend UN-Organisationen und Einrichtungen, die auf internationale Umweltverhandlungen zurückgehen.

Beispiel: Die WHO in Genf wird mit A „Schweiz“ und B „zwischenstaatlich“ codiert.

- B4) Ein Sonderfall multinationaler und zwischenstaatlicher Seiten sind **EU**-bezogene Seiten: Für Einrichtungen, Programme oder Informationsseiten der EU sowie für internationale EU-Forschungsprojekte, länderübergreifende Regionalförderung durch die EU o.ä. wird unter B EU und unter A ggf. der angegebene Sitz eingetragen.

Beispiel: Für die Seite des europäischen Parlaments in Strassbourg wird unter A „Frankreich“ und unter B „EU“ eingetragen.

Ein Hinweis auf ein mehrstaatliches EU-Projekt ist beispielsweise in dessen Förderung nach dem EU-Rahmenprogramm oder dessen Bezeichnung als Concerted Action oder als Network of Excellence zu sehen. Einstaatliche Projekte u.ä. werden nicht mit „EU“ klassifiziert, auch wenn auf der Seite ein Bezug hergestellt wird.

Beispiel: Ein Forschungsprojekt an der Uni Hamburg wird u.a. aus Mitteln des Europäischen Sozialfonds gefördert, hat aber keine internationalen Partner. Es fällt somit nicht in die Kategorie B4, weil es nicht multinational ist.

- B7-B9) Diese Kategorien entsprechen A7-A9 (**Sprache unverständlich / keine interpretierbaren Inhalte / keine Hinweise**).

C) Quelle der Information

Oft sind die relevanten Informationen in der **Kopfzeile (Header)**, im seitlichen **Rahmen (Frame)** sowie **ganz oben** oder **ganz unten** (z.B. im Copyright) auf der Seite enthalten.

Es geht hier vorwiegend um die Quelle der Information zur nationalen Anbindung. Wenn zusätzlich ein Hinweis auf eine mehrstaatliche Anbindung vorliegt, die ein andere Quelle hat, ist sie zu vernachlässigen.

- C1) **Anschriften** mit Ort sind ein deutlicher Hinweisgeber. Allerdings gilt es zu prüfen, ob hier eine mehrstaatliche Organisation/Veranstaltung repräsentiert wird.
- C2) Viele Organisationen tragen ihren **Sitzort im Namen** (*Beispiel*: Universität Bielefeld). Dies wird als Hinweis gewertet, sofern die Bezeichnung authentisch erscheint.
- C3) Organisationen, die die Organisationsform **LLC** aufweisen oder nach **§501(c)(3)** als gemeinnützig registriert sind, können den **USA** zugeschrieben werden.
- C4) **Andere Hinweise** im Text oder in Symbolen werden akzeptiert, wenn sie hinreichend eindeutig sind. Auch Telefonnummern mit nationalem Vorwahlkürzel werden als Hinweis akzeptiert.

Beispiel: Auf der Seite eines Hotels wird die Lage in einer bestimmten Gegend beschrieben. Eine Informationsseite enthält an prominenter Stelle das Logo eines Ministeriums, so dass hier eine Trägerschaft unterstellt werden kann.

- C5) Wenn eine Seite keinen der vorangegangenen Hinweise enthält, die repräsentierte **Organisation/Person/Veranstaltung** etc. der Kodiererin aber eindeutig **bekannt** ist, kann die Seite in den Kategorien A und B klassifiziert werden. Für C wird dann der Code C5 verwendet. Auch eine Übertragung von Informationen zwischen Dateien ist zulässig.

Beispiel: Wenn eine Informationsseite des Familienministeriums keinen Hinweis auf den Standort enthält und das „Bundes“ allein nicht eindeutig deutsch ist, kann ein bekanntes Logo oder ein Foto von Frau von der Leyen eine für die Codierung geeignete Referenz sein.

Beispiel: Datei X enthält ein Logo mit dem Akronym BUND, dessen Bedeutung in der Datei Y als Bund für Umwelt- und Naturschutz Deutschland ausgeführt wird. Datei X kann somit unter „Deutschland“ eingeordnet werden.

Ebenso gelten die zehn erhobenen Institute als bekannt:

- Bigelow Lab / USA
- Harbor Branch Institution (HBOI) / Harbor Branch, USA
- Monterey Bay Aquarium Research Institute (MBARI) / Moss Landing, USA
- IfM-Geomar / Kiel
- Institut für Ostseeforschung (IOW) / Warnemünde
- Mauritius Oceanographic Institute (MOI) / Quatre Bornes, Mauritius
- Kenyan Marine Fisheries Research Institute (KMFRI) / Mombasa et al., Kenia
- Oceanographic Research Institute (ORI) / Durban, South Africa
- National Institute of Oceanography (NIO) / Goa, India
- National Institute of Oceanography (NIO) / Karachi, Pakistan

- C9) Enthält eine Seite keine expliziten Bezüge zu einem Ort und sie auch nicht mit hoher Sicherheit bekannt, wird sie mit **kein Hinweis** kodiert.

Allgemeine Hinweise mit Beispielen:

Prinzipiell geht es um **Urheberschaft** und nicht um den inhaltlichen Schwerpunkt der Seite. *Beispiel:* Das Forschungsprojekt einer deutschen Forschungsgruppe in Kenia wird mit „Deutschland“, nicht mit „Kenia“ kodiert. Die private Themenseite einer Norwegerin zum Thema Mittelmeerfische wird mit „Norwegen“ gelabelt.

Eine Seite mit inhaltlichem Länderschwerpunkt wird nur dem Land zugeordnet, wenn anzunehmen ist, das tatsächlich ein lokaler Urheber dahinter steht.

Beispiel: Eine Seite einer Linkliste zum Thema Sylt Tourismus, die offensichtlich nur wahllos zusammengestellte Links zu diesen Stichworten enthält, ist nicht notwendigerweise von einer Person mit Lokalkenntnissen erstellt worden und wird mit „kein Hinweis“ klassifiziert. Enthält die Seite dagegen fundiertere Inhalte zum Thema Sylt (z.B. aktuelle Termine oder umfangreiche Touristinformation) kann von einer lokalen Redaktion ausgegangen werden und die Seite unter „Deutschland“ abgelegt werden.

Wichtig ist grundsätzlich eine **Konsistenzprüfung**. Beispielsweise ist eine Adresse im Allgemeinen ein deutlicher Indikator, aber es kann auch Fälle geben, in denen sie nicht die Trägereinrichtung der Seite beschreibt. *Beispiel:* Eine Meeresschutz-NGO ruft auf einer Seite dazu auf, im Rahmen einer Kampagne eine Postkarte an den EU-Kommissar für Fischerei zu schicken und nennt dessen Postanschrift. In diesem Fall sagt die angegebene Anschrift nichts über die NGO selbst aus.

Auch lassen Organisationen zum Teil organisationsfremde Informationen auf ihrer Website zu.

Beispiel: Ein neu gegründetes, internationales E-Journal hat keinen eigenen Internetauftritt, sondern wird in der Domain der Universität eines Herausgebers abgelegt. Gleiches kann bei organisationsübergreifenden Studiengängen, Fachgesellschaften, Konferenzen u.ä. auftreten. In diesen Fällen ist der Rahmen zu vernachlässigen und der Inhalt (in dem Beispiel: das E-Journal) zu klassifizieren.

Einige Dateien enthalten keinen Rahmen. Hier ist ebenfalls der Inhalt die Grundlage der Klassifikation.

Beispiel: Die Datei ist ein pdf-Dokument, das einen Fachartikel, deren AutorInnen deutsche und englische Adressen haben. Er wird als binational / Deutschland und UK eingeordnet, auch wenn er laut URL auf der Website einer deutschen Universität liegt. Würde die Datei den gleichen Artikel auf der Website und mit dem Rahmen eines Fachverlages enthalten, wäre der Rahmen zu klassifizieren.

Im Copyright werden oft, aber nicht immer Hinweise auf die nationale Verortung gegeben. *Beispiel:* In Themenforen ist im Copyright häufig die verwendet Forensoftware angegeben, die keinen Hinweis auf den Träger des Forums gibt.

Gedruckt auf alterungsbeständigem Papier °° ISO 9706.