



**SFB 1199**

Processes of Spatialization  
under the Global Condition

Ninja Steinbach-Hüther

Dirk Hänsgen

Thomas Efer

Ute Wardenga

**Geographiegeschichtsschreibung  
und Digital Humanities.**

**Neue Methoden für  
Zeitschriftenanalysen**

Working paper series  
des SFB 1199  
an der Universität Leipzig  
Nr. 15

Collaborative Research Center (SFB) 1199  
**Processes of Spatialization under the Global Condition**  
at the University of Leipzig

Funded by Deutsche Forschungsgemeinschaft



---

**Ninja Steinbach-Hüther, Dirk Hänsgen, Thomas Efer, Ute Wardenga**  
Geographiegeschichtsschreibung und Digital Humanities.  
Neue Methoden für Zeitschriftenanalysen  
(= Working paper series des SFB 1199 an der Universität Leipzig, Nr. 15)

© SFB 1199  
03 / 2019

Vertrieb:  
Leipziger Universitätsverlag GmbH, Oststrasse 41, 04317 Leipzig,  
info@univerlag-leipzig.de

ISBN: 978-3-96023-223-0  
ISSN: 2510-4845

---

# Inhalt

<b>1</b>	<b>Einführung und Digital Humanities am Standort Leipzig</b>	<b>4</b>
1.1	„Unser Feld ist die Welt“: Geographische Gesellschaften 1821–1914 im internationalen Vergleich im SFB 1199	4
1.2	Projektergebnisse und methodische Herausforderungen	6
1.3	Entwicklungen und Voraussetzungen der Digital Humanities	7
1.4	Digital Humanities am Standort Leipzig	9
<b>2</b>	<b>Methodische Entwicklungen im Projekt</b>	<b>12</b>
2.1	Aufbau der Datengrundlage	12
2.2	Grundfragen der Datengewinnung und -erfassung	14
2.3	Praxis der Datenverarbeitung und -kodierung	16
2.4	Formen der Datenauswertung	19
<b>3</b>	<b>Beispiele aus der Digital Humanities-gestützten Datenauswertung</b>	<b>20</b>
3.1	Vor- und Aufbereitung des Materials für die Digital Humanities-gestützte Auswertung	21
3.2	Anwendung der Visualisierungstechnik TagPies zur inhaltlichen Analyse	22
3.3	Digital Humanities-spezifische Strategien zur Bildung von Suchbegriffen	26
3.4	Anwendung der Visualisierungskomponente Vega zur Analyse zeitlicher Entwicklungen	26
<b>4</b>	<b>Fazit und Ausblick</b>	<b>29</b>

# 1 Einführung und Digital Humanities am Standort Leipzig

Das Projekt CO1 „*Unser Feld ist die Welt*“: *Geographische Gesellschaften 1821–1914 im internationalen Vergleich* des Sonderforschungsbereichs 1199 untersucht Geographische Gesellschaften als wichtige Akteure von Verräumlichungsprozessen unter Globalisierungsbedingungen. In diesem Working Paper stellen wir die dafür am Leibniz-Institut für Länderkunde (IfL) entwickelte Methodik bzw. die zugrundeliegende Methodologie vor. Sie bedient sich konventioneller Verfahren der Datengenerierung und -verarbeitung und verbindet diese mit computergestützten Analysen aus dem Bereich der Digital Humanities, die auf diese Daten angewandt werden. Die Digital Humanities befinden sich an der Schnittstelle zwischen Informatik sowie Geistes- und Sozialwissenschaften.<sup>1</sup> Sie gelten als Bindeglied zwischen beiden Disziplinen, weil sie digitale Methoden als Werkzeuge zur Bearbeitung digitalisierter Materialien der Geisteswissenschaften anwenden.<sup>2</sup> Allerdings sind, um digitale Methoden anwenden zu können, zunächst einmal digitalisierte Materialien gefragt. Zusätzlich benötigen wir nach Abwägen der Aufwand-Nutzen-Relationen nicht nur das nötige informatische Wissen über die bestmögliche Aufbereitung und Speicherung dieser Materialien, sondern auch Experten auf dem Gebiet einer fachbezogenen Auswertung. Dabei müssen alle Fachwissenschaftler/-innen aus den verschiedenen Disziplinen offen für interdisziplinäre Zusammenarbeit sein. Dieses Working Paper ist Ergebnis einer solchen Kooperationsbereitschaft zwischen Wissenschaftler/-innen des IfL und der Universität Leipzig, die im Laufe der letzten Jahre zu einer fruchtbaren Zusammenarbeit in verschiedenen Projekten geführt hat. Wir kommen aus den Bereichen Geschichte, Geographie, Global Studies, Kulturwissenschaften und Informatik und haben im Rahmen der ersten Förderphase des SFB 1199 die Möglichkeit genutzt, durch interdisziplinäre Zusammenarbeit neue Forschungshorizonte zu entwickeln.

Derzeit sind solche Kooperationen zwischen unseren Fachgebieten noch selten. Deswegen dient dieses Working Paper nicht nur dazu, exemplarisch am Beispiel von Teilergebnissen aufzuzeigen, welchen Mehrwert dieses methodische Zusammenspiel durch die Integration konventioneller und Digital Humanities-gestützter Ansätze der Datenerhebung und -bearbeitung für unser Projekt hat, sondern indirekt auch aus der Praxis zu berichten: Denn Kooperationen dieser Art benötigen einen vergleichsweise langen methodischen Vorlauf und sind in allen Arbeitsschritten zeitintensiv, was häufig zu einem gestreckten Verwertungszyklus führt.<sup>3</sup> Mit Blick auf die Ergebnisse zeigt unser Working Paper, dass diese methodische Prämisse für alle Beteiligten eine bereichernde Wirkung hat.

## 1.1 „Unser Feld ist die Welt“: Geographische Gesellschaften 1821–1914 im internationalen Vergleich im SFB 1199

Gegenstand unserer Untersuchungen sind Geographische Gesellschaften und damit ein spätestens seit Mitte des 19. Jahrhunderts weltweit beobachtbares bürgerlich-städtisches Phänomen, das erheblich zur Herausbildung der *global condition* beitrug. Geographische Gesellschaften regten zahlreiche Expeditionen und explorative Forschungsreisen, insbesondere in die Polargebiete, nach Zentral-, Ost- und Südasiens sowie nach Afrika an, sammelten Mittel zu deren Ausstattung und übernahmen die Dokumentation, Verarbeitung und/oder kartographische Visualisierung der Ergebnisse dieser Reisen in den von ihnen herausgegebenen (wissenschaftlichen) Zeitschriften und Reihen. Seit Anfang der 1870er Jahre waren sie es, die unter Nutzung ihrer vergleichsweise guten personellen und infrastrukturellen Ausstattung an wechselnden Orten internationale Geographenkongresse organisierten und so als Foren für die Herausbildung einer rasch wachsenden internationalen Scientific Community von Geographen fungierten. Zahlreiche europäische

1 W. McCarty, „Humanities computing as interdiscipline“, 1999, <http://www.iath.virginia.edu/hcs/mccarty.html> (letzter Zugriff 25. Februar 2019).

2 P. Sahle, „Digital Humanities? Gibt's doch gar nicht!“, in: C. Baum und T. Stäcker (Hrsg.), *Grenzen und Möglichkeiten der Digital Humanities*, Sonderband 1 der Zeitschrift für digitale Geisteswissenschaften, Wolfenbüttel: Forschungsverbund Marbach Weimar Wolfenbüttel, 2015, o. S.

3 Siehe hierzu auch S. Schwandt, „Digitale Objektivität in der Geschichtswissenschaft? Oder: Kann man finden, was man nicht sucht? Ein Erfahrungsbericht“, *Rechtsgeschichte. Zeitschrift des Max-Planck-Instituts für europäische Rechtsgeschichte* (2016) 24, S. 337–338.

Gesellschaften setzten sich überdies für eine Hebung der Standards des geographischen Schulunterrichts ein und nutzten, oft gezielt, ihre Verbindungen zu nationalen und /oder regionalen Eliten von Politik, Verwaltung und Wirtschaft, um die Einrichtung von geographischen Lehrstühlen an Universitäten energisch voranzutreiben.

Der weitgehend in disziplinhistoriographischen Zusammenhängen der Geographie entstandene internationale Forschungsstand zu Geographischen Gesellschaften lässt sich in zwei weitgehend unabhängig voneinander existierende Forschungslinien einteilen. Zum einen kann eine vergleichsweise umfangreiche, vorwiegend „traditional“<sup>4</sup> ausgerichtete Historiographie beobachtet werden, die in Jubiläumspublikationen und Festschriften von kleineren, oft eng mit den jeweiligen Universitätsgeographien verbundenen Gesellschaften zu finden ist, oder aber in den Historiographien von Gesellschaften, die nationale Führungsansprüche erhoben und in den jeweiligen Hauptstädten angesiedelt waren. Eine zweite Gruppe von Forschungen charakterisiert seit den 1980er Jahren vor dem Hintergrund von Professionalisierungs- und zunehmenden Vernetzungsprozessen in der internationalen Disziplinhistoriographie<sup>5</sup> sowie in zunehmender Kombination mit neu entwickelten (de-)konstruktivistischen Forschungsperspektiven der Sozial- und Kulturgeographie Geographische Gesellschaften nicht mehr als Orte der Förderung und Pflege von vermeintlich wertfreiem geographischen Wissen, sondern als wesentliche Triebkräfte einer gezielten Herstellung, machtpolitischen Instrumentalisierung, anwendungsbezogenen Inwertsetzung und Imagination von raumbezogenem Wissen für Praktiken kolonialer und imperialer Aneignung.<sup>6</sup> Im Ergebnis dieser v. a. die national agierenden großen Gesellschaften analysierenden Forschungen kann als gesichert gelten, dass Geographische Gesellschaften in z. T. hohem Maße Lobbyarbeit für expansive nationale Regierungspolitiken betrieben haben und die von den Gesellschaften massiv vorangetriebene Exploration der Erde sowie die von ihnen organisierte fortlaufende Raubeobachtung wesentlich der Kontrolle jeweiliger nationaler als auch überseeischer Gebiete und Interessensphären diente.

Durch die im SFB 1199 verfolgten Fragestellungen zu Verräumlichungsprozessen und ihre Beziehungen zu Raumformaten und Raumordnungen<sup>7</sup> ergaben sich für das Projekt neue, über den bisherigen Stand hinausgehende Analyseperspektiven. Sie bestanden in drei wesentlichen Punkten: Erstens der Frage danach, wie man Geographische Gesellschaften vergleichend untersuchen kann, ohne einem methodischen Nationalismus anheim zu fallen, da der Fokus der vorliegenden Darstellungen in der Regel auf einer einzelnen,

4 Vgl. J. Rösen, „Die vier Typen des historischen Erzählens“, in: R. Koselleck, H. Lutz und J. Rösen (Hrsg.), *Formen der Geschichtsschreibung*, München: dtv wissenschaft, 1982, S. 514–605.

5 Vgl. U. Wardenga, „Writing the History of Geography: What we have learnt – and where to go next“, *Geographica Helvetica* 68 (2013) 1, S. 27–35.

6 Vgl. z. Bsp. M. Bassin, „The Russian Geographical Society, the ‘Amur Epoch’ and the Great Siberian Expedition 1855–1863“, *Annals of the Association of American Geographers* 73 (1983) 2, S. 240–256; ders., *Imperial Visions: Nationalist Imagination and Geographical Expansion in the Russian Far East, 1840–1865*, Cambridge: Cambridge University Press, 1999; F. Driver, „Geography’s Empire: Histories of Geographical Knowledge“, *Environment and Planning D: Society and Space* 10 (1992) 1, S. 23–40; ders., *Geography Militant: Cultures of Exploration and Empire*, Oxford: Blackwell, 2001; D. Lejeune, *Les Sociétés de géographie en France et l’expansion coloniale au XIXe siècle*, Paris: Albin Michel, 1993; A. Godlewska und N. Smith (Hrsg.), *Geography and Empire*, Oxford: Blackwell, 1994; D. Gregory, *Geographical Imaginations*, Cambridge, MA: Blackwell, 1994; H. Nicolai, « Les géographes belges et le Congo », in: M. Bruneau und D. Dory (Hrsg.), *Géographies des colonisations: XVe–XXe siècles*, Paris: L’Harmattan, 1994, S. 51–65; P. van der Velde, „The Royal Dutch Geographical Society and the Dutch East Indies, 1873–1914: From Colonial Lobby to Colonial Hobby“, in: M. Bell, R. A. Butlin und M. J. Heffernan (Hrsg.), *Geography and Imperialism: 1820–1940*, Manchester: Manchester University Press, 1995, S. 80–92; J. A. Rodríguez Esteban, *Geografía y colonialismo: La sociedad Geográfica de Madrid (1876–1936)*, Madrid: UAM Ediciones, 1996; J. Nogué und J. L. Villanova, „Spanish Colonialism in Morocco and the Sociedad Geográfica de Madrid 1876–1956“, *Journal of Historical Geography* 28 (2002) 1, S. 1–20; M. J. Heffernan, „Histories of Geography“, in: S. L. Holloway, S. P. Rice und G. Valentine (Hrsg.), *Key Concepts in Geography*, London: Sage Publications, 2003, S. 3–22; C. W. J. Withers, D. Finnegan und R. Higgitt, „Geography’s other Histories? Geography and Science in the British Association for the Advancement of Science, 1831–c.1933“, *Transactions of the Institute of British Geographers. New Series* 31 (2006) 4, S. 433–451; C. Weiss, *Wie Sibirien »unser« wurde: Die Russische Geographische Gesellschaft und ihr Einfluss auf die Bilder und Vorstellungen von Sibirien im 19. Jahrhundert*, Göttingen: V&R Unipress, 2007; R. A. Butlin, *Geographies of Empire: European Empires and Colonies c. 1880–1960*, Cambridge: Cambridge University Press, 2009.

7 Siehe hierzu M. Middell, „Raumformate – Bausteine in Prozessen der Neuverräumlichung“, Working paper series des SFB 1199 an der Universität Leipzig (2019) 14, S. 3–23. Siehe außerdem M. Möhring, G. Pisarz-Ramirez und U. Wardenga, *Imaginationen*, Berlin: De Gruyter, 2019.

im Container des Nationalstaats untersuchten Geographischen Gesellschaft lag.<sup>8</sup> Zweitens wurde im Licht des SFB die von der bisherigen Forschung noch überhaupt nicht behandelte Dimension transnationaler Verflechtungen wichtig. Der dritte Punkt betraf die Frage der Wissensordnungen und damit die ebenfalls bislang noch nicht untersuchte Dimension des möglichen Wandels von Raumsemantiken im Zuge der internationalen Herausbildung einer an Universitäten institutionalisierten Geographie, die sich seit um 1900 international mit deutlichen Distinktionszielen als eine von anderen akademischen Disziplinen unterscheidbare „Raumwissenschaft“ beschrieb.<sup>9</sup>

Um im Gesamtkontext des SFB valide Ergebnisse erzeugen zu können, hat sich das Teilprojekt zunächst auf empirische Erhebungen konzentriert, die allerdings nicht die einzelne Geographische Gesellschaft in den Mittelpunkt rückten, sondern deren Vergleiche und Verflechtungen.

## 1.2 Projektergebnisse und methodische Herausforderungen

Das Teilprojekt hat in der ersten Förderphase gut 30 in Europa, den beiden Amerikas sowie in Asien, Nordafrika und Australien angesiedelte Geographische Gesellschaften im Zeitraum von 1821–1914 in ihren Struktur- und Aktivitätsprofilen sowie ihren Verflechtungsnetzwerken vergleichend untersucht. Hierzu wurde mit Hilfe einer feingliedrigen Erfassungs- und Auswertungssystematik<sup>10</sup> eine neue Methodik für die quantitative und qualitative Analyse der mittlerweile mit rund 60.000 aus Periodika von Geographischen Gesellschaften erfassten Dateneinträgen entwickelt. Im Ergebnis der Erhebungen konnten u. a., erstens, strukturelle Spezifika der einzelnen Geographischen Gesellschaften, zweitens, Verflechtungen zwischen den Standorten und, drittens, die thematischen und regionalen Schwerpunktbildungen über das gesamte 19. Jahrhundert hinweg im internationalen Vergleich ermittelt werden. Dabei zeigte sich eine an allen Standorten zentrale Bedeutung von Geographischen Gesellschaften als Raumentrepreneurinnen bei der Herausbildung von Verräumlichungslogiken, die ihrerseits wiederum zur Grundlage der Formatierung von Räumen, z. B. als Imperium, Nationalstaat oder imperialer Ergänzungsraum wurden. Bei der Funktion Geographischer Gesellschaften als Raumentrepreneurinnen geht es nach dem derzeitigen Stand der Forschung hauptsächlich um die Territorialisierung und Durchherrschaftung des eigenen Staatsgebietes bzw. um die politische und/oder wirtschaftliche Einflussnahme, Aneignung, Kolonisierung von benachbarten und/oder außereuropäischen Gebieten. Beides basiert in besonderem Maße auf einer wissenschaftsgestützten Ressourcen- und Vermessungsorientierung, die in einem engen Zusammenhang mit der Verwissenschaftlichung und Professionalisierung Geographischer Gesellschaften steht. Die Art und Weise, inwieweit Geographische Gesellschaften als Raumentrepreneurinnen agieren, unterscheidet sich anhand ihrer Aktivitäts- und Interessenprofile.

Häufig sind es gerade die spezifischen Herausforderungen eines Forschungsdesigns sowie dessen Operationalisierung innerhalb eines Projektverlaufs, die in besonderer Weise die Entwicklung methodischer Ansätze vorantreiben. So auch im Falle unseres Projektes, bei dem es, bildlich gesprochen, in hohem Maße um das sichere Navigieren zwischen der Skylla der Datenquantität und der Charybdis der Datenqualität ging. Es war deshalb erforderlich, einen Weg zu finden, der einerseits den bereits aus den Vorarbeiten zu erwartenden erheblichen Umfang global vorhandener Datenbestände zu Geographischen Gesellschaften auf ein bearbeitbares Maß eingrenzte bzw. verdichtete, der andererseits aber, mit Blick auf die an das Material zu stellenden differenzierten Fragen, hinreichend breit bzw. tief angelegt sein musste, damit entsprechende, qualitativ gehaltvolle Inhalte aus den erfassten Datenbeständen extrahiert werden konnten. Dabei musste drei methodischen Herausforderungen begegnet werden, die sich aus den oben genannten Punkten<sup>11</sup> ergeben: Erstens sollten aus dem zu erhebenden Daten- bzw. Quellenmaterial Indikatoren für einen Vergleich der Geographischen Gesellschaften generiert werden; zweitens sollten die so erzeugten komparativ aus-

8 Als Ausnahmen vgl. T. W. Freeman, "The Manchester and Royal Scottish Geographical Societies", *The Geographical Journal* 150 (1984) 1, S. 55–62.; W. H. Schneider, "Geographical Reform and Municipal Imperialism in France, 1870–80", in: J. M. MacKenzie (Hrsg.), *Imperialism and the Natural World*, Manchester: Manchester University Press, 1990, S. 90–117; Lejeune, *Les Sociétés de géographie en France et l'expansion coloniale au XIXe siècle*; J. M. MacKenzie, "The Provincial Geographical Societies in Britain, 1884–1914", in: Bell, Butlin und Heffernan (Hrsg.), *Geography and Imperialism*, S. 93–124; Butlin, *Geographies of Empire*.

9 Vgl. U. Wardenga, *Geographie als Chorologie: Zur Genese und Struktur von Alfred Hettners Konstrukt der Geographie*, Stuttgart: Steiner, 1995.

10 Diese wird in Abschnitt 2.3 erläutert.

11 Siehe Abschnitt 1.1.

wertbaren Datenbestände die Möglichkeiten mehrdimensionaler Typisierungen erlauben und drittens sollten sich aus dem erarbeiteten Datenbestand Vernetzungsbeziehungen der Geographischen Gesellschaften rekonstruieren lassen. Dazu galt es, zunächst eine geeignete Datenbasis zu finden, die sich für unser Projekt aus der Zeitschriftenproduktion von Geographischen Gesellschaften über knapp 100 Jahre ergab. Mit konventionellen Verfahren der Datenerhebung und -erfassung wurde ein Datenbestand und Analysekörper angelegt<sup>12</sup>, den wir schließlich sowohl konventionell als auch mithilfe computergestützter Anwendungen aus dem Bereich der Digital Humanities<sup>13</sup> analysieren wollten. Die methodische Kopplung beider Verfahren hatte zum Ziel, aggregierte differente Muster von Verräumlichungsprozessen und -logiken aufzuzeigen, die sich aus der bloßen Betrachtung einzelner Geographischer Gesellschaften nicht erschließen lassen, und neben konventionellen Verfahren deshalb auch die Chancen zu nutzen, die die Digital Humanities bieten

### 1.3 Entwicklungen und Voraussetzungen der Digital Humanities

In den letzten Jahren haben die Digital Humanities eine breite Entfaltung und rasante Weiterentwicklung erfahren.<sup>14</sup> Ihre Anfänge gehen allerdings bereits auf die 1940er Jahre zurück, weitere entscheidende Entwicklungsschritte werden auf die 1960er und 1980er Jahre datiert.<sup>15</sup> Mit der Durchsetzung des Internets und im Zuge der technischen Entwicklungen im Bereich der Computer Sciences haben die Digital Humanities seit den 1990er Jahren vereinfachte Anwendbarkeit und weitreichende Aufmerksamkeit erlangt und den Weg in eine größere Wissenschaftscommunity gefunden. Spätestens mit Beginn der 2010er Jahre wurden sie auch in den populären Medien diskutiert.<sup>16</sup> Zunächst konsolidierte und institutionalisierte sich das Feld im englischsprachigen Raum, was eine große Bandbreite an entsprechender Literatur belegt.<sup>17</sup> Auch im deutschsprachigen Raum existiert mittlerweile eine beinahe unüberschaubare Anzahl von Projekten, Tagungen, Zeitschriften und Blogs, die sich unter dem Dach der Digital Humanities verorten.<sup>18</sup> Darüber hinaus haben sich Verbände mit eigens eingerichteten Kompetenzzentren gegründet und eigene Professuren an Universitäten wurden eingerichtet, die differenzierte Qualifikationsmöglichkeiten für Studierende im Bereich der Digital Humanities anbieten.<sup>19</sup>

Parallel dazu laufen umfangreiche Digitalisierungsinitiativen analoger Bestände etwa in Archiven, Bibliotheken, Museen und Stiftungen, die die Grundlage für völlig neue Formen von Recherche, Zugriff und Abruf bilden.<sup>20</sup> Ursprünglich aus dem Bemühen heraus, die jeweiligen Bestände langfristig und nachhaltig zu

12 Vgl. hierzu besonders die Abschnitte 2.1 und 2.2.

13 Vgl. hierzu besonders die Abschnitte 3.2 und 3.3.

14 Vgl. z. Bsp. K. Fitzpatrick, "The Humanities, Done Digitally", in: M. K. Gold (Hrsg.), *Debates in the Digital Humanities*, Minneapolis, MN: University of Minnesota Press, 2012, S. 12–15; S. Schreibman, R. G. Siemens und J. Unsworth (Hrsg.), *A New Companion to Digital Humanities*, Chichester, West Sussex: Wiley Blackwell, 2016; F. Jannidis, H. Kohle und M. Rehbein (Hrsg.), *Digital Humanities: Eine Einführung*, Stuttgart: J.B. Metzler, 2017; P. Svensson, "The Landscape of Digital Humanities", *Digital Humanities Quarterly* 4 (2010) 1, o. S.; D. M. Berry, "The Computational Turn: Thinking about the Digital Humanities", *Culture Machine* 12 (2011), S. 1–22; P. Sahle, "Digital Humanities? Gibt's doch gar nicht!".

15 Zur Geschichte der Digital Humanities siehe u. a. S. Hockey, "The History of Humanities Computing", in: S. Schreibman, R. G. Siemens und J. Unsworth (Hrsg.), *A Companion to Digital Humanities*, Malden, MA: Blackwell Publishing, 2004, S. 3–19; Jannidis, Kohle und Rehbein (Hrsg.), *Digital Humanities*, S. 3 ff. Siehe außerdem einführend zur Geschichte und Namensgebung M. G. Kirschenbaum, "What Is Digital Humanities and What's It Doing in English Departments?", *ADE Bulletin* (2010) 150, S. 55–61 und W. G. Thomas, "The Promise of the Digital Humanities and the Contested Nature of Digital Scholarship", in: Schreibman, Siemens und Unsworth (Hrsg.), *A New Companion to Digital Humanities*, S. 524–537. Siehe außerdem Fitzpatrick, "The Humanities, Done Digitally".

16 Ebd.

17 Siehe Hinweise aus Fußnote 15 zur Geschichte der Digital Humanities.

18 Siehe hierzu vertiefend P. Sahle, "Digital Humanities? Gibt's doch gar nicht!".

19 Ebd.

20 Siehe weiterführend T. Efer, *Graphdatenbanken für die textorientierten e-Humanities*, Dissertationsschrift, Leipzig: Universität Leipzig (2016), S. 5. Siehe außerdem Deutsche Forschungsgemeinschaft, "Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung von Sprachkorpora", 2015, [https://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/informationen\\_fachwissenschaften/geisteswissenschaften/standards\\_sprachkorpora.pdf](https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/standards_sprachkorpora.pdf) (letzter Zugriff 7. Februar 2019); Deutsche Forschungsgemeinschaft, "Informationen zu rechtlichen Aspekten bei der Handhabung von Sprachkorpora", 2015, [https://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/informationen\\_fachwissenschaften/geisteswissenschaften/standards\\_recht.pdf](https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/standards_recht.pdf) (letzter Zugriff 7. Februar 2019); Deutsche Forschungsgemeinschaft, "DFG-Praxisregeln, Digitalisierung", 2016, [http://www.dfg.de/formulare/12\\_151/12\\_151\\_de.pdf](http://www.dfg.de/formulare/12_151/12_151_de.pdf) (letzter Zugriff 7. Februar 2019).

sichern, aufzubewahren/abzulegen und zugänglich zu machen, stellt das digital verfügbare Material die Arbeitsgrundlage für digitale Forschungspraktiken dar.<sup>21</sup>

Für unser Projekt waren die Verfügbarkeit digitalisierter Periodika sowie die von uns entwickelte Systematik als methodische Vorverarbeitungsschritte der Materialien entscheidend, um computergestützte Auswertungen vornehmen zu können. Denn für digitale Forschung benötigen wir zunächst adäquate Quellen. Sie entscheiden „über die Möglichkeit des effektiven Einsatzes von digitalen Werkzeugen“.<sup>22</sup> So mögen „(...) Datenbanken und digitale Tools zur Speicherung und schnellen Bereitstellung von Informationen hilfreich sein“, „der Erkenntnisfortschritt“ kann allerdings „erst in der Zusammenschau und der Kontextualisierung des Materials erreicht“ werden, „einem Prozess, der momentan noch nicht digital nachzubilden ist, da er den vollen Kontext menschlicher Erfahrungsräume erfordert – eine Komplexität also, die noch keine Ontologie oder neuronales Netzwerk außerhalb des menschlichen Gehirns heute in der Lage ist zu liefern.“<sup>23</sup> Es sind also nicht nur Experten für die Technik gefragt, die die computergestützten Anwendungen umsetzen können, sondern auch Geisteswissenschaftler/innen, die Verfahren der Digital Humanities gegenüber offen sind, über einen entsprechend theoriebezogenen Zugang verfügen und wissen, welche Fragen an digitale Ressourcen überhaupt gestellt werden können und wo die Grenzen der Vorgehensweise für das jeweilige Projekt liegen.

*„In diese Prozesse können beide Seiten ihre jeweiligen Stärken einbringen. Wichtige Kompetenzen der Informatik liegen dabei u. a. in der formell korrekten und technisch effizienten Modellierung von Daten, im Umgang mit sehr großen Datenmengen, der Vernetzung von Wissensressourcen, der Entwicklung von Analyseverfahren und -werkzeugen sowie der Visualisierung von Analyseergebnissen. Die Geisteswissenschaften bringen dagegen u. a. sorgfältige manuelle Arbeitsweisen, theoriegeleitete Auswertungs- und Interpretations-Frameworks, ein hohes Verständnis für relevante Kontexte der Quellen und Daten sowie Erfahrung im Umgang mit unscharfen und unterspezifizierten Kategorien ein.“<sup>24</sup>*

Wenn diese Komponenten nicht oder nicht ausreichend abrufbar vorhanden sind, kann es trotz des offenkundigen Trends der Hinwendung zu den sich mehr und mehr konsolidierenden Digital Humanities zu einer ganzen Reihe an Hindernissen kommen, die interdisziplinär angelegte Forschungsprojekte und -arbeiten teilweise aufwendig und kompliziert gestalten. Neuere Arbeiten beleuchten deswegen u. a. Möglichkeiten und Grenzen digitaler Methoden für die Geisteswissenschaften und gehen auf Beispiele aus unterschiedlichen Fachdisziplinen ein, in denen neue bzw. neu erschlossene Quellen auch neue methodische Vorgänge und Implikationen erfordern.<sup>25</sup> So ist mittlerweile nicht nur eine eigene Historiographie des Feldes entstanden, sondern die Literatur zu Digital Humanities reflektiert auch über ihre Anwendungen, ihren Nutzen und die Herausforderungen interdisziplinärer Kooperationen.<sup>26</sup> Dies spricht für eine neu entstandene Disziplin, selbst wenn der aktuelle Forschungsstand nach wie vor auch Bezeichnungen wie Feld, Bereich und Area

21 Efer, *Graphdatenbanken für die textorientierten e-Humanities*, S. 4 ff. Weiterführend W. McCarty, „Collaborative Research in the Digital Humanities“, in: M. Deegan und W. McCarty (Hrsg.), *Collaborative Research in the Digital Humanities: A volume in honour of Harold Short, on the occasion of his 65th birthday and his retirement, September 2010*, Farnham: Ashgate, 2012, S. 1–10 sowie ders., „Becoming interdisciplinary“, in: Schreibman, Siemens und Unsworth (Hrsg.), *A New Companion to Digital Humanities*, S. 69–83.

22 J. Wettlaufer, „Digitale Transformation der Geschichtswissenschaft: Welcher Erkenntnisfortschritt? Wann?“, *Rechtsgeschichte. Zeitschrift des Max-Planck-Instituts für europäische Rechtsgeschichte* (2016) 24, S. 344–346, Zitat S. 345.

23 Ebd., S. 344 ff.

24 Efer, *Graphdatenbanken für die textorientierten e-Humanities*, S. 5 ff. Efer verweist in diesem Zusammenhang auch auf die Praxisregeln der Deutschen Forschungsgemeinschaft (vgl. Fußnote 20) und auf D. A. Michelson, „Syriaca.org as a Test Case for Digitally Re-Sorting the Ancient World“, in: C. Clivaz, P. C. Dillely und D. Hamidović (Hrsg.), *Ancient Worlds in Digital Culture*, Leiden: Brill, 2016, S. 59–85.

25 Damit setzt sich eine Vielzahl von Aufsätzen in einer Themenrubrik der 24. Ausgabe von „Rechtsgeschichte. Zeitschrift des Max-Planck-Instituts für europäische Rechtsgeschichte“ (Rg) auseinander. Siehe etwa C. Birr, „Die geisteswissenschaftliche Perspektive: Welche Forschungsergebnisse lassen Digital Humanities erwarten?“, *Rg* (2016) 24, S. 330–334; B. Jurish, „Tools, Toys, and Filters: A Tinker’s Apology“, *Rg* (2016) 24, S. 347–348; Schwandt, „Digitale Objektivität in der Geschichtswissenschaft? Oder: Kann man finden, was man nicht sucht?“, *Rg* (2016) 24, S. 335–336; Wettlaufer, „Digitale Transformation der Geschichtswissenschaft“.

26 Zur Historiographie vgl. beispielsweise J. Nyhan und A. Flinn, *Computation and the Humanities: Towards an Oral History of Digital Humanities*, Cham: Springer, 2016; University College London, Center for Digital Humanities und J. Nyhan, „Hidden Histories: Digital Humanities 1949 – Present“, <https://hiddenhistories.github.io> (letzter Zugriff 18. Februar 2019); M. Thaller, „Geschichte der Digital Humanities“, in: Jannidis, Kohle und Rehbein (Hrsg.), *Digital Humanities*, S. 3–12. Zur Geschichte und Entwicklung der Digital Humanities vgl. O. Le Deuff, *Digital Humanities: History and Development*, London: ISTE Ltd, 2018.



als Charakterisierung für die Digital Humanities gebraucht.<sup>27</sup> Dabei sind sich die Vertreter der Digital Humanities darüber einig, dass eine allgemeingültige Definition bisher aussteht.<sup>28</sup> Der Trend weist eher in die Gegenrichtung, weil den Digital Humanities-Experten ein „Definitionen-Sport“ unterstellt wird, bei dem „jeder DH-Spezialist mindestens eine eigene Definition“ hervorbringt.<sup>29</sup> Dieses Phänomen ist für Frühphasen der Herausbildung transdisziplinärer Forschungszusammenhänge im Bemühen um Klärung von Standorten und Positionen sowie zur Rechtfertigung und Selbstbehauptung des neuen Feldes hinreichend bekannt. Deshalb wollen wir hier diese Diskussion nicht weiter fortführen, sondern anhand eines praktischen Beispiels den Mehrwert einer gelungenen Kooperation vorstellen, die schon seit Längerem besteht und in Zukunft weiter ausgebaut werden soll. Für unsere Zwecke gibt Patrick Sahle in seiner Definition von Digital Humanities einen entscheidenden Hinweis:

*„Die Digital Humanities befassen sich mit Problemen, die über die benachbarten Einzelfächer in einem doppelten Sinne hinausgehen. Zum einen betreffen sie Fragen, die für viele Fächer gleichermaßen gelten; zum anderen betreffen sie Fragen, die von den benachbarten Fächern nicht behandelt werden, weil sie aus ihrer Sicht zu speziell sind oder Kompetenzen erfordern, die in den Fächern nicht enthalten sind.“<sup>30</sup>*

Diese Kompetenzen betreffen zumeist die technischen Voraussetzungen zur Bearbeitung der Forschungsfragen. Durch computergestützte Verfahren können manuelle oder teilweise automatisierte Arbeitsschritte für die Bearbeitung von einer Vielzahl von digitalen Ressourcen in den Geistes-, Kultur- und Sozialwissenschaften ergänzt, teilweise sogar ersetzt werden.<sup>31</sup> Die Anwendungen der Digital Humanities haben aber nicht nur die Möglichkeit oder den Auftrag, konventionelle Verfahren der Datenbearbeitung einfach zu erweitern, zu erneuern oder auszutauschen, sondern sie bieten eine neue Forschungsheuristik, aus der sich neue Forschungsfragen generieren lassen. Das bedeutet, dass sich die Methoden der Digital Humanities und konventionelle Verfahren gegenseitig bereichern können. Dies setzt aber voraus, dass bei allen Beteiligten eine grundsätzliche Bereitschaft zur Kooperation und damit eine Haltung des wechselweisen Lernens besteht.

## 1.4 Digital Humanities am Standort Leipzig

Getreu ihrem Leitspruch „Aus Tradition Grenzen überschreiten“, beherbergt die Universität Leipzig zahlreiche interdisziplinäre und institutsübergreifende Initiativen, welche die reichhaltige und vielgestaltige geisteswissenschaftliche Tradition mit digitaler Kompetenz verbinden. In einem von großer methodischer Offenheit geprägten Forschungsumfeld finden sich die unterschiedlichsten Formen der Zusammenarbeit, welche oft auf das Engagement einzelner Fachdisziplinen zurückgehen.

So ist die Universität Leipzig beispielsweise seit 2009 Gastgeberin der von Elisabeth Burr, Professorin für französische/frankophone und italienische Sprachwissenschaft, durchgeführten „European Summer University in Digital Humanities“. Am Institut für Informatik ist seit 2008 im Rahmen von D-SPIN und später CLARIN-D unter der Leitung von Gerhard Heyer, Professor für Automatische Sprachverarbeitung, ein Kompetenzzentrum entstanden, welches seither als Teil einer europäischen Forschungsinfrastruktur für die Geistes- und Sozialwissenschaften neben Servicedienstleistungen und Forschungsarbeiten im Bereich von

27 Willard McCarty bezeichnet die Digital Humanities als akademisches Feld. Vgl. McCarty, „Humanities computing as interdisciplinary“. Patrik Svenssons Definition ähnelt der McCartys. Die Digital Humanities seien ein loses Feld oder eine Area an der Schnittstelle zwischen „the humanities and information technology or the digital“. Svensson, „The Landscape of Digital Humanities“, Absatz 12. In dem Artikel stellt Svensson darüber hinaus einen kritischen Überblick über die Digital Humanities und ihre spezifischen Anwendungsgebiete dar.

28 Frühere Begrifflichkeiten wie Computing in the Humanities und Humanities Computing werden heutzutage eher mit Digital Humanities und e-Humanities ersetzt. Zur Verwendung der Begrifflichkeiten vgl. u. a. Berry, „The Computational Turn“ und A. Aschenbrenner et al., „Von e-Science zu e-Humanities – Digital vernetzte Wissenschaft als neuer Arbeits- und Kreativbereich für Kunst und Kultur“, *Bibliothek. Forschung und Praxis* 31 (2007) 1, S. 11–21.

29 P. Sahle, „Digital Humanities? Gibt's doch gar nicht!“, Abschnitt 1. Vgl. außerdem M. Terras, J. Nyhan und E. Vanhoutte (Hrsg.), *Defining Digital Humanities: A Reader*, Farnham: Ashgate, 2013.

30 P. Sahle, „Digital Humanities? Gibt's doch gar nicht!“, Abschnitt 2.

31 Vgl. die ausführliche Darstellung in den Abschnitten 2 und 3.

Technologien, Forschungs- und Metadatenmanagement auch eine wichtige Transfer- und Vermittlerrolle in Bezug auf die Fachdisziplinen einnimmt.<sup>32</sup>

Über die Jahre wurden zahlreiche Projekte unter anderem in den Altertums-, Orient-, Politik-, Sprach- und Musikwissenschaften durchgeführt, oft auch in Form von Kooperationsprojekten mit lokalen, nationalen und internationalen Partnern. Solche Aktivitäten werden am Standort nicht nur als solitäre Initiativen einzelner Akteure wahrgenommen, sondern sind in der Regel auch institutionell verankert. Zudem sind sie nicht zuletzt als Teil des Forschungsprofilbereichs „Sprache und Kultur im Digitalen Zeitalter“ auch von hochschulstrategischer Bedeutung.<sup>33</sup>

Während die Fachdisziplinen unter dem Eindruck digitalisierter Quellenmaterialien und sich wandelnder Fachfragen den Weg in die Digital Humanities suchen, ist auch auf Seiten der Informatik das Interesse groß, in diesem Bereich zu forschen. Mit Gregory Crane, Humboldt-Professor für Digital Humanities, und Manuel Burghardt, Juniorprofessor für Computational Humanities, sind zwei ausgewiesene Experten auf dem Gebiet der Anwendung digitaler Methoden in den Geistes-, Kultur- und Sozialwissenschaften in die Forschungs- und Lehrtätigkeit des Instituts für Informatik eingebunden. Sie verantworten auch den Bachelor- und den Masterstudiengang Digital Humanities (B.Sc. und M.Sc.), welche deutschlandweit durch ihren expliziten Schwerpunkt auf informatische Fragestellungen eine Sonderstellung innerhalb der Digital Humanities-Studiengänge einnehmen.<sup>34</sup>

Darüber hinaus kümmert sich an der Universität Leipzig eine zentrale Kontaktstelle bestehend aus dem Forschungsdezernat, dem Rechenzentrum und der Bibliothek um Fragen zum Forschungsdatenmanagement, die auch in entsprechenden Beschlüssen zur Forschungsdatenpolicy der Universität ihren Niederschlag findet.<sup>35</sup> Ein kürzlich von der für die strukturierte Doktorandenausbildung zuständigen Research Academy Leipzig angebotener Workshop zum Thema Forschungsdatenmanagement, auf dem Teile der genannten Projekte näher vorgestellt wurden, zeigt, inwiefern auch die Wissens- und Forschungsdokumentation über solche Projekte und der Transfer von Theorie in die Praxis für Nachwuchsforscher/innen unterstützt wird.<sup>36</sup> Viele dieser Bemühungen werden seit Mai 2018 in einem neugegründeten Forum für Digital Humanities Leipzig gebündelt, welches nicht nur der Diskussion über Digital Humanities und ihre praktischen Anwendungsgebiete eine Gesprächs- und Kooperationsmöglichkeit bieten will, sondern darüber hinaus seinen Mitgliedern einen Überblick zu laufenden und bevorstehenden Initiativen bietet.<sup>37</sup>

Das Forum steht u. a. im Austausch mit der an der Sächsischen Akademie der Wissenschaften angesiedelten Arbeitsgruppe „Digital Humanities Mitteldeutschland“, welche unter Beteiligung zahlreicher Universitäten, Digital Humanities-Initiativen und außeruniversitärer Forschungseinrichtungen aus der Region eine niedrigschwellige Vernetzung anstrebt.<sup>38</sup> Ein ähnlicher Grundgedanke steht auch hinter dem vom Leipziger Universitätsarchiv aufgebauten „Mitteldeutschen Archivnetzwerk“, in dem die Nutzung moderner, insbesondere auch digitaler Arbeitsweisen und Infrastrukturen gefördert wird.<sup>39</sup> Als weitere interessante Initiative ist die interdisziplinäre Arbeitsgruppe „#digitalegegenwart“ zu nennen, in welcher der Spannungsbogen von Medien, Digitalität, Wissen und Räumlichkeit thematisiert wird.<sup>40</sup>

Wir selbst blicken auf verschiedene Formen einer ertragreichen interdisziplinären Zusammenarbeit in den letzten Jahren zurück. Ninja Steinbach-Hüthers Dissertation im Fach Global Studies über die Zirkulation

32 Siehe Universität Leipzig, „Repository CLARIN-D Centre Leipzig“, <https://clarin.informatik.uni-leipzig.de/repo/> (letzter Zugriff 26. Februar 2019). Siehe außerdem Universität Tübingen, „CLARIN-D“, <https://www.clarin-d.net/de/> (letzter Zugriff 26. Februar 2019); Universität Tübingen, „Deutsche Sprachressourcen-Infrastruktur D-SPIN“, <https://weblicht.sfs.uni-tuebingen.de/dspin.shtml> (letzter Zugriff 26. Februar 2019).

33 Universität Leipzig, „Intelligente Methoden und Materialien: Sprache und Kultur im Digitalen Zeitalter“, <https://www.uni-leipzig.de/forschung/forschungsprofil/strategische-forschungsfelder/intelligente-methoden-und-materialien/#c12554> (letzter Zugriff 1. März 2019).

34 Universität Leipzig, Fakultät für Mathematik und Informatik, „Informationen zu den Studienangeboten im Bereich Digital Humanities“, <http://studium.fmi.uni-leipzig.de/studium/digital-humanities/> (letzter Zugriff 1. März 2019).

35 Universität Leipzig, „Forschungsdatenmanagement“, <https://www.uni-leipzig.de/forschung/forschungsservice/forschungsdatenmanagement/> (letzter Zugriff 7. Februar 2019).

36 Teile der Seminarvorträge finden sich online: <https://www.ral.uni-leipzig.de/de/research-academy-leipzig/struktur/kompetenzschule/junge-wissenschaft-und-praxis/> (letzter Zugriff 7. Februar 2019).

37 Forum für Digital Humanities Leipzig, <https://fdhl.info/> (letzter Zugriff 7. Februar 2019).

38 Sächsische Akademie der Wissenschaften, „AG Digital Humanities Mitteldeutschland“, <https://www.saw-leipzig.de/de/ueber-die-akademie/arbeitsgruppen/ag-digital-humanities-mitteldeutschland> (letzter Zugriff 7. Februar 2019).

39 Universitätsarchiv Leipzig, „Mitteldeutsches Archivnetzwerk“, <http://mitteldeutschearchive.de/> (letzter Zugriff 7. Februar 2019).

40 Universität Leipzig, „#digitalegegenwart“, <http://home.uni-leipzig.de/digitalegegenwart/> (letzter Zugriff 7. Februar 2019).

afrikanischen Wissens und die Präsenz und Rezeption akademischer Literatur aus Afrika ist hier besonders hervorzuheben. In ihrer als Cotutelle zwischen der Universität Leipzig und der École normale supérieure Paris durchgeführten Dissertation integriert Steinbach-Hüther die kooperative Arbeit mit Thomas Efer aus der Informatik. Sie führt in die selbst entwickelte Methodologie ein, Bibliotheksdaten in einem bibliotheksfernen Kontext für Analysezwecke zu nutzen. Die Arbeit zeigt, dass Ansätze der Digital Humanities-gestützten Datenauswertung auch helfen, die von der qualitativen Forschung endogen erzeugten blinden Flecke wahrzunehmen und daher aufgrund datengetriebener Generierung neuer Hypothesen die vorhandenen Deutungsmuster zu ergänzen, erweitern und hinterfragen.<sup>41</sup> Thomas Efers Dissertation enthält einen Unterabschnitt zu den technisch-konzeptionellen Details des dafür entwickelten Recherchewerkzeugs.<sup>42</sup>

Ihre Kenntnisse und Erfahrungen aus der Zusammenarbeit konnten Ninja Steinbach-Hüther und Thomas Efer an verschiedensten Stellen anbringen, sei es im Rahmen von Vorträgen<sup>43</sup>, von Veranstaltungen für Studierende der European und Global Studies<sup>44</sup> oder auch anlässlich eines im SFB 1199 von Steffi Marung, Julia Oheim und Ninja Steinbach-Hüther organisierten Doktorandenworkshops, „Working with ‘Big Data’, textual and digital sources: Insights from the digital humanities“.<sup>45</sup>

All diese Formate der fächerübergreifenden Kooperation haben einen Sockel für das hier verfolgte Vorhaben gelegt, der zunächst einmal die Kultur des wechselseitigen Lernens etabliert hat und darauf aufbauend die gemeinsame Neugier auf neue Erkenntnisse stimuliert. Die hier vorzustellende Studie setzt gerade an der Schnittstelle zwischen einerseits eher konventionellen, d. h. in einer der beteiligten Disziplinen bereits seit Längerem bekannten und etablierten Verfahren, und andererseits Digital Humanities-gestützten Anwendungen der Verarbeitung von Daten aus der geistes- und sozialwissenschaftlichen Forschung an. Im Folgenden legen wir dar, wie es im Projekt gelang, Geographische Gesellschaften weltweit und international vergleichend analysierbar zu machen. Dazu beschreiben wir, wie wir zu einer geeigneten Datenbasis gefunden haben und wie wir durch ihre systematische Auswertung unseren Forschungsdatenbestand aufbauen konnten, der für spätere Analysezwecke auch Digital Humanities-geeignet sein sollte.

41 Siehe N. Steinbach-Hüther, *Zirkulation afrikanischen Wissens. Präsenz und Rezeption akademischer Literatur aus Afrika in Deutschland und Frankreich*, Dissertationsschrift, Leipzig, Paris: Universität Leipzig, École normale supérieure (2017).

42 Siehe Efer, *Graphdatenbanken für die textorientierten e-Humanities*, S. 149 ff.

43 T. Efer und N. Steinbach-Hüther, *Quantitative Analyses in Global and Area Studies using Graph-based Filtering of Heterogeneous Catalogue Data*, vorgetragen anlässlich der Tagung *INFORMATIK 2014: Big Data – Komplexität meistern*, Stuttgart, 22. bis 26. September 2014; N. Steinbach-Hüther und T. Efer, *African knowledge in BnF catalogue data. Quantitative graph-based analyses clarifying tracks of ‘published and perished’?*, vorgetragen als Keynote anlässlich der Tagung *ELIAS 9th annual meeting*, Paris, 7. Juli 2015; N. Steinbach-Hüther und T. Efer, *Datengetriebene Annäherung an Verräumlichungsprozesse. Verlage als internationale Akteure?*, vorgetragen anlässlich der Tagung *Respatialization of the World – Actors, Moments, Effects*, Leipzig, 6.–9. Juni 2016.

44 Entsprechende Veranstaltungen im Seminar “Methods for the Study of Globalization” haben in den Wintersemestern 2016 / 2017, 2017 / 2018 und 2018 / 2019 stattgefunden.

45 Universität Leipzig, “Collaborative Research Centre 1199. Processes of Spatialization under the Global Condition: Doing Research in Interdisciplinary and Global Environments. Methodologies, Sources, Research Organization”, <https://research.uni-leipzig.de/~sfb1199/event/doing-research-in-interdisciplinary-and-global-environments/> (letzter Zugriff 7. Februar 2019).

## 2 Methodische Entwicklungen im Projekt

Bei der Sichtung der umfangreichen Materialien und Archivalien zu Geographischen Gesellschaften stellte sich die Ausgangssituation so dar, dass zunächst auf der Basis historisch verlässlicher Datenzusammenstellungen<sup>46</sup> für den Beobachtungszeitraum von 1821 bis 1914 und für knapp 130 weltweit verteilte Geographische Gesellschaften Material zur Verfügung steht. In der ersten Untersuchungsphase ließen sich durch Auswertung ergänzender Datenquellen (z.B. historischer Adressbücher) bzw. einschlägiger Literatur (z.B. Jubiläums- oder Festschriften von Geographischen Gesellschaften) weitere Gesellschaften, Vereine und ähnlich geographiebezogene Institutionen mit dem Ergebnis identifizieren, dass der potenzielle Gesamtbestand global auf rund 170 Einrichtungen quellentechnisch nachverdichtet werden konnte.

Aufgrund einer extrem heterogenen und in der Überlieferung äußerst lückenhaften archivalischen Bestandssituation haben wir uns mit Blick auf die Forschungsfragestellungen des SFB entschieden, die von Geographischen Gesellschaften herausgegebenen Zeitschriften zur Grundlage unserer Analysen zu machen, weil diese (teilweise sehr umfangreichen) Kommunikationsartefakte mit ihrem ausdifferenzierten Berichts- und Beitragswesen ideale Quellen für die Bearbeitung unserer oben beschriebenen Forschungsfragen der Herstellung von Vergleichbarkeit, mehrdimensionaler Typisierung und der Aufdeckung von transnationalen Vernetzungsstrukturen sind.

### 2.1 Aufbau der Datengrundlage

Um zu einem bearbeitbaren Zeitschriftenkorpus zu gelangen, filterten wir in einem mehrstufigen Auswahlprozess solche Geographischen Gesellschaften heraus, deren Kommunikationsartefakte es erlauben, einen hinreichend differenzierten Datenbestand aufzubauen. Diesbezügliche Auswahlkriterien waren: 1) Die auszuwählende Geographische Gesellschaft musste über ein eigenes Publikationsorgan verfügen, das allerdings nicht nur episodisch, sondern periodisch, in der Regel im Jahresrhythmus, erschien. 2) Das Periodikum sollte über einen längeren Zeitraum veröffentlicht worden sein, um mit der Datenerfassung auch die valide Analyse von Zeitreihen abdecken zu können. Zeitschriften mit einem Ersterscheinen nach 1890 wurden deshalb ausgeschlossen, um mindestens eine 25-jährige Beobachtungszeit abbilden zu können. Einzig im Hinblick auf die räumliche Ausgewogenheit des Gesamtdatenbestandes haben wir diesen Grundsatz partiell verlassen und entsprechende Ausnahmen gemacht (so bei den Geographischen Gesellschaften von Lima, Philadelphia, Prag, Algier und La Paz, alle erst erschienen zwischen 1891–1898). 3) Ein standortnaher, d.h. reisekostenneutraler Bibliothekszugriff auf die Originale (z.B. in Kooperation mit der Bibliothek der Leopoldina – Nationale Akademie der Wissenschaften in Halle) oder aber ein Zugriff auf hochwertige Digitalisate in einschlägigen elektronischen Repositorien (z.B. über Gallica der Bibliothèque Nationale de France) sollte möglich sein. 4) Die im Projekt bzw. am Universitätsstandort Leipzig verfügbare Sprachkompetenz (die z.B. über Masterarbeiten oder Werkverträge genutzt wurde) bildete das abschließende Auswahlkriterium für die Aufnahme eines Periodikums in den Datenbestand.

Auf diese Weise haben wir eine Datenbasis (Abbildung 1) aus den Zeitschriften von 34 Geographischen Gesellschaften aus 22 Staaten auf fünf Kontinenten in 13 Sprachen (Tabelle 2) aufbauen können. Auch in dieser Auswahl spiegelte sich die historisch gewachsene Dominanz europäischer Geographischer Gesellschaften wider, denn nach Kollm<sup>47</sup> waren fast die Hälfte (44,4 %) aller 1908 weltweit bekannten Geographischen Gesellschaften in Frankreich und dem Deutschen Reich angesiedelt, ein knappes Drittel (30,2 %) in weiteren europäischen Staaten und in Summe nur ein gutes Viertel in der gesamten außereuropäischen Welt (16,7 %) sowie im Russischen Reich (8,7 %).

46 G. Kollm, „Geographische Gesellschaften, Zeitschriften, Kongresse und Ausstellungen“, *Geographisches Jahrbuch* 32 (1909), S. 409–438.

47 Ebd., S. 419.

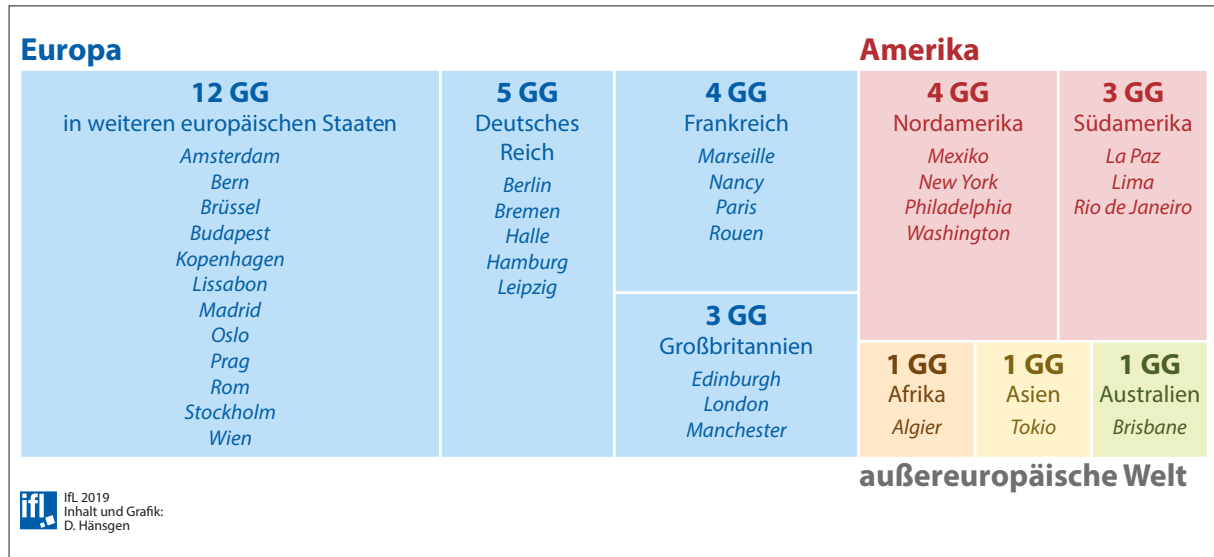


Abbildung 1: Räumliche Verteilung der 34 ausgewählten Geographischen Gesellschaften

Da die materiellen Publikationsformen der Zeitschriften sowohl vom jeweiligen Umfang, dem Druckformat als auch von der Erscheinungsfolge in Jahressbänden, Teilbänden, Heften oder entsprechenden Mischformen stark variierten, führten wir zur groben Darstellung und Kalkulation der Mengengerüste die fiktive Maßeinheit der *Berichtsjahre* einer Geographischen Gesellschaft ein, die vom Jahr des Ersterscheinens bis zum Ende des Beobachtungszeitraumes 1914 die rein rechnerische Spanne zwischen diesen beiden Referenzpunkten repräsentieren. Global gelangt man auf diese Weise zu einem Berichts- bzw. Beitragsumfang aus den Periodika, der einer kumulierten Zeitspanne von 1.380 Berichtsjahren entspricht (vgl. Tabelle 1).

Ausgewählte Geographische Gesellschaften (GG)		Berichtsjahre (Bj)
<b>Europa</b>	<b>(24 GG)</b>	<b>1.048</b>
Deutsches Reich	(5 GG)	247
Frankreich	(4 GG)	204
Großbritannien	(3 GG)	145
in weiteren Staaten	(12 GG)	452
<b>Amerika</b>	<b>(7 GG)</b>	<b>247</b>
Nordamerika	(4 GG)	176
Südamerika	(3 GG)	71
<b>Afrika</b>	<b>(1 GG)</b>	<b>19</b>
<b>Asien</b>	<b>(1 GG)</b>	<b>36</b>
<b>Australien</b>	<b>(1 GG)</b>	<b>30</b>
Insgesamt		<b>1.380</b>

Tabelle 1: Grobes Mengengerüst der 34 ausgewählten Geographischen Gesellschaften (GG) nach Berichtsjahren

Sprache	GG	Bj	Sprache	GG	Bj	Sprache	GG	Bj
deutsch	7	341	italienisch	1	47	schwedisch	1	34
englisch	7	287	ungarisch	1	40	norwegisch	1	25
französisch	6	261	niederländisch	1	39	tschechisch	1	20
spanisch	4	144	dänisch	1	38			
portugiesisch	2	68	japanisch	1	36	<b>Insgesamt</b>	<b>34</b>	<b>1.380</b>

Tabelle 2: Sprachenverteilung innerhalb der 34 ausgewählten Geographischen Gesellschaften (GG) nach Berichtsjahren

Im Fortgang galt es, diesen Zeitschriftenkorpus methodisch so zu durchmustern, dass daraus im Ergebnis ein globalhistorisch wertvoller Forschungsdatenbestand aufgebaut werden konnte.

## 2.2 Grundfragen der Datengewinnung und -erfassung

Nach der formalstrukturellen Fixierung des Korpus auszuwertender Gesellschaftszeitschriften lag die besondere methodische Herausforderung auf Fragen der Datengewinnung. Denn es galt, ein Auswahl- bzw. Selektionsmodell zu finden, das es erlaubt, trotz der sehr heterogenen inhaltsbezogenen Strukturierung der verschiedenen Kommunikationsartefakte, dennoch entsprechend vergleich- und verknüpfbare Dateninhalte zu extrahieren. Da es nach der Ausgangsfragestellung zunächst um die globalhistorische Typisierung von Geographischen Gesellschaften auf der Basis ihrer Strukturen, Aktivitäten und Interessen ging, ließen wir uns zur Lösung dieses methodischen Problems vom modernen Modell der Evaluierung von Wissenschaftseinrichtungen inspirieren. Denn bei solchen Evaluierungen geht es – analog zu unserer Problemstellung – ebenfalls um die analytische Operationalisierung einer vergleichend bewertenden Bilanzierung von Infrastrukturen der Wissensproduktion und -distribution sowie deren spezifischer Leistungsfähigkeit. So identifizierten wir in Anlehnung an die zentralen Bewertungskriterien solcher Evaluierungen Indikatoren bzw. kategoriale Einheiten im Berichts- und Beitragswesen der Gesellschaftszeitschriften, anhand derer sich entsprechende Daten methodisch kontrolliert extrahieren und in eine systematisch organisierte Erfassungsform überführen lassen. Dieser Ansatz (vgl. Abbildung 2) erlaubte es, Struktur-, Aktivitäts- und Interessenprofile der Geographischen Gesellschaften aus ihren Kommunikationsartefakten herauszuarbeiten und datentechnisch zu fixieren. Dabei zeigte sich im praktischen Umgang mit dem Quellenmaterial, dass man gerade im Berichtswesen der Geographischen Gesellschaften mit eher unstrukturierten und heterogenen Daten konfrontiert wird, die einen besonderen Einblick in das Strukturprofil einer Geographischen Gesellschaft geben, während man es im Beitragswesen der Geographischen Gesellschaften mit stärker strukturierten, homogenen Daten zu tun hat, die es ermöglichen, das Aktivitäts- bzw. Interessenprofil einer Geographischen Gesellschaft analytisch aufzubereiten. Um diesem strukturellen Dualismus und den unterschiedlichen Anforderungen an die Arbeit mit unstrukturierten und strukturierten Daten adäquat zu begegnen, wählten wir im Projekt zwei passende Teilmodelle der Datenerfassung. Die sogenannte *textuelle Erfassung* bereitet in besonderem Maße die Inhalte des Berichtswesens auf. Eine zusätzliche *tabellarische Erfassung* konzentriert sich vornehmlich auf die Auswertung des Beitragswesens der Zeitschriften. Die



Abbildung 2: Datenerfassungsmodelle im Teilprojekt C01 des SFB 1199

Benennung der beiden Erfassungsmodelle verweist bereits auf die technischen Formate der jeweiligen Datenerfassung. Die unstrukturierten, heterogenen Daten wurden möglichst offen und flexibel mit einer Textverarbeitung erfasst, während sich die strukturierten, homogenen Daten besser in einer Tabellenkalkulation systematisch organisieren ließen (Abbildung 2).

Eine zusätzliche methodische Problemstellung bei der Datengewinnung lag im Umgang mit der schier unermesslichen Menge des systematisch zu durchmusternden Ausgangsmaterials. Je nach Laufzeit einer Zeitschrift sind im groben Durchschnitt 10.000 bis 30.000 Seiten zu sichten und die entsprechenden Daten daraus zu extrahieren. Hier erwiesen sich besonders die Arbeiten Franco Morettis<sup>48</sup> zu dem von ihm entwickelten Ansatz des *distant reading* als äußerst anregend. Anstatt sich dem Korpus des Materials in klassisch hermeneutischer Weise mit einem zeitaufwändigen Verfahren des *deep reading* einzelner, unter Umständen subjektiv ausgewählter Texte zuzuwenden, werden kursorisches Lesen von Titeln, Gliederungen und Zwischenüberschriften sowie Einleitungen und Zusammenfassungen oder auch der analytische Blick auf die spezifische Verwendung von Abbildungen und Karten zu zentralen Elementen einer auf schnelle und sichere Überblicksgewinnung ausgerichteten Erfassungssystematik. Dieser Ansatz erlaubte es, je nach Umfang und routinierter Einarbeitung, durchschnittlich rund ein bis maximal zwei Berichtsjahre pro Arbeitstag auswerten zu können.

Die *textuelle Erfassung* gestaltete sich nach dem gegebenen Modell relativ frei, denn die Erfassungsvorlagen sind nur grob durch Zwischenüberschriften vorstrukturiert. Die Struktur orientiert sich dabei an den ausgewählten Indikatoren bzw. kategorial gebildeten Einheiten (vgl. Abbildung 2). Unterhalb dieser Ebene erfolgte eine freie, den jeweiligen individuellen Gegebenheiten einer Geographischen Gesellschaft angepasste Dateneingabe, aus Gründen der Arbeitseffektivität z.T. auch nur als Hinterlegung von Verweisungsstrukturen auf z.B. gesicherte Digitalisate von Vorstands- und Mitgliederlisten, Bilanzierungen materieller Infrastrukturen und Vermögenswerten etc. Diese Angaben können bei Bedarf in späteren Arbeitsschritten oder zur tiefergehenden Analyse eines Strukturprofils fallbezogen aufgerufen werden.

Die *tabellarische Erfassung* wirkt allein schon durch ihre formale Anordnung von Spalten und Zeilen deutlich strukturierender auf die Datenerfassung. Bei der hauptsächlichen Auswertung des Beitragswesens wurden zunächst formale, i.d.R. bibliographische, Daten zeilenweise erfasst, die wir dann über verschiedene Kodierebenen mit wichtigen Metadaten angereichert haben.<sup>49</sup> Aus arbeitsökonomischen Gründen wurden auch die über den Indikator des Schriftentauschs angezeigten Vernetzungsbeziehungen der Geographischen Gesellschaften untereinander ebenfalls in der tabellarischen Erfassung festgehalten.

In Anbetracht der erheblichen Datenquantitäten und des breiten Spektrums der Datenqualitäten, die in dieser Form erfasst wurden, ergibt sich berechtigterweise die Frage, warum dies nicht „professioneller“ mit einer ambitionierten Datenbanklösung aufgefangen wurde. Hier stellten sich als projektentscheidende Faktoren allerdings die kombinierten Ressourcen Zeit, Kapazität und Flexibilität gegen eine solche Form von Professionalisierung. Die Entwicklungszeit einer adäquaten Datenbanklösung mit sukzessive aufgebauten Versionen und entsprechenden Test- und Prüfphasen, wäre deutlich zu Lasten der eigentlichen Auswertungs- und Erfassungszeiten gegangen bzw. hätte diese empfindlich zum Projektende hin verschoben. Aufgrund der Sprachvielfalt des Zeitschriftenkorpus war besonders die Einbindung von sprachkompetenten Hilfskräften, Bachelor- bzw. Masterstudierenden und Werkvertragsnehmern zwingend nötig, die ihrerseits häufig ortsverteilt und auch nur auf einem EDV-technischen Mindeststandard von Textverarbeitung und Tabellenkalkulation arbeiten konnten. Ferner hätte die Einführung eines spezifischen Datenbanksystems zusätzliche Lern- und Einarbeitungszeiten erfordert. Unter diesen limitierenden Bedingungen erwies sich die funktional einfache und pragmatische Datenerfassungstechnik als optimale Praxislösung, die eine datentechnische „Professionalisierung“ – besonders auch im Hinblick auf den Einsatz von Methoden und Verfahren der Digital Humanities – im Nachgang des Projektes bzw. in seiner Fortführung nicht ausschloss, da gerade die ubiquitären Standardformate Text und Tabelle für die verschiedensten inhaltlichen und technischen Nachnutzungsformen im Vergleich zu einem spezifischen Datenbankformat, zumindest aus fachwissenschaftlicher Nutzersicht, deutlich offener, transparenter und anschlussfähiger sind.

48 F. Moretti, *Kurven, Karten, Stammbäume: Abstrakte Modelle für die Literaturgeschichte*, Frankfurt a. M.: Suhrkamp, 2009; F. Moretti, *Distant Reading*, Konstanz: Konstanz University Press, 2016.

49 Dies wird im Detail im Folgeabschnitt 2.3 erläutert.

## 2.3 Praxis der Datenverarbeitung und -kodierung

Im praktischen Vollzug der Verarbeitung und Kodierung der Daten wurden von den Bearbeiter(n) / innen der ausgewählten Geographischen Gesellschaften die Berichtsjahre der jeweiligen Publikation(en) in chronologischer Folge durchgesehen und für jedes Berichtsjahr eine Einzeldatei für die *textuelle Erfassung* angelegt sowie für die *tabellarische Erfassung* ein zentrales Tabellendokument je Geographische Gesellschaft mit mehreren tausend Zeilen. Besonders die *tabellarische Erfassung* ist im Sinne des *distant reading* als ein innovatives Prospektions- bzw. Explorationsinstrument zu verstehen, mit dem ein umfassender Überblick über den Korpus substantieller Texte gewonnen werden kann, der, je nach Fragestellung bzw. Analyseperspektive im nachgeordneten Gang der Forschung, themenflexible Hinweise auf mögliche interessante Stellen für eine tiefergehende Analyse im klassischen Modus des *deep reading* bieten kann.

Aufgrund der darin enthaltenen methodischen Innovation soll die im Projekt angewandte Kodierpraxis der *tabellarischen Erfassung* hier noch etwas eingehender erläutert werden. Ein zentrales Kodierprinzip ist, dass im vorgegebenen Erfassungssystem sowohl frei als auch kontrolliert kodiert werden kann. Die freie Kodierung erlaubt es gerade bei inhaltlichen und räumlichen Bezügen, dicht an den originalen Begrifflichkeiten der Quellen zu bleiben. Beispielsweise können darüber historisch genutzte, inzwischen fast unbekannte Topo- bzw. Choronyme (z. B. alte endonymische oder indigene Benennungen von Orten, Gebieten, Regionen etc.) so für eine individuelle, fallbezogene Recherche im Datenbestand gesichert werden. Gleichzeitig erlaubt die parallele Kodierung unter Verwendung eines kontrollierten Begriffsapparates eine übergreifend systematisierte Auswertung des Gesamtdatenbestandes. Insgesamt stehen für die kontrollierte Kodierung 81 feste Kategorien zur Verfügung, die sich in zwei Hauptgruppen mit zusammen 5 Untergruppen ausdifferenzieren und so eine dichte, feinstrukturierte Anreicherung mit beschreibenden Metadaten erlauben (Abbildung 3).

Bei der Basisaufnahme eines Eintrages in die Datentabelle wurden zunächst die grundlegenden bibliographischen Angaben (z. B. Autor / en, Titel, Fundstelle) des in den Korpus aufzunehmenden Textes erfasst. Dann erfolgte auf dieser Grundlage die freie, quellennahe Kodierung und abschließend wurde eine Tite-lautopsie nach Grundsätzen des *distant reading* durchgeführt, die wiederum die Basis für die kontrollierte Kodierung der neu angelegten Datenzeile bildete. Das von uns entwickelte Kodierschema berücksichtigt dabei drei zentrale Bereiche der Datenbeschreibung. Zunächst erfolgt eine *strukturbezogene Kodierung*, die sowohl nach bibliographischen Formen (z. B. Beiträge / Aufsätze, Dokumentenabdrucke oder Nekrologe) als auch nach formalen Inhaltstypen (z. B. Forschungsarbeiten, Forschungsreisen, Metatexten zur Geographie oder programmatischen Leitbildern) differenziert.

Der zweite Bereich konzentriert sich auf eine *sach- bzw. themenbezogene Kodierung*, bei der zunächst grob zwischen Beiträgen zur Physischen Geographie, der Anthropogeographie, der Kartographie sowie geographischen Darstellungen allgemeinerer bzw. landeskundlicher Art unterschieden wird. Ergänzend wird

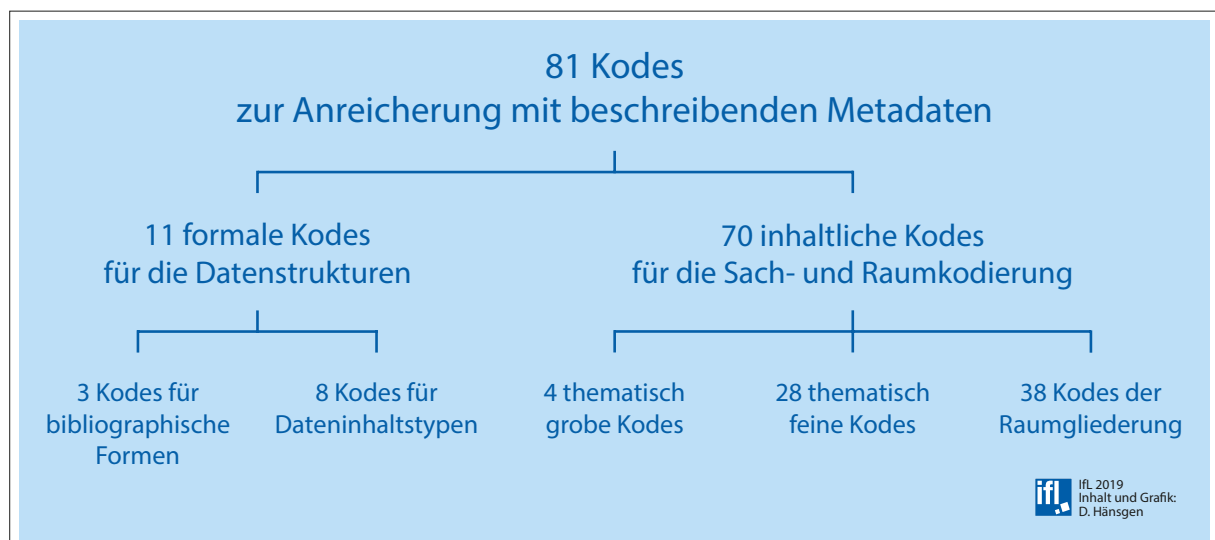


Abbildung 3: Schema der kontrollierten Kodierung (Version 2.0)



diese Kodierung auf einer zweiten, feineren Stufe deutlicher ausdifferenziert, indem die klassischen geographischen Felder des Naturraums (z. B. Geologie, Oberflächenformen, Gewässer, Klima, Pflanzen- und Tierwelt), der Infrastrukturen (z. B. Siedlungen, Städte, Verkehr, Kommunikation), der Wirtschaft (z. B. Ressourcen, Landwirtschaft, Industrie, Handel, Finanzen) sowie der Gesellschaft (z. B. Bevölkerung, Geschichte, Sprache, Bildung, Kultur, Politik) kategorienbildend wirken.

Aus geographischer Sicht stellt schließlich der dritte Bereich, die *raumbezogene Kodierung*, eine besondere methodische Herausforderung dar. Erwartungsgemäß finden sich in den zeitgenössischen Landes- und Länderkunden eine Vielzahl unscharfer, kulturalistisch grundierter Regionalisierungen, wie etwa Mittel- oder Zentraleuropa, Balkan, Orient etc. Damit erwiesen sie sich als Quellen bzw. Ausgangsmaterial für die Entwicklung einer global umfassenden Raumgliederung als gänzlich ungeeignet, denn unreflektierte raumbezogene Reifikationen, welcher Art auch immer, sollten im Prozess des Kodierens möglichst vermieden werden. Angeregt durch Helmut Klüters<sup>50</sup> Begriff des *Adressraums* ergab sich jedoch eine elegante Möglichkeit, zu einer handhabbaren Raumgliederung zu gelangen. *Adressräume* sind chiffrierte, flächenbezogene Adressangaben, die als ordnende Raumabstraktion fungieren (wie z. B. Postleitzahlengebiete) und sind somit ideal zum Aufbau systematisch strukturierter Notationssysteme für raumbezogene (Meta-)Daten geeignet. Sie dienen der Datenablage bzw. -zuordnung, können aber dank ihrer Chiffrierung jederzeit wieder zerlegt oder rekonfiguriert werden. Ausgehend von diesem Ansatz erwies sich die planetarische Verteilung von Land und Meer sowie das mathematisch konstruierte geodätische Netz der Erde als Schlüssel für den Aufbau einer entsprechenden Kodierung. Was fehlte, war ein passendes globales Referenzsystem, das zur Raumgliederung herangezogen werden konnte. Die Lösung für dieses Problem fand sich in der möglichen Nutzung der definierten Blattschnitte von Weltkartenwerken. Die naheliegende Bezugnahme auf die bereits im Übergang vom 19. zum 20. Jahrhundert konzipierte Internationale Weltkarte (IWK)<sup>51</sup> war allerdings nicht praktikabel, da sie mit ihrem Maßstab von 1:1.000.000 noch zu großmaßstäbig war, denn sie bezieht sich aufgrund des gewählten Maßstabes auf ein sehr kleinteiliges Raster von rein theoretisch gerechnet 2.130 Kartenblättern, um den kompletten Globus abbilden zu können. Wesentlich geeigneter erschien daher das in der zweiten Hälfte des 20. Jahrhunderts entwickelte sozialistische Gemeinschaftskartenwerk der Karta Mira<sup>52</sup> (KM) mit dem Maßstab von 1:2.500.000, bei dem sich das Adressraster auf nur 224 Kartenblätter in brauchbarer Weise reduziert. Bei der methodischen Umwidmung eines solchen Referenzsystems geht es in erster Linie gar nicht so sehr um die geodätische „Korrektheit“ des Kartenwerkes an sich, sondern um den für Ordnungszwecke praktikablen Zuschnitt der Kartenblätter, der auf den dafür verwendeten Längen- und Breitengraden basiert. Im Ergebnis wurden einzelnen Blattschnitte auf der Basis der Land-Meer-Verteilung zu zusammenhängenden Raumeinheiten aggregiert, wobei diese je nach vermuteter Datendichte unterschiedlich groß angelegt sind, wie beispielsweise im Minimum mit 3 Kartenblättern für Europa (Süd) und Europa (Nord) und im Maximum mit 33 Kartenblättern für Pazifik (Süd). Nach dieser Methode entstand eine Raumkodierung, die sich in 23 Großräume gliedert (vgl. Abbildung 4), die ihrerseits wiederum zu 11 Einheiten auf kontinentaler bzw. ozeanischer Ebene zusammengefasst werden können. Ferner gibt es weitere Codes, die auf gesamtglobale und transregionale Zusammenhänge verweisen, sowie auch auf explizit nichträumliche Bezugnahmen.

Die Verteilung von Land und Meer zur Grundlage der Kodierung zu machen, entbehrt natürlich auch nicht der impliziten Willkür eines geographischen Mythos.<sup>53</sup> Dennoch bemüht sich die hier eingeführte Raumkodierung darum, eine bewusst abstrahierende Raumsprache zu benutzen, um raumbezogene Reifikationen bzw. „Raumfallen“ klassischer geographischer Regionalisierungsverfahren möglichst zu vermeiden. So wird beispielsweise die Fläche zwischen 36° und 48° nördlicher Breite sowie 18° westlicher bis 36° östlicher Länge in der Projektterminologie mit *KM-Area 08 Eu(S)* bzw. *Europa (Süd)* kodiert und nicht etwa das all-

50 H. Klüter, *Raum als Element sozialer Kommunikation*, Gießen: Selbstverlag Geographisches Institut der Universität Gießen, 1986.

51 U. Schneider, „Fluchthilfe: Die Internationale Weltkarte (IWK) und die europäische Vermessung der Welt“, *Unikate/Universität Duisburg-Essen: Berichte aus Forschung und Lehre* (2009) 34, S. 92–101; F. Grenacher, „Die Internationale Weltkarte 1:1 000 000 im Zeitgeschehen“, *Geographica Helvetica* 2 (1947) 2, S. 112–122.

52 C. Lotz und C. Gohr, *Anti-imperialism or new imperialism? Examining the production and content of the World Map / Karta Mira 1:2 500 000 (1956–1989)*, Marburg: Herder-Institut für historische Ostmitteleuropaforschung – Institut der Leibniz-Gemeinschaft, 2017; K.-H. Meine, „KAPTA МИПА – World Map – Weltkarte – 1:2 500 000“, *Allgemeine Vermessungs-Nachrichten* 78 (1971) 1, S. 12–23.

53 M. W. Lewis und K. E. Wigen, *The Myth of Continents: A Critique of Metageography*, Berkeley: University of California Press, 1997.

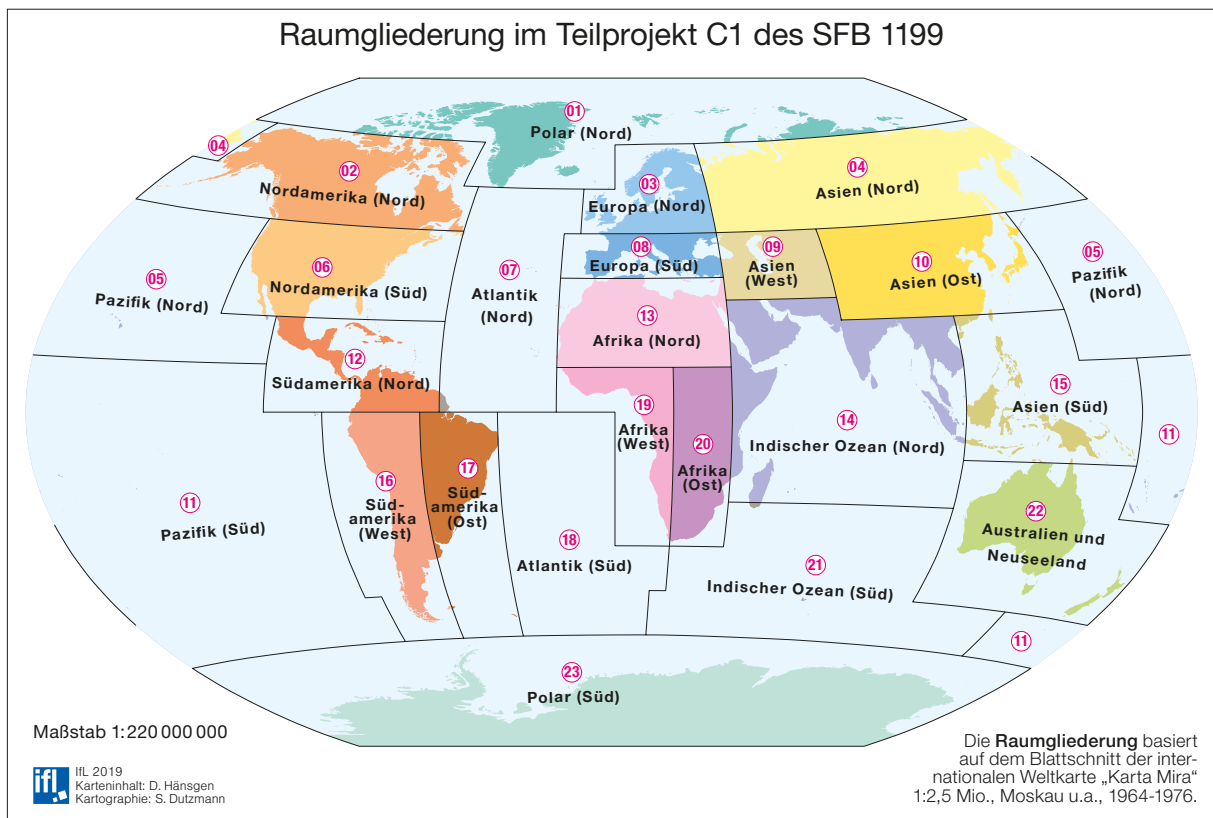


Abbildung 4: Karte der Raumgliederung im Teilprojekt C01 des SFB 1199

gemein gebräuchliche Choronym „Südeuropa“ dafür verwandt, denn die drei betreffenden Kartenblätter 52 Madrid, 53 Rom und 54 Sofia der Karta Mira, die einen zusammenhängenden Adressraum chiffrieren, schließen eben auch Teile der nordafrikanischen (Gegen-)Küste und weite Teile der asiatischen Türkei sowie Teile des Schwarzen Meeres bis zur Krim in diese datentechnische Raumabstraktion eines südlichen Europas mit ein (Abbildung 5).

Abschließend lässt sich festhalten, dass die Kombination der drei Formen der Struktur-, Sach- bzw. Themen- und Raumkodierung ein hinreichend differenziertes System bilden, um die aus den Kommunikationsartefakten der Geographischen Gesellschaften erfassten Basisdaten systematisch zu beschreiben und zu klassifizieren.

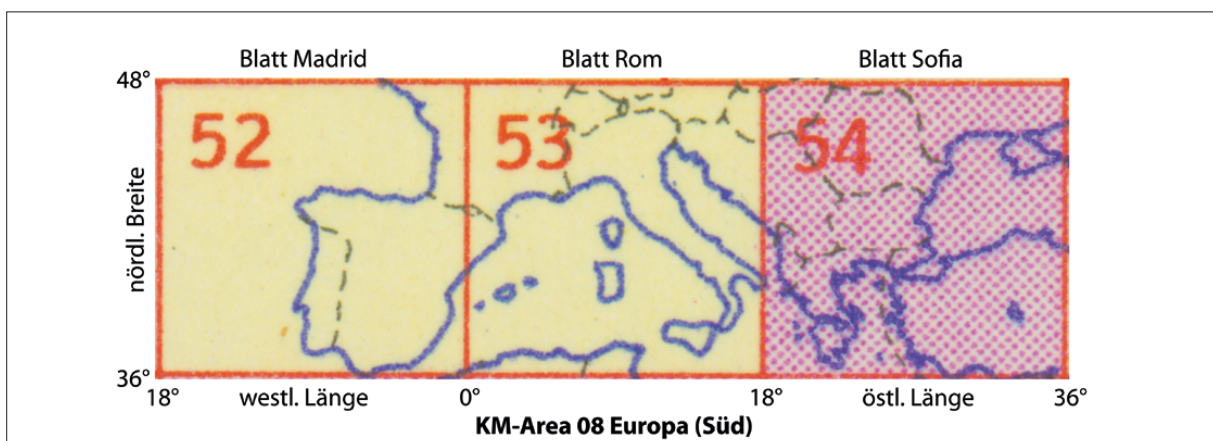


Abbildung 5: Vergrößerter Auszug aus dem Gesamtschnittplan der Karta Mira<sup>54</sup>

54 Meine „KAPTA MIPA – World Map – Weltkarte – 1:2 500 000“, Faltbeilage. Die verschiedenen Einfärbungen verweisen auf die unterschiedliche Provenienz: Blätter 52 und 53 wurden in der DDR, Blatt 54 in Rumänien produziert.

## 2.4 Formen der Datenauswertung

Die Gesamtheit der erfassten Daten aus den Periodika der 34 ausgewählten Geographischen Gesellschaften erlaubt vielfältige Formen von Auswertungen und Wissensgenerierung. Die *textuellen Erfassungen* bilden vor allem die Basis für Kurzbeschreibungen bzw. -profile, die besonders die Unterschiede und Gemeinsamkeiten der Strukturen von Geographischen Gesellschaften, auf der Basis eines schematischen Leitfadens, möglichst vergleichbar erschließen. Sie geben dabei Auskunft über Gründungsgeschehen, Satzungsfragen und Aufgabenwandel, Zusammensetzung und Strukturen von Vorständen und Mitgliedern, materielle Infrastrukturen (bsd. Sammlungen, Bibliotheken, Gebäude), Förderaktivitäten und Preisvergaben sowie besondere räumliche und thematische Interessen und Aktivitäten. Die *tabellarischen Erfassungen* dienen, wie bereits geschildert, vornehmlich als Explorationsinstrument. Sie erlauben aufgrund der differenzierten Erfassungsmethodik zum einen qualitative bzw. spezifische Einzelfallzugriffe (z. B. über Autoren- / Personenangaben) oder über die freie, quellennahe Kodierung Zugriffe auf individuelle geographische Bezeichnungen und Themen. Zum anderen gestatten sie gleichzeitig über geschachtelte Filterauswertungen der kontrollierten Kodierung stärker quantitativ orientierte Zugriffe nach Themen und Regionen, auch in Kombination mit der Filterung des historischen Verlaufs. Auf diese Weise lassen sich aus dem Gesamtkorpus diverse themen- und raumbezogene Teilkorpora generieren, die Basis erweiterter bzw. neuer Forschungen sind bzw. werden können.

Der so für einen Untersuchungszeitraum von fast 100 Jahren erarbeitete Bestand von rund 60.000 Datenzeilen, der Korpus von über 25.000 substantiellen Texten in 13 Sprachen sowie ebenfalls die über 25.000 dokumentierten Schriftentauschbeziehungen bieten ausgezeichnete Möglichkeiten, aus der Perspektive und mit den Werkzeugen der Digital Humanities weiter analysiert zu werden. Am Beispiel von ersten Teilergebnissen werden im Folgenden die Möglichkeiten vorgestellt, die sich bisher aus der computergestützten Datenanalyse ergaben und auf denen aufbauend die zweite Projektphase geplant ist.

### 3 Beispiele aus der Digital Humanities-gestützten Datenauswertung

Nach der Erhebung und Verarbeitung des Datenmaterials konnten Vergleich, Typisierung und Vernetzung ausgearbeitet werden. Geleitet wurden wir dabei durch ein breites Spektrum an Fragen: Interessieren sich beispielsweise die Geographischen Gesellschaften als Raumentrepreneurinnen<sup>55</sup> für dieselben Räume und Themen oder nicht? Kann man von nationalen Perspektiven und Interessenmustern sprechen? Wenn ja, stehen die Geographischen Gesellschaften im intranationalen Vergleich als Akteure stellvertretend für die nationalen Interessen ihrer Herkunftsländer? Wie unterscheiden sie sich wiederum im internationalen Vergleich zu anderen Geographischen Gesellschaften? Existieren gemeinsame, ggf. sogar globale, Interessengebiete bzw. Diskrepanzen? Wenn ja, welche Themen tauchen hier besonders bzw. kaum oder gar nicht auf? Bekommt beispielsweise ein Raum, der von „Franzosen“ vorzugsweise mit Handel und Wirtschaft verbunden wird, eine andere Bedeutung durch „Engländer“, „Dänen“, „Südamerikaner“ zugewiesen und wenn ja, warum? Welche Rolle spielen bei allen diesen Fragen die Zeit und der zeitliche Wandel?

Im Ergebnis haben wir beispielsweise erfahren, inwiefern die Geographischen Gesellschaften verstärkt national bzw. international agierten und vernetzt waren, ob sie in ihrer Berichterstattung eher beobachtend vorgingen oder selbst ausführende Akteure darstellten, respektive ob sich das Interesse aktiv (beispielsweise anhand von durch sie subventionierten Forschungsreisen und eigener Verräumlichungserfahrungen), reaktiv oder passiv (beispielsweise allein durch die Beobachtung fremder Forschungsreisen und Verräumlichungsprozesse) gestaltete, für welche Räume sie sich im Zeitverlauf überhaupt interessierten etc. Dadurch ließen sich Muster über die Aktivitätsprofile der Geographischen Gesellschaften erkennen, die nicht durch Einzelbeispiele hätten aufgedeckt werden können.

Im Hinblick auf die übergeordneten Fragen des SFB 1199 stellen diese Ergebnisse wichtige Einsichten zur Bedeutung und zum Handeln von Akteuren in Verräumlichungsprozessen bereit. In Bezug auf den oben aufgeworfenen Fragenhorizont interessieren uns fortan nicht mehr nur ihre Interessen- und Aktivitätsprofile, sondern auch, wie sich diese semantisch manifestieren, kurzum: Was verrät uns die Sprache der Akteure? So hat sich unser Interesse nunmehr auf die Verwendung bestimmter sach- und raumbezogener Semantiken im Zeitverlauf verlagert. Betrachten wir Geographische Gesellschaften als Raumentrepreneurinnen im oben genannten Sinne<sup>56</sup>, gelten ihre Zeitschriften als Medien der kommunikativen Selbstdarstellung und -vermarktung sowie der Dissemination von Wissen. Dadurch erlangt für unsere Zwecke die dort gewählte Semantik eine besondere Bedeutung. Dafür betrachten wir unser Material aus anderen Blickwinkeln. Bewusst haben wir uns dafür entschieden, zu unserem Untersuchungsgegenstand für diesen Bereich der Analyse weiter Distanz einzunehmen. Verfahren der Digital Humanities bieten hierfür einen vielversprechenden Zugang.

Im Folgenden sollen deshalb erste Digital Humanities-gestützte Auswertungen des Materials vorgestellt werden, die sich am Beispiel der Analyse der Aufsatztitel in den Zeitschriften zeigen: Als Datenbasis für das Working Paper wählten wir aus der erstellten Übersicht strukturiert erfasster Informationen im Tabellenformat<sup>57</sup> die Titelinformationen aller erfassten Texte für die Geographischen Gesellschaften von Rio de Janeiro (BR04), Leipzig (DE04), Algier (DZ02), Paris (FR01), Marseille (FR07), Rouen (FR13), Nancy (FR14), London (GB01), Manchester (GB02), Rom (IT02), Amsterdam (NL02), Lissabon (PT01), Washington D. C. (US03) und Philadelphia (US05). Dies ergab ein ausreichend großes Sample von insgesamt 10.753 Tabellenzeilen, welches mehrere Sprachen abdeckt (deutsch, englisch, französisch, niederländisch, portugiesisch, spanisch), einen Untersuchungszeitraum von insgesamt 90 Jahren vorgibt (1825–1914) und somit Vergleiche auf verschiedenen Ebenen (gesellschaftsindividuell, gesellschaftsübergreifend, semantisch, zeitlich) zulässt. Die genaue zeitliche Aufschlüsselung der Berichtsjahre wird nachfolgend noch anhand von Grafiken illustriert.

Die eingetragenen Artikelüberschriften sollten im Sinne der Digital Humanities programmgestützt ausgewertet, zeitlich kontextualisiert und untereinander kontrastiert werden. Dafür wurde ein interaktives Visua-

55 Zum Begriff „Raumentrepreneurinnen“ vgl. Abschnitt 1.2.

56 Vgl. Abschnitt 1.2.

57 Vgl. die Ausführungen zum Modus der tabellarischen Erfassung in Abschnitt 2.2.

lisierungssystem als Recherche- und Analysewerkzeug konzipiert und prototypisch umgesetzt, welches einen Überblick über die Struktur des vorhandenen Materials gibt. Durch Auswertung der Auftretenshäufigkeiten (sog. Frequenzen) einzelner Terme ist es möglich, auf Muster in der verwendeten Fachsprache zu schließen, welche sich als erste Hinweise auf eine systematisch unterschiedliche Themensetzung oder gar Verräumlichungspraxis interpretieren lassen können.<sup>58</sup> In der durchgeführten Analyse der Titeldaten wurden generische, sprachunabhängige Verfahren angewendet, die es erlauben, bereits mit einfachen statistischen Betrachtungen aussagekräftige Einblicke in die Nutzung von Vokabular und Nennung von Entitäten in den Titeln der Artikel zu erhalten.

### 3.1 Vor- und Aufbereitung des Materials für die Digital Humanities-gestützte Auswertung

Die notwendigen Vorverarbeitungsschritte auf Seiten der automatischen Sprachverarbeitung beliefen sich auf eine simple Worttrennung an Leerstellen (Whitespace-Tokenisierung) und eine sehr einfache Form der sogenannten „Normalisierung“: Alle Zeichen eines Wortes werden, falls möglich, in Kleinbuchstaben umgewandelt. Alle Sonderzeichen werden entfernt, das heißt für die untersuchten Texte, dass neben regulären lateinischen Buchstaben nur noch die Zeichen aus der Liste „ä ü ö ß é é è à ù â ê î ô û ë ï ç“ beibehalten werden. Zudem wurden sämtliche Numerale durch das Zeichen # ersetzt. Dadurch werden redundante Konstrukte wie „1. Teil“ und „2. Teil“ auf „#.Teil“ abgebildet und Formulierungen wie „1882 bis 1883“ und „1872 bis 1880“ fallen beide auf „##### bis #####“. Dadurch werden sich oft in ähnlicher Form wiederholende, zu Themenausrichtung und Verräumlichung nicht beitragende Phrasen automatisch zusammengefasst.

Eine solche automatische Verarbeitung birgt neben einer allgemeinen Arbeitserleichterung auch Potential für unerwünschtes Verhalten. Ein italienischer Titel lautet beispielsweise: „Itinerario e note del viaggio Brava-Egherta-Lugh-Bardera-Brava“. Er enthält die Stationen einer Rundreise durch Somalia, mit Bindestrichen zusammengefügt. Nach einer Tokenisierung und Normalisierung wie oben beschrieben erhält man das „Wort“ bzw. besser die Zeichenkette „bravaeghertalughbarderabrava“ mit genau einem Vorkommen im Titeldatenkorpus. Ein solches Verhalten wirkt unschön und auch eine Beibehaltung des Bindestrichs bei der Normalisierung löst das Problem nicht. Eine Worttrennung am Bindestrich würde helfen, jedoch würde dies bedeuten, dass z. B. „Bismarck-Archipel“ künftig in zwei wenig bedeutungstragende Einzelwörter zerfiel bzw. dass es zu einer nicht beabsichtigten Bedeutungsverschiebung käme. Letztlich sind solche Fehlbehandlungen seltener Ausnahmefälle für eine erste explorierende Analyse der Daten aller Erfahrung nach jedoch verschmerzbar. Unser Fokus liegt hauptsächlich in der Betrachtung starker, auffälliger Muster, die sich im Bereich von Vokabular mit großer Auftretenshäufigkeit zeigen.

Spätestens mit der Formulierung des Zipf'schen Gesetzes ist bekannt, dass häufige Wörter in einem Korpus exponentiell öfter auftreten als seltene.<sup>59</sup> Da eine Korrelation zwischen der Häufigkeit und der Kürze von Wörtern existiert (vermutlich aus sprachökonomischen Gründen), sind also kurze Wörter wie Artikel, Präpositionen und Partikel überproportional häufig vertreten. Diese besitzen (einzeln für sich gesehen) keinen hohen Aussagewert über ein Fachvokabular und sollten daher aus aggregierten Ansichten entfernt werden (sprachunabhängige Stoppwortbehandlung). Im Prototypen wurde hierfür die Möglichkeit zum Herausfiltern aller Wörter mit weniger als vier Zeichen geschaffen.

Vokabularnutzung ist weiterhin nicht nur durch die Verwendung einzelner Begriffe, sondern insbesondere auch durch Phrasen und Mehrwortterme charakterisiert. Um diesem Umstand Rechnung zu tragen, wurde auch der Übertrag der Analysen von tokenbasierter Zählung zu größeren Einheiten wie Bi- und Trigrammen vorgenommen. Hierbei werden die Effekte der eben beschriebenen Kurzwortfilterung deutlicher sichtbar. Einige Phrasen besitzen etwa charakteristische Verwendungsmuster von Präpositionen, welche unter Umständen für die Analysen von Interesse sind. Deshalb wurden für die Visualisierung Varianten mit und ohne Anwendung dieser Filterung erstellt.

58 Vgl. M. Middell, „Raumformate – Bausteine in Prozessen der Neuverräumlichung“.

59 Vgl. G. K. Zipf, *Selected Studies of the Principle of Relative Frequency in Language*, Cambridge, MA: Harvard University Press, 1932.

Insgesamt ergeben sich zahlreiche Nuancen möglicher Interpretationen aus den unterschiedlichen Ausgabetensätzen der dafür geschaffenen flexiblen Verarbeitungs-Prozesskette. Im Folgenden sollen anhand einiger Beispiele die dadurch möglichen Erkenntnisgewinne skizziert werden.

### 3.2 Anwendung der Visualisierungstechnik TagPies zur inhaltlichen Analyse

Zunächst werden Visualisierungen vorgestellt, welche unter Verwendung der Visualisierungstechnik TagPies erstellt wurden.<sup>60</sup> Diese Form der Darstellung erlaubt verknüpfte Sichtweisen auf das Material, nämlich eine quantitative Einschätzung relativer Größenverhältnisse von gruppierten Daten (wie in Tortendiagrammen, den sogenannten „Pie Charts“) bei gleichzeitiger Darstellung der Häufigkeiten von gemeinsamem und spezifischem Vokabular in Bezug auf die Gruppierung. Als Gruppen wurden die Geographischen Gesellschaften gewählt und das Vokabular wurde wie oben beschrieben aufbereitet.

Diese Visualisierungen ermöglichen einen induktiven Zugriff auf das Datenmaterial. Einen ersten vergleichenden Blick auf die Zusammensetzung der darin verzeichneten Titel erhalten wir durch die gemeinsame Visualisierung aller verwendeten Wörter in den Einträgen (raw\_token\_counts). Es ist daneben auch möglich, die Ergebnisse in verschiedenen Schritten zu spezifizieren und dadurch präzisere Visualisierungen zu erhalten, beispielsweise indem der nicht zur Lesbarkeit der Visualisierung beitragende Unterschied zwischen Groß- und Kleinschreibung wie oben beschrieben aufgehoben wird (normalized\_token\_counts). Abbildung 6 zeigt die dadurch erzeugte Gesamtansicht aller Geographischen Gesellschaften über Sprachgrenzen hinweg. Einzelnen Geographischen Gesellschaften sind dabei jeweils Farben zugeordnet. Vokabular, das in dieser Farbe um die Kennziffer der jeweiligen Geographischen Gesellschaft angeordnet ist, wird ausschließlich dort im Titel verwendet. Schwarz dargestelltes Vokabular taucht in mehreren Geographischen Gesellschaft-

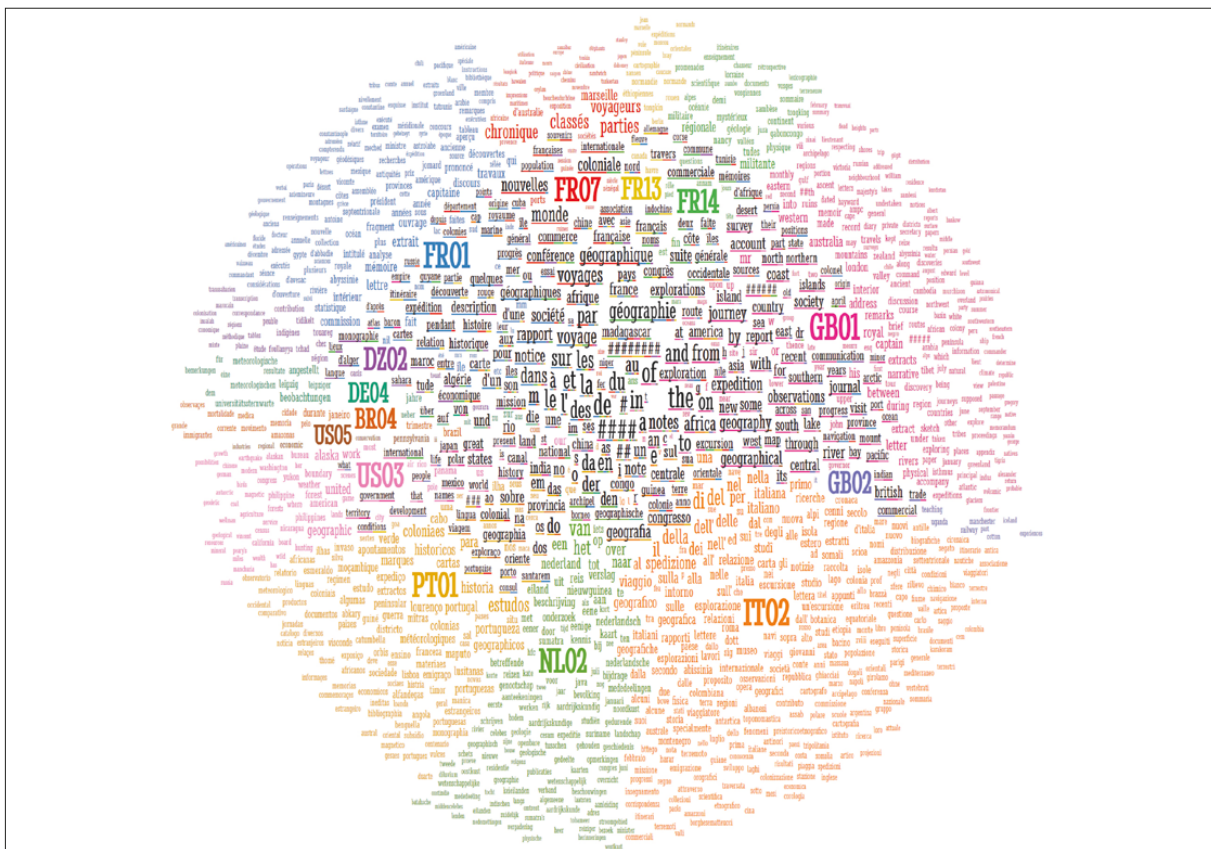


Abbildung 6: TagPie der normalisierten Wörter aus den Titeln aller Geographischen Gesellschaften

60 S. Jänicke, „TagPies“, <http://www.tagpies.vizcovery.org/> (letzter Zugriff 7. Februar 2019); S. Jänicke et al., „TagPies: Comparative Visualization of Textual Data“, in: A. Telea, A. Kerren und J. Braz (Hrsg.), *VISIGRAPP 2018: Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Band 3, Setúbal: SCITEPRESS – Science and Technology Publications Lda, 2018, S. 40–51.

ten auf. Ein mehrfarbiger Balken unter dem Wort gibt Aufschluss über die relativen Anteile der Verwendung. Die Farben sind dabei absteigend nach Häufigkeit von links nach rechts sortiert.

Bereits in dieser Ansicht lässt sich wichtiges Vokabular ablesen, jedoch verwirrt neben der Vermischung von Sprachen auch die vordergründige Präsenz häufiger kurzer Wörter (im Wesentlichen sogenannte Stoppwörter). Im Visualisierungswerkzeug ist eine Filterung nach Land oder Sprache möglich sowie die Auswahl eines Datensatzes, indem alle Wörter der Länge 3 und kürzer herausgefiltert wurden. In Abbildung 7 ist eine entsprechende Darstellung für die beiden britischen Geographischen Gesellschaften zu sehen, in welcher die kurzen Wörter nicht enthalten sind.

Hier treten die inhaltsbestimmenden Terme visuell viel deutlicher hervor. Das von den Farben der Geographischen Gesellschaften gebildete Tortendiagramm zeigt in einer solchen reduzierten Ansicht viel klarer das relative Verhältnis der Titelanzahlen innerhalb der Jahresschriften. Es sind interessante Muster erkennbar, etwa ein Fokus der Geographischen Gesellschaft Manchester auf „trade“, „development“, „commercial“ und „railway“. Zumindest für letzteren Begriff gibt ein Blick auf die zeitliche Anordnung der Jahresschriften beider Gesellschaften, wie er noch später in einer weiteren Visualisierungsform vorgestellt werden wird, die logische Erklärung.

Neben der Betrachtung einzelner Wörter ist die Bi- und Trigrammsicht ein aufschlussreiches Werkzeug für die Analyse der in den Titeln verwendeten Fachsprache. Jeweils in Folge erscheinende und somit zusammengehörende Worte sind dabei visuell und technisch mit einem „\_“ verbunden worden. Abbildung 8 zeigt Ausschnitte aus der nach englischer Sprache gefilterten Bigramm-Ansicht einmal in der Variante mit und einmal ohne kurze Wörter.

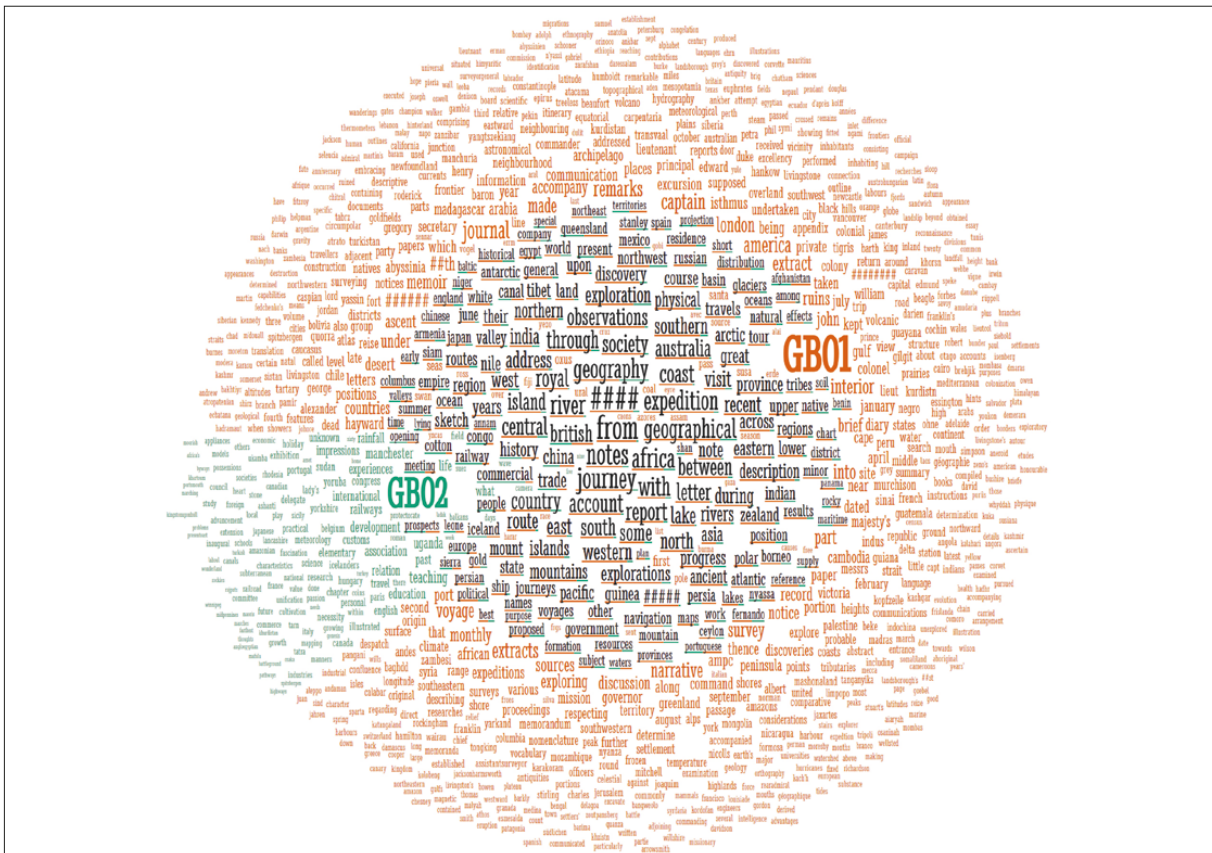


Abbildung 7: TagPie der normalisierten Wörter aus den Titeln der britischen Geographischen Gesellschaften



Abbildung 8: Ausschnitte der TagPies von Bigrammen aus den Titeln englischsprachiger Geographischer Gesellschaften

Während in der linken Ansicht insbesondere Kombinationen von Präpositionen und Artikeln untereinander und mit anderen Begriffen auffallen, werden diese Effekte rechts nicht mehr abgebildet. Obwohl solche einfachen Kollokationen und idiomatische Ausdrücke nicht sonderlich interessant scheinen, sind sie wohl geeignet, anderes Vokabular besser zu qualifizieren, etwa dass es sich bei „notes“ um „notes\_of“ und nicht etwa „notes\_from“ handelt, bei „explorations“ um „explorations\_in“ und nicht „explorations\_of“. Während links noch von „map\_of“ zu lesen ist, tauchen beide Worte ob ihrer Kürze rechts nicht mehr auf. Zumindest „map“ könnte sehr wohl als Fachausdruck vermisst werden. Tendenziell ist der Aufbau und die Verwendung von eigens erzeugten Stoppwortlisten oder umgekehrt Positivlisten für kurze Wörter in allen Sprachen wünschenswert. Für ein solches, sehr akribisches Vorgehen bestand jedoch zunächst keine Notwendigkeit. In den meisten Fällen können auch kurze Wörter gut im Geiste durch den Leser ergänzt werden. Bei „coast\_africa“ ist klar und ohne besonderen kognitiven Zusatzaufwand ersichtlich, dass in den Originaltiteln ein dazwischen befindliches „of“ (oder Ähnliches) existiert hat. Die Filterung bringt dabei durchaus große Vorteile mit sich: Über sie werden beispielsweise die mit „central“, „east“ oder „south“ spezifizierten Teilregionen von „america“ und „africa“ besser aufgeschlüsselt. Auch zu erwartende Teile von Eigennamen, wie „national\_geographic“ und „royal\_geographical“ können prominent abgelesen werden, wie auch Bezeichner von Subdisziplinen wie „physical\_geography“.

Im Ergebnis all dieser Schritte wird auf einen Blick für alle Daten vergleichend, für einzelne Sprachen sowie sprachübergreifend sichtbar, welche Begriffe besonders häufig und welche weniger häufig im Ausgangsmaterial (einzeln und in Folge) vorkommen. Gleichzeitig lassen sich Hypothesen über die Verwendung von Begriffen in den verschiedenen Ausgangssprachen prüfen und Wörter explizit suchen. Dadurch eignet sich das Visualisierungswerkzeug hervorragend für die in den Digital Humanities als „explorative Analyse“ bezeichnete Herangehensweise, anhand des konkreten Datenbestandes seine Zusammensetzung zu untersuchen und daraus Ideen für interessante, genauer abzuklärende Detailfragen zu gewinnen.<sup>61</sup>

Neben der rein visuellen Analysemöglichkeit besteht die Option zum gezielten Durchgriff auf die dahinterliegenden statistischen Auswertungen. Neben dem eigentlichen TagPie als statischer Visualisierung öffnet sich, sobald man sich per Mouseover über einem der Ergebniswörter befindet und dieses anschließend anklickt, ein Fenster, welches alle diese Begriffe sowie Buchstabengruppen, die mit den aus den Begriffen vorkommenden Wörtern teilweise übereinstimmen, anzeigt. So generiert das Ergebnisfenster auch Begriffe, die die Wörterketten der Beispielsuchwörter beinhalten, aber noch mit Prä- und/oder Suffixen versehen sind, oder aber andere Begriffe, die sich aus denselben und ggf. weiteren Buchstaben zusammensetzen. Unterschiedliche solche Suchvorgänge führen dazu, dass wir mit dem Ausgangsmaterial vertraut werden

61 Zu „explorative Analyse“ siehe beispielsweise M. Rehbein, „Informationsvisualisierung“, in: Jannidis, Kohle und Rehbein (Hrsg.), *Digital Humanities*, S. 328–342, besonders S. 332; oder auch N. Bubenhofer und M. Kupietz (Hrsg.), *Visualisierung sprachlicher Daten: Visual Linguistics – Praxis – Tools*, Heidelberg: Heidelberg University Publishing, 2018.



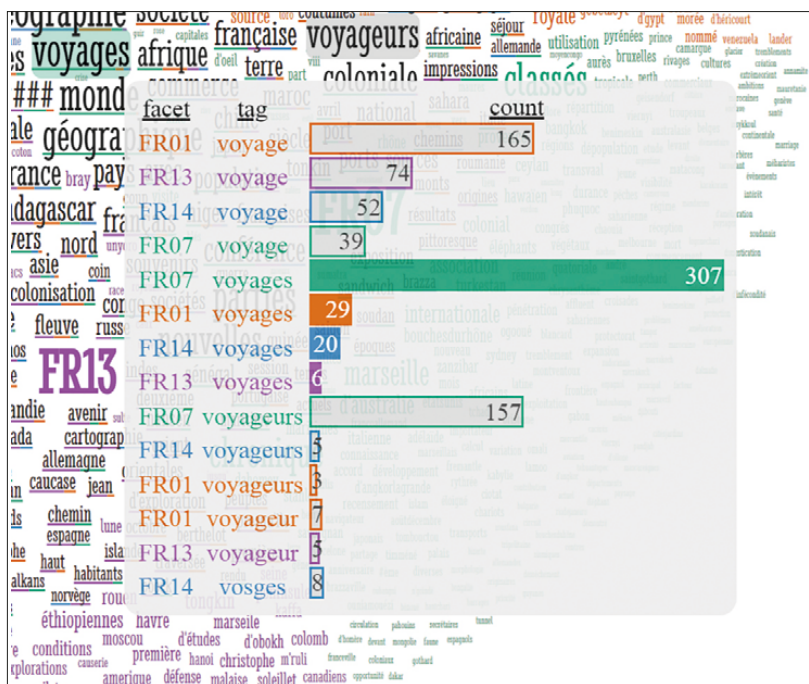


Abbildung 9:  
Detailansicht für den Begriff „voyages“  
in einer TagPie-Visualisierung

und ein Gefühl dafür bekommen, was im Untersuchungszeitraum thematisch für die unterschiedlichen Geographischen Gesellschaften, aber auch in den verschiedenen einzelnen Sprachräumen, von Interesse war.

In Abbildung 9 ist eine solche Detailansicht für den im Französischen auffallend dominanten Term „voyages“ zu sehen. Darüber zeigte sich, dass in der Geographischen Gesellschaft mit dem Code FR07 dieses Wort extrem häufig vorkommt. In den Daten selbst konnte dies auf die sehr häufige unspezifische Titelergabe „Voyages classés par parties du monde“ zurückgeführt werden. Um diesen Effekt zu mindern, wurde für die Datensätze jeweils eine „unique“-Version erstellt (z. B. normalized\_token\_counts\_unique), die solche mehrfach in identischer Form vorkommenden Titel nur einmalig zählt.

Im Ergebnis der Betrachtung aller Daten sind allgemeingeographische Begriffe, Topo- bzw. Choronyme und Begriffe, die auf Methodik schließen lassen, besonders auffällig. Die Titel der Aufsätze der für den Untersuchungszeitraum ausgewählten Periodika ausgewählter Geographischer Gesellschaften sind demnach häufig an die Verwendung von Choronymen wie „Afric(k)a“ gebunden und/oder fassen die methodische Herangehensweise an den Themengegenstand wie, hier nur für Englisch und Französisch aufgeführt, „description(s)“, „report“, „rapport“, „observations“, „explorations“, „discovery“, „découverte(s)“ zusammen, die dem Verfassen des jeweiligen Textes vorausging. Im Vergleich der unterschiedlichen Ausgangssprachen zeigen sich durchaus Unterschiede in Bezug auf die geographischen Orientierungen, auf die die Topo- bzw. Choronyme verweisen, was weiterer Prüfung mithilfe unseres Kodierschemas bedarf (Abbildung 3). Auch was die Begriffe betrifft, die auf die in den Texten behandelten Themen hinweisen, besteht weiterer Forschungsbedarf. Im Englischen und Französischen (s. o.) scheinen sie sich zu ähneln, für die anderen Sprachen sollte dies weiter überprüft werden.

Trotz dieser Auffälligkeiten lässt sich erkennen, dass die Titel offenbar unspezifischer sind als vermutet. Seit der Herausbildung von Erdkunde als Schulfach und später von der Geographie als Hochschuldisziplin teilten sich geographische Fragen spätestens gegen Ende unseres Untersuchungszeitraums auf die groben Bereiche Physische Geographie, Humangeographie (bzw. Anthropogeographie), Geodäsie und Kartographie auf. Anders als eingeschätzt kommen explizite Hinweise auf diese Bereiche in den Aufsatztiteln allerdings nicht vor.

Darüber hinaus lassen die Visualisierungen, wie bereits am Beispiel des Begriffes „voyages“ aufgezeigt, die vergleichsweise häufige Verwendung bestimmter Begriffe durch bestimmte Geographische Gesellschaften erkennen. Dies sollte nicht dazu verleiten, allzu voreilige methodische Schlüsse auf Basis der durch TagPies erzeugten Ergebnisse zu ziehen. Vielmehr sollten diese Auffälligkeiten tiefergehend geprüft werden, zumal die Periodika der verschiedenen Geographischen Gesellschaften als Ausgangsmaterial unterschiedliche individuelle Komponenten mitbringen, die wir berücksichtigen müssen. So stellte sich bei näherer Prüfung der Ergebnisse zu „voyages“ (s. o.) mithilfe der vorliegenden tabellarischen Erfassung und

der darin enthaltenen sachbezogenen Kodierungen<sup>62</sup> heraus, dass die Häufigkeit durch eine bestimmte Rubrizierung der Zeitschrift und einer ausgeprägten Orientierung an Forschungs- und touristischen Reisen der Geographischen Gesellschaft Marseille zustande kommt.

Für die Geographische Gesellschaft Nancy wiederum fällt beispielsweise der vergleichsweise häufige Gebrauch des Wortes „militaire“ in den Aufsatztiteln auf, was die Hypothese zulässt, es handele sich hier um eine Auseinandersetzung mit den Folgen des deutsch-französischen Krieges. Tatsächlich bezieht sich die Häufung der Begriffe aber in der Regel auf eine bestimmte militärgeologische Studie zu der Schweiz und den französischen Alpen.

Der auf diese Weise ermöglichte tiefere Einstieg in das Datenmaterial verdeutlicht zudem, dass französischsprachige Texte, teilweise aber auch anderssprachige Texte, zu physischer Geographie unter Abhandlungen zu regionaler Geographie subsumiert werden, u. a. weil ihre Überschriften von Topo- bzw. Choronymen gekennzeichnet sind. Hier bietet sich eine durch Text Mining-Verfahren begleitete Analyse an, die es ermöglicht, die zu den Überschriften zugehörigen Texte in die Analyse mit einzubeziehen und auf Basis aller Aufsätze ähnliche Suchanfragen wie die oben beschriebenen durchzuführen. Dies wird ein Bestandteil der zweiten Projektphase sein, um die von Geographischen Gesellschaften benutzten Raumsemantiken über einen Zeitraum von mehr als 100 Jahren vergleichend aufzudecken und zu analysieren.

### 3.3 Digital Humanities-spezifische Strategien zur Bildung von Suchbegriffen

Unabhängig von der inhaltlichen Ausgestaltung der Titel und den durch einzelne Geographische Gesellschaften besonders häufig verwendeten Termini, die sich dank der Visualisierungen direkt ableiten und erkennen lassen, interessieren wir uns im Rahmen unseres Forschungsprojekts dafür, auch weniger häufig vertretene Begriffe zu betrachten, wenn wir uns durch sie weitere Hinweise auf Verräumlichungsprozesse unter Globalisierungsbedingungen und Geographien geographischer Gesellschaften erhoffen. Hierfür wählen wir einerseits Begriffe, die sich aus der Visualisierung ergeben. Andererseits wählen wir Termini, die sich aus unseren Vorkenntnissen durch die Datenerfassung und die Kodierung des Textkorpus ableiten lassen, wie beispielsweise „global“, „mondial“, „weltweit“, „Welt“, „Globus“, „Land“, „Region“, „terrain“, „terre / a“, „territoire“, „territorium“, „k / colonial“, „meteo“, „Klima“, „climate“, „humaine“, „human“. Die Möglichkeit des Herunterladens aller Termini in Tabellenform erleichtert diese Anfrage zusätzlich. Über die Bordmittel von Tabellenkalkulationssoftware lassen sich Sortierung nach Häufigkeit in einzelnen Geographischen Gesellschaften, Filterung und die Suche nach Termen und Wortbestandteilen leicht realisieren. Es bietet sich so ein direkter Blick auf die konkreten Zahlen, insbesondere auf die Häufigkeiten sprachlicher Varianten (Flexionsformen) oder sprachübergreifender Phänomene. Einige der Begriffe kommen gar nicht, andere selten vor. Suchen wir beispielsweise nach Texten, die bereits im Titel auf ein globales Themenspektrum verweisen, bietet sich eine Suche nach damit inhaltlich verwandten Begriffen an. Keiner der Begriffe wie „global“ oder „mondial“ taucht allerdings in den Daten auf, sodass wir die Suche alternativ gestalten und lediglich nach ihrem jeweiligen Wortstamm suchen können. Hierbei ergeben Stichproben für die Sprachen Englisch und Französisch häufig Substantivierungen anstelle von Adjektivierungen, sodass die Suche nach vollständigen Adjektiven weniger Nutzen und Erkenntnisgewinn hat als die Eingrenzung auf Wortstämme (im Beispiel „glob“ oder „mond“), um auf die verwendeten Substantive zu schließen.

### 3.4 Anwendung der Visualisierungskomponente Vega zur Analyse zeitlicher Entwicklungen

Daneben ist für unseren Fall die Integration einer zeitlichen Komponente bei der Untersuchung des Auftretens bestimmter Begriffe notwendig, weil wir im Beispiel Raumsemantiken im zeitlichen Verlauf sprachübergreifend beobachten wollen. Hierfür wird eine Form der Visualisierung vorgestellt, die zunächst einen quantitativen Überblick über die jeweils pro Jahr (ggf. innerhalb mehrerer Ausgaben) erschienenen Titel pro Geographische Gesellschaft gibt. Daraus ist ein zeitliches Aktivitätsprofil ablesbar, welches zwar nicht tatsächliche textliche Produktivität widerspiegelt, weil beispielsweise Artikellängen nicht berücksichtigt sind,

62 Siehe hierzu Abschnitt 2.3.

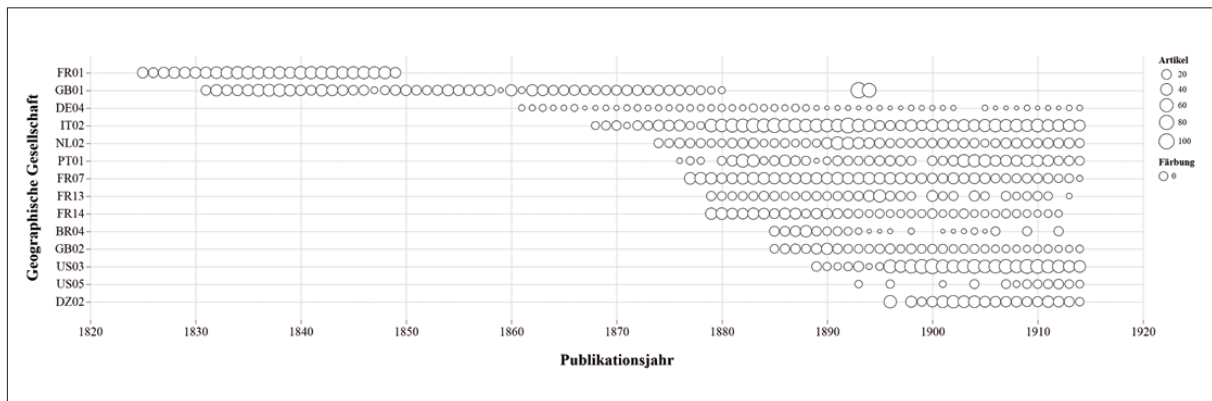


Abbildung 10: Publikationsverlauf aller Geographischen Gesellschaften über die Zeit

welches aber dennoch geeignet ist, um etwa aktivere von weniger aktiven Publikationsjahren zu unterscheiden. Bereits anhand dieser in Abbildung 10 wiedergegebenen Übersicht lassen sich einige der in den TagPie-Visualisierungen auftretenden Effekte besser verstehen.

Diese Darstellung bildet weiterhin das Rahmenwerk für eine Suchfunktion im Titelkorpus. Durch die Eingabe von Suchausdrücken werden die entsprechend nach Jahr aggregierten Ergebnisanzahlen pro Geographische Gesellschaft in eine Färbung der jeweiligen Kreise übersetzt. Schwarz steht dabei stets für das jeweilige Maximum an Suchtreffern, Weiß für keine Treffer, Graustufen für einen (linearen) Anteil am Maximum. Kleine schwarze Kreise sind dabei also ggf. noch „überraschender“ als große. Als Suchausdrücke können sogenannte „Reguläre Ausdrücke“ („Regular Expressions“, kurz „RegEx“) verwendet werden, die in der Informatik sehr üblich für die Mustersuche in Textdaten sind. Durch sie werden trunkierte Suche und eine standardmäßige Teilwortsuche ermöglicht. Abbildung 11 zeigt die Ergebnisse für eine sprachübergreifende Suche nach dem jeweiligen Stamm des Wortes Afrika.



Abbildung 11: Publikationsverlauf aller Geographischen Gesellschaften über die Zeit mit Hervorhebungen nach zum Suchmuster passenden Titeln

Der dafür verwendete komplexe Ausdruck „afri[kqc][ua]e?“, welcher zum gleichzeitigen Auffinden z.B. von „afrika“, „african“ und „afrique“ konstruiert werden kann, liefert 496 Ergebnisse und damit nur eins weniger als der sehr viel allgemeinere Ausdruck „afri“. Damit wird lediglich der portugiesische Titel „Codigo dos milandios cafriaes“ zusätzlich zu den anderen gefunden, da er zufällig auch diese vierbuchstabile Zeichenfolge enthält. Es zeigte sich sehr häufig, dass einfache Ausdrücke genügen, um einen hinreichend guten Überblick über die Verwendungsprofile von Begriffen zu erlangen.

Dieses zweite Visualisierungswerkzeug wurde unter Zuhilfenahme der webbasierten Visualisierungskomponente Vega von Grund auf selbst entwickelt.<sup>63</sup> In der Konsequenz bedeutete dies auch eine stetige Weiterentwicklung der Funktionalität anhand von sich ergebenden Erfordernissen bei der Analyse. So wurde

63 University of Washington, Interactive Data Lab, „Vega – A Visualization Grammar“, <https://vega.github.io/vega/> (letzter Zugriff 7. Februar 2019).

die Ausgabe beispielsweise um eine Gesamtanzahl an „Suchtreffern“ ergänzt und die verwendete Farbskala so neu implementiert, dass bei fehlenden Ergebnissen stets ein weißer Kreis dargestellt wird, unabhängig von der Farbwertinterpolation der Visualisierungskomponente. Eine solche parallele Entwicklung von Werkzeugen zur Herausbildung der Arbeitsmethodik bei ihrer Anwendung hat sich als sehr hilfreich für die Erzeugung einer ergonomischen Oberfläche herausgestellt. Es zeigte sich, dass die Handhabung der besonderen Eingabesprache anfangs nicht ganz einfach für die Fachwissenschaftler außerhalb der Informatik war. Jedoch war das Erlernen der „RegEx“-Grundlagen nicht nur lohnenswert im Kontext des Visualisierungswerkzeuges sondern mit einem generellen Zugewinn an technischem Wissen verbunden.

Eine methodische Diskussion der Verwendung beider Visualisierungsansätze liefert die Feststellung, dass sich beide zueinander auf mehreren Ebenen komplementär verhalten, indem sie gewissermaßen Blicke auf die Daten in unterschiedlichen „Maßstabs- und Betrachtungsebenen“ bieten. Der Vorteil des ersten Tools liegt im detaillierten Zugriff auf die begriffliche Ebene und die damit verbundene sprachübergreifende Synopse von zusammenhängenden bzw. unterschiedlichen Begriffslinien, der Nachteil liegt in der kognitiv nur begrenzt verarbeitbaren Information der „Ergebnisfenster“ (vor allem, wenn man viele hintereinander öffnet und sich nicht mehr genau an die vorherigen erinnern kann). Der Vorteil des zweiten Tools liegt neben der Darstellung der zeitlichen Dimension in der überschaubaren und kognitiv verarbeitbaren synoptischen Darstellung der absoluten Mengenverhältnisse und implizit auch den relativen Verhältnissen von einzelnen „Begriffen“, der Nachteil liegt aber gleichzeitig im sehr selektiven Blick auf nur einzelne Begriffe bzw. Begriffslinien und dass man aktiv suchen muss und keine konkrete(n) Anreize / Leitung zur weiteren Exploration des Vokabulars erhält.

Die Verschränkung eines induktiven Zugriffs inklusive eines möglichen Serendipitätseffekts<sup>64</sup> im ersten Werkzeug mit dem deduktiven Zugriff durch die Suche im zweiten führt zur Kopplung von unbekanntem und bekanntem Komponenten der Analyse. Aus einem vermeintlichen Informationsverlust beim Wechsel der Maßstabs- bzw. Betrachtungsebene wird – durch besseren Überblick – ein informatorischer Gewinn möglich (wie gleichfalls durch den vermeintlichen Gewinn von zusätzlichen detailreichen Informationen der Überblick oder die klaren Linien bzw. Strukturen im Kleinklein der Details verloren gehen kann).

Im Überblick haben uns diese ergänzenden Verfahren bei der Analyse von Geographischen Gesellschaften neue Horizonte eröffnet, weil wir nicht nur ihre Aktivitäts- und Interessenprofile vergleichend beleuchten konnten, sondern vielversprechende Hinweise auf die in bestimmten Zeiten und Sprachen vorherrschenden raum- und sachbezogenen semantischen Felder erhalten haben.

---

64 L. McCay-Peet und E. G. Toms, „The Process of Serendipity in Knowledge Work“, in: N. J. Belkin (Hrsg.), *Proceedings of the Third Symposium on Information Interaction in Context*, New York: Association for Computing Machinery, 2010, S. 377–382.

## 4 Fazit und Ausblick

Mit Abschluss der ersten Phase des SFB 1199 haben wir im Projekt C01 u. a. einen Bestand von rund 60.000 Datenzeilen erarbeitet, der als Referenzinstrument über einen Zeitraum von annähernd 100 Jahren auf einen Korpus von über 25.000 substantiellen, verschiedensprachigen Texten von Geographischen Gesellschaften verweist sowie über 25.000 Tauschbeziehungen zwischen Geographischen Gesellschaften weltweit dokumentiert. Diese umfangreichen Datenbestände erlauben es, den eingangs geschilderten methodischen Herausforderungen von Vergleichbarkeit, mehrdimensionaler Typisierung und Vernetzung in der bisherigen und weiter folgenden Forschung in adäquater Weise begegnen zu können. Im Ergebnis haben wir neben den aus der Datenauswertung gewonnenen Einsichten und Erkenntnissen zu den spezifisch projektbezogenen Fragestellungen einen methodisch ausdifferenzierten, globalhistorischen Forschungsdatenbestand geschaffen, dessen besonderer Wert darin liegt, nachhaltig zu sehr unterschiedlichen, fachthematischen und raumbezogenen Fragestellungen des langen 19. Jahrhunderts befrag- bzw. nachnutzbar zu sein.

Dabei haben wir darauf geachtet, dass sich innovative Zugriffe auf unser Datenmaterial insbesondere aus der Perspektive und mit den Werkzeugen der Digital Humanities bieten, weil sie uns ermöglichen, zum Untersuchungsgegenstand Distanz einzunehmen. Die entsprechenden Analysen hatten die Entdeckung von bisher unbekanntem Mustern zum Ziel, die für uns im Hinblick auf unsere Fragestellung von besonderer Bedeutung sind.

Die hier vorgestellten ersten Auswertungen sind Ergebnis des Zusammenspiels von Informatik, Geographie und Geisteswissenschaften und zeigen exemplarisch auf, worauf solche Kooperationen in der Praxis hinauslaufen können. So trivial Wörterzählen zunächst erscheinen mag, konnte das Working Paper verdeutlichen, wie viele Schritte der Vorverarbeitung des Ausgangsmaterials und welche detaillierte Aufarbeitung der Daten etc. notwendig sind, um sie auf einer abstrahierenden Ebene zu verhandeln<sup>65</sup> und zu den durchgeführten Analysen zu gelangen. Diese stellen für unser Projekt einen Mehrwert dar, weil sie den Weg für weitere induktiv angelegte Forschung ebnen. So haben sich in der Zusammenarbeit neue Forschungsfragen ergeben. Was würde beispielsweise passieren, wenn wir nicht nur die Aufsatztitel mit Hilfe von Wortanalysen im zeitlichen Verlauf untersuchen würden, sondern wenn wir uns direkt das gesamte digitalisierte Textmaterial vornehmen? Ziel des Folgeprojektes ist es deshalb, genau solche Untersuchungen durchzuführen und seriell auswertbare Texte durch Text Mining-Verfahren<sup>66</sup> nach einer Vielzahl von Worten und Begriffen zu durchsuchen, um den Gebrauch und den Wandel von semantischen Bedeutungen sprachübergreifend und vergleichend systematisch zu verfolgen. Dies wird helfen, um der im Fortsetzungsantrag des SFB prominent gestellten Frage nach dem Wandel von Raumsemantiken im Zuge von Prozessen der Neuverräumlichung der Welt noch besser auf die Spur zu kommen. Denn derartige Analysen, wie wir sie hier erst einmal nur explorativ durchgeführt haben, können in Zukunft empirisch dicht für den Fall Geographischer Gesellschaften dreierlei zeigen: Erstens, wie aus bloßen Verräumlichungen von vielen einzelnen Akteuren durch Prozesse der Verdichtung sich schließlich kollektiv und transnational Konzepte entwickeln, die bereits implizit Handlungsanweisungen darüber enthalten, wie „man“ „richtig“ verräumlicht. Die daraus resultierenden empirisch beobachtbaren „Raumformate“ wie z. B. „Empire“, „Nationalstaat“, „Imperialer Ergänzungsraum“ oder „Region“ bauen ihrerseits wiederum, zweitens, in wechselnden Kombinationen synchron und diachron wandelbare Vorstellungen von jeweils adäquat erscheinenden „Raumordnungen“ auf.

Gerade die Geographie als ein in hohem Maße visuell und visualisierungsgestützt arbeitendes Fach produziert und nutzt in diesen Zusammenhängen jede Menge Imaginationen. Sie sind nicht nur bloße Repräsentationen, sondern ihrerseits wiederum in komplexen Aushandlungs- und Rezeptionsprozessen entstehende hoch verdichtete Artefakte geographischer Einbildungskraft.<sup>67</sup> Um sowohl Mechanismen ihrer Verstetigung als auch ihres Wandels transnational erkennen zu können, bedarf es, und das ist wohl, drittens, das erstaunlichste Ergebnis unseres Projekts, einer genaueren Analyse von Raumsemantiken. Denn Raumsemantiken werden – das zeigen die Zeitschriften der Geographischen Gesellschaften in den verschiedenen Sprachen

65 Siehe weiterführend u. a. F. Moretti, *Kurven, Karten, Stammbäume*, 12 ff.

66 G. Heyer, U. Quasthoff und T. Wittig, *Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse*, Berlin: Springer Campus, 2017.

67 Möhring, Pizarz-Ramirez und Wardenga, *Imaginationen*.

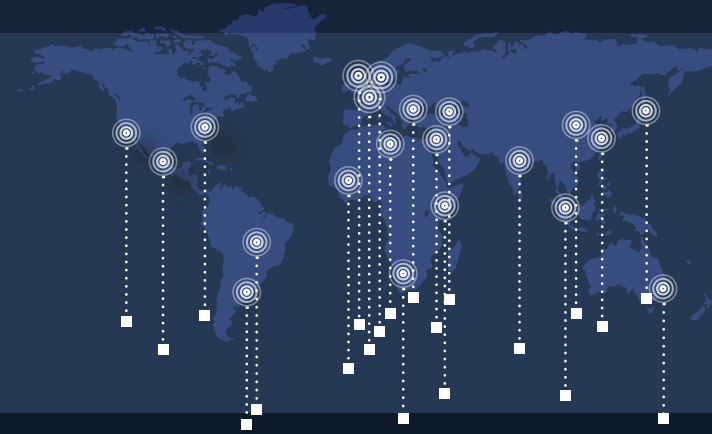
deutlich auf – vor allem durch die Sprache der Akteure von Verräumlichungsprozessen konstituiert und vermittelt. Für unser Projekt spielen dabei die unterschiedlichen „Sprachen“ der Geographischen Gesellschaften als Akteure eine besondere Rolle, unabhängig davon, ob sie aktiv, reaktiv, beobachtend, bewertend, aus der Distanz heraus etc. handeln und berichten.

Welche Semantiken, sowohl sach- als auch raumbezogen, waren für ausgewählte Geographische Gesellschaften entscheidend? Sind sie über sprachliche Grenzen hinweg ähnlich oder unterschiedlich, verändern sie sich im Zeitverlauf, werden sie übersetzt, übertragen etc.? Die Ergebnisse und Teilergebnisse zu diesen Fragen können, wie wir oben gesehen haben, Digital Humanities-gestützt schichtweise betrachtet werden, wobei ausgewählte Schichten auch visuell zur räumlichen und inhaltlichen Analyse „übereinander gelegt“ werden können. Dadurch ergibt sich für uns die Möglichkeit, Verräumlichungsprozesse zu kartieren und kartographieren, wodurch beispielsweise die jeweiligen nationalen Blickwinkel der Akteure auf ein und denselben Raum vielseitig sachthematisch und sprachübergreifend sichtbar werden. Stellen wir uns beispielsweise vor, dass wir die französische Auseinandersetzung mit Afrika abbilden wollen, ist dies zunächst pro französische Gesellschaft möglich, aber auch für alle französischen Gesellschaften zusammengenommen, dies wiederum lässt sich mit britischen Gesellschaften vergleichen etc. Die vorgestellten Visualisierungsinstrumente eignen sich wunderbar für ein solches Vorgehen und helfen, die verwendete Semantik sprachübergreifend vergleichend für ausgewählte Akteure sichtbar zu machen. So erhalten wir ein besseres Verständnis dafür, was für eine ausgewählte Gruppe von Akteuren in einem Zeitraum, in dem sie großen Einfluss auf die Vermittlung geographischen Wissens, auf die Geographie als sich entwickelnde wissenschaftliche Disziplin und auf ihre Umgebung ausübte, Verräumlichungsprozesse mit prägte und versprachlichte.

Unter einer vergleichsweise dünnen Schicht von Normierung(sversuchen) lassen sich einerseits durch Verflechtungs-, Übersetzungs- und Transferprozesse teilweise erhebliche Bedeutungsverschiebungen in unterschiedlichen Kontexten beobachten. Andererseits spielen wahrscheinlich Prozesse der Metaphorisierung eine wichtige Rolle. Genau an diesem Punkt können Verfahren der Digital Humanities in exzellenter Weise helfen. Denn mit und durch sie kann nicht nur ein systematischer Überblick über (transnationale) Entwicklungen der *longue durée* gewonnen werden. Sie helfen auch (jenseits tradierter Vorstellungen bisheriger (geographischer) National-Historiographien), die zeitliche Verteilung etwaiger Brüche oder inhaltsbezogener Veränderungen vergleichend zu identifizieren. Dies erlaubt wiederum, Schlüsseltexte für eine eingehende hermeneutische Analyse methodisch reflektiert auszuwählen, sie dann einem bereits kanonisierten Sample kritisch an die Seite zu stellen, um neue Dimensionen einer Geographisierung der Welt zu entdecken. Uns ist bewusst, dass es sich auch in der sich anschließenden Forschung um ein aufwändiges kollaboratives Arbeiten jenseits der Fakultätsgrenzen handeln wird, die sich jedoch, wie wir oben exemplarisch darstellen konnten, allemal lohnt.



# CENTRE FOR AREA STUDIES



Working paper series des SFB 1199 an der Universität Leipzig No. 15

ISBN: 978-3-96023-223-0

ISSN: 2510-4845

Universität Leipzig  
SFB 1199

E-Mail: [sfb1199@uni-leipzig.de](mailto:sfb1199@uni-leipzig.de)

<http://research.uni-leipzig.de/~sfb1199>

Funded by Deutsche Forschungsgemeinschaft

**DFG** Deutsche  
Forschungsgemeinschaft



UNIVERSITÄT  
LEIPZIG

Centre for Area Studies

Leibniz-Institut  
für Länderkunde **ifl**

**ÓWZO** Leibniz-Institut für  
Geschichte und Kultur  
des östlichen Europa