



universität
wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

**Datenintensiver Journalismus:
Erforschung und redaktionelle Routinen einer
technisch spezifizierten journalistischen Praxis**

verfasst von / submitted by

Mag. (FH) Julian Ausserhofer

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Doktor der Philosophie (Dr. phil.)

Wien, 2017 / Vienna, 2017

Studienkennzahl lt. Studienblatt /
Degree programme code as it appears on
the student record sheet:

A 092 301

Dissertationsgebiet lt. Studienblatt /
Field of study as it appears on the student
record sheet:

Publizistik- und
Kommunikationswissenschaft

Betreut von / Supervisor:

emer. o. Univ.-Prof. Dr. Thomas A. Bauer

Für meine Eltern.

Danksagung und Anmerkungen zur Form

Dieses Buch trägt zwar meinen Namen am Cover, aber es würde nicht existieren ohne die unzähligen Gespräche, Begegnungen, Einsichten, Empfehlungen und Unterstützungen, die mir viele Menschen und Institutionen im Laufe der Jahre zuteilwerden ließen.

Thomas A. Bauer möchte ich für die inspirierende, vertrauensvolle und ermutigende Betreuung der Dissertation meinen tiefen Dank aussprechen. Ein weiterer großer Dank gebührt Robert Gutounig, Katharina Serles und Lena Wanner, die den Text vor der Abgabe vollständig gegengelesen haben und wichtige Hinweise gegeben haben.

Ein Dankeschön von Herzen ergeht an meine Eltern und deren Partner, meine Freundin Lena, meine Großeltern, meine Brüder, an alle Verwandten, FreundInnen und Bekannten, die mich in den vergangenen Jahren unterstützt haben.

Im Laufe der Jahre waren zahlreiche weitere Personen eine große Unterstützung und Inspiration – unter anderem als Co-AutorInnen von Publikationen, die in diese Arbeit eingeflossen sind, als deren KorrekturleserInnen, oder schlicht als GesprächspartnerInnen. Dazu gehören (in alphabetischer Reihenfolge): Wolfgang Aigner, Mercedes Bunz, Paul Duguid, Eva Goldgruber, Markus Hametner, Cori Hayden, Christian Katzenbach, Wiebke Loosen, Noura Maan, Axel Maireder, Annette Markham, Sarah Matiasek, Christina Niederer, Michael Oppermann, Cornelius Puschmann, Alexander Rind, Joseph Robinson, Elena Rudowsky, Christina Schumann, Michael Sedlmair, Monika Taddicken, Heinz Wittenbrink und Thomas Wolking. Auch hier zu nennen sind die InterviewpartnerInnen, die mir so bereitwillig Auskunft gegeben haben und in Tabelle 5 aufgeführt sind.

Des Weiteren möchte ich den institutionellen UnterstützerInnen und BereitstellerInnen von Forschungsinfrastruktur danken. Die Forschung, die dieser Arbeit zugrunde liegt, wurde ermöglicht durch Mittel der *Universität Wien* und deren *HochschülerInnenschaft (ÖH)*, der *FH Joanneum*, der *Österreichischen Forschungsförderungsgesellschaft* bzw. dem *Bundesministerium für Verkehr, Innovation und Technologie* (Projekt: 845598), der *Internet Privatstiftung Austria* sowie der *Volkswagenstiftung*. Beherbergt wurde ich unter anderem vom *Alexander von Humboldt Institut für Internet und Gesellschaft* in Berlin, vom *Center for Science, Technology, Medicine and Society* an der *University of California, Berkeley*, von Alessandro Bennati in München, von Wilfried Lipp in Altaussee. An diesen Orten haben neben Wien und Graz längere Phasen der Forschung und Textproduktion stattgefunden.

Viele der hier genannten Personen und Institutionen werden von der öffentlichen Hand finanziert. So verhält es sich auch mit den Universitäten, an denen ich studiert habe, und mit den meisten der Bibliotheken, die ich für diese Arbeit genützt habe. Sie alle haben viel zum Zustandekommen dieser Arbeit beigetragen. Deshalb möchte ich mich hiermit bei all jenen bedanken, die direkt oder indirekt mitwirken, dass diese Infrastrukturen unserer Gesellschaften funktionieren.

Dieser Text profitierte weiters stark von frei zugänglichen Publikationen und Methodenerläuterungen, von offenen Daten und Softwareanwendungen. Deshalb gilt schließlich mein aufrechter Dank all jenen, die Open Science¹ und Open Source unterstützen.

In Formfragen folgt die Arbeit so weit wie möglich den Empfehlungen der American Psychological Association (2010). Deren Anleitung ist zwar umfassend, kann jedoch nicht alle Fragen beantworten, die bei einer Textproduktion auftauchen. Etwa sind Aspekte wie gendersensible Sprache oder der Umgang mit fremdsprachigen Ausdrücken unzureichend geregelt. Deshalb habe ich parallel zu dieser Arbeit *gspusi* (Great Scientific Publication Style and Information) geschrieben. Dabei handelt es sich um Gestaltungsrichtlinien für wissenschaftliche Publikationen. *gspusi* ist unter einer offenen Lizenz auf <https://github.com/julianaus/gspusi/> veröffentlicht und wird auch in Zukunft erweitert werden.

Diese Publikation wurde in *Markdown*² geschrieben und so weit wie möglich mit plattformunabhängigen Open-Source-Applikationen produziert. Dafür wurde ein Setup weiterentwickelt, das unter anderem auf einem Texteditor, dem Literaturverwaltungsprogramm *Zotero*³ und dem Dokumentenkonvertierwerkzeug *Pandoc*⁴ beruht. Die Anleitung für dieses Setup wurde ebenfalls auf *GitHub* im *gspusi*-Repository publiziert.

Zahlreiche Datensätze, die im Forschungsprozess generiert wurden, sind unter <https://github.com/julianaus/data/journalism/> abrufbar.

Julian Ausserhofer, Wien, April 2017

¹ Open Science ist Thema eines Unterabschnitts in Kapitel 4, der methodologischen Verortung der Arbeit.

² Markdown ist eine für Menschen einfach lesbare maschinelle Auszeichnungssprache.

³ <https://www.zotero.org/>

⁴ <http://pandoc.org/>

Inhaltsverzeichnis

1	Einleitung.....	12
1.1	Kontexte der Datafizierung der journalistischen Praxis	14
1.2	Forschung zu datenintensivem Journalismus.....	18
1.2.1	„Erste Welle“ der Datenjournalismusforschung.....	20
1.2.2	Forschung abseits der Journalistik.....	21
1.2.3	Forschung und Reflexionen über datenjournalistische Ausbildungen.....	24
1.3	Ziele der Arbeit	24
1.4	Zentrale Fragestellungen.....	26
1.5	Forschungszusammenhang und Verwertungsperspektive	28
1.6	Aufbau der Arbeit	29
2	Zum Begriff des datenintensiven Journalismus.....	30
2.1	Datenintensiver Journalismus als Prozess	33
2.1.1	Zum Datenbegriff im datenintensiven Journalismus	36
2.2	Datenintensiver Journalismus als Produkt.....	39
2.3	Zwischenresümee: Datenintensiver Journalismus als Hybrid	41
3	Wissenschaftstheoretische Grundannahmen.....	45
3.1	Theorie-Theorie und Theorie-Typen	45
3.2	Ludwik Flecks Lehre vom Denkstil und Denkkollektiven	48
3.3	Thomas Kuhns Struktur wissenschaftlicher Revolutionen	51
3.4	Alfred Schütz' Lebensweltkonzept.....	55
3.5	Die Konstruktion gesellschaftlicher Wirklichkeit nach Thomas Berger und Peter Luckmann.....	59
3.6	Die Animation des Objekts: Karin Knorr-Cetina, Bruno Latour und die soziotechnische Konstruktion von Fakten.....	62
3.7	Zwischenresümee.....	66
4	Methodologische Verortung und Forschungsdesign	68
4.1	Zur Konstruktion eines Forschungsdesigns	68
4.1.1	Forschungsprozess-Schemata	68
4.1.2	Forschungsdesign-Einflussfaktoren.....	74
4.1.3	Methodologische Herausforderungen digitaler Methoden	76
4.1.3.1	Vermeintliche Verlässlichkeit.....	77
4.1.3.2	Vermeintliche Repräsentativität.....	78
4.1.3.3	Vermeintliche Objektivität.....	78
4.1.4	Open Science	79
4.1.5	Triangulation.....	82

4.2	Das Forschungsdesign dieser Arbeit.....	84
5	Systematische Literaturanalyse: Datenintensive journalistische Arbeitsabläufe aus Forschungssicht.....	88
5.1	Methode: Structured Literature Reviews	88
5.2	Dokumentensuche und -auswahl	89
5.3	Quantitative und qualitative Inhaltsanalysen	92
5.4	Ergebnisse der systematischen Dokumentenanalyse	94
5.4.1	Korpus-Charakteristiken	94
5.4.1.1	Einflussreiche Publikationen	96
5.4.2	Produktionsstätten der Datenjournalismusforschung	101
5.4.3	Theorien der empirischen Datenjournalismusforschung	102
5.4.4	Forschungsdesigns der Datenjournalismusforschung.....	104
5.4.4.1	Qualitative Interviews.....	104
5.4.4.2	Inhaltsanalysen.....	105
5.4.4.3	Andere Methoden.....	105
5.4.4.4	Örtlicher Fokus der Untersuchungen	107
5.4.5	Forschungslücken der Datenjournalismusforschung	109
5.4.6	Datenintensive Arbeitsabläufe in Redaktionen.....	110
5.4.6.1	Themenfindung, Auftragserteilung und Datensammlung.....	111
5.4.6.2	Datensäuberung, -verifikation, -analyse	114
5.4.6.3	Produktion und Visualisierung	116
5.4.6.4	Veröffentlichung und Evaluierung	119
5.5	Diskussion der Ergebnisse der systematischen Literaturanalyse.....	121
5.5.1	Agenda zur weiteren Erforschung des datenintensiven Journalismus.....	121
5.5.2	Datenintensive Arbeitsabläufe im Überblick.....	124
6	Interviewstudie: Die Praxis datenintensiver journalistischer Arbeitsabläufe	127
6.1	Methode: Semistrukturierte Interviews und teilnehmende Beobachtungen	128
6.1.1	Sampling	129
6.1.2	Anbahnung und Ablauf der Datenerhebung	131
6.1.3	Auswertung	133
6.2	Ergebnisse der Interviewstudie	134
6.2.1	Interdisziplinäre Kollaborationen	134
6.2.2	Routinen und Arbeitsabläufe	136
6.2.3	Akteure digitaler datenjournalistischer Routinen	139
6.2.4	Digitale datenjournalistische Methoden	141
6.3	Diskussion der Ergebnisse der Interviewstudie	142

7	Resümee und Ausblick	146
7.1	Forschungssynthese	146
7.2	Datenintensive journalistische Arbeitsabläufe in der Forschungsliteratur	148
7.3	Datenintensive journalistische Arbeitsabläufe in der Praxis	149
7.4	Die Konsequenzen des datenintensiven Journalismus für mediale Vertrauensbildung und Medienkompetenz	152
7.4.1	Zur Faktizität des datenintensiven Journalismus	152
7.4.2	Menschlich-algorithmisch-co-konstruierte Öffentlichkeiten.....	154
7.4.3	Vertrauensbildung mit datenintensivem Journalismus	157
7.4.4	Vergesellschaftete data literacy	159
8	Literaturverzeichnis	162
9	Anhang.....	205
9.1	Codebuch	205
9.1.1	Codierprinzipien	205
9.1.2	Codetabelle	205
9.2	Interviewleitfaden	214
9.2.1	Ziele der Datenerhebung.....	214
9.2.2	Leitprinzipien.....	214
9.2.3	Eröffnung des Interviews.....	214
9.2.4	Hintergründe zur Person	215
9.2.5	Arbeitsabläufe bei der Produktion von datenintensiven Beiträgen.....	215
9.2.5.1	Teamebene	215
9.2.5.2	Organisationsebene	216
9.2.5.3	Methoden- und Technikebene.....	217
9.2.5.4	Abschluss	217
9.3	Zusammenfassung.....	218
9.4	Abstract	219
9.5	Lebenslauf.....	220

Abbildungsverzeichnis

Abbildung 1. Erwähnungen von ‚computer-assisted reporting‘ in englischsprachiger Literatur 1970–2008.....	31
Abbildung 2. Suchinteresse online für ‚computer-assisted reporting‘ und ‚data journalism‘ .32	
Abbildung 3. Beispiele für Prozess-Visualisierungen des datenintensiven Journalismus.....	34
Abbildung 4. Beispiel für eine personalisierbare, interaktive Online-Visualisierung	40
Abbildung 5. Schema eines zirkulären Forschungsdesigns	70
Abbildung 6. Die Forschungsspirale.....	72
Abbildung 7. Der Spiralprozess theoriegenerierender Forschung.....	73
Abbildung 8. Das Forschungsdesign dieser Arbeit.....	85
Abbildung 9. Schritte in einer systematischen Dokumentenanalyse	89
Abbildung 10. Ablauf der Dokumentensuche und -auswahl	92
Abbildung 11. Entwicklung der Literatur über die Zeit	95
Abbildung 12. Publikationen nach Typ und Anzahl an Zitationen	96
Abbildung 13. Referenzen nach Anzahl und Veröffentlichungsjahr.....	98
Abbildung 14. Zitationsnetzwerk.....	99
Abbildung 15. Netzwerk von Korpus-Publikationen und Referenzen	100
Abbildung 16. Netzwerk von Korpus-Publikationen und Referenzen (Ausschnitt).....	101
Abbildung 17. Institutionelle Anbindungen und Kooperationen in der Datenjournalismusforschung	102
Abbildung 18. Mit datenintensivem Journalismus assoziierte Medienorganisationen.....	109
Abbildung 19. Akteure einer datenjournalistischen Produktion.....	139

Tabellenverzeichnis

Tabelle 1. Synonyme Suchbegriffe für die Dokumentensuche	90
Tabelle 2. Durchsuchte wissenschaftliche Datenbanken	90
Tabelle 3. Die meistzitierten Referenzen im Korpus.....	97
Tabelle 4. Forschungsmethoden, geografischer Fokus und Themen der Datenjournalismusforschung	106
Tabelle 5. Übersicht über die geführten Interviews.....	130
Tabelle 6. Codebuch	205

1 Einleitung

Die vernetzte Informationsgesellschaft, in der wir leben, bringt stetig mehr und mehr Daten hervor (Lyman & Varian, 2003). Innerhalb weniger Jahrzehnte sind diese zu einer zentralen Ressource für Unternehmen, Regierungen, die Wissenschaft und die Zivilgesellschaft avanciert und haben zu einer tiefgreifenden Transformation sämtlicher gesellschaftlicher Bereiche beigetragen (Castells, 2001, 2010). Der globale Kapitalismus ist zu einem datengesteuerten Kapitalismus geworden (Zwick & Denegri Knott, 2009). Debatten über ‚big data‘, soziale Medien und Algorithmen bestimmen heute massenmediale Diskurse und verändern kulturelle Konventionen. Die Grenzen zwischen Öffentlichem und Privatem werden in vielen Bereichen neu gezogen. Daten werden als ‚Evidenz‘ herangezogen, um ‚Fakten‘ von ‚fakes‘ zu trennen.

Mit Metaphern wird versucht, die Veränderungen, die sich durch die Datenproliferation ergeben, zu fassen. Daten seien etwa der „digitale Treibstoff des 21. Jahrhunderts“ (Kundra, 2012), „das neue Gold“ (Kroes, 2011), ein „Berg“ (Bunz, 2011), den es zu erklimmen gilt, oder ein „Ozean“ (Lupton, 2013), in dem man sich zurechtfinden müsse. Bei diesen bildhaften Ideen von bislang ungenutzten Ressourcen und unerforschten Naturphänomenen schwingt stets eine gewisse Euphorie mit (Puschmann & Burgess, 2014). Nicht selten ist diese begleitet vom Glauben an einen bevorstehenden Paradigmenwechsel, in dem theoriebasierte Methodologien von „datengetriebenen“ Methoden abgelöst werden würden (z.B. von C. Anderson, 2008; Mayer-Schönberger & Cukier, 2013) – was wiederum vielfach kritisiert wurde (argumentiert z.B. von boyd & Crawford, 2012; N. Heise, 2015; Mahrt, 2015).

Zugleich greifen Organisationen und Individuen in allen gesellschaftlichen Bereichen und wirtschaftlichen Branchen vermehrt auf quantitative datenintensive Methoden zurück, um verschiedene Phänomene unserer Zeit einzuordnen. Der Journalismus bildet hier keine Ausnahme. Auch dort zeigt sich die zunehmende Orientierung an Daten auf unterschiedlichen Ebenen:

- Im Bereich der Steuerung von Medienunternehmen, wo wirtschaftliche und redaktionelle Entscheidungen vermehrt auf Signalen und Erkenntnissen basieren, die durch Datenanalysen erlangt werden – etwa im Bereich der „editorial analytics“ (C. W. Anderson, 2011a; Cherubini & Nielsen, 2016; Tandoc, 2014).

- In der Absatzebene, wo Medienorganisationen dazu übergehen, neben Inhalten in Form von Texten, Bildern oder Tönen auch kuratierte Daten anzubieten – etwa mittels Schnittstellen (Aitamurto & Lewis, 2013).
- In der Berichterstattung, in der Themen mit Datenbezug und deren gesellschaftliche Implikationen öffentlich erläutert und debattiert werden (Puschmann & Burgess, 2014).
- Und schließlich auf einer journalistisch-praktischen Ebene – und darum geht es in dieser Arbeit: als Quelle und Darstellungsform in der Berichterstattung, im sogenannten Datenjournalismus.

In den vergangenen Jahren hat das, was allgemein unter Datenjournalismus verstanden wird, eine starke Popularisierung erfahren. Das zeigt sich einerseits am Umstand, dass immer mehr Redaktionen auf datenintensive Methoden und Darstellungsformen zurückgreifen, andererseits an der Zunahme von einschlägigen Stellenausschreibungen, Ausbildungsmöglichkeiten, Interessensgruppen und Diskussionen. Letztere finden sowohl online – oft in Social Media – als auch offline – etwa auf Konferenzen – statt. Vom ‚Hype‘ um den Datenjournalismus ist selbst die Wissenschaft nicht gefeit, die in vielen ersten Veröffentlichungen die optimistischen Erfolgsnarrative der PraktikerInnen übernommen und mit der fundierten Untersuchung des Phänomens erst begonnen hat. Die vorliegende Arbeit versteht sich als Beitrag zu einer solchen fundierten wissenschaftlichen Analyse des Datenjournalismus und einer kritischen Auseinandersetzung mit der Erforschung des Phänomens.

1.1 Kontexte der Datafizierung⁵ der journalistischen Praxis⁶

Die Popularität des Datenjournalismus ist verbunden mit verschiedenen Kontexten, von denen einige einander auch reziprok bedingen. Die Entwicklungen sind dabei sowohl technischer wie auch sozialer Natur.

Der offensichtlichste Kontext ist die eingangs bereits erwähnte zunehmende Verfügbarkeit von Daten: Verwaltungen und Institutionen in allen Teilen der Welt sind in den vergangenen Jahrzehnten dazu übergegangen, ausgewählte Daten zu veröffentlichen. Unter dem Stichwort ‚open government data‘ haben sich seit Mitte der 2000er Jahre⁷ BeamtInnen, EntwicklerInnen, PolitikerInnen und BürgerInnen dafür eingesetzt, dass Daten der öffentlichen Hand zu offenen Daten werden, also frei und strukturiert zur Verfügung stehen – vom Parlamentsprotokoll bis zum Firmenbuch, von der Bevölkerungsstatistik bis zur Feinstaubmessung (Janssen, Charalabidis & Zuiderwijk, 2012; Ubaldi, 2013). In der Regel sind personenbezogene Daten, geistiges Eigentum sowie Daten, die kritische nationale Infrastruktur oder Betriebs- und Geschäftsgeheimnisse betreffen, von diesen Datenveröffentlichungen ausgeschlossen. Es haben sich bestimmte Standards und Gütekriterien etabliert, die etwa Formate, Lizenzen, Metadaten und Schnittstellen regeln (Tauberer, 2014). NGOs, Unternehmen und (internationale) Institutionen sind auch dazu übergegangen, Informationen in strukturierter Form zu veröffentlichen. So wie auch

⁵ Der Begriff ‚Datafizierung‘ kann aufgrund seiner Endung im Deutschen negativ konnotiert verstanden werden; dies ist hier aber nicht intendiert. Es handelt sich dabei um eine Übersetzung des englischen Begriffs „datafication“ (Mayer-Schönberger & Cukier, 2013) und bezeichnet hier anschließend an den Medialitäts-, Medialisierungs- bzw. Mediatisierungsbegriff (Bauer, 2008b, 2008a; Krotz, 2008; Krotz & Hepp, 2012) die zunehmende Orientierung von Alltagszusammenhängen, Lebenszusammenhängen und gesellschaftlich relevanten Organisationszusammenhängen an Daten (vgl. Bauer, 2014a, S. 27). Dieses Phänomen wird zunächst weder als positiv, negativ noch als neutral gesehen (Kranzberg, 1986), sondern in weiterer Folge im Kontext der Verknüpfung mit journalistischer Praxis diskutiert. Erst in Verbindung mit konkreten Praktiken lassen Konsequenzen und Funktionen dieses sozio-technischen Phänomens bestimmen. Das hier verfolgte Verständnis des Datenbegriffs wird in Kapitel 2 dargelegt.

⁶ Teile dieses Unterabschnitts basieren auf einem andernorts veröffentlichten Text (Ausserhofer, 2012), der für diese Arbeit adaptiert und aktualisiert wurde.

⁷ Die ideengeschichtlichen Ursprünge von Open Government Data lassen sich zumindest bis in die 1950er Jahre zurückverfolgen (Emerson, 1976; Parks, 1957). Ähnlich wie der datenintensive Journalismus war die Open-Data-Bewegung jedoch ab 2008, 2009 von einer bemerkenswerten Popularisierung ergriffen und konnte zahlreiche neue Akteure für die Idee gewinnen. Die Popularisierung hängt wohl auch mit Ersetzen des Begriffs der ‚public sector information (PSI)‘ durch den Begriff ‚open (government) data (OGD)‘ zusammen.

Verwaltungen haben sie sich dabei speziell an JournalistInnen als Zielgruppe gewandt. Nicht zuletzt fungieren auch manche WissenschaftlerInnen als DatenlieferantInnen für JournalistInnen. Sie teilen ihre Forschungsdaten, um ihre Arbeit offener, transparenter und reproduzierbarer zu machen.

Einen relevanten Beitrag zur besseren Verfügbarkeit von Daten, die journalistisch genutzt werden, haben darüber hinaus Transparenzinitiativen verschiedenster Natur geleistet: Zum einen wurden vielerorts Gesetze eingeführt, die den Zugang zu Informationen und Daten von staatlichen Stellen erleichtern und rechtlich durchsetzbar machen (Stichwort ‚Informationsfreiheitsgesetz‘ oder IFG bzw. ‚freedom of information act‘ oder FOIA)⁸. Zum anderen wurden einzelnen Medien bzw. Medienverbänden größere Datensätze von InformantInnen ‚zugespielt‘ (Stichwort ‚Whistleblowing‘). Vielfach werden diese größeren Datenmengen unter dem Stichwort ‚leak‘ diskutiert. Zu den prominentesten Beispielen, die auch datenjournalistisch aufbereitet wurden, zählen etwa die Analysen der von *Wikileaks* veröffentlichten Geheimdokumente der US-Armee über die Kriege im Irak und Afghanistan (New York Times, 2010) oder die *Panama Papers*, die die Steuervermeidung und -hinterziehung tausender Personen aufdeckte (Obermayer & Obermaier, 2016).

Weitere Praktiken, die das Datenangebot für datenintensiven Journalismus ausweiten, sind Eigenerhebungen durch JournalistInnen unter Einbeziehung von RezipientInnen, im Fachdiskurs oft als „crowdsourcing“ (Aitamurto, 2016; Muthukumaraswamy, 2010) „co-creation“ (Aitamurto, 2013) oder „participatory journalism“ (Domingo u. a., 2008; Singer u. a., 2011) bezeichnet. In einigen Fällen werden die Methoden auch zur Verbesserung der Datenqualität eingesetzt, indem etwa automatisiert erfasste Daten durch Laien verifiziert werden (Vehkoo, 2013).

Die verstärkte Datenverfügbarkeit ist nicht allein Grund für die weite Verbreitung des datenintensiven Journalismus innerhalb weniger Jahre. Ein weiterer Kontext ist das

⁸ Aus juristischer Perspektive sind in vielen Ländern Informationsfreiheit und Informationsweiterverwendung zwei unterschiedliche Materien, die auch in unterschiedlichen Gesetzen geregelt sind – mit dem Argument, dass der Zugang zu Informationen nicht gleichzeitig zur Weiterverwendung berechtigt. In Österreich wird die Informationsweiterverwendung von Dokumenten öffentlicher Stellen durch das Informationsweiterverwendungsgesetz (IWG) geregelt (für eine Übersicht der jüngsten Novelle siehe Huter, 2015). Ein Informationsfreiheitsgesetz (IFG) in Form einer Weiterentwicklung des aktuellen Auskunftspflichtgesetzes, das das in der Verfassung geregelte Amtsgeheimnis abschwächen würde, ist seit Jahren Gegenstand politischer Diskussionen.

Aufkommen von Technologien, die datenintensive redaktionelle Prozesse erleichtern. Zahlreiche Applikationen und Programmiersprachen der vergangenen Jahre haben sich zunehmend einsteiger- und anwenderfreundlich entwickelt. Oft können neue (Versionen von) Programme(n) größere Datenmengen besser verarbeiten, zuweilen liefern sie präzisere Analysen; neue Datenquellen können nutzbar gemacht, oder heterogene Daten miteinander kombiniert werden. Auch redaktionelle Kollaborationen und Projektmanagement werden durch Software zunehmend besser unterstützt. Ebenso haben sich die Technologien, um datenbasierte Stücke im Web zu präsentieren, weiterentwickelt. Ursache für den verstärkten Technologieeinsatz im datenintensiven Journalismus ist unter anderem, dass sich in den vergangenen Jahren Personen mit technischem Hintergrund – Informatik, Statistik, Informationsvisualisierung, etc. – zunehmend für den Journalismus interessieren (Usher, 2016, S. 37–69). Das wird etwa am Aufkommen der internationalen Graswurzel-Bewegung *Hacks/Hackers* evident, bei der regelmäßig JournalistInnen und TechnikerInnen verschiedenster fachlicher Provenienz zusammenkommen, um über technologische Innovationen im Journalismus zu diskutieren (Lewis & Usher, 2014). Aber auch die akademische technische Forschung entdeckt zunehmend den datenintensiven Journalismus als ein Anwendungsgebiet (siehe auch den später in diesem Kapitel folgenden Unterabschnitt zu Forschung abseits der Journalistik sowie Diakopoulos, 2012). Sie trägt unter anderem dazu bei, Technologien, die für andere Bereiche entwickelt wurden, in den Journalismus überzuführen.

Einen bedeutenden Anteil haben nicht zuletzt Institutionen, die die Entwicklung von neuen Technologien und Praktiken, die im Journalismus eingesetzt werden können, finanziell fördern – überregional etwa die *Knight Foundation* (S. C. Lewis, 2010; Lewis, 2011) oder *journalismfund.eu*, in Deutschland zum Beispiel diverse Stiftungen (z.B. *Augstein*⁹, *Bosch*¹⁰, *Volkswagen*¹¹), oder in Österreich die öffentliche Hand¹². Neben Förderinstitutionen haben noch weitere „Infrastrukturen des Journalismus“ (Russ-Mohl, 2006) in ihrer Ausrichtung einen „quantitative turn“ (Coddington, 2015) vollzogen. In der Journalistenausbildung etwa werden vermehrt Kurse für datenintensiven Journalismus angeboten (Splendore u. a., 2016). In

⁹ <http://www.rudolf-augstein-stiftung.de/augsteinstiftung/home.nsf/Navigation/BE812189298CF6A4C1257BF700333D57>

¹⁰ <http://www.bosch-stiftung.de/content/language1/html/39434.asp>

¹¹ <https://www.volkswagenstiftung.de/datenjournalismus.html>

¹² z.B. <http://www.validproject.at/>

ähnlicher Weise haben der Medienjournalismus und journalistische NGOs datenintensive Operationen als Thema entdeckt, was sich wiederum auf eine Verstärkung der Praxis auswirkt (Gynnild, 2014).

Ein weiterer Baustein ist die Entwicklung der journalistischen Praxis selbst: In den vergangenen Jahren wurden zahlreiche investigative Recherchen veröffentlicht, die auf internationale, datenintensive Zusammenarbeit von verschiedenen Medienorganisationen zurückgehen. In vielen Fällen wurde die Zusammenarbeit dabei vom *International Consortium of Investigative Journalists (ICIJ)* koordiniert. Beispiele für derartige Kollaborationen sind die bereits erwähnten Berichterstattungen über die *Panama Papers* und diverse Veröffentlichungen, die auf *Wikileaks* zurückgehen, sowie Projekte wie *Offshore Leaks*, *Swiss Leaks*, *Luxembourg Leaks*, *Football Leaks* oder *The Migrants' Files*. Meist werden die Daten dabei an einer zentralen Koordinationsstelle zusammengetragen, wo die datenintensiven Operationen konzentriert durchgeführt werden. Jedes Medium im Kooperationsverbund ist selbst für die Gestaltung der eigenen Berichterstattung verantwortlich – wiewohl es permanent enge Abstimmungen zum Beispiel in Bezug auf den Veröffentlichungszeitpunkt gibt (Léchenet, 2014). Derartige Leuchtturmprojekte des datenintensiven Journalismus strahlen wiederum auf die tagtägliche Praxis aus.

In einer Zeit, in der insbesondere der Qualitätsjournalismus unter enormem finanziellen Druck steht und Medienunternehmen Finanzierungsmodelle suchen, werden verschiedene Anstrengungen unternommen, um Onlinejournalismus durch direkte Zuwendungen von RezipientInnen zu finanzieren (Mahle, 2016). Datenintensiver Journalismus wird dabei von vielen Medienhäusern als „premium content“ gesehen, schließlich wurde er mit viel Aufwand von Domänen- und MethodenexpertInnen kollaborativ produziert (Batsell, 2014). Verglichen mit ‚gewöhnlichem‘ Textjournalismus, stellt er sich für RezipientInnen hochwertiger dar; auch die Verweildauer ist bei datenintensiven Projekten in der Regel höher und sie werden häufiger mit anderen geteilt (Gartner, 2014). Es besteht deshalb bei Medienunternehmen die Hoffnung, mittels Datenjournalismus mehr Einnahmen für den Onlinebereich zu lukrieren oder zumindest einen Imagegewinn beim Publikum zu erzielen (Kayser-Bril, 2013). Diese Hoffnung wird freilich von den hohen Kosten, die mit der Produktion datenintensiver Stücke verbunden sind, beeinträchtigt. Es existiert jedoch eine Reihe von Geschäftsmodellen für datenintensiven Journalismus (Gartner, 2014, S. 36–37; A. B. Howard, 2014, S. 24–28; Lorenz, 2012), wenn auch noch keine wissenschaftlich fundierten Evaluierungen über deren Erfolg.

Zusammenfassend lässt sich feststellen: Die Kontexte, die zu einer Popularisierung des datenintensiven Journalismus geführt haben, sind vielfältig und breit gefächert. Wichtige Aspekte sind dabei die bessere Verfügbarkeit von verwertbaren Daten, die Entwicklung von (neuen) Technologien zur Datenanalyse und -darstellung, die zunehmende Datenorientierung von traditionellen Infrastrukturen des Journalismus (Medienjournalismus, Journalistenausbildung, Förderinstitutionen) und die Etablierung von neuen Infrastrukturen für datenintensiven Journalismus (transnationale Plattformen und Kollaborationen). Diese Entwicklungen geschehen in einem für den Onlinejournalismus wirtschaftlich schwierigen Umfeld. Auch die journalistische Autorität befindet sich in einer Krise (Usher, 2016, S. 6). Diese und weitere Kontexte rahmen die Bedingungen, in denen datenintensiver Journalismus heute produziert wird.

1.2 Forschung zu datenintensivem Journalismus¹³

Nachdem verschiedene gesellschaftliche Rahmenbedingungen zu datenintensivem Journalismus im Groben abgesteckt wurden, stellt sich im Folgenden die Frage nach dem (kommunikations-)wissenschaftlichen Beitrag zu diesem Phänomen: Wie begegnet die akademische Forschung dem Thema datenintensiver Journalismus und dessen steigender Relevanz bislang?

Wie viele andere Felder ist auch der Journalismus seit jeher ein Bereich, der permanentem Wandel unterworfen ist. Die Ursachen für diesen Wandel sind vielfältig und selten eindeutig identifizierbar, liegen jedoch zu einem nicht unbedeutenden Teil in der technologischen Entwicklung begründet (Eisenstein, 1979/2005; Habermas, 1962/1990). Technologie ist weder gut noch schlecht; noch ist sie neutral (Kranzberg, 1986). Um ihre Folgen einschätzen zu können, kommt es auf die Untersuchung der sozialen Interaktion mit der Technologie an – dies wird an anderer Stelle in dieser Arbeit noch ausführlicher argumentiert. Die Folgen der Etablierung von Technologie in einem gesellschaftlichen Bereich gehen dabei über dessen Grenzen hinaus: „Technology’s interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices

¹³ Teile dieses Unterabschnitts basieren auf andernorts veröffentlichten Texten (Ausserhofer, 2015; Ausserhofer, Gutounig, Oppermann, Matiassek & Goldgruber, 2017), die für diese Arbeit adaptiert und aktualisiert wurden.

themselves (...)“ (Kranzberg, 1986, S. 545). In unterschiedlichen Kontexten oder unter Umständen eingeführt, kann die gleiche Technologie unterschiedliche Resultate zur Folge haben.

Nicht zuletzt deshalb verhandelt die Forschung seit Beginn der 2010er Jahre die Konsequenzen der Einführung datenintensiver Technologien in den Journalismus neu – wiewohl das für andere gesellschaftliche Bereiche wie etwa die industrielle Produktion schon unzählige Male und umfassend durchdekliniert und aus unterschiedlichen Perspektiven beleuchtet worden ist. Ein Gutteil der frühen Publikationen zu datenintensivem Journalismus scheint dabei dem zu unterliegen, was Münch und Schmidt (2005, S. 202) bei der Beschreibung des metamedialen Diskurses den „Fehlschluss der neuen Medien“ genannt haben: Die Medieninnovation wird als „revolutionär, einzigartig, qualitativ neuartig“ interpretiert. Mit der Charakterisierung als „revolutionär“ gehen utopische oder dystopische Prognosen über zu erwartende soziale Auswirkungen einher, die sich im Verlauf der Zeit stark relativieren (Lievrouw, 2002; Marvin, 1988; Münch & Schmidt, 2005; Passig, 2013).

Ähnlich wie bei der Forschung zu partizipativem Journalismus (Borger, van Hoof, Costera Meijer & Sanders, 2013) positioniert sich die frühe Forschung zu datenintensivem Journalismus mehrheitlich als Fürsprecherin der Praxis. Sie setzt viel Hoffnung darin, dass technologieinduzierte Innovation dem Journalismus aus seiner Krise helfen kann (C. W. Anderson, 2011b, 2013b). Paradigmatisch für diese Position ist der sogar als „The promise of computational journalism“ betitelte Zeitschriftbeitrag von Flew et al. (2012), der in abgeänderter Form bereits zuvor als Konferenzbeitrag erschienen ist (Flew, Daniel & Spurgeon, 2010). So heißt es etwa dort: „Computational journalism may assist news providers by generating new ways of producing quality news at greater speed and with reduced costs to active and participative audiences“ (Flew u. a., 2012, S. 168). Ähnlich argumentieren Hamilton und Turner (2009, S. 2), die durch datenintensive Praktiken in Redaktionen den investigativen Journalismus gestärkt sehen: „Computational journalism (...) can create new tools that may reduce the cost of watchdog reporting in certain circumstances, take better advantage of the new information environment, and ultimately help sustain watchdog work during the technological sea change now under way.“ Gemeinsam mit Cohen (Cohen, Hamilton & Turner, 2011a, S. 69) formulieren sie ihre These noch gewagter: „Being able to analyze and visualize interactions among entities (...) would give stories more depth, reduce the cost of reporting, and expand the potential for new stories and new leads“. Pointiert fasst Anderson (2013b, S. 1007) diese optimistischen Narrative zusammen: Was die Technologie (als Auslöserin der

Strukturkrise der Medien) fortgenommen hat, gibt sie zurück – nur anders, und vielleicht besser als zuvor („What technology taketh away, technology giveth – only differently, and perhaps even better than before.“).

Als Alternative zu diesen ersten konzeptionellen Publikationen, die dem Datenjournalismus vor allem viel Potenzial in verschiedene Richtungen zuschrieben, schlug Anderson (2013b) verschiedene empirisch fundierte Ansätze vor, um datenintensiven Journalismus zu erforschen. Dabei baute er einerseits auf Schudsons (2005) Typologie für eine Nachrichtensoziologie auf, andererseits auf Bourdieus Feldtheorie. Anderson empfahl, Studien mit verschiedenen „Linsen“ durchzuführen: Durch den Fokus auf ökonomische, politische, kulturelle, organisatorische, technologische oder feldtheoretische Aspekte ließe sich das Phänomen besser begreifen.

1.2.1 „Erste Welle“ der Datenjournalismusforschung

Zwei Jahre nach Andersons Vorschlag stellte sich das Feld bereits stark gewandelt dar: Gemeinsam mit Fink (Fink & Anderson, 2015, S. 476) sprach Anderson von einer Explosion von Datenjournalismus-orientierter Forschung. Und auch Lewis (2015, S. 322) sah eine schnell wachsende Ansammlung an Forschungsarbeiten, die die Natur von Automatisierung und Quantifizierung im Journalismus analysierten (zit. n. Loosen, Reimer & Schmidt, 2015 bzw. Loosen, Reimer & De Silva-Schmidt, 2016). Sieht man von vereinzelt frühen Publikationen und den eingangs in diesem Abschnitt diskutierten Publikationen mit optimistischem Grundton und ohne große empirische Basis ab, so können die ab 2012 nachfolgenden sozialwissenschaftlichen Publikationen durchaus als „first wave of data journalism research“ (Uskali & Appelgren, 2015) bezeichnet werden. Darunter fällt eine Reihe von kommunikationswissenschaftlichen Studien, die die nationalen Szenen und „Datentische“¹⁴ von Newsrooms näher beleuchteten – wobei insbesondere die Vorgänge in westlichen Ländern im Mittelpunkt des Interesses standen: USA (z.B. Boyles & Meyer, 2016; Fink & Anderson, 2015; Parasie, 2015; Parasie & Dagiral, 2013; Royal, 2012; Weber & Rall, 2013; Young & Hermida, 2015), England (z.B. Borges-Rey, 2016, 2017; Hannaford, 2015; Knight, 2015; Tandoc & Oh, 2015), Schweden (Appelgren & Nygren, 2014a, 2014b), Deutschland (Weinacht & Spiller, 2014), Norwegen (Karlsen & Stavelin, 2014; Stavelin, 2013), Kanada (Tabary,

¹⁴ Datentisch, der eingedeutschte Ausdruck für ‚data desk‘, meint jenen Ort in einer Redaktion, an dem für gewöhnlich mit Datenjournalismus befasste Personen sitzen.

Provost & Trottier, 2016), Finland (Uskali & Kuutti, 2015) und Belgien (De Maeyer, Libert, Domingo, Heinderyckx & Le Cam, 2015).

Viele dieser Publikationen beschäftigten sich mit der Praxis des datenintensiven Journalismus in einer oder mehreren Medienorganisationen. Unter anderem untersuchten sie die Organisationskulturen, die Redaktionsstrukturen, die Epistemologien von DatenjournalistInnen oder die Merkmale datenintensiver Stücke. Nicht untypisch für eine ‚erste Forschungswelle‘, waren die Fragestellungen und Ergebnisdarstellungen eher deskriptiv angelegt. In den Arbeiten zeigte sich unter anderem, dass zwar so manche Unterschiede in den analysierten Häusern und Ländern bestehen –Hierarchie und Organisationskultur etwa –, im Hinblick auf Epistemologie, Praktiken und technologische Entwicklungsstandards aber durchaus viele Ähnlichkeiten festzustellen sind: So existieren einerseits einige wenige datenjournalistische Leuchtturmprojekte und Redaktionen, die auch häufig im Blickpunkt der Analysen stehen, und andererseits viele (Individuen in) Organisationen, die mit Ressourcenknappheit kämpfen. Diese Dichotomie wird auch in der Art der realisierten Projekte offenbar – zum einen gibt es jenen Datenjournalismus, der von einer Einzelperson im ‚Tagesgeschäft‘ produziert wird, zum anderen Investigativprojekte, mit denen Teams in Arbeitsteilung oft wochen- oder monatelang beschäftigt sind (Fink & Anderson, 2015). Ähnlich wie im ‚traditionellen‘ Investigativjournalismus bilden sich dabei die interdisziplinären Teams je nach Projektanforderung neu, oft auch unter Einbeziehung von externen DienstleisterInnen. Deadlines treten in den Hintergrund, im Vordergrund stehen die Vollständigkeit des Beitrags und der Wunsch, mit jedem Projekt das technologische Potenzial etwas weiter auszuloten (Royal, 2012). Dementsprechend gelten viele DatenjournalistInnen in ihren Redaktionen als technologische Avantgarde, als ‚Nerds‘: Sie zeigen Affinität zu Transparenz-Ideologien aus der IT-Kultur und versuchen Methodologien wie Hacking in die Redaktionen zu tragen (Baack, 2015; A. B. Howard, 2014; Parasie & Dagiral, 2013; Weinacht & Spiller, 2014). Diese Arbeitskulturen sind mit jenen im ‚traditionellen‘ Journalismus nicht völlig kompatibel, weshalb sich Redaktionen durchaus schwer damit tun, Routinen zu etablieren (Karlsen & Stavelin, 2014).

1.2.2 Forschung abseits der Journalistik

Datenintensiver Journalismus ist ein interdisziplinäres Handwerk – nicht anders steht es mit dessen Erforschung. Parallel zu den journalismuswissenschaftlichen Explorationen der datenjournalistischen Praxis beschäftigen sich auch Forschende in anderen Disziplinen mit

dem Thema. Neben den Sozialwissenschaften haben sich in den vergangenen Jahren vor allem die Informatik und verwandte Disziplinen wie Statistik und Visualisierungsforschung dem Thema gewidmet. Eine wichtige Rolle spielen dabei auch die transdisziplinären Labels bzw. deren Foren, an denen Forschende unterschiedlicher wissenschaftlicher Disziplinen zusammenkommen: Computational Social Science, Web Science, Digital Methods und Digital Humanities (Puschmann, 2015).

Die meisten Publikationen abseits der Journalistik beschäftigen sich mit konkreten Anwendungsszenarien bzw. der Darstellung von Lösungen und Prototypen, mithilfe derer bestimmte Datentypen (besser) gesammelt, analysiert oder visualisiert werden können. Oft sind es Ergebnisse der Kooperation zwischen Forschenden und JournalistInnen, in denen die Forschungsseite die komplexeren Datenoperationen durchführt und an die JournalistInnen liefert. Rein quantitativ erschienen weit mehr Publikationen in diesem Kontext als in der klassischen Journalismusforschung, was wohl mit der unterschiedlichen Größe der Fächer sowie deren finanziellen Mitteln zu tun hat. Im Folgenden werden einige Beispiele für derartige Ansätze dargestellt, um einen Eindruck über diese Art der Forschung zu geben. Einen systematischen Überblick, aufgefächert nach informatischen Konzepten, liefert Diakopoulos (2012).

Rusch, Hofmarcher, Hatzinger und Hornik (2013) etwa beschäftigten sich mit den *Afghanistan War Logs*. Dabei handelt es sich um Daten, die den Einsatz der US-Armee im Krieg in Afghanistan zwischen 2004 und 2010 dokumentierten, und 2010 durch die Internetplattform *Wikileaks* veröffentlicht wurden. Die darauf aufbauenden journalistischen Datenanalysen wurden mehrfach als ein Meilenstein in der Entwicklung des Datenjournalismus genannt – insbesondere das Mapping von Kriegsopfern durch den *Guardian* (erwähnt u.a. von Baack, 2013; Knight, 2015; S. Rogers, 2013, S. 69–77; Splendore u. a., 2016) – und führten bei der Nachrichtenagentur *Associated Press* zur Entwicklung der Software *Overview*, einem weit verbreiteten Werkzeug zur Analyse von großen Dokumentenkorpora (Brehmer, Ingram, Stray & Munzner, 2014). Rusch und KollegInnen jedenfalls lieferten einen alternativen Ansatz: Sie schlugen Topic Modeling vor, um Sinn aus den großen Textmengen zu ziehen (zu Topic Modeling im kommunikationswissenschaftlichen Kontext siehe auch Puschmann & Scheffler, 2016).

Eine weitere beliebte Datenquelle der Datenjournalismus-bezogenen, technischen Forschung und Entwicklung sind Social-Media-Daten. Im Kontext von Ausschreitungen in

England im Sommer 2011, bei denen soziale Medien eine wichtige Rolle spielten, stellen etwa Procter, Vis und Voss (2013) einen Mehrmethoden-Ansatz vor, der Natural Language Processing mit traditionelleren Inhaltsanalysen kombiniert. Seine Anwendung fand dieser Ansatz schließlich in einer *Guardian*-Serie mit dem Titel *Reading the Riots*, die sich aus mehreren interaktiven Beiträgen zusammensetzte.

Social-Media-Daten standen ebenfalls im Zentrum einer Kooperation zwischen der Wiener Tageszeitung *Der Standard* und dem Berliner *Alexander von Humboldt Institut für Internet und Gesellschaft*, an der ich beteiligt war: In enger Abstimmung wurde die Kommunikation von rechtspopulistischen Empörungsbewegungen in Europa auf verschiedenen Social-Media-Plattformen untersucht. Zum Einsatz kamen verschiedene Methoden: Unter anderem wurden Topic-Modeling und Netzwerkanalyse mit Experteninterviews kombiniert. Der Output erfolgte dabei sowohl in journalistischen Formaten (u.a. Gartner, Maan, Schmid & Hametner, 2016; Maan & Schmid, 2016; Maan, Schmid & Hametner, 2016; Schmid, Maan, Hametner & Šlerka, 2016) wie auch über wissenschaftliche Veröffentlichungskanäle (Puschmann & Ausserhofer, 2017; Puschmann, Ausserhofer, Maan & Hametner, 2016).

Eine große Anzahl an Publikationen liefert Visualisierungslösungen für bestimmte Datentypen. Dazu gehören etwa die Entwicklung von Visual-Analytics-Anwendungen, die DatenjournalistInnen erste Einsichten in größere, heterogene oder dynamische Datenmengen liefern. Ziel ist dabei weniger, die Daten für eine Visualisierung in einem Medium aufzubereiten; vielmehr können diese Lösungen dazu beitragen, RedakteurInnen visuell unterstützt einen Überblick über einen oder mehrere Datensätze zu geben und potenzielle Fehler oder Geschichten in den Daten finden zu lassen. In Projekten, an denen ich beteiligt war, wurden etwa Prototypen für die Exploration der Ausgaben öffentlicher Hand für heimische Medienunternehmen (Niederer u. a., 2016; Rind, Pfahler, Niederer & Aigner, 2016) sowie geo-basierte Visualisierungen von Diversitätsdaten (Andrews u. a., 2016; Andrews, Traunmüller, Wolking, Gutounig & Ausserhofer, 2015; Gutounig u. a., 2016) entwickelt. Ein weiteres Beispiel für Visual-Analytics-Anwendungen sind Dashboards zur Sammlung und Evaluation von Social-Media-Daten (Backholm u. a., in Druck; Diakopoulos, Naaman & Kivran-Swaine, 2010).

Neben Visual-Analytics-Lösungen, die insbesondere den Prozess der Datenexploration visuell unterstützen sollen, entsteht auch viel Forschung, die datenbasierte journalistische

Darstellungsformen und Visualisierungen entwickelt und evaluiert. Überblicke liefern etwa Segel und Heer (2010) für den Bereich der „narrativen Visualisierungen“, Niederer, Aigner und Rind (2015) im Bereich dynamischer, gewichteter und gerichteter Graphen sowie Weber (2013) für den Bereich interaktiver Infografiken.

1.2.3 Forschung und Reflexionen über datenjournalistische Ausbildungen

Einen weiteren Strang der Forschung zu datenintensivem Journalismus bilden Publikationen, die sich mit der Implementierung von datenintensiven Methoden in der Journalistenausbildung auseinandersetzen. Diese Auseinandersetzung passiert zum einen aus einer Panoramaperspektive, hauptsächlich in Form von Umfragen, die erfragen, wo und wie Datenjournalismusbildung angeboten wird (L. D. Davenport, Fico & DeFleur, 2002; K. C. Lee & Fleming, 1995; Splendore u. a., 2016; Yarnall, Johnson, Rinne & Ranney, 2008), zum anderen in Form von Reflexionen der eigenen Lehrpraxis sowie Vorschlägen zur Gestaltung von Curricula (Feigenbaum, Thorsen, Weissmann & Demirkol, 2016; Hewett, 2013, 2016; Plau & Cook, 2015; Pulimood, Shaw & Lounsberry, 2011; Tankard & Lasorsa, 2000; Treadwell, Ross, A. Lee & Lowenstein, 2016).

1.3 Ziele der Arbeit

Der vorangegangene Abschnitt stellte dar, wie die akademische Forschung in den vergangenen Jahren dem Thema ‚datenintensiver Journalismus‘ begegnet ist. Dabei wurden verschiedene Stränge der Forschung zu datenintensivem Journalismus diskutiert: Zunächst wurden Veröffentlichungen der „first wave of data journalism research“ besprochen – eine Reihe von Publikationen, die klassische Fragestellungen der Journalismusforschung verfolgen und die Praxis und Produkte des datenintensiven Journalismus in ausgewählten westlichen Ländern untersuchen. Anschließend wurden weitere Forschungsstränge beleuchtet: Veröffentlichungen aus technischen Disziplinen und darüber hinaus, die entwickelte Lösungen und Prototypen für datenjournalistische Verfahren vorstellen. Die Konzeptionen unterstützen verschiedene Prozesse der datenintensiven journalistischen Arbeit – von der Datensammlung und deren Evaluierung bis hin zur Analyse und Visualisierung. Einen dritten Strang stellten Publikationen dar, die die Vermittlung von Datenjournalismus an Studierende thematisieren.

Zum Zeitpunkt der Niederschrift dieser Arbeit stellt sich die Publikationslandschaft zu datenintensivem Journalismus jedenfalls als fragmentiert dar. Das hat erstens damit zu tun, dass in den vergangenen Jahren viele Publikationen fast zeitgleich erschienen, dabei ähnliche Ergebnisse zutage förderten und sich wenig aufeinander bezogen. Die Ergebnisse der meisten Publikationen können allein Gültigkeit für Aussagen über einen Fall wie etwa eine Redaktion reklamieren, der zumeist auch noch rekonstruktiv erhoben wurde. Zweitens scheint die Fragmentierung der Forschung durch die Interdisziplinarität des Feldes befördert worden zu sein: Viele Publikationen erschienen in disziplin-spezifischen Kanälen, die bei disziplin-fremden Forschenden wenig Aufmerksamkeit fanden. In Kombination führten diese zwei Entwicklungen zu diesem Zustand der Fragmentierung. Hier setzt diese Arbeit zunächst an.

Um dieser Fragmentierung etwas entgegen zu setzen, definiert diese Arbeit als **erstes Ziel, eine Forschungssynthese zu datenintensivem Journalismus zu entwickeln**, die über den hier gegebenen ersten Literaturüberblick hinausgeht. Der Fokus soll dabei auf zweierlei liegen: Einerseits soll die Entwicklung der Forschung zu datenintensivem Journalismus beschrieben werden, andererseits soll zusammengefasst werden, was bereits zu den datenjournalistischen Akteuren und Arbeitsabläufen geforscht wurde.

Der erste, kursorische Literaturüberblick zeigte darüber hinaus, dass die Erforschung des Datenjournalismus im deutschsprachigen Raum ein weitestgehend blinder Fleck auf der Forschungslandkarte ist – mit Ausnahme der Arbeit von Weinacht und Spiller (2014) und wenigen Qualifikationsarbeiten. Dieser Umstand verwundert, schließlich sind mit der *Zeit*, dem *Spiegel*, der *Süddeutschen Zeitung* dem *Standard* und der *NZZ* einige der europäischen Datenjournalismus-Leuchttürme im deutschsprachigen Raum zu finden. In Anbetracht des starken Interesses der bisherigen Forschung an den datenjournalistischen Produktionsbedingungen anderswo, soll die vorliegende Arbeit dazu beitragen, diesen blinden Fleck auf der Forschungslandkarte zu eliminieren. Somit kann als **zweites Ziel dieser Forschung definiert werden, die datenintensiven, redaktionellen Arbeitsabläufe im deutschsprachigen Raum zu erkunden**.

Mit der so angestellten Diskussion des Forschungsstands und der redaktionellen Arbeitsabläufe des datenintensiven Journalismus, ist der Wissensakkumulation durch Beschreibung vorerst hinreichend Genüge getan. Es gilt dann, die deskriptive Ebene zu verlassen und zu reflektieren. Datenintensiver Journalismus ist eingebettet in durch Menschen und Algorithmen co-konstruierte, vernetzte Öffentlichkeiten und datafizierte Gesellschaften,

in denen ‚data literacy‘ als zentrale Kulturtechnik genannt wird. **Das dritte Ziel dieser Arbeit ist demnach eine Reflexion darüber, was die Erforschung und Praxis des datenintensiven Journalismus für die datafizierten Gesellschaften und vernetzten Öffentlichkeiten und ihren Kommunikationssinn bedeuten.**

1.4 Zentrale Fragestellungen

Um die im vorangehenden Abschnitt formulierten Ziele zu erreichen, ist es sinnvoll, die Ziele in Form von forschungsleitenden Fragestellungen zu operationalisieren. Ziel 1, eine Forschungssynthese unter besonderer Berücksichtigung von Feldentwicklung und Arbeitsabläufen zu formulieren, soll durch Beantwortung der unter FF1 genannten Fragen erreicht werden. Nach dem Vorbild anderer Forschungssynthesen (z.B. Fecher, Friesike & Hebing, 2015; Jungherr, 2016; Massaro, Dumay & Garlatti, 2015) beschäftigt sich die erste Frage mit der Entstehung und dem aktuellen Zustand des Felds:

- FF1.1: Wie entwickelt sich die Forschungsliteratur zu datenintensivem Journalismus, unter anderem in Bezug auf Publikationsaktivität, Veröffentlichungen und Zitationen?

Die Beantwortung dieser Frage hilft anderen Forschenden dabei, die Entwicklung des Felds zu beurteilen und Werke zu identifizieren, die den größten Einfluss auf die Konstruktion des wissenschaftlichen Objekts ‚datenintensiver Journalismus‘ hatten. Von besonderer Bedeutung sind in diesem Zusammenhang die verwendeten Theorien und Methoden sowie die wahrgenommenen Forschungslücken der Datenjournalismusforschung, weshalb diesen drei Aspekten eigene Forschungsfragen gewidmet sind:

- FF1.2: Was sind die Theorien der Datenjournalismusforschung?
- FF1.3: Was sind die Methoden der Datenjournalismusforschung?
- FF1.4: Welche Forschungslücken haben Datenjournalismus-Forschende identifiziert und was sind ihre Vorschläge für zukünftige Forschung?

Nach der Entwicklung einer Metaperspektive für das ‚wissenschaftliche Objekt‘ datenintensiver Journalismus, wendet sich die Forschungssynthese der leitenden Frage der redaktionellen Arbeitsabläufe im datenintensiven Journalismus zu:

- FF1.5: Was hat die Datenjournalismusforschung bislang zu datenjournalistischen Arbeitsabläufen festgestellt?

Diese Synthese ist sowohl für ForscherInnen wie auch DatenjournalistInnen selbst relevant: Erstere können damit unerforschte Aspekte identifizieren, zweitere erhalten eine Übersicht, wie anderswo datenjournalistische Arbeitsabläufe funktionieren. Darüber hinaus bereitet die Beantwortung der Forschungsfrage den Boden für die Erreichung des zweiten Ziels auf, die Exploration von datenintensiven redaktionellen Arbeitsabläufen im deutschsprachigen Raum. Dieses Forschungsinteresse wird in den drei Forschungsfragen FF2.1-2.3 operationalisiert. Wie dies ausführlicher im Kapitel zu wissenschaftstheoretischen Grundannahmen argumentiert wird, verfolgt diese Arbeit die Perspektive, dass verschiedene menschliche und nicht-menschliche Akteure zur Produktion eines datenintensiven journalistischen Beitrags kooperieren (vgl. auch Primo & Zago, 2015 sowie Steensen, 2016). Will man also die Arbeitsabläufe des datenintensiven Journalismus erkunden, gilt es, den Beitrag der beteiligten Akteure zu ebendiesen Arbeitsabläufen zu deklarieren. Dazu muss man zunächst die Abläufe aus Sicht der Akteure innerhalb und außerhalb der Redaktionen beschreiben (FF2.1), deren Zusammenarbeit charakterisieren (FF2.2) und schließlich deren einzelnen Beitrag herausarbeiten (FF2.3).

- FF2.1: Wie stellen sich die redaktionellen Produktionsbedingungen für mit datenintensivem Journalismus befasste JournalistInnen im deutschsprachigen Raum dar?
- FF2.2: Wie stellen sich redaktionelle Routinen und Arbeitsabläufe dar?
- FF2.3: Welche menschlichen Akteure und nicht-menschlichen Entitäten beteiligen sich an der Entstehung datenjournalistischer Produkte?

Jedem der Ziele ist ein einzelnes Kapitel dieser Arbeit gewidmet, wobei die einzelnen Forschungsfragen in eigenen Abschnitten beantwortet und im letzten Kapitel zusammengefasst werden. Sie sind deskriptiv angelegt und füllen verschiedene Lücken in der Forschungslandschaft. Zugleich legen sie die empirische Basis zur Erreichung des letzten Ziels dieser Arbeit, die Praxis des datenintensiven Journalismus innerhalb menschlich-algorithmisch co-konstruierter vernetzter Öffentlichkeiten und zunehmend von Datenkompetenz abhängigen Gesellschaften konzeptionell zu verorten. Entsprechend offen ist auch die Forschungsfrage 3 dieser Arbeit formuliert:

- FF3: Welche Konsequenzen hat die Praxis des Datenjournalismus für zunehmend von Datenkompetenz abhängige Gesellschaften und deren menschlich-algorithmisch co-konstruierte Öffentlichkeiten?

Die Beantwortung dieser Frage erweitert den theoretischen Diskurs zu Datenjournalismus um eine bisher wenig verfolgte Perspektive. Wie der erste Literaturüberblick gezeigt hat, ist die Anzahl an Arbeiten, die sich theoretisch mit der Rolle von datenintensiven Operationen und Journalismus in unserer Gesellschaft auseinandergesetzt haben, überschaubar. Diese Fragestellung ermöglicht es, der Theoriearmut etwas Abhilfe zu schaffen.

1.5 Forschungszusammenhang und Verwertungsperspektive

Wozu diese Forschung zu diesem Thema? Vorab soll angemerkt sein, dass der Frage nach einer wirtschaftlichen Verwertungsperspektive meines Erachtens keine Priorität einzuräumen ist. In einer Zeit der zunehmenden Ökonomisierung sämtlicher Lebensbereiche genießen manche Teile der Forschung den Luxus, sich nicht nach den Gesetzen des Markts orientieren zu müssen. Will man dem Ökonomisierungsdiktat entgegenreten, muss es möglich sein, ‚unpraktische‘ Forschung zu betreiben, die keinen unmittelbaren wirtschaftlichen Anwendungsfall kennt.

Anders sieht es mit einer an Wissenschaft oder Gemeinwohl orientierten Verwertungsperspektive aus, der sich Forschung meines Erachtens sehr wohl unterordnen muss. Der Beitrag dieser Arbeit zur Kommunikationswissenschaft und insbesondere zur Journalismusforschung wird vielfach im Text thematisiert. Es wird an verschiedenen Stellen klar herausgearbeitet, wo die Forschungslücken zu datenintensivem Journalismus liegen und welche davon die dieser Arbeit zugrundeliegende Forschung füllt: Eine systematische Forschungssynthese zu Datenjournalismus wurde bislang noch nicht durchgeführt. Ebenso ist bislang wenig über die datenjournalistischen Arbeitsabläufe im deutschsprachigen Raum bekannt. Und auch die Bausteine zu einer die vernetzten Öffentlichkeiten berücksichtigenden Theorie des Datenjournalismus wurden bislang wenig erörtert.

Diese Analysen und Konzeptionen nützen sowohl Forschenden wie auch PraktikerInnen zum besseren Verständnis des Phänomens. Eine konkrete wissenschaftliche Verwertungsperspektive, die diese Arbeit entwickelt, ist die Schaffung einer Grundlage für eine große quantitative Erhebung zu Datenjournalismus in Europa. Eine weitere Verwertungsperspektive dieser Arbeit betrifft die Journalistenausbildung: Auch wenn mittlerweile viele Aus- und Weiterbildungsinstitutionen Kurse für datenintensiven Journalismus anbieten, scheint eine Professionalisierung der Ausbildung noch in weiter Ferne

(Splendore u. a., 2016). Die Perspektive, die diese Arbeit auf die datenjournalistische Praxis liefert, kann dazu beitragen, diese Professionalisierung voranzutreiben.

1.6 Aufbau der Arbeit

Das auf die Einleitung folgende Kapitel ist der Definition der wichtigsten Begriffe gewidmet. Der Begriff des datenintensiven Journalismus wird erläutert und Charakteristiken datenjournalistischer Beiträge werden dargestellt.

Im daran anschließenden Kapitel werden die wissenschaftstheoretischen Grundannahmen dieser Arbeit diskutiert. Aufbauend auf verschiedenen epistemologischen Ansätzen wird für ein holistisches Konstruktivismuskonzept argumentiert, in dem historische, soziale, kommunikative und technische Faktoren wichtige Rollen bei der Realitätskonstruktion einnehmen.

Das Epistemologie-Kapitel bildet die Grundlage für die methodologische Verortung und die Darstellung des Forschungsdesigns, das im darauffolgenden Kapitel diskutiert wird. Insbesondere gehe ich aber auf die Kontexte von Mixed-Method-Forschung und digitalen Methoden ein.

Die wiederum nächsten zwei Kapitel sind der Ergebnisdarstellung der empirischen Forschung gewidmet. Unter Anwendung einer systematischen Dokumentenanalyse wird eine Forschungssynthese zu Arbeitsabläufen im datenintensiven Journalismus vorgelegt. Diese wird kontextualisiert durch eine Interviewstudie, die die Produktionsbedingungen für datenintensiven Journalismus im deutschsprachigen Raum beschreibt.

Das Abschlusskapitel schließlich bietet einerseits eine Zusammenfassung aus den beiden vorangegangenen Kapiteln, andererseits eine Theoretisierung der Praxis des datenintensiven Journalismus vor dem Hintergrund menschlich-algorithmisch co-konstruierter Öffentlichkeiten.

2 Zum Begriff des datenintensiven Journalismus¹⁵

Der Begriff des ‚datenintensiven Journalismus‘ kennt zahlreiche Definitionen, die sich mindestens in drei Gruppen einteilen lassen: Während manche Definitionen datenintensiven Journalismus vor allem als Produktionsprozess begreifen, gibt es andere, die ihn als Darstellungsform charakterisieren. Eine dritte Gruppe an Definitionen wiederum integriert Prozess- und Produktvorstellungen. Dieser Perspektive schließt sich auch meine Arbeit an.

‚Computer-assisted reporting‘, ‚data-driven journalism‘, ‚computational journalism‘, ‚database journalism‘, ‚Datenjournalismus‘, ...: Die Liste an Ausdrücken für „technologically specific forms of work“ (Powers, 2012), die die datenintensive Berichterstattung im Journalismus beschreiben, ist lang. Viele Ausdrücke sind verbunden mit unterschiedlichen Communities, historischen Hintergründen, Epistemologien und Visionen von Öffentlichkeit (Borges-Rey, 2017; Bounegru, 2012; Coddington, 2015). Auch sind die Ausdrücke zu unterschiedlichen Zeiten unterschiedlich populär und meinen unterschiedliche Praktiken bzw. Produkte. Dies sei im Folgenden am Verhältnis von ‚computer-assisted reporting‘ und ‚data journalism‘, den zwei populärsten Begriffen, erläutert.

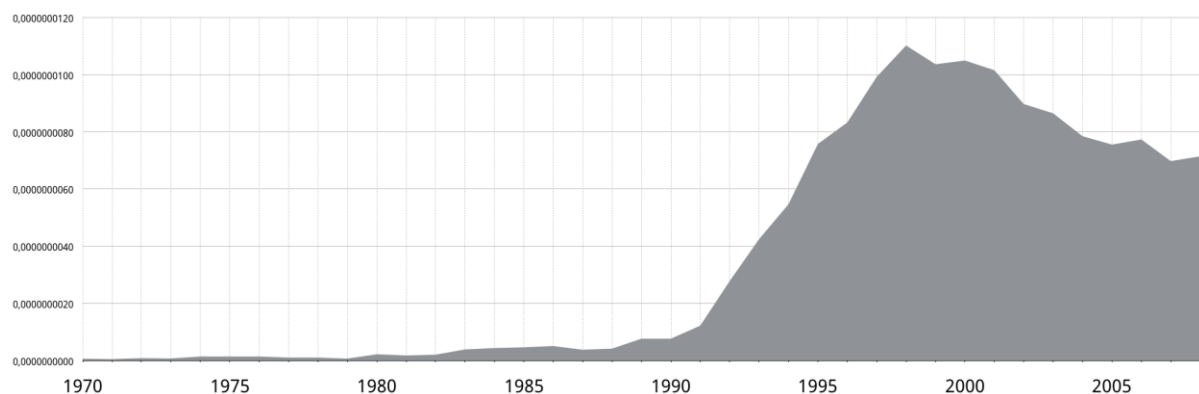
‚Computer-assisted reporting‘, kurz CAR, ist der älteste spezifische Ausdruck im Kontext des datenintensiven Journalismus. Cox (2000) verortet die Ursprünge von CAR in den 1950er Jahren und definiert es als „anything that uses computers to aid in the news-gathering process“ (S. 2). Diese sehr allgemeine Definition inkludiert heute wohl sämtlichen Journalismus, denn die computerlose Recherche wird nirgendwo mehr praktiziert. Doch von den 1950er Jahren bis zur allgemeinen Technologisierung des Journalismus war der Einsatz von Computern in der Recherche ein Distinktionsmerkmal, das zur Herausbildung eines spezifischen Subgenres führte. Das zeigt sich einerseits an rückblickenden Reflexionen aus dieser Zeit (Poynter Institute for Media Studies, 1999), andererseits an spezifischer Journalismusforschung, die beispielsweise erhob, wie stark E-Mail in einer Redaktion eingesetzt wurde (Garrison, 1996). Im Laufe der Zeit erfolgte jedoch eine Begriffsverschiebung: CAR bezog sich nun meist auf den Einsatz von Datenbanken sowie auf die Verwendung von quantitativen sozialwissenschaftlichen Methoden, insbesondere Statistik,

¹⁵ Teile dieses Kapitels basieren auf einem andernorts veröffentlichten Text (Ausserhofer u. a., 2017), der für diese Arbeit übersetzt, adaptiert und aktualisiert wurde.

Umfragen und Inhaltsanalysen (Coddington, 2015). Die Forschungsliteratur schreibt diesen Paradigmenwechsel insbesondere Philip Meyer (1973/2002) zu, dessen Buch *Precision Journalism*, „an effort to make journalism more accountable and scientific“ (Stavelin, 2013, S. 28), von vielen als Meilenstein gesehen wird. Die CAR-Bewegung ist besonders stark in den USA und eng mit dem investigativen Journalismus verbunden (Coddington, 2015; Gynnild, 2014). Ein wichtiges Fundament der CAR-Community ist das in Missouri (USA) ansässige *National Institute for Computer-assisted Reporting*, dessen Mailingliste¹⁶ und jährliche Konferenz *NICAR* (Parasie & Dagiral, 2013).

Seit den frühen 2000er Jahren, mit der breiten Technologisierung der Redaktionen, verlor der Begriff des ‚computer-assisted reporting‘ jedoch an Popularität (Coddington, 2015). Dieser Trend zeigt sich auch in der Literatur (siehe Abbildung 1): Mit Ende der 80er, Beginn der 90er Jahre fand der Ausdruck ‚computer-assisted reporting‘ seine ersten Erwähnungen in der Literatur. Zehn Jahre später nahm das Vorkommen des Begriffs wieder ab. Nichtsdestotrotz ist der Begriff heute immer noch gebräuchlich und die Community rund um CAR ist vor allem in den USA sehr aktiv.

Abbildung 1. Erwähnungen von ‚computer-assisted reporting‘ in englischsprachiger Literatur 1970–2008



Die Werte der y-Achse geben die Häufigkeit des Begriffs relativ zum gesamten englischsprachigen Korpus von Google Books 1970–2008 wieder. Datum der Abfrage: 2. Februar 2017. Datenquelle: Google Books Ngram Viewer, 2017 (<https://goo.gl/mPtAAE>).

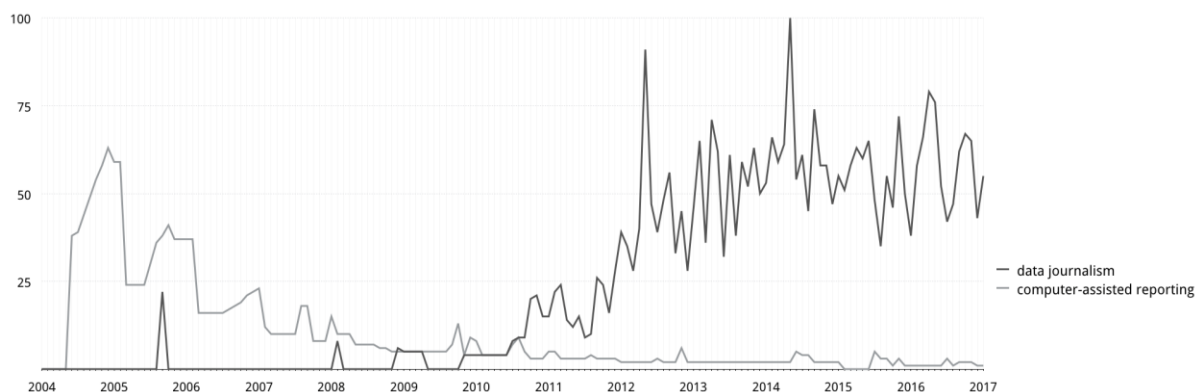
Die Ursprünge des Ausdrucks ‚data journalism‘ sind nicht vollständig geklärt (Hirst, 2014). Eine der ersten dokumentierten öffentlichen Erwähnungen des Begriffs erfolgte Ende 2008 im *Guardian* (Knight, 2015). Dort schrieb dessen Redakteur Simon Rogers (2008) in einem Blogbeitrag:

¹⁶ <https://ire.org/resource-center/listservs/subscribe-nicar-1/>

As of yesterday, our development team has come up with an application which takes the raw data and turns it into an editable map. Which meant that we could produce a fantastic interactive graphic based on these figures. It's data journalism — editorial and developers producing something technically interesting and that changes how we work and how we see data.

Rogers beschreibt hier bereits, dass datenintensiver Journalismus die Arbeitsabläufe in einer Redaktion ändert („changes how we work“). Ob Rogers zu diesem Zeitpunkt die vor allem in den USA praktizierte, ältere Tradition des computer-assisted reporting bekannt war, ist nicht klar. Jedenfalls arbeiteten er und einige andere in der Folge intensiv an der Popularisierung des Begriffs. Im März 2009 eröffnete der *Guardian* seinen *Datablog* (Hirst, 2009), der als Referenz in weiterer Folge einen großen Anteil an der Bekanntmachung des Konzepts haben sollte. Erst Mitte 2010, im Zuge der datenintensiven Berichterstattung über die Kriegstagebücher (*War Logs*) des Irak- und Afghanistankriegs wurde der Begriff des Datenjournalismus bzw. ‚data journalism‘ einer breiteren Öffentlichkeit bekannt. Dies zeigt auch Abbildung 2, die das Suchinteresse nach ‚computer-assisted reporting‘ und ‚data journalism‘ bei *Google* nachzeichnet: 2010 markiert jenes Jahr, in dem kaum mehr nach dem Begriff ‚computer-assisted reporting‘ gesucht wurde. Zugleich ist großes Interesse an ‚data journalism‘ zu verzeichnen. Andere verwandte Begriffe erzielten keine aussagekräftigen Ergebnisse.

Abbildung 2. Suchinteresse online für ‚computer-assisted reporting‘ und ‚data journalism‘



Die Werte der y-Achse geben das Suchinteresse bei *Google* relativ zum höchsten Punkt im Diagramm an. Der Wert 100 steht für die höchste Beliebtheit des Suchbegriffs, der Wert 50 bedeutet, dass der Begriff halb so beliebt war. Datum der Abfrage: 2. Februar 2017. Datenquelle: Google Trends, 2017 (<https://goo.gl/Bcuq8a>).

Wie es häufig bei sich im Wandel befindlichen Phänomenen der Fall ist, gibt es keine allgemein akzeptierte Definition von Datenjournalismus. Es existieren zumindest mehrere Dutzend Definitionen, geäußert bzw. veröffentlicht sowohl von PraktikerInnen als auch von Forschenden; und laufend kommen neue hinzu. Royal und Blasingame (2015) haben 63 Definitionen zusammengetragen und analysiert. Dabei arbeiten sie unterschiedliche

Dimensionen bzw. Perspektiven in den Definitionen heraus. Diese werden auch im Folgenden herangezogen, um das Phänomen abzustecken.

2.1 Datenintensiver Journalismus als Prozess

In zahlreichen Arbeiten etwa wird Datenjournalismus als ein spezifischer Prozess in der Produktion von Nachrichten charakterisiert (z.B. bei Appelgren & Nygren, 2014b; Bounegru, 2012; L. D. Davenport, Fico & Detwiler, 2000; Diakopoulos, 2011; Felle, 2016; Flew u. a., 2012; Karlsen & Stavelin, 2014; Parasie & Dagiral, 2013; Tandoc & Oh, 2015; Uskali & Kuutti, 2015; Weber & Rall, 2012; Weinacht & Spiller, 2014; Young & Hermida, 2015). Diese Definitionen unterstreichen die Rolle der Daten als zusätzliche Quelle in der Recherche und der Produktion von Nachrichten. Für die Datenverarbeitung werden spezielle Fertigkeiten und Software vorausgesetzt.

Prozessvorstellungen laden dazu ein, in Schemata gedacht zu werden.¹⁷ Dementsprechend haben viele versucht, datenintensiven Journalismus in einzelne Schritte bzw. Phasen aufzuteilen. Aitamurto, Sirkkunen und Lehtonen (2011, S. 11) etwa nennen sieben verschiedene Phasen in der Produktion:

1. Identifying the gist for the story and the potential role for the data in the story.
2. Identifying and obtaining the right data sets to respond to journalists' questions.
3. Modifying the data to make them ready for analysis—e.g., correcting errors in the data sets.
4. Analysing the data with the right tools, and mashing the data with other data sets if relevant.
5. Producing the story: text, visualisations, interactive elements.
6. Publishing the data sets that were used in the analysis.
7. Inviting readers to participate by reusing the data, commenting on and sharing the story through applications in social media, and submitting more content through applications like Flickr, Facebook and Twitter.

Ähnlich gelagert ist auch das Schema von Weinacht und Spiller (2014, S. 426), die zwischen „Datenjournalismus im engsten Sinne“ (Datensammlung, Datensicherung, Datenbereinigung, Datenauswertung, Datenaufbereitung) und „Datenjournalismus im weiteren Sinne“ (bereits

¹⁷ Zur Kritik an Schemata siehe auch die Diskussion zu Konsequenzen von Schematisierungen in der Forschung in Kapitel 4.

erwähnte Tätigkeiten plus Projektkonzeption, Schreiben, Redigieren, Kontrolle und Freigabe) unterscheiden. Hinzu kommen in ihrem Modell noch Vorarbeiten wie Ideengebung und die Durchsetzung des Themas in der Redaktionskonferenz, sowie Nacharbeiten wie Kommunikation zum Veröffentlichungstermin oder ‚Lobbying‘ für den Datenjournalismus.

Auch einige PraktikerInnen haben Schemata entwickelt, die teils weit verbreitet sind und vielfach für die Vermittlung von datenintensivem Journalismus herangezogen wurden. Abbildung 3 zeigt drei populäre Modelle.

Abbildung 3. Beispiele für Prozess-Visualisierungen des datenintensiven Journalismus

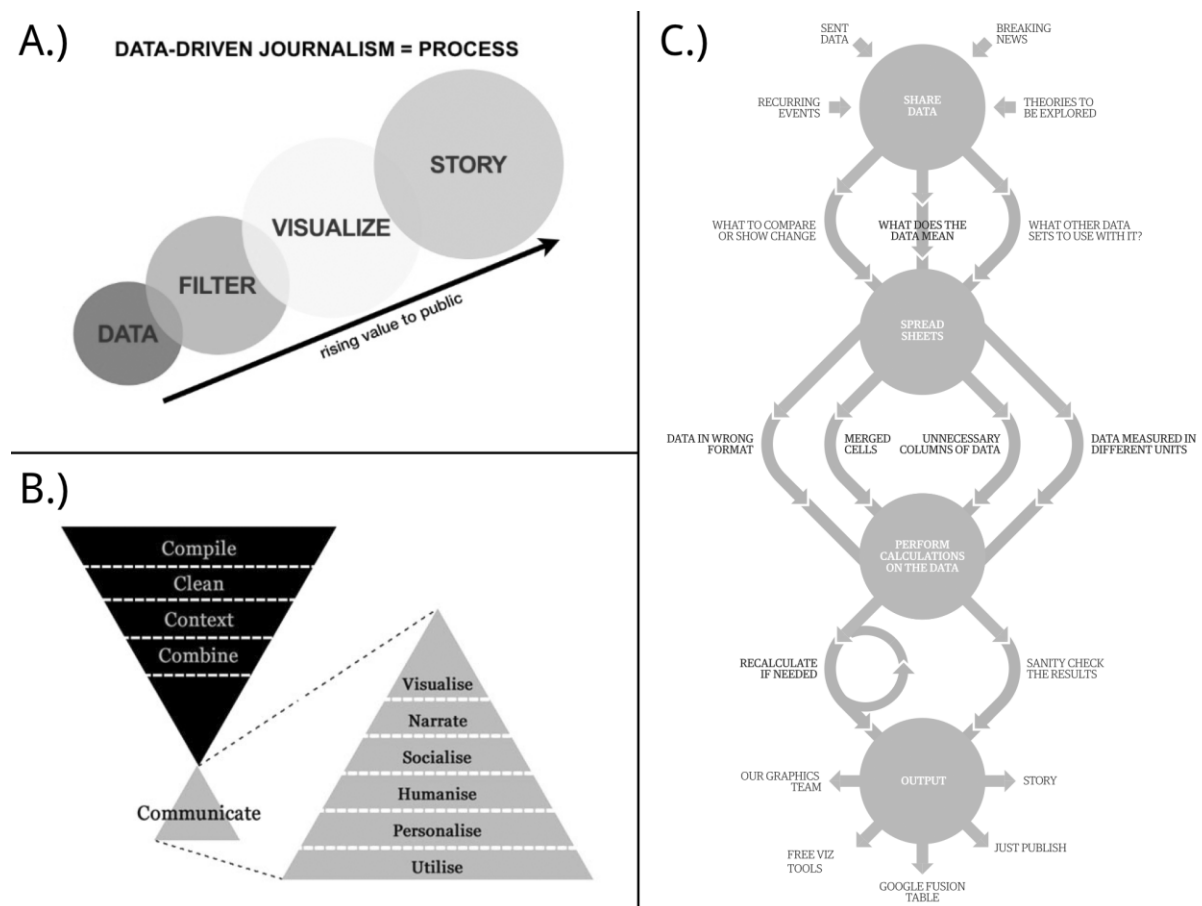


Abbildung 3A. „The data-driven journalism process,“ von M. Lorenz, 2010 (https://en.wikipedia.org/wiki/Data-driven_journalism#/media/File:Data_driven_journalism_process.jpg). CC BY-SA 3.0.

Abbildung 3B. „The inverted pyramid of data journalism,“ von P. Bradshaw, 2011 (<https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>). CC BY-NC 2.0.

Abbildung 3C. „A data journalism workflow,“ von M. McCormick und S. Rogers, 2013 (<https://simonrogers.net/2013/01/27/a-data-journalism-workflow/>). CC BY-NC-SA 3.0.

Abbildung 3A (Lorenz, 2010), beschreibt einerseits die Veränderung, die Daten erfahren, bis sie für die RezipientInnen vorliegen: Daten werden gefiltert, visualisiert und mit einem Narrativ assoziiert. Andererseits thematisiert das Schema den steigenden ‚Wert‘ für die ‚Öffentlichkeit‘: Mit fortschreitender Arbeit an den Daten würde das Gemeinwohl verbessert

bzw. gesteigert werden, weil die Daten gleichsam zugänglicher werden und mehr Menschen die Inhalte hinter den Daten erfassen würden. Abbildung 3B (Bradshaw, 2011) visualisiert den Prozess der Datensammlung und Bearbeitung etwas detaillierter. In der Wortwahl ist durchaus eine Tendenz zur Alliteration festzustellen. Bradshaw teilt den Prozess in eine Produktions- und eine ‚Kommunikations‘-Komponente. Interessant ist ‚Context‘ (im Bereich der Produktion), weil diesen Aspekt nur wenige Modelle berücksichtigen. Der Begriff meint hier die Suche nach Unregelmäßigkeit und Bias in den Daten, aber auch die Erhebung der UrheberInnen und ursprünglichen Zwecke der Daten. Unter ‚Communicate‘ fallen verschiedene Ansätze, Daten erfahrbar zu machen: Visualisierung, Erzählung, ‚Sozialisierung‘ (z.B. Daten zur Verfügung zu stellen und Weiterverarbeitung zu fördern, ‚Crowdsourcing‘), Vermenschlichung (d.h. den Aspekt hinter den Daten anhand von konkreten Beteiligten bzw. Betroffenen darzustellen), Personalisierung (Beiträge an die individuellen Bedürfnisse der RezipientInnen anzupassen) sowie Nutzbarmachung (d.h. Servicejournalismus zu produzieren, der den RezipientInnen nützt). Abbildung 3C, ursprünglich am *Guardian Datablog* veröffentlicht (S. Rogers, 2011), visualisiert die Teilprozesse, die hinter einem datenintensiven Beitrag beim *Guardian* stehen. Das Schema lässt sich jedoch auch auf Praktiken außerhalb des Mediums übertragen. Das Modell konzentriert sich hier auf die Bearbeitungen, die die Daten durchlaufen, und blendet jene Aspekte aus, die für den datenintensiven Journalismus nicht spezifisch sind (vgl. z.B. die Beschreibung von Weinacht und Spiller zuvor, die Aspekte beschreiben, die auch den klassischen Textjournalismus betreffen). Rogers beschreibt darin weiters Fragen, die sich die RedakteurInnen stellen, und Probleme, die bei der Bearbeitung auftreten, etwa falsche Datenformate und -einheiten.

Zusammenfassend kann vorerst festgehalten werden, dass zahlreiche unterschiedliche Prozessdefinitionen und -modelle für datenintensiven Journalismus existieren. Später in dieser Arbeit, im Rahmen der systematischen Literaturanalyse und der Interviewstudie werden die Arbeitsabläufe im datenintensiven Journalismus noch genauer beschrieben werden. Was die Definitionen eint, ist das Verständnis von Daten als Quelle in einer journalistischen Produktion. Der Begriff der Daten ist jedoch, wie Bradshaw (2012) schreibt, ein „troublesome term“, der vielfach zu undifferenziert verwendet wird. Was hat es also mit dem Datenbegriff auf sich und von welchem Datenbegriff ist im datenintensiven Journalismus die Rede? Welche Daten werden im datenintensiven Journalismus verwendet? Das soll im Folgenden geklärt werden.

2.1.1 Zum Datenbegriff im datenintensiven Journalismus

So wie viele andere Begriffe hat der Datenbegriff im Laufe der Zeit zahlreiche Bedeutungsverschiebungen erfahren. Im Zuge der sozial- und geisteswissenschaftlichen Kritik an datenintensiven Verfahren in verschiedenen Kontexten wurde die Entwicklung des Datenbegriffs aufgearbeitet. Der Begriff der ‚Daten‘ kommt ursprünglich aus dem lateinischen ‚dare‘, ‚zu geben‘ (Puschmann & Burgess, 2014). ‚Daten‘, die deutsche Mehrzahl des lateinischen Perfektpartizips (‚datum‘), kann demnach allgemein als ‚die Gegebenen‘ übersetzt werden. Im 18. Jahrhundert veränderte sich das Verständnis des Begriffs ‚data‘, weg von etwas ‚Gegebenem‘ hin zum Resultat eines empirischen Experiments, einer Entdeckung oder einer Sammlung (Rosenberg, 2013, S. 33). Im Laufe des 19. Jahrhunderts war der Begriff nur in wenigen Bereichen präsent und nicht von größerer kultureller Bedeutung – auch nicht im Kontext der großen wissenschaftlichen Errungenschaften dieser Zeit (S. 34). Diese weitestgehende ‚Unversehrtheit‘ mag zu dessen Allgegenwart im 20. Jahrhundert beitragen haben, wie Rosenberg (2013, S. 33) beschreibt:

Ironically, this long period of latency may partly account for the great usefulness of the term in the twentieth century. In the twentieth century, when ‘data’ reached its point of statistical takeoff, it was already a well-established concept, but it remained largely without connotative baggage.

Mit der in der zweiten Hälfte des 20. Jahrhunderts einsetzenden Informatisierung sämtlicher Lebensbereiche wurde der Datenbegriff inflationär. In vielen Fällen, insbesondere in der Management- und Informatikliteratur, tritt der Datenbegriff dabei als ein Drittel des epistemologischen Dreiergespanns – ‚Daten‘, ‚Information‘, ‚Wissen‘ – auf (Gutounig, 2015, S. 57–64). Auch die oben diskutierten Schemata betten den Datenbegriff in dieses Tripel ein. Daten werden darin als prozessierbare Repräsentationen von (Zuständen von) Entitäten betrachtet; Informationen sind Daten, ausgestattet mit Relevanz und Sinn; Wissen ist wertvolle Information (z.B. bei T. H. Davenport, 1997; oder Drucker, 1988). Daten sind also Teil einer Wertschöpfungskette, an dessen Ende Wissen steht. Manche Modelle stellen den Daten noch ‚Zeichen‘ voran oder stellen ‚Wissen‘ als Basis für ‚Können‘, ‚Handeln‘ und ‚Kompetenz‘ dar (z.B. das Modell der Wissenstreppe von North, 2016; Stingl de Vasconcelos Guedes, 2011, S. 213–214).

Diese Hierarchie – Daten zuunterst, Wissen als höchstes Gut – wird auch kritisiert. Willke (1998, S. 7) argumentiert aus einer konstruktivistischen Perspektive heraus, dass ‚es keine Daten an sich gibt, sondern nur beobachtungsabhängige, also qua Beobachtung erzeugte

oder konstruierte Daten“. Besonders wichtige Instrumente der Beobachtung seien Ideen, Konzeptionen, Vorurteile, Ideologien, Theorien, etc., insgesamt die kognitiven ‚Landkarten‘ in den Köpfen der BeobachterInnen. Sie würden bestimmen, was gesehen wird und was nicht. In dieselbe Kerbe schlägt Tuomi (1999, S. 107), der darauf hinweist, dass derartige Hierarchiemodelle eigentlich verkehrt sind und umgedreht werden müssten: „There are no ‚isolated pieces of simple facts‘ unless someone has created them using his or her knowledge. Data can emerge only if a meaning structure, or semantics, is first fixed and then used to represent information.“ Mit anderen Worten: Hinter jeder Generierung von Daten stehen Informationen und Wissen. Daten würden also nicht zu Information nach der Zugabe von Bedeutung werden (S. 108), sondern Informationen sind der Ausgangspunkt, um Daten zu verstehen. Daten sind folglich ein Produkt einer zunehmend an Wissen orientierten Gesellschaft. Im folgenden Kapitel zu den wissenschaftstheoretischen Grundannahmen, insbesondere im Abschnitt zu Ludwik Fleck wird dieser Aspekt zur sozialen Konstruktion von Daten detaillierter ausgeführt werden.

Für die hier verfolgte erste Konzeption des datenintensiven Journalismus als Prozess ist Tuomis Ansatz insofern relevant, als damit klar wird, dass Daten nicht mehr den alleinigen Ausgangspunkt des datenintensiven Journalismus darstellen. Die Ausdrücke ‚datengetrieben‘ bzw. ‚data-driven‘ sind deshalb nicht wirklich adäquat (siehe unten dazu mehr, wenn ‚datenintensiver Journalismus‘ als Hybrid definiert wird).

Zuvor soll jedoch noch die Frage beantwortet werden, woher die Datenquellen für datenintensiven Journalismus kommen. Dies wurde in verschiedenen Inhaltsanalysen erhoben (u.a. Knight, 2015; Loosen u. a., 2016; Tandoc & Oh, 2015). Zu den häufigsten DatenlieferantInnen zählen offizielle Institutionen, (nicht-kommerzielle) Organisationen und Agenturen, redaktionelle Eigenerhebungen, Unternehmen und schließlich Forschungsinstitutionen. Der Anteil an Beiträgen, die selbst generierte Daten verarbeiten, die auf einer Erhebung durch ein Medium basieren, ist relativ gering. Etwa haben rund 20 % der Arbeiten, die bei den internationalen *Data Journalism Awards* zwischen 2013 und 2015 gelistet waren, Daten selbst gesammelt (z.B. mittels einer Umfrage oder Archivrecherche; s. Loosen u. a., 2016, S. 21). Geht man davon aus, dass es sich bei diesen Arbeiten um den „Goldstandard“ des Datenjournalismus“ (S. 21) handelt, kann man annehmen, dass der Anteil an Beiträgen mit eigener Erhebung durchschnittlich noch weit geringer sein wird. Im Umkehrschluss bedeutet dies: Datenintensiver Journalismus ist massiv auf externe DatenlieferantInnen angewiesen.

Loosen, Reimer und De Silva-Schmidt (2016) haben darüber hinaus noch erhoben, welche Art von Daten verarbeitet werden: Am häufigsten wird in der Produktion von datenintensivem Journalismus auf Geodaten, Finanzdaten und (Sensoren-)Messwerte sowie soziodemografische Daten zurückgegriffen (S. 24). Weitere zuweilen herangezogene Datentypen sind personenbezogene Daten, Metadaten, also sowie Befragungsdaten. Gelegentlich bilden auch Daten aus sozialen Medien eine Quelle für datenintensiven Journalismus (Ausserhofer, 2015). Die Daten unterscheiden sich in Format und Zugang: Gängige im Datenjournalismus eingesetzte Formate sind etwa XLS(X)-, CSV- oder TSV-Dateien¹⁸, aber auch komplexere Dateiformate wie JSON¹⁹ oder Dateien nach XML-Spezifikationen²⁰ (Bradshaw, 2013). Für besondere Anwendungsfälle (Geoinformationssysteme, Datenbanken, Netzwerke, etc.) werden andere Datenformate eingesetzt (Boyer, Keefe, Lindenberg, Park & Wu, 2012). Bezogen werden die Daten unter anderem als Download über das offene Web, automatisiert über diverse Schnittstellen (so genannte APIs, siehe Puschmann & Ausserhofer, 2017) sowie durch Scraping. Diese Prozesse und deren Akteure werden sowohl in einem Unterabschnitt der Literaturanalyse als auch in einem Unterabschnitt der Interviewstudie detaillierter erläutert.

Mit der in diesem Abschnitt erfolgten Erläuterung der Prozessmodelle und Darstellung der Ergebnisse von Inhaltsanalysen sollte der Datenbegriff des datenintensiven Journalismus etwas klarer abgesteckt sein. Zwar werden nach wie vor die Omnipräsenz von Daten in unserer Gesellschaft und die zunehmende Durchdringung sämtlicher Lebensbereiche mit Daten als Erklärungskontexte für den datenintensiven Journalismus herangezogen (siehe auch das Einleitungskapitel). Und insbesondere in nichtakademischen Beschreibungen von datenintensivem Journalismus ist immer wieder ein verklärender Datenbegriff zu finden, in dem Daten als *pars-pro-toto* für sämtliche Informations- und Kommunikationstechnologien stehen. Doch das soll der hier skizzierten Vorstellung von Daten keinen Abbruch tun. **Der Ausdruck ‚Daten‘ im Kontext des datenintensiven Journalismus meint**

¹⁸ *CSV* steht für ‚comma-separated values‘, *TSV* für ‚tab-separated values‘: Die Bezeichnungen beziehen sich auf den Aufbau einer Daten beinhaltenden Textdatei, die durch Komma bzw. Tabulator getrennt werden. *CSV* und *TSV* gelten gemeinhin als sehr simples Datenformat.

¹⁹ *JSON*, in Langform *JavaScript Object Notation*, ist ein Datenformat spezialisiert zum Datenaustausch zwischen verschiedenen Anwendungen, unabhängig von der Programmiersprache.

²⁰ *XML* ist Auszeichnungssprache und steht für *Extensible Markup Language*. Mit ihr lassen hierarchisch strukturierte Textdateien darstellen.

(maschinenlesbare) Informationen, die in einer strukturierten Form wie einer Tabelle oder Datenbank organisiert sind und einen Aspekt eines nachrichtenrelevanten Themas beschreiben. Datenintensiven Journalismus zu betreiben bedeutet unter anderem, ortsbezogene, soziodemografische, und andere Daten zu sammeln, zu organisieren und durchsuchbar zu machen. Welche Besonderheiten der auf diesen Daten aufbauende journalistische Beitrag hat, wird im nächsten Abschnitt diskutiert.

2.2 Datenintensiver Journalismus als Produkt

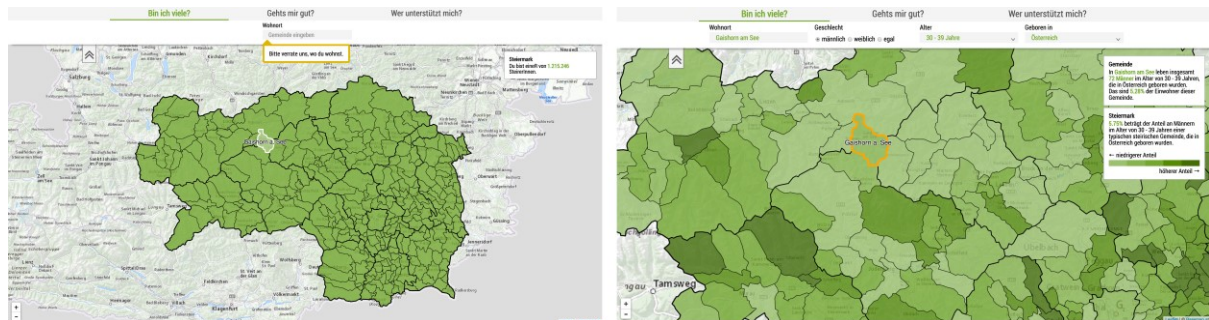
Einige andere Definitionen wiederum beschreiben Datenjournalismus als Produkt bzw. Resultat eines Produktionsprozesses (z.B. bei Baack, 2013; Hullman, Diakopoulos, Momeni & Adar, 2015). Sie betonen dabei die Besonderheit der Nachrichtenbeiträge, die auf Datenanalyse basieren und meist eine Form der interaktiven Visualisierung²¹ wie etwa Karten oder Diagramme beinhalten (Loosen u. a., 2016). Auch wenn es nur selten vorkommt, ist das Vorhandensein einer Visualisierung kein ‚Muss-Kriterium‘ für die Qualifizierung eines Beitrags als ‚datenintensiver Journalismus‘ (Loosen u. a., 2016; Tandoc & Oh, 2015). Es gibt datenintensiven Journalismus auch in Kanälen mit keinen oder wenigen Interaktionsmöglichkeiten wie dem Fernsehen, dem Radio oder Printprodukten (z.B. Siebenhofer & Kanya, 2015). In vielen Fällen wird die Berichterstattung in diesen Kanälen mit interaktiv aufbereiteten Onlinebeiträgen kombiniert.

Mit der Implementierung der interaktiven Visualisierung ist vielfach auch ein Bruch des klassischen linearen Narrativs des Textjournalismus verbunden (Tandoc & Oh, 2015). Ein Beispiel dafür sind „drillable interfaces“ (Karlsen & Stavelin, 2014, S. 40), also Datenvisualisierungen, die je nach Eingabe durch die NutzerInnen verschiedene Ausschnitte zeigen, etwa eine Meta- und eine Mikroperspektive auf ein Phänomen präsentieren (siehe Abbildung 4). Die Inputmöglichkeiten für den/die NutzerIn werden durch die Gestaltung der interaktiven Visualisierung vorgegeben und sind in der Regel stark eingeschränkt. Etwa kann aus einer Liste mit unterschiedlichen Optionen eine Auswahl getroffen werden, über die dann eine Detailsicht geboten wird. In zahlreichen Beiträgen wird diese Funktion dazu genutzt, um

²¹ Bucy (2004, S. 374) folgend wird Interaktivität als Inhaltskonzept verstanden und definiert als „control that users exercise over the selection and presentation of online content, whether story text, audiovisuals, or multimedia, and other aspects of the interface“ (zit. n. Usher, 2016, S. 19).

RezipientInnen die Möglichkeit zu bieten, die Geschichten besser an sich anzupassen, gleichsam zu ‚personalisieren‘. Durch die Auswahl können sich die NutzerInnen selbst innerhalb einer Gruppe verorten (z.B. bei Carter, Cox & Quealy, 2009; Carter, Cox, Quealy & Schoenfeld, 2009). Weil die Auswahlmöglichkeiten für NutzerInnen oft eingeschränkt sind und durch die ProduzentInnen vorformatiert werden, ist das Argument, dass es sich dabei um eine paternalistische, virtuelle Interaktivität handelt (Appelgren, 2017), durchaus nachvollziehbar.

Abbildung 4. Beispiel für eine personalisierbare, interaktive Online-Visualisierung



Die NutzerInnen werden über ein Dialogfenster gebeten, den Wohnort bzw. einen Ort in der Steiermark zu nennen. Anschließend können sie weitere soziodemografische Merkmale (Geschlecht, Alter, Geburtsregion) auswählen. Aufbauend auf diesen Angaben wird den NutzerInnen per Karte und Info-Kasten erläutert, welchen bevölkerungsmäßigen Anteil die gewählte Kohorte in der Gemeinde und der Steiermark hat. Durch die Möglichkeit für NutzerInnen, sich selbst über verschiedene, beliebig kombinierbare Merkmale innerhalb einer gewissen Gruppe zu verorten, wird soziodemografische Statistik leichter erfahrbar (Andrews u. a., 2016, 2015; Gutounig u. a., 2016). Screenshot: Steirische Vielfalt visualisiert, 2017 (<http://www.diversitaet.steiermark.at/steirische-vielfalt/>).

Die Datenvisualisierungen im datenintensiven Journalismus werden in der Regel mit anderen Formaten kombiniert, etwa der Schilderung von Einzelfällen oder Zitaten von individuell Betroffenen und ExpertInnen (Parasie & Dagiral, 2013). Dies passiert sowohl in Textform als auch mittels Fotos, Videos und multimedialer Formate. Oft dienen diese Darstellungsformen der Erläuterung, Kontextualisierung und Illustration der Datenvisualisierung. Die Mehrzahl der Projekte bietet retrospektiv einen Blick auf eine Situation oder einen vergangenen Zeitraum bzw. -punkt, repräsentiert durch statische Datensätze. Es gibt jedoch auch Beiträge, die regelmäßig manuell aktualisiert werden oder automatisiert neue Datensätze einspielen (z.B. Los Angeles Times Data Desk, 2017; behandelt in Young & Hermida, 2015). Die Instandhaltung dieser Projekte stellt Redaktionen vor große Herausforderungen (Appelgren & Nygren, 2014b).

2.3 Zwischenresümee: Datenintensiver Journalismus als Hybrid

Eine große Zahl an Definitionen gibt schließlich an, dass Datenjournalismus beides ist – ein Prozess und ein Produkt (Aitamurto u. a., 2011; Appelgren & Nygren, 2014a; Coddington, 2015; Diakopoulos, 2011; Gynnild, 2014; Hannaford, 2015; A. B. Howard, 2014; Knight, 2015; Loosen u. a., 2016; Lorenz-Meyer, 2011; Radchenko & Sakoyan, 2014; Stavelin, 2013; Tabary u. a., 2016; Usher, 2016). Während jede Beschreibung andere Schwerpunkte setzt, lassen sich doch gemeinsame Elemente festmachen.

Aus den bisherigen Überlegen ergibt sich hiermit folgende Definition:

Die synonymen Begriffe ‚Datenjournalismus‘ und ‚datenintensiver Journalismus‘ bezeichnen (die manuelle Produktion von) Nachrichten, die überwiegend auf der konzentrierten Sammlung, Analyse und Visualisierung von strukturierten Informationen unter Zuhilfenahme von Algorithmen und quantitativen sozialwissenschaftlichen Methoden beruhen.

Die Begriffe ‚Datenjournalismus‘ und ‚datenintensiver Journalismus‘ werden in dieser Arbeit synonym verwendet. Sie fungieren hier als Sammelbezeichnung für verschiedene Manifestationen jenes Phänomens (‚computer-assisted reporting‘, ‚computational journalism‘, etc.), das in diesem Abschnitt abgesteckt wurde. Die teils unterschiedlichen Kontexte der einzelnen Begriffe können in Hinblick auf das artikulierte Forschungsinteresse in den Hintergrund rücken. Wenn die unterschiedlichen Kontexte einen Ausschlag geben, dann wird gesondert darauf hingewiesen werden.

Der Ausdruck ‚(die manuelle Produktion von) Nachrichten‘ in der obigen Definition berücksichtigt, dass, wenn von ‚datenintensivem Journalismus‘ die Rede ist, ein journalistischer Produktionsprozess, ein mediales Produkt oder beides gemeint ist. Damit soll auch der hybriden Natur des Ausdrucks ‚Journalismus‘ Rechnung getragen werden, der sich über verschiedene Dimensionen konstituiert, insbesondere aber über die professionelle Identität von JournalistInnen (Deuze, 2005). ‚Manuell‘ in diesem Kontext zielt auf die Unterscheidung zwischen Datenjournalismus und automatisiertem bzw. algorithmischem Journalismus ab. Während bei ersterem JournalistInnen computerunterstützt Beiträge

produzieren, zielt letzterer auf die massenhafte automatisierte Generierung von (meist) singulären Texten ohne (oder nur mit geringem) menschlichen Beitrag.²²

Die Erläuterung hinsichtlich der Basis von datenintensivem Journalismus („konzentrierte Sammlung, Analyse und Visualisierung von strukturierten Informationen unter Zuhilfenahme von Algorithmen und quantitativen sozialwissenschaftlichen Methoden“) schließlich inkorporiert die Gemeinsamkeiten der verschiedenen hier ausgeführten Definitionen von datenintensivem Journalismus. „Sammlung“, „Analyse“ und „Visualisierung“ von strukturierten Informationen sind Elemente, die so gut wie jede Prozessdefinition von datenintensivem Journalismus nennt. Die letzten beiden Elemente der Definition schließlich sind von Gynnilds (2014, S. 715) Definition der „computational explorations in journalism“ übernommen, die von „the experimental use of algorithms (...) and social science methods in the news media“ spricht.

Die Bezeichnung ‚datenintensiver Journalismus‘ bzw. ‚data-intensive journalism‘ ist dabei eine eigene Wortkreation. Sie entstand aufgrund der Unzufriedenheit mit dem Begriff ‚data-driven‘ bzw. ‚datengetrieben‘. Der Ausdruck ‚data-driven journalism‘ wird seltener im akademischen Kontext verwendet, aber ist doch präsent (z.B. bei Baack, 2013; A. B. Howard, 2014). Im Praxisdiskurs ist er stark prägend, etwa in Form des Hashtags ‚#ddj‘, mit dem auf *Twitter* einschlägige Nachrichten gebündelt werden. Dem Begriff ‚data-driven‘ wohnt eine gewisse technikdeterministische Perspektive inne. Er erweckt den Eindruck, dass Daten den Journalismus ‚an-‘ bzw. ‚vor sich hertreiben‘ würden. Damit impliziert ‚data-driven‘ eine Abkehr vom klassischen epistemologischen Modell, das sich sowohl Sozialwissenschaften und Journalismus zur Grundlage nehmen (Haas, 1999, S. 21–44). Datengetriebene Verfahren können eben nicht gleichzeitig theorie- oder hypothesengeleitet sein (Mahrt, 2015). Wie noch an verschiedenen Stellen in dieser Arbeit herausgearbeitet werden wird, geht man in der Praxis seltener Fälle so vor, dass zuerst Daten vorliegen und dann alles weitere ‚antreiben‘. Am Beginn steht zumindest eine Frage. ‚Datenintensiv‘ charakterisiert die Operationen da schon besser. Der Vorteil am Ausdruck ‚datenintensiv‘ ist darüber hinaus, dass er bislang im Kontext der Journalismusforschung noch völlig ‚unverbraucht‘ ist. Bislang wurde er allein eingesetzt,

²² Dörr (2016, S. 702) definiert algorithmischen Journalismus als „the (semi)-automated process of NLG [natural language generation] by the selection of electronic data from private or public databases (input), the assignment of relevance of pre-selected or non-selected data characteristics, the processing and structuring of the relevant data-sets to a semantic structure (throughput), and the publishing of the final text on an online or offline platform with a certain reach (output)“.

um die Bandbreitenauslastung von Internetdiensten zu beschreiben. Ein alternativer Begriff wäre noch der Ausdruck des ‚datengestützten Journalismus‘ (auf Norwegisch: „Datastøttet journalistikk“, siehe Andersen, 2013).

In der Einleitung und dem an dieser Stelle abzuschließenden definitorischen Kapitel wurde das Feld des datenintensiven Journalismus abgesteckt. Es wurden verschiedene Kontexte der Datafizierung der journalistischen Praxis diskutiert und die Forschungslandschaft zum Thema eingeteilt und beschrieben. Darüber hinaus wurden die Ziele der Arbeit und zentrale Fragestellungen formuliert. Anschließend wurde die Praxis des datenintensiven Journalismus näher beleuchtet. Diese Arbeiten lieferten ein erstes Fundament für das weitere Vorgehen.

Bevor jedoch die durchgeführten empirischen Untersuchungen ausgeführt werden können, bedarf es noch der epistemologischen und methodologischen Fundierung. Diese Schritte sind vor dem Hintergrund der Komplexität der erforschten Materie notwendig.

An verschiedenen Stellen wurde bereits deutlich, dass in quantitativen Verfahren, wie sie im datenintensiven Journalismus häufig vorkommen, vielfach positivistische Perspektiven vorherrschen. Im Mittelpunkt scheinen ‚Fakten‘ zu stehen, die unter anderem durch Daten repräsentiert werden und so etwas wie der ‚heilige Gral‘ des datenintensiven Journalismus sind. Offenbar wird dies etwa an drei in der Praxis populären Titeln: *Facts are sacred: The power of data* (S. Rogers, 2013), *Precision journalism* (P. Meyer, 1973/2002) und *The data journalism handbook: How journalists can use data to improve the news* (Gray, Bounegru & Chambers, 2012). Vielerorts scheint demnach die Annahme zu bestehen, Datenjournalismus sei eine ‚genauere‘, ‚bessere‘ und ‚faktentreuere‘ Spielart des traditionellen Journalismus (kritisiert unter anderem von Heuguet & Langlais, 2014; Lambie, 2001). In diesem Kontext macht es Sinn, sich mit der Genese und Natur von Fakten auseinanderzusetzen und die Bedingungen der Konstruktion von Erkenntnis zu reflektieren. Das soll im folgenden Kapitel zu den wissenschaftstheoretischen Grundannahmen geschehen.

Die Komplexität von Datenjournalismus und dessen Erforschung zeigt sich nicht nur an dessen Faktizität – sie wird auch an dessen Produktionsbedingungen offenbar. Menschliche und nichtmenschliche Akteure aus verschiedensten Sphären tragen dazu bei, dass ein datenintensiver journalistischer Beitrag entsteht. Diese agieren sowohl innerhalb als auch außerhalb einer Redaktion, teils gleichzeitig und teils zeitversetzt. Nur über eine methodologische Reflexion lässt sich ein Forschungsdesign entwickeln, das eine solcherart

komplexe Praxis adäquat erfassen kann. Deshalb diskutiert das übernächste Kapitel die methodologischen Grundlagen dieser Arbeit.

3 Wissenschaftstheoretische Grundannahmen

Dieses Kapitel behandelt das wissenschaftstheoretische Fundament der vorliegenden Arbeit. Diese epistemologische Fundierung ist aus zwei Gründen notwendig: Erstens wird durch sie klar, warum im Verlauf des Texts bestimmte Aspekte der zu erforschenden sozialen Wirklichkeit in den Fokus gerückt, andere wiederum ausgeblendet werden. „Der (...) Theoretisierungsprozess über Bedingungen der Möglichkeit von Erkenntnis klärt das Verhältnis von Reflexion und Forschung“, fasst es Bauer (2014b, S. 79) zusammen. Zweitens liefert diese Fundierung die Grundlage für Dekonstruktion der positivistischen Perspektiven, die vielfach im Datenjournalismus vorherrschen. Daten sind eben nicht Fakten. Warum das so ist, klärt dieses Kapitel.

Dazu bewegt sich der Fokus weiter weg von der Praxis des datenintensiven Journalismus und thematisiert zunächst die Rolle von Theorien im Forschungsprozess. Anschließend werden verschiedene wissenstheoretische Positionen, auf die sich diese Arbeit beruft, diskutiert. Ziel ist dabei nicht die enzyklopädische Wiedergabe der Ansätze, sondern die Absteckung einer Position, die sich aus den verschiedenen Perspektiven konstituiert. Im Zentrum stehen zunächst Ludwik Fleck und Thomas S. Kuhn, deren Perspektiven auf die Genese von Erkenntnis in der Wissenschaft ein fruchtbarer Ausgangspunkt sind. Daran anschließend werden die Erkenntnistheorien von Alfred Schütz sowie von Berger und Luckmann diskutiert, deren Vorstellungen von ‚Lebenswelt‘ beziehungsweise der ‚sozialen Konstruktion von Wirklichkeit‘ wichtige Eckpfeiler für eine konstruktivistische Perspektive bilden. Die epistemologischen Verortungen werden schließlich mit Bruno Latour und Karin Knorr-Cetina abgeschlossen, die die Rolle von nichtmenschlichen Akteuren thematisieren. Das Kapitel schließt mit einem Zwischenresümee.

3.1 Theorie-Theorie und Theorie-Typen

In einem ersten Schritt soll hier das Verständnis über Theorie dargelegt werden, – mit anderen Worten – die leitende „Theorie der Theorien“ (Popper, 1935, S. 26) bzw. Meta-Theorie deutlich gemacht werden, schließlich gibt es in der Geschichte der Wissenschaft unzählige verschiedene Auffassungen des Theoriebegriffs. Wie Krotz (2005, S. 67) herausgearbeitet hat, ist eine „wichtige Gemeinsamkeit aller Theoriekonzeptionen (...), dass sich Theorien oder theorieartige Aussagen in aller Regel nicht darauf beschränken, Aussagen über die

Gegenstände zu machen, anhand derer sie gewonnen werden“. Vielmehr würden sie einen breiteren Geltungsbereich beanspruchen. Auch stellt Krotz klar, dass es sich bei Theorien letztlich immer um Texte handelt, „die sich auf Sachverhalte beziehen, die prinzipiell außerhalb der theoretischen Aussagen existieren, die außerhalb dieser Theorien erlebbar sind und die Bedeutung für jemanden haben“ (S. 68). Damit endet aber schon bald die Schnittmenge verschiedener Auffassungen der Rolle von Theorien.

Lindemann (2011, S. 94) differenziert, aufbauend auf Simmel (1908, S. 16–17), zwischen drei Typen sozialwissenschaftlicher Theorien: Gesellschaftstheorien, Theorien begrenzter Reichweite und Sozialtheorien. Diese klassische Dreiteilung findet sich etwa auch bei Merton (1949/1968), der sich insbesondere auf die Ausformulierung der Theorien mittlerer Reichweite konzentriert. Auch wurden diese Ebenen in Form von makro-, meso- und mikrosozialwissenschaftlicher Forschung operationalisiert. Die drei Theorietypen unterscheiden sich insbesondere durch ihre Nähe zu empirischen Daten. Werden sie für empirische Forschung eingesetzt, haben sie auch einen Einfluss auf die konkrete Fragestellung und die angewandte Methode (Krotz, 2005, S. 69).

Gesellschaftstheorien werden innerhalb der Kommunikationswissenschaft besonders stark in der Journalismusforschung referenziert (Scholl, 2011, S. 16). Sie beschreiben in der Regel breit angelegte historische Formationen der Gesellschaft (Lindemann, 2011, S. 94). In einzelnen Teilen ist der Bezug zur konkreten Empirie noch nachvollziehbar, vielfach ist er jedoch verborgen, nicht vorhanden oder allein zu Illustrationszwecken gegeben (vgl. Krotz, 2005, S. 73–75). Empirische Forschungen, die sich auf Gesellschaftstheorien beziehen, lassen sich vielfach mithilfe von ‚Übersetzungstheorien‘ durchführen. Damit sind Theorien gemeint, die Gesellschaftstheorien für die empirische Forschung operationalisieren – zum Beispiel Essers (1991) Rekonstruktion der Handlungstheorie von Alfred Schütz in eine Rational-Choice-Theorie (zit. n. Scholl, 2011, S. 17).

Theorien begrenzter Reichweite, die zweite Theorie-Kategorie, behandeln spezifische soziale oder kulturelle Phänomene (Lindemann, 2011, S. 94). Merton (1949/1968) skizziert diese Theorien als Kompromiss zwischen Gesellschaftstheorien und Sozialtheorien:

Theories that lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization and social change. (S. 38)

Im Vergleich zu Gesellschaftstheorien lässt sich in Theorien begrenzter Reichweite die Verbindung zur empirisch erfahrbaren (Konstruktion von) Wirklichkeit größtenteils noch nachvollziehen und damit überprüfen. Dieser Typ Theorie entsteht aus „einem methodisch angelegten Prozess der inhaltlichen Auseinandersetzung mit der Wirklichkeit, und zwar vor dem Hintergrund inhaltlicher Fragen und unter Berücksichtigung von Text-Kontext-Bezügen“ (Krotz, 2005, S. 71). Der Gültigkeitsanspruch, den eine solche Theorie begrenzter Reichweite formuliert, erstreckt sich jedoch räumlich wie auch zeitlich über die Untersuchung hinaus.

Sozialtheorien, um in der Dreiteilung und Terminologie Lindemanns bzw. Simmels zu bleiben, beschreiben zeitlich und räumlich (relativ) eng gesteckte Phänomene. Dieser Theorietyp arbeitet die zentralen Konzepte und Kategorien, die Phänomene ausmachen, heraus und trägt damit dazu bei, dass Phänomene erst als ebensolche identifiziert werden können. Die Beschreibungen von Sozialtheorien sind – verglichen mit jenen von Theorien begrenzter Reichweite – dichter und detaillierter angelegt. Die Fokussierung auf Detailstrukturen bringt zwangsläufig auch eine Einschränkung des Geltungsbereichs mit sich. Der Bezug zur Empirie ist im Vergleich zu den zwei anderen Theorietypen noch am stärksten evident.

Alle drei hier kurz skizzierten Theorietypen sind von zentraler Bedeutung für die sozialwissenschaftliche Forschung. Klar ist, dass diese Unterscheidung nicht-taxativ und idealisiert ist. Die wenigsten Ansätze werden sich perfekt in dieses Kategoriensystem einfügen lassen. Dennoch kann es hilfreich sein, diese Unterscheidung zu treffen. Denn die drei Theorietypen sind wie verschieden große Zahnräder in einem Uhrwerk: Sie greifen ineinander und ermöglichen so die Übersetzung von Details zum Großen; also die Übersetzung von der lokalen Empirie hin zu einer generellen Aussage. Ohne Sozialtheorien gibt es keine Theorien mittlerer Reichweite und ohne Theorien mittlerer Reichweite sind Gesellschaftstheorien kaum zu entwickeln. Unzulänglichkeiten in Gesellschaftstheorien wiederum erfordern zunächst Anpassungen in den empirienahen Theorietypen. Simmel (1908) charakterisiert ihr Verhältnis folgendermaßen:

Diese drei Gesichtspunkte verschlingen sich fortwährend, die methodische Notwendigkeit, sie auseinander zu halten, wird immer wieder von der Schwierigkeit, jedes in eine von dem andern unabhängige Reihe zu ordnen, und von der Sehnsucht nach einem, alle Standorte umfassenden Gesamtbilde der Wirklichkeit gekreuzt. (S. 16)

Das von Simmel als „Sehnsucht nach einem (...) Gesamtbilde der Wirklichkeit“ bezeichnete Verlangen der Wissenschaft, Erkenntnis über die Realität bzw. das, was darunter verstanden wird, zu erlangen, haben die Wissenschaftsforschung und andere Disziplinen in den letzten

hundert Jahren stark relativiert. Deren VertreterInnen haben sich unter anderem mit der Frage beschäftigt, unter welchen Bedingungen überhaupt ‚Erkenntnis‘ möglich ist; oder wie ‚Theorien‘, ‚Fakten‘, ‚wissenschaftlicher Fortschritt‘, ‚Wahrnehmung‘ oder ‚Realität‘ entstehen. Die folgenden Abschnitte dieses Kapitels behandeln einige zentrale epistemologische TheoretikerInnen, die sich mit diesen Aspekten auseinandergesetzt haben. Die Argumentation wird dabei einerseits (weitgehend) chronologisch vollzogen, andererseits an Personen orientiert.

3.2 Ludwik Flecks Lehre vom Denkstil und Denkkollektiven

Die Ansätze des Mediziners und Wissenschaftstheoretikers Ludwik Fleck²³ sind ein sinnvoller Ausgangspunkt für mein Unterfangen, gilt er doch mit seinem „bahnbrechenden Buch“ (Latour, 2005, S. 112) *Entstehung und Entwicklung einer wissenschaftlichen Tatsache* als „Begründer der Wissenschaftssoziologie“ (S. 112). Nicht zuletzt hat Fleck diese Aufmerksamkeit Thomas S. Kuhn zu verdanken, der im Vorwort zur *Struktur wissenschaftlicher Revolutionen* beschreibt, dass Fleck viele seiner eigenen Ideen vorweggenommen habe (Kuhn, 1962/1970, S. vii).

Flecks zentraler Beitrag für die Wissenschaftstheorie ist die soziale und historische Kontextualisierung der Entwicklung von wissenschaftlichen Erkenntnissen. Zur Anschauung seines Ansatzes, der im Folgenden in seinen wichtigsten Grundzügen beschrieben wird, führt er unter anderem die Begriffe des ‚Denkstils‘ und des ‚Denkkollektivs‘ ein. Mit Letzterem bezeichnet er „die Gemeinschaft der Menschen, die im Gedankenaustausch oder in gedanklicher Wechselwirkung stehen“ (Fleck, 1935/1980, S. 54). Er bezieht sich damit in erster Linie auf eine „soziale Einheit der Gemeinschaft der Wissenschaftler eines Fachs“ (Schäfer & Schnelle, 1980, S. XXV), aber meint auch andere Entitäten wie etwa die Mitglieder

²³ Fleck hatte ein bemerkenswertes Leben, das Schäfer und Schnelle (1980) recherchiert haben und nicht unerwähnt bleiben soll: Geboren wurde der später als Immunologe ausgebildete Fleck 1896 in Lwów (Lemberg) im heutigen Polen, damals ein Teil von Österreich-Ungarn. Neben seiner medizinischen Forschung im Bereich der Mikrobiologie setzte er sich intensiv mit Wissenschaftstheorie, -philosophie und -geschichte auseinander. Im Zweiten Weltkrieg wurde er von den Nationalsozialisten in die Konzentrationslager Auschwitz und Buchenwald deportiert und zur Herstellung von Typhus-Impfstoff gezwungen. Er selbst, seine Frau und sein Sohn überlebten den Holocaust; alle seine übrigen Familienangehörigen wurden ermordet. Nach dem Krieg führte Fleck seine medizinischen Forschungen bis zu seinem Tod 1961 weiter. Flecks epistemologische Schriften wurden zu seinen Lebzeiten kaum rezipiert und erlangten erst posthum größere Bekanntheit.

einer Religionsgemeinde, von Parteien oder eines gesellschaftlichen Standes (Fleck, 1935/1980, S. 61). „Ein Denkkollektiv ist immer dann vorhanden, wenn zwei oder mehrere Menschen Gedanken austauschen“ (Fleck, 1935/1980, S. 60). Jedes Individuum gehört vielen unterschiedlichen sich überlappenden Denkkollektiven unterschiedlicher Größen an. Während manche nur von kurzer Dauer sind und wieder verschwinden, stabilisieren sich andere und bestehen längerfristig (Fleck, 1935/1980, S. 135). Soziale Rituale wie etwa Bildung ermöglichen die individuelle Initiation und sichern die Beständigkeit eines Denkkollektivs. Zwischen den einzelnen Individuen herrscht ein solidarisches Zusammengehörigkeitsgefühl. Jedes Denkkollektiv formiert sich um ein Denkgebilde – etwa eine wissenschaftliche Idee, einen künstlerischen Gedanken oder ein Glaubensdogma –, das ins Zentrum des Kollektivs rückt. Der innere Zirkel des Kollektivs, bei Fleck genannt ‚der esoterische Kreis‘, besteht aus wenigen eingeweihten Individuen. Sie vermitteln die Inhalte des Denkgebildes an den äußeren, den exoterischen Kreis, der sich aus einer größeren Anzahl an Individuen zusammensetzt. Zugleich sind die „Eingeweihten (...) bewußt oder unbewußt von der ‚öffentlichen Meinung‘, d.h. der Meinung des exoterischen Kreises abhängig“ (Fleck, 1935/1980, S. 139).

Für die nun zu erarbeitende epistemologische Position dieser Arbeit ist insbesondere Flecks Idee des Denkstils maßgeblich. Die Individuen, die einem Denkkollektiv angehören, teilen ein „gerichtetes Wahrnehmen, mit entsprechendem gedanklichen und sachlichen Verarbeiten des Wahrgenommenen“ (Fleck, 1935/1980, S. 130). Erkenntnis ist laut Fleck nicht individuell determiniert, sondern vielmehr vom Kollektiv abhängig. Der Denkstil wird durch intra- und interkollektive Kommunikation laufend angepasst, wobei ein Ausbrechen aus dem Denkstil innerhalb des Kollektivs unmöglich ist. Der Denkstil ist zwanghaft, wie Fleck (1935/1980, S. 56-57) beschreibt:

Obwohl das Denkkollektiv aus Individuen besteht, ist es nicht deren einfache Summe. Das Individuum hat nie, oder fast nie das Bewußtsein des kollektiven Denkstiles, der fast immer einen unbedingten Zwang auf sein Denken ausübt und gegen den ein Widerspruch undenkbar ist.

Für die wissenschaftliche Arbeit im Allgemeinen und die Theoriebildung im Speziellen bringt dieser Ansatz wichtige Implikationen mit sich. Fleck korrigiert das in der Wissenschaftsgeschichte über lange Zeit hinweg hochgehaltene Ideal des genialen (zumeist männlichen) Individuums, das sich allein für eine Erkenntnis verantwortlich zeichnet. Schließlich können Forschung und Entwicklung nur arbeitsteilig und aufeinander aufbauend geschehen:

Gedanken kreisen vom Individuum zum Individuum, jedesmal etwas umgeformt, denn andere Individuen knüpfen andere Assoziationen an sie an. Streng genommen versteht der Empfänger den Gedanken nie vollkommen in dieser Weise, wie ihn der Sender verstanden haben wollte. Nach einer Reihe solcher Wanderungen ist praktisch nichts mehr vom ursprünglichen Inhalte vorhanden. Wessen Gedanke ist es, der weiter kreist? Ein Kollektivgedanke eben, einer der keinem Individuum angehört. (Fleck, 1935/1980, S. 58)

Fleck rückt den Prozess des Scheiterns von Kommunikation, von sozialer Verständigung, in den Mittelpunkt des Erkenntnisprozesses. Wenn etwa Beschreibungen von Experimenten unter WissenschaftlerInnen zirkulieren, dann entstehen neue Erkenntnisse aus Missverständnissen bei der Rezeption dieser Beschreibungen, aus Zurückweisungen und Überarbeitungen des Publizierten, und aus der Assoziation mit anderen Denkkollektiven (Sady, 2012).

Neben einer sozialen Dimension durch Verständigung betont Fleck die historische Komponente von Erkenntnisgewinn, den historischen Charakter von Wissen:

A truly isolated investigator is impossible, and so also is an ahistoric discovery, or a styleless observation. An isolated investigator without bias and tradition, without forces of mental society acting upon him, and without the effect of the evolution of that society, would be blind and senseless. Thinking is a collective activity, just as choral singing or conversation. It is subjected to specific changes in time, and displays a historic continuity of these changes. (Fleck, 1935/1986, S. 77)

Welche Auswirkungen hat der dargelegte Ansatz Flecks auf die in diesem Text vertretene Auffassung von Theorien? Theorien werden hier in einem moderat-sozialkonstruktivistischen Sinn als ein Gefüge von zusammenhängenden, einander nicht widersprechenden Beschreibungen über Ausschnitte der kollektiv erfahrbaren Welt verstanden. Die Poppersche Metapher vom „Netz, das wir auswerfen, um ‚die Welt‘ einzufangen, – sie zu rationalisieren, zu erklären und zu beherrschen“ (Popper, 1935, S. 26) ist dabei durchaus hilfreich – auch wenn Popper alles andere als ein Sozialkonstruktivist ist. Jedoch ist die Entwicklung von Theorien im besonderem Maße abhängig von Denkkollektiven und den damit verbundenen Denkstilen, in die die VerfasserInnen von Theorien eingebettet sind. Soziale Verständigung und historische Entwicklung der Kollektive sind *Conditio sine qua non* für Theorien. Insofern ist die Urheberschaft von Theorien auch nie allein den VerfasserInnen zuzuschreiben. Durch ihre Entwicklung wiederum co-konstituieren Theorien die erfahrbare Welt und (damit) auch ihren Denkstil und ihr Kollektiv. Theorien sind also nicht als ein neutrales, über der erfahrbaren Welt schwebendes Gebilde zu sehen, sondern ein integrierter Teil eines sozialen Kommunikationsprozesses in der Wissenschaft.

Heute gilt Fleck als ein „Klassiker der Epistemologie und historischen Wissenschaftssoziologie“ (R. S. Cohen & Schnelle, 1986, S. ix) – wenn auch als ein (zu) spät Entdecker. Schließlich sollte es, selbst nachdem Thomas Kuhn ihn 1961 in seinem Vorwort zu *Die Struktur wissenschaftlicher Revolutionen* genannt hatte, noch bis Anfang der 1980er Jahre dauern, bis Fleck wieder stärker rezipiert wurde. Flecks Fokus auf den von kollektiver Verständigung abhängigen Charakter von (wissenschaftlichen) Erkenntnisprozessen legte den Grundstein für darauf aufbauende, wichtige Theoriegebäude, etwa Knorr-Cetinas (1999) epistemische Kulturen oder das Konzept der Wissensregime (vgl. Schützeichel, 2012, S. 23; Wehling, 2007).

3.3 Thomas Kuhns Struktur wissenschaftlicher Revolutionen

Thomas Kuhn entwickelt die Konzepte Flecks bedeutend weiter. Die Ausdrücke, die er in *Die Struktur wissenschaftlicher Revolutionen* (mit)prägte – ‚Paradigma‘ und ‚Paradigmenwechsel‘, ‚Inkommensurabilität‘ und ‚wissenschaftliche Revolution‘ – sind mittlerweile auch außerhalb wissenschaftlicher Diskurse weit verbreitet. Die Veröffentlichung des Buchs, das zu den meistzitierten wissenschaftlichen Büchern überhaupt zählt, löste ähnliche Prozesse aus wie sie Kuhn in seinem Text beschrieben hatte. Für die nun folgende Darstellung von Kuhns Ansatz stütze ich mich neben der deutschsprachigen Übersetzung der zweiten Auflage (Kuhn, 1969/1976) insbesondere auf die Aufarbeitungen und Vorstrukturierungen von Kuhns Werk durch Hoyningen-Huene (1989) sowie Hoyningen-Huene mit Lohse (Hoyningen-Huene & Lohse, 2012).

Auf Grundlage historiografischer Studien entwirft Kuhn ein zyklisches Phasenmodell für wissenschaftliche Entwicklung. In diesem Modell spielen Theorien eine zentrale Rolle. Am Anfang steht die „Proto-Wissenschaft“ (Kuhn, 1970) oder „vornormale Wissenschaft“ (Hoyningen-Huene, 1989, S. 185): ein Stadium, in dem es keinen Konsens bezüglich des Fachgebiets gibt und die Forschung „im Vergleich zu einem Forschungsmodus, in dem ein solcher Konsens besteht, viel oberflächlicher, diffuser, spekulativer, un schlüssiger in der Wahl des Wesentlichen, weniger zielgerichtet, leichter von außen beeinflussbar und stärker mit ihren eigenen Grundlagen und konkurrierenden Ansätzen beschäftigt [ist]“ (Hoyningen-Huene & Lohse, 2012, S. 76). Es existieren unterschiedliche Schulen, die untereinander um eine Vormachtstellung konkurrieren und kaum voneinander abgeschottet sind. Sie bearbeiten ähnliche Themen unter Zuhilfenahme verschiedener Theorien.

Schafft eine dieser Schulen eine bedeutende Entwicklung in Grundlagenfragen und ist in der Lage, die konkurrierenden Schulen von der Qualität und Abschlussfähigkeit des Ansatzes zu überzeugen, dann wird die Phase der vornormalen Wissenschaft von jener der ‚normalen Wissenschaft‘ abgelöst. Von zentraler Bedeutung ist in diesem Prozess, dass die Schule mit der Rekrutierung neuer Mitglieder, insbesondere nachfolgender Wissenschaftlergenerationen, erfolgreich ist und sie sich gegen die anderen Schulen durchsetzen kann. Der Forschungsprozess in dieser Phase läuft routiniert ab. „Bestehende Theorien werden lange Zeit unhinterfragt verwendet, (...) weil man von der prinzipiellen Fruchtbarkeit des Ansatzes überzeugt ist“ (Joas & Knöbl, 2004, S. 30). Die Entität, die die Forschenden verbindet, nennt Kuhn das „Paradigma“ – und meint damit mehr als nur einen gemeinsamen Theoriestrang:

Wenn der Wissenschaftler ein Paradigma erlernt, erwirbt er sich Theorien, Methoden und Normen, gewöhnlich in einer unentwirrbaren Mischung. Wenn Paradigmata wechseln, gibt es deshalb normalerweise bezeichnende Verschiebungen der Kriterien, welche die Zulässigkeit von Problemen und den sich anbietenden Lösungen bestimmen. (Kuhn, 1969/1976, S. 122)

Im Kern des Paradigmas ist der Konsens der Forschenden. Gemeint ist damit die „Forschung, die fest auf einer oder mehreren wissenschaftlichen Leistungen der Vergangenheit beruht, [und] Leistungen, die von einer bestimmten wissenschaftlichen Gemeinschaft (...) als Grundlagen für ihre weitere Arbeit anerkannt werden“ (Kuhn, 1969/1976, S. 25). Kuhn vergleicht die Forschung innerhalb eines Paradigmas mit dem Lösen von Rätseln: Innerhalb der vorgegebenen Grenzen, die auch nicht infrage gestellt werden, werden Probleme und Fragen aufgeworfen und Lösungen gesucht. „Der Wissenschaftler [ist] weniger durch das Ergebnis seiner Arbeit selbst motiviert als vielmehr dadurch, seine Fähigkeiten zur produktiven Arbeit unter Beweis zu stellen“ (Hoyningen-Huene & Lohse, 2012, S. 76).

Im Rahmen des Forschungsprozesses treten immer wieder unerwartete Ergebnisse und neue Entdeckungen auf, die nicht in die Vorstellungen des vorherrschenden Paradigmas passen. Kuhn nennt diese Phänomene ‚Anomalien‘. Diese können beiseite geschoben oder durch Hilfhypothesen notdürftig in das Theoriegebäude integriert werden. In diesem Fall nimmt die Theoriekomplexität jedoch zu und kann einen Punkt erreichen, an dem der Entwurf nicht mehr praktikabel ist. Oder: Andere Theorien werden als Beschreibungsvehikel herangezogen, um das Phänomen zu klären. In diesem Moment beginnt der Zustand der Krise – auch genannt: ‚außerordentliche Wissenschaft‘.

Die Krise ist gekennzeichnet von „offen geäußelter Unzufriedenheit mit dem Funktionieren der bisher leitenden Theorie“ (Hoyningen-Huene, 1989, S. 228) und Unsicherheit innerhalb der Forschungsgemeinschaft. Das Paradigma wird hinterfragt. In der Forschungsarbeit kommen zwar die Grundregeln des Paradigmas weiter zur Anwendung, sie werden jedoch regelmäßig aufgeweicht. Es werden häufiger Experimente durchgeführt, „ohne daß eine genaue Erwartung hinsichtlich ihrer Ergebnisse vorhanden ist“ (Hoyningen-Huene, 1989, S. 228). Wenn insbesondere junge Forschende in großer Zahl die alte Theorie infrage stellen und einem anderen Ansatz bessere Problemlösungskapazität attestieren, kann das den Grundstein für die nächste Phase, die ‚wissenschaftliche Revolution‘ darstellen. Persönliche Dispositionen spielen bei der Theoriewahl eine wesentliche Rolle:

Aber die Krise allein ist nicht genug. Es muss auch eine Basis des Vertrauens zu dem gewählten Kandidaten vorhanden sein, wenn sie auch nicht rational oder endgültig richtig zu sein braucht. Er muß wenigstens einigen Wissenschaftlern das Gefühl geben, daß der neue Gedanke auf dem richtigen Wege ist, und manchmal sind es nur persönliche und unartikulierte ästhetische Erwägungen, die das können. (Kuhn, 1969/1976, S. 168)

Die individuelle Entscheidung für eine Theorie, ein neues Paradigma, fußt also bei den wenigsten ForscherInnen auf rein rationalen Fundamenten. Jedoch, so beschreibt es Kuhn, steht hinter der Gesamtentscheidung des Kollektivs ein tiefgründigeres Abwägen, das insbesondere auf der Problemlösungskompetenz der neuen Theorie basiert. Mit der Umwälzung, dem ‚Paradigmenwechsel‘, sind jedoch massive Veränderungen verbunden (vgl. Hoyningen-Huene & Lohse, 2012, S. 80): Erstens verschiebt sich die Perspektive auf den Untersuchungsgegenstand – „sowohl der Bereich der notwendig als auch der Bereich der legitim zu bearbeitenden Probleme“ ist davon betroffen (Hoyningen-Huene & Lohse, 2012, S. 80). Zweitens erfährt die Sprache, insbesondere die Bezeichnungen der untersuchten Objekte, eine Bedeutungsverlagerung; und drittens – die beträchtlichste Veränderung – führt der Paradigmenwechsel zu einem Wandel der erfahrbaren Erscheinungswelt. Hier treten die konstruktivistischen Ideen in Kuhns Theorie deutlich hervor: Denn es gibt eine rein „objektseitige Welt, die unseren Erkenntnisbemühungen nicht zugänglich ist“ und eine Erscheinungswelt, die sich „auch durch Beiträge der Erkenntnissubjekte konstituiert“ (Hoyningen-Huene & Lohse, 2012, S. 80). Neue Theoriekonstrukte würden eben auch die erfahrbare Erscheinungswelt ändern, weil sich mit ihnen auch die Wahrnehmung verändern würde (vgl. Hanson, 1958).

Ist der Paradigmenwechsel vollzogen, dann tritt die Forschungsgemeinschaft wieder in eine ‚normale Phase‘ ein. Kuhn argumentiert aus den oben genannten Gründen, dass zwei

sukzessive Paradigmen nicht bzw. kaum miteinander verglichen werden können. Er führt dazu den Begriff der ‚Inkommensurabilität‘ ein (Kuhn, 1969/1976, S. 116), der ihm von verschiedenen Seiten den Vorwurf des Relativismus einbrachte (z.B. Fodor, 1984; Weinberg, 1998). Jedoch sei dieser nicht aufrechtzuerhalten, wie Hoyningen-Huene (1989, S. 212–217) argumentiert, schließlich würde Kuhn zwar eine direkte Gegenüberstellung von Theorien aus verschiedenen Paradigmen (aus den erwähnten Gründen) als nicht zielführend erachten; jedoch könne man, so der Ansatz Kuhns, Theorien sehr wohl hinsichtlich ihrer empirischen Leistungsfähigkeit miteinander in Beziehung setzen.

Kuhn illustriert seine Theorie ausschließlich mit historischen Beispielen aus der Naturwissenschaft, insbesondere aus Physik und Chemie. Der Hang zur narrativen Historisierung der epistemischen Kontexte und die Idee von Wissenschaft als Wettbewerb sind seinem Konzept nicht abzusprechen. Sein Modell der zyklischen Entwicklung von Wissenschaft wurde vielfach zur Beschreibung der vermeintlich ‚weichen‘ Wissenschaften verwendet – wenn auch zuweilen ungenügend und unangebracht (vgl. Hoyningen-Huene & Lohse, 2012, S. 82). Nichtsdestotrotz ist sein Phasenmodell (vornormale Wissenschaft, normale Wissenschaft, Anomalie, Krise, Revolution, normale Wissenschaft, ...) für die in diesem Abschnitt verfolgte Darlegung des Theorieverständnisses meiner Arbeit eine fruchtbare Perspektive. Kuhns Schema des wissenschaftlichen Entwicklungszyklus, in dem die leitende Theorie eines Paradigmas in unterschiedlichen Phasen unterschiedliche Rollen einnimmt, ist durchaus ergiebig. Schließlich haben auch gewisse sozialwissenschaftliche Paradigmen oder Theorien der Journalismusforschung zu gewissen Zeiten Konjunktur und werden zu anderen Zeiten als ungenügend abgetan (Steensen & Ahva, 2015). Die Gründe für den Wandel sind dabei in den seltensten Fällen rein rationaler Natur. Vielmehr spielen einerseits individuelle Dispositionen wie etwa Alter, ästhetisches Empfinden und Karriere eine zentrale Rolle, andererseits erfüllt die Gemeinschaft der WissenschaftlerInnen einen wichtigen Part. Darüber hinaus besteht die Auffassung, dass sich mit der zunehmenden Mediatisierung von sozial- und geisteswissenschaftlicher Forschung ein Kuhnscher Paradigmenwechsel vollzieht (Rieder & Röhle, 2012). Umgelegt auf die journalistische Praxis kann das Aufkommen des datenintensiven Journalismus auch als ein Teil eines Paradigmenwechsels im Journalismus gedeutet werden.

Die Parallelen von Kuhns Ansatz zu Flecks Werk sind bei der Erläuterung der Rolle der wissenschaftlichen Gemeinschaft besonders evident. Im Grunde stellt sich in Hinblick auf die Theoriewahl und -entwicklung die Frage: Wird die Theorie von den FachkollegInnen

akzeptiert und sind sie bereit, sie mitzutragen? Gerade die kollektiven interdependenten Entscheidungen entscheiden über Erfolg oder Misserfolg einer Theorie. Wie jedoch Löffelholz und Rothenberger (2016, S. 13) anmerken, verdiene „jede theoretische Bemühung Beachtung – sei es, indem sie die Erkenntnis innerhalb eines bestehenden Paradigmas kumulativ erweitert; sei es, indem sie Erkenntnis außerhalb eines bestehenden Paradigmas ermöglicht und damit zu dessen Ablösung beiträgt“. Die Paradigma-Metapher hilft insbesondere beim Verständnis, dass Theorien Teil der zu erforschenden Welt sind und immer auch die Sicht auf das Forschungsfeld verengen. Kuhn vergleicht ein Paradigma mit einer Landkarte, die

dem Wissenschaftler sagt, welche Entitäten es in der Natur gibt und welche nicht, und wie sie sich verhalten. Durch diese Informationen entsteht eine Landkarte, deren Einzelheiten durch reife wissenschaftliche Forschung aufgeheilt werden. Und da die Natur viel zu komplex und vielfältig ist, um auf gut Glück erforscht zu werden, ist diese Landkarte genauso wichtig für die kontinuierliche Weiterentwicklung der Wissenschaft wie Beobachtung und Experiment. (Kuhn, 1969/1976, S. 121)

Theorien sind somit gleichsam die Projektionsmodelle einer solchen Landkarte, indem sie die Strukturen der abzubildenden Welt formatieren (vgl. Carey, 2009; zit. n. Krotz, 2005, S. 69).

3.4 Alfred Schütz' Lebensweltkonzept

Wenn bisher, bei Kuhn und Fleck, der Erkenntnisprozess und die Theoriegenese im wissenschaftlichen Kontext im Mittelpunkt standen, dann ist es nun an der Zeit, einen Schritt zurückzutreten und die Bedingungen menschlicher Erfahrung zu thematisieren. Schließlich sind Wissenschaft (und Journalismus) zwar spezielle Teilbereiche einer Gesellschaft mit eigenen Regeln und Strukturierungen, jedoch auch den menschlichen Prämissen der Erkenntnis und Wissensaneignung unterworfen. John Law (1992, S. 381) – wie auch viele andere – argumentiert:

So much for science. But I have already suggested that science isn't very special. Thus what is true for science is also said to be true for other institutions. Accordingly, the family, the organization, computing systems, the economy and technologies – all of social life – may be similarly pictured.

Und auch Friedrich Krotz (2005) weist darauf hin, dass wissenschaftliche Methoden nur Ausdifferenzierungen, Weiterentwicklungen und Standardisierungen von Verfahren sind, die im Alltag angewendet werden; jedoch:

Wissenschaft überprüft ferner ihr Vorgehen und ihre Theorien immer wieder, setzt neues Wissen in Bezug zu bereits vorhandenem Wissen und gibt sich auf Dauer nicht mit Antworten auf Teilfragen zufrieden, sondern versucht möglichst große Wirklichkeitsbereiche deskriptiv und theoretisch zu erfassen. (S. 90)

Die sozialphänomenologische Wissenssoziologie von Alfred Schütz ist ein willkommener erster Referenzrahmen, schließlich hat sich Schütz intensiv mit der Frage auseinandergesetzt, wie ein Individuum die Wirklichkeit erfährt. Auch für Schütz stellt sich die Wissenschaft als ein Sinnbereich neben vielen anderen in der gesellschaftlichen Realität dar (Schütz, 1953/1971, S. 5). Im Zentrum seiner Abhandlungen steht jedoch weniger die Wissenschafts- als vielmehr die Alltagswelt, „jener Wirklichkeitsbereich, an der der Mensch in unausweichlicher, regelmäßiger Wiederkehr teilnimmt“; es ist jener Teil der Wirklichkeit, „den der wache und normale Erwachsene in der Einstellung des gesunden Menschenverstandes als schlicht vorgegeben vorfindet“ (Schütz & Luckmann, 1973/2003, S. 29). Schütz übernimmt den Begriff der ‚Lebenswelt‘ von Edmund Husserl, in dessen Phänomenologie dieser Ausdruck als „die raumzeitliche Welt der Dinge, so wie wir sie in unserem vor- und außerwissenschaftlichen Leben erfahren“ (Husserl, 1936/1976, S. 141) definiert ist. Neben Husserl ist Henri Bergson ein zentraler Ankerpunkt im frühen Werk von Schütz. Abels (2007, S. 64) fasst das Lebensweltkonzept von Schütz treffend zusammen: „An ihr nimmt der Mensch regelmäßig und unausweichlich teil. Es ist seine Welt, in der er sich immer schon befindet, und zugleich die Welt, in der er immer mit den anderen gemeinsam lebt.“ Schütz’ ursprünglicher Impetus ist es, Max Webers Konzept von Sinn, das er, Schütz, ursächlich mit seiner Theorie des sozialen Handelns verknüpft, auszubauen. Sein Anspruch ist kein geringer, will er doch mit diesem Vorhaben die „Wurzeln der sozialwissenschaftlichen Problematik bis zu den fundamentalen Tatsachen des Bewußtseinslebens“ zurückverfolgen (Schütz, 1932, S. iii).

Schütz beschreibt einen reziproken, interdependenten Erkenntnisprozess: Er legt dar, wie ein Individuum in seinem Sinn die Außenwelt konstruiert und wie wiederum diese äußere, physische Welt den Sinn eines Individuums konstituiert. Schütz unterscheidet folglich zwischen einer inneren Realität und einer äußeren Wirklichkeit (bzw. vielen verschiedenen äußeren Wirklichkeiten, vgl. Schütz, 1945/1971). Ist eine Person mit einer neuen Situation konfrontiert, so reflektiert sie diese und entwickelt in weiterer Folge Erfahrungen, die sie als Lösungs- bzw. Deutungsmuster für zukünftige Interaktionen heranzieht. Diese Erfahrungen werden abgelegt und gemeinsam mit den durch Dritte vermittelten Erfahrungen in einem Wissensvorrat zusammengefasst (Schütz & Luckmann, 1973/2003, S. 33, S. 331–443). Schütz verwendet für die Herausbildung und Festigung des Wissensvorrats die Metapher der

„Sedimentierung“. Der Wissensvorrat ist bereits vorhanden, wenn eine Person geboren wird. Das Individuum wächst damit auf und kann diesem „nicht entgehen. Insofern setzt die Wirklichkeit des Alltags ihm auch einen Rahmen des Denkens und Handelns“ (Abels, 2007, S. 70).

Für den Erkenntnisprozess sind in Schütz' Ansatz ‚Typisierungen‘ notwendig. Dabei handelt es sich um abstrahierte Erfahrungen, die das Individuum in einer ihm bisher unbekanntem Situation zum Vergleich heranzieht. Die Assoziation einer Typisierung mit einer neuen Gegebenheit kann als Herstellung eines Sinnzusammenhangs verstanden werden. Klar ist, dass Sprache in diesem Prozess eine zentrale Rolle spielt:

Der weitaus größte Bereich lebensweltlicher Typisierungen ist sprachlich objektiviert. Das, was für den einzelnen typisch relevant ist, war meist schon für seine Vorgänger typisch relevant und hat folglich in der Sprache semantische Entsprechungen abgelagert. Kurzum, die Sprache kann als die Sedimentierung typischer Erfahrungsschemata, die in einer Gesellschaft typisch relevant sind, aufgefaßt werden. Der Bedeutungswandel der Sprache kann folglich als eine Folge von Veränderungen in der sozialen Relevanz gegebener Erfahrungsschemata betrachtet werden. (Schütz & Luckmann, 1973/2003, S. 319)

Die solcherart konstruierten Alltagsrealitäten sind von zentraler Wichtigkeit. Denn nur innerhalb dieser kann sich das Individuum mit seinen Mitmenschen austauschen – Voraussetzung für eine intersubjektive Grundstruktur der Lebenswelt. Neben der Lebenswelt des Alltags gibt es in den Auffassungen Schütz' noch eine Reihe weiterer anderer so genannter ‚geschlossener Sinnbereiche‘, die allein dem Individuum zugänglich sind. Das sind etwa Phantasie, Träume oder wissenschaftliche Theoriekomplexe. Sie zeichnen sich unter anderem durch eine „spezifische Bewußtseinsspannung, (...) eine vorherrschende Form der Selbsterfahrung, (...) [und] eine spezifische Zeitperspektive“ aus (Schütz, 1945/1971, S. 265). Neben der intersubjektiven alltäglichen Lebenswelt prägen diese Bereiche das Bewusstsein eines Individuums fundamental.

Auf diesen Thesen aufbauend argumentierte Schütz in weiterer Folge für einen Perspektivenwechsel im sozialwissenschaftlichen Forschungsprogramm: „Die Wissenschaften, die menschliches Handeln und Denken deuten und erklären wollen, müssen mit einer Beschreibung der Grundstrukturen der vorwissenschaftlichen, für den – in der natürlichen Einstellung verharrenden – Menschen selbstverständlichen Wirklichkeit beginnen“ (Schütz & Luckmann, 1973/2003, S. 29). Das Verständnis für die Lebenswelt der Menschen müsse zunächst mithilfe qualitativer Methoden erlangt werden. Denn, bevor man den Raum

der sozialen Realität nicht gründlich ausgelotet habe, könne man nicht sicher sein, ob die Methoden der Naturwissenschaften diesen Raum überhaupt angemessen beschreiben würden (Barber, 2014). Schütz' (1960/1972) sozialphänomenologisches Forschungsprogramm richtet sich unter anderem gegen den Behaviorismus und die quantitativen Wirtschaftswissenschaften, deren sogenannte Objektivierungen die Perspektive der Akteure vernachlässigen würden. Schütz spricht sogar vom „vergessenen Menschen“ der Sozialwissenschaften“. Dieser Stelle soll aufgrund des programmatischen Charakters für die vorliegende Arbeit etwas mehr Platz eingeräumt werden.

Was geht uns das alles an, uns, wissenschaftliche Beobachter? Wir können ein raffiniertes Abstraktionssystem entwickeln und für diesen Zweck anwenden, das mit Absicht den Handelnden in der sozialen Welt draußen läßt, und zwar mit allen seinen subjektiven Gesichtspunkten, und wir können dies sogar tun, ohne daß wir mit den Erfahrungen in Konflikt geraten, die wir aus der sozialen Wirklichkeit gewonnen haben. Meister dieser Technik werden sich immer davor hüten, den Boden zu verlassen, auf dem man diese Technik anwenden kann, und sie werden deshalb ihre Probleme entsprechend eingrenzen – sie sind auf allen Gebieten der sozialwissenschaftlichen Forschung zahlreich.

Dies alles ändert aber nichts an der Tatsache, daß dieser Typ der Sozialwissenschaften sich nicht direkt und unmittelbar mit der sozialen Lebenswelt befaßt, die uns allen gemeinsam ist, sondern mit geschickt ausgewählten Idealisierungen und Formalisierungen der sozialen Welt, soweit diese ihren Tatsachen nicht widersprechen. Ebenso wenig bezieht sich dieser Typ der Sozialwissenschaften auf den subjektiven Standpunkt auf anderen Abstraktionsebenen, wenn sich das ursprünglich betrachtete Problem ändert. Jedoch – und das ist hier wichtig – kann dieser Bezug auf den subjektiven Standpunkt immer hergestellt werden und sollte es auch. Da die soziale Welt unter jedem beliebigen Aspekt stets ein komplizierter Kosmos von menschlichen Tätigkeiten bleibt, können wir stets zum ‚vergessenen Menschen‘ der Sozialwissenschaften zurückkehren, zum Handelnden in der sozialen Welt, dessen Tun und Fühlen dem ganzen System zugrunde liegt. Wir können dann sein Tun und Fühlen versuchen zu verstehen und den Bewußtseinszustand, der ihn dazu brachte, die jeweiligen Einstellungen gegenüber seiner sozialen Umwelt anzunehmen. (Schütz, 1960/1972, S. 7)

Die Sozialwissenschaft hat sich also der ‚Re-Konstruktion‘ der individuellen Lebenswelt-Konstruktionen zu widmen, wie Schütz herausstreicht. Diese Perspektive gilt auch programmatisch für die vorliegende Arbeit. Schütz selbst veröffentlichte, im Vergleich zum Umfang seiner theoretischen Schriften, relativ wenige Texte, die auf empirischer Forschung beruhten. Jedoch stimulierte seine soziologisch geprägte Phänomenologie eine Reihe weiterer Theoriegebäude und Forschungsprogramme, unter anderem die Ethnomethodologie von Harold Garfinkel und Harvey Sacks. Die bekannteste Weiterentwicklung von Schütz' Werk ist jedoch bei Peter L. Berger und Thomas Luckmann zu finden – beide Schüler von Schütz (so wie auch Garfinkel). Luckmann hatte nach Schütz' Tod dessen bislang nur fragmentarisches

Buch *Strukturen der Lebenswelt* zusammengesetzt und vervollständigt. Gemeinsam mit Berger veröffentlichte er den „modernen Klassiker“ (Schützeichel, 2012, S. 22) *Die gesellschaftliche Konstruktion von Wirklichkeit* (Berger & Luckmann, 1966/2009). Im folgenden Abschnitt sollen, hierauf Bezug nehmend, die wissenschafts- und erkenntnistheoretisch relevanten Argumente dieser Arbeit kompakt diskutiert werden.

3.5 Die Konstruktion gesellschaftlicher Wirklichkeit nach Thomas

Berger und Peter Luckmann

Ähnlich wie Schütz rücken Berger und Luckmann die Konstitution des allgemeinen Wissens und die gesellschaftliche Konstruktion von alltäglicher Wirklichkeit in den Mittelpunkt ihrer Ausführungen. Denn das „Wissen“ eben bildet die Bedeutungs- und Sinnstruktur, ohne die es keine menschliche Gesellschaft gäbe“ (Berger & Luckmann, 1966/2009, S. 16). Zu Beginn schreiben sie:

Die Schlüsselbegriffe (...) sind „Wirklichkeit“ und „Wissen“ (...). Für unsere Zwecke genügt es, „Wirklichkeit“ als Qualität von Phänomenen zu definieren, die ungeachtet unseres Willens vorhanden sind – wir können sie ver- aber nicht wegwünschen. „Wissen“ definieren wir als die Gewißheit, daß Phänomene wirklich sind und bestimmbare Eigenschaften haben. (S. 1)

Berger und Luckmann kritisieren die Wissenssoziologie von Max Scheler und Karl Mannheim, die sich vor allem mit der Rolle von Theorien, Ideengeschichte und Ideologien auseinandergesetzt hatten. Dabei seien diese Bereiche nur Teile von dem, was allgemein unter Wissen verstanden wird: „Nur ein begrenzter Kreis von Leuten ist zum Theoretisieren berufen, zum Geschäft mit ‚Ideen‘ bestellt, zur Fabrikation von Weltanschauungen. Aber jedermann in der Gesellschaft hat so oder so Teil an Wissen“ (Berger & Luckmann, 1966/2009, S. 16). Die beiden versuchen in weiterer Folge nichts Geringeres als eine Synthese von komplexen, heterogenen Theoriesträngen aus Europa und den Vereinigten Staaten von Amerika: der Realitätsidee von Émile Durkheim, den anthropologisch geprägten Frühschriften von Karl Marx, der philosophischen Anthropologie von Helmuth Plessner und Arnold Gehlen, der Sozialpsychologie von George Herbert Mead und der Verstehenden Soziologie von Max Weber. Der wichtigste theoretische Anker bleibt jedoch das Werk von Alfred Schütz, an dessen Terminologie sich die Autoren auch stark halten (Berger & Luckmann, 1966/2009, S. 16). Schützsche Begriffe wie ‚Typisierung‘, ‚Alltagswelt‘ oder ‚Sinnprovinzen‘ durchziehen den Text.

Berger und Luckmann streichen die Funktion der direkten Interaktion bei der Herausbildung von Wissen und der Konstruktion der gesellschaftlichen Wirklichkeit hervor. „Die fundamentale Erfahrung des Anderen ist die von Angesicht zu Angesicht. Die Vis-à-vis-Situation ist der Prototyp aller gesellschaftlichen Interaktion. Jede andere Interaktionsform ist von ihr abgeleitet“ (Berger & Luckmann, 1966/2009, S. 31). Ähnlich wie Schütz beschreiben Berger und Luckmann den Prozess der sogenannten Typisierung. Diese Kategoriebildung hilft bei der für den Alltag erforderlichen Orientierung und Komplexitätsreduktion. In der Interaktion mit anderen kommt es zu einer fortwährenden Verhandlung über diese Typisierungen. In Folge werden diese weiterentwickelt und gefestigt. Wenn Handlungsmuster erfolgreich Probleme lösen können, dann werden sie wiederholt angewandt. Berger und Luckmann sprechen in diesem Zusammenhang von ‚Habitualisierung‘. Durch die reziproke Interaktion zwischen Habitualisierungen und Typisierungen stabilisiert sich laufend die Konstruktion der Wirklichkeit in Form von Institutionalisierungen:

Die gemeinsamen Habitualisierungen und Typisierungen von [Individuum] A und B, die bislang noch den Charakter von ad hoc-Konzeptionen zweier Individuen hatten, sind von nun an historische Institutionen. (...) Institutionen sind nun etwas, das seine eigene Wirklichkeit hat, eine Wirklichkeit, die dem Menschen als äußeres, zwingendes Faktum gegenübersteht. (Berger & Luckmann, 1966/2009, S. 62)

In weiterer Folge kommt es zur Legitimierung der Institutionen. Bei komplexen, innerlich widersprüchlichen Institutionen passiert diese Legitimierung über sogenannte symbolische Sinnwelten. Beispiele solcher fortgeschrittener Legitimationen sind Begründungszusammenhänge wie Religionen, der Sozialismus oder die Postmoderne (Abels, 2007, S. 105). „Gesellschaftliche Institutionen sind auch und vor allem Einrichtungen der Vermittlung besonderen Wissens, und das besondere Wissen über symbolische Wirklichkeiten erfordert spezielle Wissensträger, die dann eine privilegierte Stellung annehmen, wenn ihr Wissen Anerkennung findet (...)“ (Knoblauch, 2014, S. 160). In fortgeschrittenen, ausdifferenzierten Gesellschaften mit hochgradiger Arbeitsteilung gibt es viele unterschiedliche ebensolche Wissensträger. Jede symbolische Sinnwelt hat ihre eigenen Theorien und Ideologien, die den Unterbau für die symbolische Sinnwelt darstellen. Zwischen verschiedenen institutionalisierten Sinnwelten treten auch regelmäßig Konflikte um die Vorherrschaft auf. Die herrschenden Gruppen versuchen stets, „ihre Vorstellungen der Wirklichkeit durchzusetzen“ (Knoblauch, 2014, S. 161). Institutionalisierte Sinnwelten etablieren auch Strukturen der Wissensvermittlung, um ihre Kontinuität sicherzustellen.

Die gesellschaftliche Konstruktion von Wirklichkeit ist bei Berger und Luckmann kein festes Gebilde, sondern einer permanenten Dynamik unterworfen. Identität und gesellschaftliche Realität werden durch dialektische Prozesse – Externalisierung, Objektivierung und Internalisierung – konstruiert und konstituiert (vgl. Berger, 1967). Ein Individuum wird einerseits in eine gesellschaftliche Wirklichkeit mit stabilisierten Sinnwelten geboren und von diesen stark beeinflusst, andererseits ist es auch an der Aushandlung der Konstruktionen beteiligt. Mit anderen Worten, wie Loenhoff (2011, S. 156–157) es zusammenfasst: „Wirklichkeitskonstruktionen sind emergente Phänomene. Sie gehen aus der alltäglichen, durch Vergegenständlichungen gestützten sozialen Interaktions- und Kommunikationspraxis hervor. (...) [Die] Dialektik von individueller Sinnkonstruktion und der Erzeugung kollektiver Gewissheiten“ ist dementsprechend in einem besonderen Maße von Sprache abhängig. Mehr noch: Berger und Luckmann entwerfen eine Theorie der Wirklichkeitskonstruktion, die Empathie voraussetzt. Ohne das Vermögen, sich in andere hineinzusetzen, kann keine gesellschaftliche Idee von Realität entstehen.

Anders als Schütz liefern Berger und Luckmann nur bedingt Anhaltspunkte, wie sich ihre Epistemologie empirisch für die Sozialwissenschaften operationalisieren ließe. Sie verwehren sich sogar gegen eine methodologische Umdeutung:

In unserer ganzen Studie haben wir jede erkenntnistheoretische oder methodologische Frage nach den Möglichkeiten soziologischer Analyse entschlossen unterlassen (...). Unser spezielles Vorhaben ist zwar theoretischer Natur. Aber unsere Theorien gehören in das empirische Fach und zu seinen konkreten Problemen und haben nichts mit der Frage nach den Grundlagen des empirischen Faches zu tun. Summa summarum betreiben wir theoretische Soziologie, nicht Methodologie der Soziologie. (Berger & Luckmann, 1966/2009, S. 15)

In erster Linie setzen Berger und Luckmann diese Argumentation jedoch ein, um sich in ihren weiteren Ausführungen auf die Realitätskonstruktion des Alltags und nicht auf die Epistemologie in den Wissenschaften konzentrieren zu können. Eberle (1992, S. 497) meint, dass das Ignorieren der empirischen Operationalisierung aus aufmerksamkeitsstrategischer Perspektive ein kluger Schachzug war, schließlich hätten so viele SozialwissenschaftlerInnen unterschiedlicher methodologischer Provenienz die Theorie aufgreifen können. Dennoch finden sich an manchen Stellen Hinweise zu ihrer Vorstellung eines empirischen Programms, etwa wenn Berger und Luckmann (1966/2009) davon sprechen, dass die Forschung besonders jenen Umständen Beachtung schenken sollte, die der „Entverdinglichung entgegenkommen“ (S. 98). Mit anderen Worten: Es geht darum, den „Mensch als Mensch“ (S. 201) zu sehen.

Anderswo unterstützen sie die Webersche Maxime, dass es die Aufgabe der Forschenden sei, soziale Handlungen und ihre Konsequenzen zu beschreiben, aber keine politische Haltung einzunehmen (Eberle, 1992, S. 496).

Wie auch oben hervorgeht, sehen Berger und Luckmann die Wissenschaft nicht außerhalb der gesellschaftlichen Wirklichkeitskonstruktion. Dennoch erachten sie es als möglich, die Dialektik der individuellen und gesellschaftlichen Wirklichkeitskonstruktion auf Grundlage empirischer Forschung zu beschreiben. Niklas Luhmann, erkenntnistheoretisch ein Vertreter des radikalen Konstruktivismus (Luhmann, 1997, S. 156), kritisiert diese Position scharf: „(...) als ob es einen Kompromiß zwischen Objektivismus und Subjektivismus geben könne, einen halben, sozialen Konstruktivismus, der den Soziologen ins Geschäft bringt“ (vgl. Knoblauch & Schnettler, 2007; Luhmann, 1996, S. 25). Die Synthese dieser beiden Positionen, von sozialem und radikalem Konstruktivismus, wird mit dem kommunikativen Konstruktivismus (Keller, Reichertz & Knoblauch, 2013; Knoblauch, 1995) versucht. Berger und Luckmann demonstrierten die empirische Operationalisierung ihres Ansatzes unter anderem mit religionssoziologischen Arbeiten. Luckmann widmete sich etwa, nachdem er Schütz' Nachlass bearbeitet hatte, verstärkt der Erforschung von interpersonaler Kommunikation und massenmedial vermittelter Sprache (z.B. Luckmann, 2002, 2013). Insbesondere in der deutschsprachigen Soziologie und Kommunikationswissenschaft hatte ihr Ansatz zahlreiche qualitative Forschungsarbeiten zur Folge, die sich mit interpersonaler Kommunikation in verschiedensten Kontexten auseinandersetzten (vgl. Eberle, 1992, S. 501).

3.6 Die Animation des Objekts: Karin Knorr-Cetina, Bruno Latour und die soziotechnische Konstruktion von Fakten

Eine weitere epistemologische Perspektive soll die wissenschaftstheoretische Verortung dieser Arbeit abschließen. Die Kritik am sozialen Konstruktivismus von Berger und Luckmann, die weiter oben im Text bereits angeklungen ist, wird auch von Karin Knorr-Cetina (1989, S. 88) vorgebracht: Zwar arbeite der sozialkonstruktivistische Ansatz in „beispielhafter Weise (...) die verschiedenen *Medien der Objektivierung* (...) (Habitualisierung, Typisierung, Symbolisierung, Sprache, etc.)“ der sozialen Wirklichkeit heraus, jedoch ließe er die Frage nach dem „empirischen Nachweis der Konstruiertheit von ‚Tatsachen‘ sowie die Frage nach dem Status der vom Konstruktivismus gelieferten Erkenntnis“ offen. Knorr-Cetina schlägt deshalb das ‚empirische Programm des Konstruktivismus‘ vor, das „die

Konstruktionsmaschinerie von Wirklichkeit und (...) [die] Konstruktionsprozesse der Teilnehmer“ (S. 91) zum Gegenstand der Untersuchung macht. In Fortführung der Annahme von Berger und Luckmann, dass Typisierung ein zentraler Prozess zur Konstruktion von gesellschaftlicher Wirklichkeit ist, gilt es, ebendiesen Prozess zu untersuchen. Forschung dieser Art müsse nach dem ‚Wie‘ der Konstruktion fragen: „WIE wird eine Kategorisierung, eine Unterscheidung oder Segmentierung der Welt von den Teilnehmern operationalisiert? (...) [Die Kategorisierungen durch TeilnehmerInnen], deren Genealogie, semantische Einbettung sowie pragmatische Operationsform [können] Gegenstand konstruktivistischer Analysen sein“ (S. 92). Im Rahmen des empirischen Konstruktivismus könne es auch keine substantiellen Gesellschaftstheorien geben. „Wohl aber kann es Phänomenbereiche geben, in denen Ereignisse als ‚Kommunikationen‘ oder ‚Handlungen‘ instrumentiert und elaboriert erscheinen“ (S. 93). Mit anderen Worten: Theorien, hier insbesondere verstanden als mikrosoziologische Erklärungsrahmen, müssen auch als Teil einer gesellschaftlichen Konstruktion betrachtet werden.

Knorr-Cetina liefert mit dieser Argumentation die methodologische Fundierung und Einordnung der ersten Ansätze einer konstruktivistischen Wissenschaftsforschung (etwa Knorr-Cetina, 1977, 1981; Latour & Woolgar, 1979/1986; Lynch, 1985). Diese Schule rückt nicht mehr die Ideengeschichte in den Fokus (wie etwa bei Kuhn), sondern untersucht die wissenschaftlichen Praktiken selbst – unter anderem mit ethnografischen Methoden. Rückblickend wird dieser Ansatz sogar als „konstruktivistische Wende“ (Hofmann & Hirschauer, 2012) der Wissenschaftsforschung eingestuft – gemeinsam mit anderen parallelen Bewegungen der jüngeren Science and Technology Studies (für einen Überblick s. Collin, 2011), etwa dem Edinburgher *Strong Programme der Sociology of Scientific Knowledge (SSK)* (Bloor, 1976/1991) oder dem *Empirical Programme of Relativism (EPOR)* (u.a. Collins, 1981; Collins & Pinch, 1998).

Analysen nach dem empirischen Konstruktivismus untersuchen, „wie soziale Gruppierungen sich selbst choreographieren und ihre Choreographien tanzen“ (Knorr-Cetina, 1989, S. 93).

Die Untersuchungen „öffnen ein Territorium (das einer in bestimmter Weise choreographierten Realität), in dem wir [Forschenden] uns aufgrund der Analyse bewegen können sollten. Der epistemische Profit einer konstruktivistischen Analyse sollte sein, ein Terrain so zu erschließen, daß die geordnete Bewegung in diesem Terrain möglich wird. (S. 94)

In dieser Beschreibung des zu erforschenden Terrains werden die Parallelen des Programms des empirischen Konstruktivismus zu ethnografischen Methodologien besonders evident. Knorr-Cetina sieht *Laboratory Life*, die bekannte ‚Laborstudie‘ von Bruno Latour und Steve Woolgar (1979/1986), als beispielhaft für den empirischen Konstruktivismus. Aufbauend auf Datenmaterial eines annähernd zweijährigen Aufenthalts von Latour am Kalifornischen *Salk Institute*, beschreiben die Autoren die, wie sie sie auch im Untertitel des Buchs nennen, „(soziale) Konstruktion wissenschaftlicher Fakten“²⁴. In der Tradition anthropologischer Studien zeichnen sie nicht nur die Praktiken und Konversationen im Laborbetrieb auf, sondern analysieren Dokumente und führen qualitative Interviews mit den Beteiligten durch. Anders als beispielsweise Ludwik Fleck untersuchen sie die Prozesse der Tatsachenkonstruktion als Fachfremde, oder wie es die Autoren nennen, mit exotischer anthropologischer Fremdheit („anthropological strangeness“, S. 29). Sie dokumentieren das alltägliche ‚Handwerk‘ des Forschens im Labor. Die Praxis scheint dabei vor allem aus Codieren, Markieren, Modifizieren, Korrigieren, Lesen und Schreiben zu bestehen (S. 49), mit dem Ziel, wissenschaftliche Artikel zu fabrizieren. Sie argumentieren, dass Fortschritt weniger aus logischen Schlussfolgerungen entsteht, sondern vielmehr auf den kombinierten Einsatz von materiellen, technischen und menschlichen Ressourcen zurückzuführen ist.

Was sich in *Laboratory Life* abzeichnet, arbeitet Latour in späteren Arbeiten noch klarer heraus. Gemeinsam mit Michel Callon, John Law, Madeleine Akrich und anderen begründet er die Akteur-Netzwerk-Theorie (ANT). Dieser Ansatz ist einerseits eine epistemologische Position, andererseits aber eine Methodologie, die sich zum Teil auch von Knorr-Cetinas empirischem Konstruktivismus unterscheidet. Im Kern geht es Latour darum, herauszuarbeiten, dass „Handeln, Fortschritt und Beziehungen entstehen, (...) indem sich Akteure – menschliche und nicht-menschliche Entitäten – zu einem heterogenen Hybriden, einem Netzwerk, zusammenschließen“ (Ausserhofer, 2008, S. 4–5). Latour (2007) fordert für die sozialwissenschaftliche Forschung nichts weniger Fundamentales als eine Aktualisierung der Subjekt-Objekt-Dichotomie und – darauf aufbauend – eine Neudefinition des Begriffs des sozialen Handelns. Er entwickelt ein erweitertes Symmetriepostulat, das auch nichtmenschliche Entitäten als Handelnde in der Beschreibung von

²⁴ Während die erste Auflage von *Laboratory Life* 1979 noch den Untertitel „The social construction of scientific facts“ hatte, wurde bei der zweiten Auflage 1986 das Attribut „social“ gestrichen, um die soziotechnische Co-Konstruktion von Fakten stärker in den Vordergrund zu rücken (Katzenbach, 2016, S. 190).

Wirklichkeitskonstruktionen berücksichtigt. Schließlich wird „[soziales] Handeln nicht nur von Aliens weitergetragen, sondern es wird auf verschiedene Akteurstypen verlagert oder delegiert, die fähig sind, das Handeln durch andere Aktionsmodi, andere Typen von Materialien zu transportieren“ (S. 122; vgl. auch Law, 1992). Bereits in *Laboratory Life* haben Latour und Woolgar (1979/1986) von „inscription devices“ gesprochen (S. 51): Gemeint sind damit Laborapparate, die Substanzen analysieren und in visuelle Repräsentationen wie Zeichen oder Diagramme übersetzen, die dann von den ForscherInnen weiterverwendet werden können. Schließlich kann eine Substanz nicht direkt in ein Forschungspapier (im wahrsten Sinne des Wortes) ‚einfließen‘, sondern wird durch Daten repräsentiert, die von Apparaten generiert wurden. Genau diesem Übersetzungsprozess von Handlungen schenken die VertreterInnen der Akteur-Netzwerk-Theorie ihre Aufmerksamkeit (vgl. Callon, 1984).

Der Begriff des ‚Akteurs‘, der in vielen (vor allem früheren) ANT-Arbeiten auch als ‚Aktant‘ bezeichnet wird, ist der linguistischen Narrationsforschung entlehnt (Greimas, 1971). Mit dem in der Journalismusforschung aus der Rational-Choice-Theorie entlehnten und zur Operationalisierung systemtheoretischer Ansätze gebrauchten Akteursbegriff (z.B. bei Fengler, 2016; Fengler & Russ-Mohl, 2005; Gerhards, 1994; Neuberger, 2004; Raabe, 2005) hat die Akteursidee der ANT wenig gemein. Ein Akteur ist der Handlungsträger in einer narrativen Struktur; jene Entität, die den ‚Akt‘ ausführt oder erträgt (vgl. Ausserhofer, 2008, S. 10–11). Ein Akteur kann ein Mensch sein, aber ebensogut kann ein technisches Artefakt wie eine Datenbank (Ausserhofer, in Druck) oder ein Kühlschrank (Brives & Latour, 2007) zu einem zentralen Akteur in einem Forschungsbericht avancieren.

[Ein] Akteur ist alles, was einen anderen in einem Versuch verändert; von Akteuren läßt sich nur sagen, daß sie handeln; ihre Kompetenz leitet sich aus ihren Performanzen ab; die Handlung ihrerseits wird stets im Verlauf eines Versuchs in einem Versuchsprotokoll – wie rudimentär auch immer – aufgezeichnet. (Latour, 2001, S. 285)

Auch wenn es vielfach anders dargestellt wurde: Latour will nicht Menschen verdinglichen und Dinge vermenschlichen. Ihm geht es darum, dass sozialwissenschaftliche Forschung dem Handlungsprogramm von nicht-menschlichen Entitäten Raum gibt und nicht vorab ignoriert. Symmetrisch forschen heie, zurückzuweisen, dass menschliche Akteure einen speziellen Status haben; stattdessen müssen die Aktivitäten von Dingen und Menschen gleichermaßen in der Untersuchung der Konstruktion sozialer Realität in Betracht gezogen werden (Lindemann, 2011, S. 96). Technik ist weit mehr als ein Werkzeug für den Menschen, schließlich lässt sie sich nie losgelöst von ihm analysieren (Kranzberg, 1986). Im Herstellungsprozess kann Technik Handlungspotenzial inskribiert werden, so die Perspektive der ANT. Eine

Schusswaffe erweitert etwa die Handlungsmacht ihres Trägers, zugleich schränkt sie diesen aber auch ein (Latour, 1994). Latour entwickelt mit diesem handlungstheoretischen Ansatz einen Kompromiss zwischen den sozial- und technikdeterministischen Positionen. Zwar umfasst Latours Werk weit mehr als die hier kompakt dargestellten Ideen, dies soll jedoch an dieser Stelle nicht weiter ausgeführt werden, geht es hier doch um die Argumentation des erkenntnis- und wissenschaftstheoretischen Standpunkts. Die Perspektive, die die ANT hier liefert, also die Rolle der nichtmenschlichen Entitäten in der rekonstruktiven Sozialforschung nicht auszublenden und zu beschreiben, ist eine brauchbare Ergänzung zum bisher Herausgearbeiteten. Wie Knoblauch (2013) betont, thematisierte auch Thomas Luckmann (1970) in seiner späteren Arbeit, dass sich soziales Handeln auf Objekte bezieht. Der epistemologische Unterschied zwischen der ANT und den anderen referierten Ansätzen ist also gar nicht so fundamental wie viele (u.a. auch Knoblauch, 2013 selbst; sowie Latour, 2010) argumentieren – insbesondere dann, wenn man den Blick den Gemeinsamkeiten in den Ansätzen zuwendet und die Distinktionsarbeit ignoriert, die vielfach nur ein Scheingefecht darstellt und der Erhöhung des eigenen wissenschaftlichen ‚Marktwerts‘ dient (vgl. Franck, 1998). Für die epistemologische Grundierung eines kommunikationswissenschaftlichen Forschungsvorhabens jedenfalls ist die von Latour vertretene Perspektive, dass Menschen und Objekte kollektiv die gesellschaftliche Realität konstruieren, äußerst fruchtbar.

3.7 Zwischenresümee

Dieses Kapitel thematisierte aus verschiedenen Perspektiven das Zustandekommen von Erkenntnis in einem Forschungsprozess und der Gesellschaft. Nach einer Diskussion über die Rolle von Theorien wurde zunächst Ludwik Flecks Lehre von Denkstil und Denkkollektiven vorgestellt. Letztere werden unter anderem durch soziale Rituale und Tradition geprägt. Auf Fleck aufbauend wurde das Kuhnsche Bild einer zyklischen Wissenschaft nachgezeichnet, die Phasen unterschiedlicher Stabilität durchläuft. Gesicherte Erkenntnis ist niemals von Dauer, denn vielfach können vorherrschende Paradigmen hinterfragt und schließlich abgelöst werden. Im nächsten Schritt wurden nicht mehr die Wissenschaft allein, sondern die Bedingungen der menschlichen Erfahrung im Alltag anhand von Alfred Schütz' Lebensweltkonzept diskutiert. Schütz spricht von Typisierungen, abstrahierten Erfahrungen, die in neuen Situationen zum Vergleich herangezogen werden und über die Erkenntnis konstruiert wird. Der Ansatz von Peter L. Berger und Thomas Luckmann, die unter anderem auf Schütz aufbauen, unterstreicht

die Rolle der direkten sozialen Interaktion bei der Wissensgenerierung. Der letzte diskutierte Ansatz, jener von Karin Knorr-Cetina und Bruno Latour, thematisierte schließlich die außergewöhnliche Rolle der Technik als aktiv Beitragende zu Erkenntnisprozessen. Menschliche und nichtmenschliche Akteure agieren gemeinsam und kooperativ bei der Konstruktion von Erkenntnis.

Die Rahmenbedingungen, die die verschiedenen Ansätze für die Konstruktion von wissenschaftlicher Realität formulieren, sind verschiedener Natur. Aber jede dieser Bedingungen hat ihre Berechtigung: Sowohl historische, soziale, kommunikative als auch technische Aspekte spielen im Prozess der Realitätskonstituierung von Bedeutung eine essentielle Rolle. Das gilt sowohl für die hier präsentierte Forschung als auch für den datenintensiven Journalismus selbst, der ja nicht einfach nur ‚Fakten‘ in Form von Daten präsentiert.

4 Methodologische Verortung und Forschungsdesign

Das vorangegangene Kapitel zu wissenschaftstheoretischen Grundannahmen thematisierte aus verschiedenen Perspektiven das Verhältnis von Theorie und Empirie im sozialwissenschaftlichen Forschungsprozess. Aufbauend auf der Diskussion der Ansätze von Fleck, Kuhn, Schütz, Berger und Luckmann sowie Knorr-Cetina und Latour wurde ein Konstruktivismuskonzept argumentiert, in dem historische, soziale, kommunikative und technische Faktoren wichtige Rollen bei der Realitätskonstruktion einnehmen. Nachdem das epistemologische Fundament nun abgesteckt ist, ist das Ziel dieses Kapitels einerseits, diese Arbeit methodologisch zu verorten, und andererseits, das Forschungsdesign zu diskutieren. Im ersten Abschnitt des Kapitels werden zunächst Theorien, die den Forschungsprozess schematisieren, besprochen. Auch gilt es Einflussfaktoren, die sich auf die Erstellung eines qualitativen Forschungsdesigns auswirken, zu benennen. Weiters wird auf die methodologischen Spezifika, die sich beim Einsatz digitaler Methoden ergeben, hingewiesen. Als Maßnahmen der Qualitätssicherung werden schließlich die Konzepte von Open Science sowie Triangulation diskutiert. Der zweite Abschnitt dieses Kapitels widmet sich dem konkreten Forschungsdesign der Arbeit und beschreibt das Vorgehen beim systematischen Literaturüberblick sowie der Interviewstudie.

4.1 Zur Konstruktion eines Forschungsdesigns

4.1.1 Forschungsprozess-Schemata

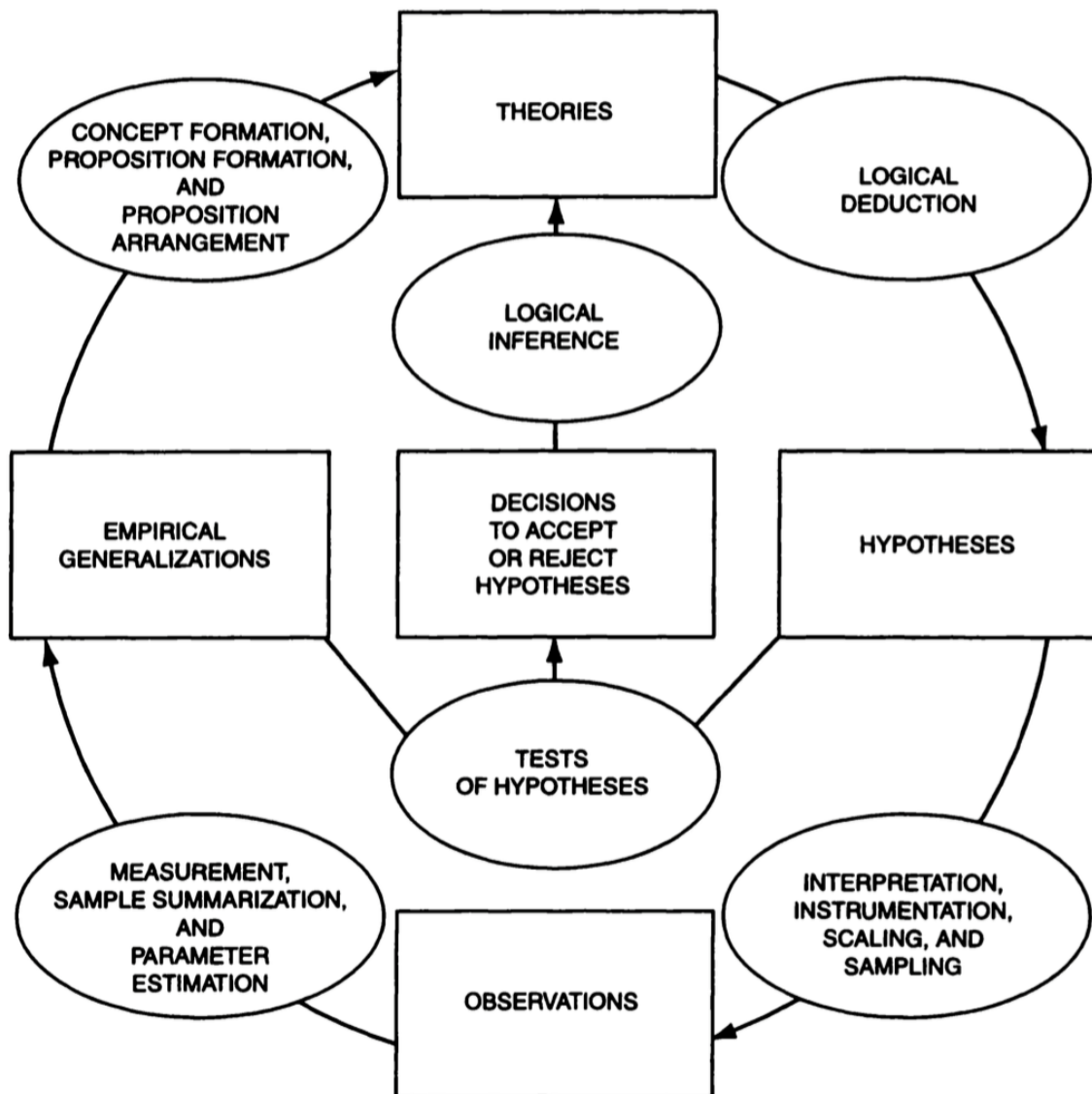
Die Argumentation eines gewissen epistemologischen Ansatzes impliziert immer auch eine bestimmte Axiologie und Ontologie, die zwar nur selten offengelegt werden, aber doch im Hintergrund präsent sind. Ebenso zwingt die Entscheidung für eine bestimmte Epistemologie in letzter Konsequenz die Forschenden dazu, sich für gewisse Fragestellungen, Theorien, Methodologien und Methoden zu entscheiden; oder zumindest gewisse Ansätze auszuschneiden, weil sie mit dem Überbau nicht kompatibel sind (Becker, 1996). Im Kontext der Journalismusforschung spricht Scholl (2011) sogar vom „unauflösbaren Zusammenhang von Fragestellung, Theorie und Methode“ und vom „Prinzip der wechselseitigen Bedingtheit“ (S. 17), schließlich legen „bestimmte Theorien (...) bestimmte Ergebnisinterpretationen nahe, und bestimmte empirische Ergebnisse haben bestimmte Folgen für oder implizieren bestimmte

Herausforderungen an die jeweilige gewählte Theorie“ (S. 18). Wer systemtheoretisch argumentiert, wird selten Daten mittels qualitativer Interviews sammeln. Wer quantitative Erhebungen durchführt, kann damit wohl kaum eine Cultural-Studies-geprägte Theorie weiterentwickeln. Mit anderen Worten: Gutes Forschungsdesign zeichnet sich durch die sinnvolle Kombination von miteinander vereinbaren Ansätzen verschiedener Dimensionen wissenschaftlicher Arbeit aus. Epistemologie, Theorie, Fragestellung, Methoden, etc. müssen schlüssig kombiniert und verbunden sein.

Damit stellt sich also die Frage, wie eine solche stringente Kombination dieser Ansätze und Dimensionen in einem Forschungsdesign funktionieren kann. Flick (2009, S. 90) unterscheidet zwischen Forschungsmodellen mit linearen und zirkulären Prozessvorstellungen. Beim traditionellen, linearen Forschungsprozess beginnen die Forschenden für gewöhnlich mit einer Theorie über ein Phänomen und stellen davon ausgehend Hypothesen auf, die es zu operationalisieren gilt; Entscheidungen über das Sampling werden getroffen, Daten erhoben und analysiert; die Hypothesen validiert bzw. falsifiziert. Am Ende steht eine belastbare Aussage über das Phänomen. Dieses (hier sehr vereinfacht dargestellte) Prozessschema findet sich vor allem in den quantitativ orientierten (Sozial-)Wissenschaften. Es wurde unzählige Male angewandt, weiterentwickelt, erprobt und kritisiert; mit Argumenten, von denen ich einige weiter unten anführe. Zuvor jedoch ein erster Ansatz eines zirkulären Forschungsprozesses, der zweiten Kategorie.

Einen Vorschlag für ein Schema eines zirkulären Forschungsprozesses liefert Wallace (1971) (siehe Abbildung 5).

Abbildung 5. Schema eines zirkulären Forschungsdesigns



Wallaces „Wheel of Science“. Aus *The logic of science in sociology* (S. 18), von W. L. Wallace, 1971, Chicago, IL: Aldine and Atherton.

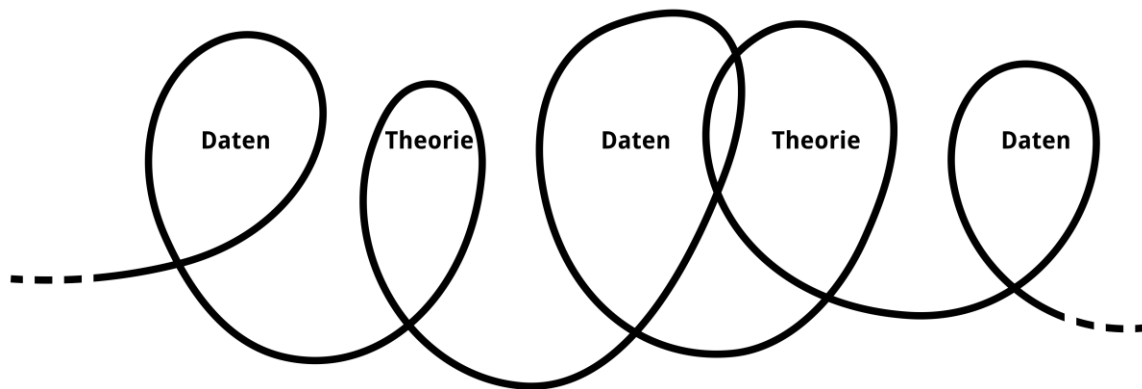
Der Forschungsprozess nach Wallace besteht demnach aus fünf sogenannten informationellen Komponenten: Theorien, Hypothesen, Beobachtungen und empirischen Verallgemeinerungen. Diese repräsentieren verschiedene Stadien des Wissens und können mittels, wie es Wallace nennt, methodologischer Kontrollen, die Informationstransformationen nach sich ziehen, in ein anderes Stadium weiterentwickelt werden. Das Modell berücksichtigt sowohl induktive – auf der linken Seite – wie auch deduktive Ansätze – auf der rechten Seite. Dabei integriert es sowohl Methoden aus dem quantitativen wie dem qualitativen Paradigma (Adler & Clark, 2011, S. 34–35). Die Konzeption erscheint aus verschiedenen Gründen ansprechend, wie van den Hoonaard (2004, S. 38) zunächst argumentiert: Die Form eines Kreislaufs lässt das Schema ganzheitlich wirken; die unparteiische Gegenüberstellung von deduktiven und induktiven

Verfahren vermittelt Toleranz gegenüber verschiedenen Epistemologien; und schließlich ‚macht es Sinn‘ – in der doppelten Bedeutung.

Zugleich ist das *Wheel of Science* insbesondere für qualitative Forschung unzureichend, wie van den Hoonaard in weiterer Folge deutlich macht: Er kritisiert, dass die Theorie im Modell oben steht, sie gewissermaßen den Ausgangspunkt des Forschungsprozesses darstellt – und nicht die Beobachtung eines Phänomens, die ein Forschungsinteresse hervorbringt. Im Hinblick darauf, dass in einem Modell immer irgendetwas oben sein muss und es sich um einen Kreislauf handelt, lässt sich dieser Einwand noch vernachlässigen. Weiters sieht van den Hoonaard qualitative Forschungsstrategien in Wallaces Modell nicht ausreichend repräsentiert: Wo sind Erhebungsmethoden wie Interviews oder Textanalyse in diesem Schema abgebildet? Unter Beobachtung lassen sich diese induktiven Typologien, die empirische Erkenntnisse in breiteren Kontexten organisieren, wohl kaum subsumieren. Schließlich – und das ist der wohl zentralste Einwand – stellt das Modell eine Abfolge dar, die der Praxis qualitativer Forschung nur in seltenen Fällen nahekkommt.

Forschung mit qualitativen Schwerpunkten folgt keinem stringenten Ablaufprozess. Darauf wurde vielfach hingewiesen: Etwa argumentiert Kleining (1982), dass das Vorverständnis der Forschenden über das zu untersuchende Phänomen nur ein vorläufiges ist und im Forschungsprozess einer wiederholten Aktualisierung unterliegt. Schließlich liefert „erst die Untersuchung (...) eine tragfähige Antwort auf die Forschungsfrage und ermöglicht es, das Ausgangsphänomen zu verstehen“ (Krotz, 2005, S. 124). Der Forschungsprozess lässt sich also vielmehr als repetitive Zirkularbewegung schematisieren (Flick, 2009, S. 92; van den Hoonaard, 2004, S. 40), in der einzelne Schritte wiederholt reflektiert werden, wie Abbildung 6 illustriert:

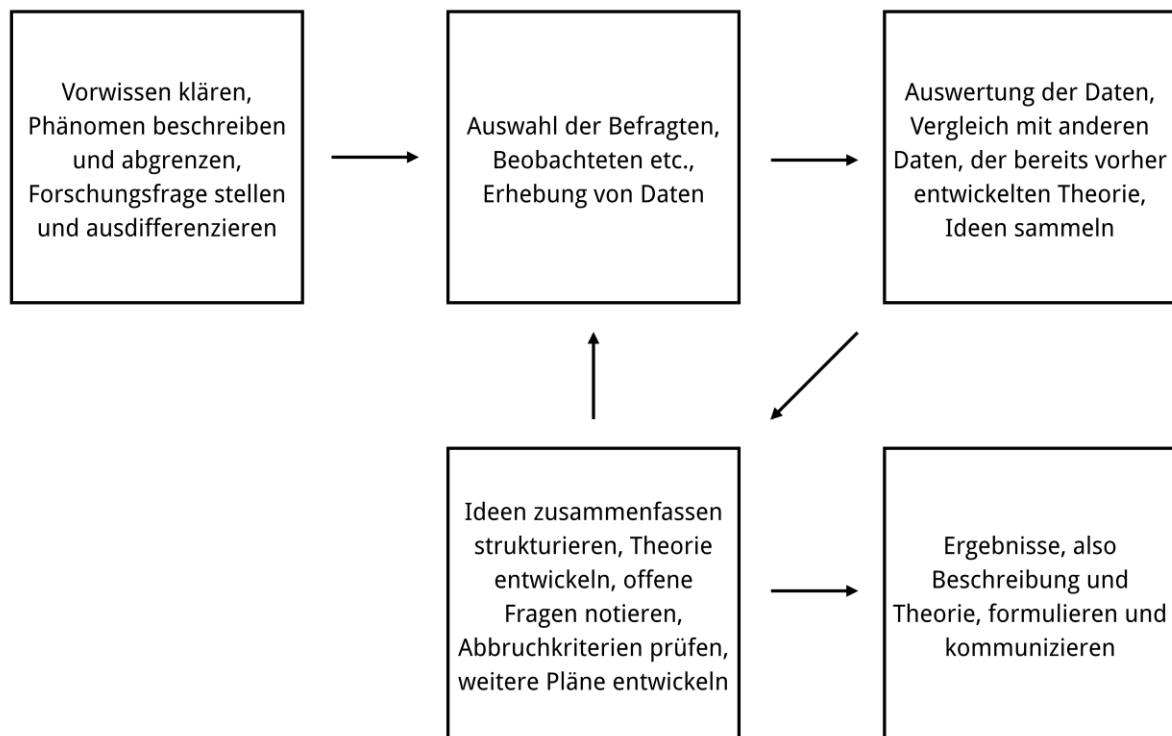
Abbildung 6. Die Forschungsspirale



Adaptiert von *The epistemological framework of qualitative research: Is it still a Cinderella to quantitative research?* (S. 40), von W. C. van den Hoonaard, 2004. In J. Fikfak, F. Adam & D. Garz (Hrsg.), *Qualitative research: Different perspectives, emerging trends* (S. 35–49). Ljubljana: Založba ZRC.

Formalisiert und operationalisiert wurde diese Idee des wiederholten zirkulären Forschungsprozesses insbesondere in der Methodologie der Grounded Theory (sowie anderer theoriegenerierender Forschungsansätze). Demnach finden Prozesse wie Datenerhebung, -analyse und Theoriebildung zeitlich parallel statt und sind stark voneinander abhängig (Ragin, 1994, S. 86; Strübing, 2008, S. 14). Die Forschenden wechseln dabei ständig zwischen Handeln und Reflexion (Strübing, 2008, S. 15). Sie vergleichen das gesammelte Datenmaterial permanent mit ihren Kategorien, aktualisieren diese und passen ihre Erhebungsstrategien an (siehe Abbildung 7) (Glaser, 1965; Glaser & Strauss, 1967, S. 101–116). Das Zirkulieren zwischen diesen Phasen passiert so lange, bis die Berücksichtigung weiterer Fälle keinen theoretischen Mehrwert mehr generiert; erst mit der Reflexion über den Forschungsprozess ist die Forschung abgeschlossen (praktisch erläutert z.B. bei Charmaz, 2006). Dieses „iterativ-zirkuläre Prozessmodell“ (Strübing, 2008, S. 30) – hier stark verkürzt am Beispiel der Grounded Theory wiedergegeben – unterscheidet sich also grundlegend von einer linearen Auffassung von Forschungsprozessen.

Abbildung 7. Der Spiralprozess theoriegenerierender Forschung



Adaptiert von *Neue Theorien entwickeln: Eine Einführung in die Grounded Theory, die Heuristische Sozialforschung und die Ethnographie anhand von Beispielen aus der Kommunikationsforschung* (S. 135), von F. Krotz, 2005, Köln: Herbert von Hellem.

Forschung lässt sich aber nicht allein als iterativ-spiralförmiger Prozess schematisieren: Insbesondere in den Sozialwissenschaften macht es Sinn, Forschung auch als Dialoge zu betrachten, in die die Forschenden „mit dem Phänomen bzw. Forschungsgegenstand in der sozialen Wirklichkeit, mit den Daten und natürlich mit sich selbst“ (Krotz, 2005, S. 132) treten. Eine solche Rahmung ist eine Konsequenz des iterativ-zirkulären Forschungsprozesses, schließlich ist dialektische Kommunikation für dessen Realisierung konstitutiv (Bauer, 2014b, S. 86). Die Metapher des Dialogs wird in verschiedenen Ebenen des Forschungsprozesses evident, wie Krotz (2005, S. 133–134) erläutert:

- Auf der Ebene der Datenerhebung begeben sich die Forschenden in einen (virtuellen) Dialog mit den Forschungspersonen und dem Forschungsgegenstand;
- bei der Dateninterpretation interagieren die Forschenden dialektisch mit den Daten, befragen sie, erhalten Antworten, befragen sie weiter, ...;

- permanent befinden sich die Forschenden in einem Status der Selbstbeobachtung und sind damit auch mit sich selbst in einem inneren Dialog, der sich etwa in einem Forschungstagebuch materialisiert;
- und schließlich geht es auch in der Ergebnisvermittlung um Dialog, insofern als die Forschenden ihre Arbeit präsentieren und zur Diskussion stellen.

Die Idee, dass in einer Forschung verschiedene Prozesse dialektisch ablaufen, Dialog also gleichsam zum „Muster forschenden Handelns“ (Krotz, 2005, S. 133) wird, beeinflusst insbesondere das Verhältnis zwischen den Forschenden und ihrer Umwelt: Damit die Forschenden in einen Dialog mit Personen, Gegenständen, Daten, Öffentlichkeiten, etc. treten können, müssen sie mit diesen Akteuren die soziale und kulturelle Vorstellung von Wirklichkeit teilen. Die Forschenden können sich nicht außerhalb dieser Sphäre als neutrale BeobachterInnen positionieren. Für einen gelungenen Dialog muss das Verhältnis zwischen Forschenden und Beforschten möglichst symmetrisch und hierarchiefrei sein. Und schließlich sind die Forschenden ihren DialogpartnerInnen verpflichtet, was sich wiederum auf die Forschungsethik auswirkt, die weiter unten besprochen wird.

Bislang wurde in diesem Kapitel die in dieser Arbeit vertretene Perspektive auf Forschungsprozesse dargelegt, nämlich, dass gute Forschung iterativ-zirkulären Mustern folgt, in denen Datenerhebung, Analyse und Theoriebildung wiederholt durchlaufen werden. Als normatives Grundmuster forschenden Handelns wurde der Dialog genannt, dessen konsequente Anwendung die Forschenden auf eine Ebene mit dem zu untersuchenden Phänomen stellt. Bevor jedoch das konkrete Forschungsdesign dieser Arbeit dargelegt werden kann, ist es nötig, zu benennen, welche Faktoren die Konstruktion eines Forschungsdesigns beeinflussen. Das soll im folgenden Abschnitt geschehen.

4.1.2 Forschungsdesign-Einflussfaktoren

Die Artikulation von Einflussfaktoren, die die Konstruktion eines Forschungsdesigns berühren, dient nicht nur der eigenen Reflexion („Wovon hängt meine Forschung ab?“); sie ist auch konstitutiv dafür, dass Forschung ihre gesellschaftliche Verantwortung wahrnehmen kann („Wer steht hinter einer Forschung?“). Schließlich kann die Nicht-Artikulation von externen Einflüssen, was im Umkehrschluss eine Verschleierung darstellen kann, in der Außenwirkung selbst hehre Forschung kompromittieren. Zugleich kann die Offenlegung dieser Einflussfaktoren (zumindest in Ansätzen) nachvollziehbar machen, welche Interessen hinter

einer Forschung stehen und auf welcher Grundlage Entscheidungen für den gewählten Forschungsweg gefällt werden.

Flick (2007, S. 38–44; Flick, Kardorff & Steinke, 2004) nennt zahlreiche Faktoren, die die Konstruktion eines Forschungsdesigns beeinflussen: Die Wahl von Zielen, Forschungsfragen, Forschungsperspektiven, Theorien und Methoden ist beispielsweise eine grundlegende Entscheidung in einem Forschungsdesign. Die Hintergründe für diese Entscheidungen wurden bzw. werden an anderen Stellen in dieser Arbeit erläutert.

Zahlreiche VertreterInnen der Science and Technology Studies (STS) weisen auf weitere Faktoren hin, die Einfluss auf das Forschungsdesign haben. Folgt man deren Auffassung, dass wissenschaftliches Handeln über heterogene Akteure weitergetragen und vermittelt wird (siehe hierzu das Kapitel zu den wissenschaftstheoretischen Grundannahmen sowie Latour, 1988; Law, 2004; Marres, 2012a; Rheinberger, 1997), so ist auch ein Forschungsdesign von deren Beitrag abhängig. Marres (2012b) bezeichnet diesen Perspektivenwechsel als ‚Umverteilung der Forschung‘ („redistribution of research“):

It highlights how scientific research tends to involve contributions from a broad range of actors: researchers, research subjects, funders, providers of research materials, infrastructure builders, interested amateurs, and so on. Scientific research, according to this notion, must be understood as a shared accomplishment of a diverse set of actors. (S. 140)

Aus Sicht der Forschenden stellen die von Marres genannten Akteure wie etwa Finanziere oder Infrastruktur-BereitstellerInnen einerseits verschiedene Ressourcen dar, die den Forschungsprozess unterstützen bzw. ermöglichen. Und auch Software hat einen essentiellen Anteil am Zustandekommen von Forschung (Manovich, 2013; Siles & Boczkowski, 2012). Neues Wissen und Innovation in Forschungsprozessen werden durch komplexe Aushandlungsprozesse arbeitsteilig zwischen ebendiesen Akteuren entwickelt. Andererseits haben diese Ressourcen natürlich einen maßgeblichen Einfluss auf das Forschungsdesign. Ein Geldgeber etwa bevorzugt bestimmte Methoden oder Formate der Ergebnispräsentation; Forschungsbeteiligte verweigern die Zusammenarbeit usw. Das Forschungsdesign wird sich diesen Vorgaben anzupassen haben.

4.1.3 Methodologische Herausforderungen digitaler Methoden²⁵

Bislang wurden Einflussfaktoren beschrieben, die auf sozialwissenschaftliche Forschungsvorhaben in unterschiedlicher Intensität einwirken. Externe Ressourcen wie Infrastrukturen und Finanzen, die ein Vorhaben beeinflussen – oft schon bevor es begonnen hat – kennt jedes Forschungsprojekt, unabhängig von Forschungsgegenstand, Perspektive oder Geltungsanspruch. Sobald, wie im Fall dieser Arbeit, bestimmte Fragestellungen computervermittelte Kommunikation betreffen und auch mithilfe digitaler Methoden beantwortet werden, treten weitere vielschichtige Probleme auf, die die Gestaltung eines Forschungsdesigns maßgeblich beeinflussen.

Formuliert als Methodologie für die Sozial- und Geisteswissenschaften unterscheidet Rogers (2013) zwischen digitalisierten und digitalen Methoden: Erstere umfassen ursprünglich analoge Forschungsansätze, die für das Internet adaptiert wurden – Ethnografien oder Fragebögen etwa. Zumeist beschäftigen sich Forschungsprojekte, die diese Methoden zur Grundlage haben, mit netzkulturellen Phänomenen oder sie nutzen das Web schlicht als Kommunikationsmedium im Forschungsprozess. Digitale Methoden hingegen sind Teil des „Computational Turn“ (Berry, 2011), sie sind „nativ digital“ (R. Rogers, 2014), sie wurden also speziell für Online-Forschungsumgebungen entwickelt und haben kein direktes analoges Vorbild. Auch die Forschungsperspektive ist bei digitalen Methoden eine andere: Das Web wird nicht als virtueller Raum gesehen, dessen Kultur durch Forschung durchdrungen wird; die Analyse von Online-Umgebungen soll vielmehr dazu beitragen, Aspekte gesamtgesellschaftlicher Kommunikation und Kultur zu verstehen: „Rather, in studying the online, we make and ground findings about society and culture with the Internet. Thus, the Internet is a research site where one can ground findings about reality.“ (R. Rogers, 2010, S. 243)

Digitale Methoden haben in den vergangenen Jahren einen großen Aufschwung erlebt. Diese Popularisierung hängt nicht zuletzt mit dem starken Plattformfokus des Ansatzes zusammen. Die Konzentration auf Plattformen kommt auch uns Erforschenden dieser Dynamiken gelegen, weil ihre Infrastrukturen und Metriken leichter zu operationalisieren sind als die Akteure, die an diesen Kommunikationsprozessen beteiligt sind. Zugleich bringt dieser

²⁵ Teile dieses Unterabschnitts basieren auf andernorts veröffentlichten Texten (Ausserhofer, 2015, in Druck; Schumann, Ausserhofer, Maireder & Taddicken, 2015), die für diese Arbeit adaptiert und aktualisiert wurden.

Fokus auch einige Probleme mit sich. Die mit digitalen Methoden einhergehenden methodologischen Herausforderungen für die Sozialwissenschaften wurden anderswo bereits ausführlich erörtert (etwa bei boyd & Crawford, 2012; Burrows & Savage, 2014; Puschmann & Ausserhofer, 2017). Auf drei Aspekte soll im Folgenden trotzdem noch hingewiesen werden, weil sie bei datenintensiven computerunterstützten Verfahren so schwer wiegen: die vermeintliche Verlässlichkeit, Repräsentativität sowie Validität digitaler Methoden.

4.1.3.1 Vermeintliche Verlässlichkeit

Ein erster Aspekt ist die vermeintliche Verlässlichkeit von Daten und deren undurchsichtiger Entstehungsprozess: Informationen haben nur Bestand, wenn sie in Maschinensprache übersetzt werden können. Alles andere wird aussortiert und folglich im Forschungsprozess nicht berücksichtigt. Darauf weist Lyotard (1979/2015) bereits Ende der 1970er Jahre hin:

Man kann daher die Prognose stellen, dass all das, was vom überkommenen Wissen nicht in dieser Weise übersetzbar ist, vernachlässigt werden wird, und dass die Orientierung dieser neuen Untersuchungen sich der Bedingung der Übersetzbarkeit etwaiger Ergebnisse in die Maschinensprache unterordnen wird. (S. 30–31)

Dieses Aussortieren, das Übersetzbar-Machen, basiert dabei in der Regel auf Normen, die unterschiedlich sichtbar sind. Zum einen gibt es Normen, die allein zwischen Personen die Verwendung von Datenbanken regeln und beispielweise in Form von Gebrauchsanweisungen, Geschäftsbedingungen oder Codebüchern niedergeschrieben sind. Diese Normen lassen sich – zumindest hypothetisch – verändern oder brechen: Ich kann mir für den Eigengebrauch etwa die Inhalte einer Datenbank kopieren, auch wenn die Nutzungsbedingungen das eigentlich untersagen; oder ich kann andere Bezeichnungen wählen, als es mir die Plattform vorschlägt.

Zum anderen gibt es Normen, die im Prozess der technischen Ausgestaltung in die Datenbank inskribiert werden. Die ArchitektInnen der Datenbank treffen dabei Annahmen über die spätere Nutzung und gestalten die Datenbank entsprechend. Manchmal können Patente, Pflichtenhefte, Protokolle oder die Datenbank-ArchitektInnen selbst noch Zeugnis über diese Gestaltungsentscheidungen ablegen. Doch meist sind diese Normen opak. Bei der Nutzung stehen die UserInnen vor einer ‚Blackbox‘, die ihnen genau die vorgegebene Wahl lässt. Ein Beispiel aus einer Literaturdatenbank: Warum kann ich die Ergebnisse einer Datenbankabfrage nicht im Volltext sehen? Oder: Wieso wurden die ÜbersetzerInnen der Werke in keinem Feld erfasst?

Selten sind die Designentscheidungen, die der Inskription von Normen und Algorithmen zugrunde liegen, für die Nutzenden einseh- oder nachvollziehbar. Die Deskription, das Reverse Engineering des technischen Endprodukts ist meist äußerst aufwändig oder sogar unmöglich (vgl. hierzu die Arbeiten rund um ‚algorithmic accountability‘ und ‚reverse engineering‘ – Diakopoulos, 2014, 2015; Friesinger & Herwig, 2014).

4.1.3.2 Vermeintliche Repräsentativität

Ein zweiter Aspekt ist die Frage nach der Repräsentativität der Daten. Damit ist gemeint, dass sich auf Grundlage der Analyse von Daten in den meisten Forschungsszenarien nur Aussagen über die Daten selbst treffen lassen. Über die Menschen, Objekte oder Eigenschaften, die durch Daten repräsentiert werden und die eigentlich im Mittelpunkt des Forschungsinteresses stehen, lässt sich erst etwas aussagen, wenn die Beziehung zwischen (repräsentierenden) Daten und (repräsentierter) Entität geklärt ist. Ein Beispiel: Verweisen zwei durch Entitätenerkennung identifizierte gleiche Wörter inhaltlich wirklich auf dasselbe Konzept? Durch ‚distant reading‘ (Moretti, 2013), also das computergestützte Auslesen von Textstrukturen, wird sich das kaum herausfinden lassen; erst durch kontextorientierte Lektüre wird man zu einer befriedigenden Antwort kommen.

Ähnlich gelagert ist ein weiteres Problem im Zusammenhang mit der Repräsentation von Daten: Nur selten sind Daten wirklich so hochwertig, wie es für das Forschungsvorhaben wünschenswert wäre. Die Probleme können hier vielschichtig gelagert sein: Die Daten können mangelhaft sein, beispielsweise unvollständig, unsauber, unzugänglich, nicht maschinenlesbar oder unzureichend verlinkt, etc. Des Weiteren kann es im Vergleich zu konventioneller Forschung schwerer zu kontrollieren sein, ob wirklich die gewünschte Gesamtheit der Daten einbezogen wurde (Schumann u. a., 2015). Und schließlich treten bei der softwareunterstützten Analyse regelmäßig ‚Glitches‘ auf, also unerwartete Irregularitäten und Fehlfunktionen. Dabei ist oftmals nicht klar, wo der Fehler liegt: in der Software, im Datenset, im Input oder der Erwartungshaltung der NutzerInnen?

4.1.3.3 Vermeintliche Objektivität

Als dritte Herausforderungsdimension gilt es, den Objektivitätsanspruch, den viele Projekte formulieren, zu hinterfragen. Die jahrzehntelangen Kontroversen zwischen VertreterInnen qualitativer und quantitativer Verfahren erleben im Zeitalter datenintensiver Methoden eine Renaissance – fast so, als hätte es diese Auseinandersetzungen nie gegeben. Zu verlockend ist

die Aussicht auf Wissensgenerierung, in welcher der Algorithmus die Arbeit viel ‚objektiver‘ verrichtet; zu ansprechend ist die Perspektive einer Forschung, die mit den Quantifizierungen der ‚harten‘ Naturwissenschaften mithalten kann. Programmcode basiert in der Regel auf Algorithmen, die von Menschen geschrieben wurden. Ihre Ideologien, Erwartungen und Vorstellungen finden sich auch im Programmcode wieder (Ausserhofer, 2013). In jedem datenintensiven Forschungsprozess wird an vielen Stellen ausgesondert, entschieden und – vor allem – interpretiert (Bowker & Star, 1999a).

Viele der in diesem Abschnitt formulierten Probleme und Herausforderungen bei Forschung mit Datenbanken sind in ihrer Argumentationslinie nicht völlig neu. Ähnliche methodologische Bedenken wurden bereits in anderen Kontexten formuliert: Sie gelten etwa auch für viele quantitative Verfahren, die sich nicht mit digitalen Inhalten beschäftigen. Die Wichtigkeit, diese Reflexionen zu artikulieren, nimmt jedoch bei der Forschung mit digitalen Objekten und Datenbanken zu, weil Projekte leichter skaliert werden können: Weil die Maschine die Rechenleistung übernimmt, kann ich etwa statt einer Stichprobe einfach auf die Gesamtheit des Materials zugreifen. Wenn ich beispielsweise einen Korpus von mehreren Millionen Tweets untersuche, muss das Klassifizieren von Inhalten nicht mehr ausschnittsweise und händisch erfolgen, sondern kann durch Machine-Learning-Algorithmen geschehen.

Nachdem in diesem Kapitel bislang der Ablauf in einem Forschungsprozess, Einflussfaktoren auf die Planung und die Herausforderungen bei digitalen Methoden diskutiert wurden, sollen im Folgenden noch zwei Strategien zur Qualitätssicherung angesprochen werden: einerseits die Öffnung des Forschungsprozesses (‚Open Science‘), andererseits – im darauffolgenden Abschnitt – die Triangulation.

4.1.4 Open Science²⁶

Vor allem für die Naturwissenschaften wurden unter dem Stichwort ‚Open Science‘ in den vergangenen Jahren methodologische Prinzipien entwickelt, die sich die digitalen Sozialwissenschaften zunutze machen können. Die Anwendung der Open-Science-Prinzipien in den (ehemals) ‚weichen‘ Disziplinen scheint umso relevanter vor dem Hintergrund, dass

²⁶ Teile dieses Unterabschnitts basieren auf einem andernorts veröffentlichten Text (Ausserhofer, in Druck), der für diese Arbeit adaptiert und aktualisiert wurde.

sich ebendiese Disziplinen immer stärker der naturwissenschaftlichen Methodenrepertoires bedienen. Mit anderen Worten: Wenn Teile der Sozial- und Geisteswissenschaften in ihren Analysen immer ‚naturwissenschaftlicher‘ werden, dann sollten sie auch deren Ansätze zur methodologischen Reflexion und Offenheit aus den Naturwissenschaften adaptieren.

Die Prinzipien von Open Science sind in allen Phasen des Forschungsprozesses anwendbar – von der initialen Konzeption bis zur Dissemination (Whyte & Pryor, 2011). Das soll unter anderem die genaue Nachvollziehbarkeit und die Reproduzierbarkeit erleichtern; etwas, das mit den knappen Methodenbeschreibungen in vielen Forschungspublikationen von Computational-Social-Science-Arbeiten oft schwer bis gar nicht möglich ist. Neben dem Nutzen für die Nachvollziehbarkeit und Reproduzierbarkeit der Forschung ergeben sich durch die Anwendung von Open-Science-Prinzipien viele weitere Effekte: Dazu zählen unter anderem die Weiterverwendung der Daten durch Dritte (und damit verbundene Folgen), die Entstehung neuer Fragestellungen, die stärkere Sichtbarkeit der Forschung und der Forschenden in der Öffentlichkeit sowie die Generierung von neuen Dienstleistungen (Fecher u. a., 2015; Fry, Lockyer, Oppenheim, Houghton & Rasmussen, 2008).

In der Forschung selbst geht es unter anderem darum, den Fortgang bestmöglich online zu dokumentieren: Bibliografien, Notizen, Daten, Berichte über erreichte und verfehlte Meilensteine und vieles andere sollen veröffentlicht werden. Wie auch Whyte und Pryor (2011) beschreiben, ist Offenheit dabei kein dichotomes Konzept. Forschung kann leichter und schwerer zugänglich sein: Etwa können Forschungsinhalte unterschiedlich großen Kreisen zugänglich gemacht werden – von KollegInnen bis hin zur globalen Öffentlichkeit. Ein bemerkenswertes Beispiel für eine sozialwissenschaftliche Veröffentlichung nach Open-Science-Prinzipien legt Heise (2016) vor, der seine Dissertation bereits im Erstellungsprozess offenlegte.

Die methodologische Offenlegung beinhaltet auch die Dokumentation der Entscheidungen bei der Datensammlung und Datenmodifikation. Diese Transparenz radikal weitergedacht führt zum Konzept der ‚Open Notebook Science‘, in der neben sämtlichen relevanten Aufzeichnungen auch sämtliche Daten noch vor der Veröffentlichung der Ergebnisse offengelegt werden (Bradley, Owens & Williams, 2008). Klarerweise wirft dieser Ansatz einige technische Fragen auf, etwa nach den Möglichkeiten zur Sicherstellung der Datenqualität oder dem Einfluss auf Peer-Review-Verfahren.

Der Öffnung von Daten kommt in Open-Science-Projekten eine besondere Rolle zu. Die strategische Datenveröffentlichung wird zu einem wichtigen Kapitel in der Disseminationsarbeit. Sowohl Rohdaten als auch bereits prozessierte Daten können geteilt werden – gemeinsam mit Erläuterungen zur vorgenommenen Datenbearbeitung. Um die Weiterverwendung der Daten zu befördern, haben sich für Fragen, die sich viele Forschende bei der Datenveröffentlichung stellen, bestimmte Standards und Gütekriterien etabliert, die etwa Formate, Lizenzen, Metadaten, Schnittstellen oder Verlinkungen von Datenbanken und Datensätzen regeln. Offene Standards für (Meta-)Datenformate sind etwa die *Data Protocols* von *Open Knowledge*²⁷, die Übereinkunft *Dublin Core*²⁸, das *Data Catalog Vocabulary (DCAT)* des *W3C*²⁹ oder die Arbeiten des *Federal Geographic Data Committee*³⁰.

Auch gibt es eine Reihe von Plattformen, die die Veröffentlichung von Forschungsdaten erleichtern und die langfristige Verfügbarkeit garantieren. Dass Publikationen als Open Access veröffentlicht werden, ist ein weiterer Teil des Open-Science-Konzepts („Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities“, 2003, „Budapest Open Access Initiative“, 2012; Suber, 2012). Auch ist es üblich, den Quellcode für im Forschungsprozess entwickelte Anwendungen zu veröffentlichen (Kraker u. a., 2016). Programme beziehungsweise Programmteile wie Scripts, Crawler oder Scraper, die (Web-)Inhalte automatisiert zu strukturierten Daten verarbeiten, können so leicht für andere Szenarien angepasst werden.

Zusammenfassend lässt sich feststellen, dass vieles für die Anwendung von Open-Science-Prinzipien in der empirischen Forschungsarbeit spricht. Der erforderliche Zeit- und Ressourcenaufwand ist dabei relativ groß. Im Resultat nützt Open-Science-Praxis unter anderem der Qualität einer Forschungsarbeit. In Kombination mit der Berücksichtigung einer weiteren Grundregel nachhaltiger wissenschaftlicher Arbeit, der Triangulation, die im folgenden Abschnitt thematisiert wird, lassen sich durchaus belastbare Ergebnisse erzielen.

²⁷ <http://dataprotocols.org/>

²⁸ <http://dublincore.org/>

²⁹ <http://www.w3.org/TR/vocab-dcat/>

³⁰ <https://www.fgdc.gov/>

4.1.5 Triangulation

Im sozialwissenschaftlichen Kontext bezeichnet Triangulation „die Einnahme unterschiedlicher Perspektiven auf einen untersuchten Gegenstand“ (Flick, 2011, S. 12): Zur Beantwortung von gemeinsamen Forschungsfragen werden verschiedene Methoden, Datensorten und/oder theoretische Zugänge etc. miteinander kombiniert. Diese Verschränkung soll möglichst theoretisch oder methodologisch begründet erfolgen. Die „Perspektiven sollten so weit als möglich gleichberechtigt und gleichermaßen konsequent behandelt und umgesetzt werden“ (S. 12). Das Ziel von Triangulation ist Erkenntnis, die im Vergleich zu Forschung, die auf einer einzigen Perspektive fußt, gesicherter ist bzw. darüber hinausgeht.

Wie Flick (2011, S. 8–9) beschreibt, zeichnet zahlreiche sozialwissenschaftliche Klassiker aus, dass sie verschiedene Datensorten verwenden, die mit unterschiedlichen Methoden erhoben wurden. Er nennt die *Arbeitslosen von Marienthal* (Jahoda, Lazarsfeld & Zeisel, 1933/2014) oder Glasers und Strauss' (1967) Grounded-Theory-Einführung als Beispiele, die auf heterogenen Perspektiven und Daten beruhen. Letztere begründen dies folgendermaßen:

Different kinds of data give the analyst different views or vantage points from which to understand a category and to develop its properties; (...) While the sociologist may use one technique of data collection primarily, theoretical sampling for saturation of a category allows a multi-faceted investigation, in which there are no limits to the techniques of data collection, the way they are used, or the types of data acquired. (S. 65)

Aus erkenntnistheoretischer Perspektive passt die Triangulation auf den ersten Blick mit dem in Kapitel 3 argumentierten Konstruktivismuskonzept zusammen, schließlich kann es nicht ‚den einen‘ Forschungsgegenstand geben, der aus verschiedenen Perspektiven beleuchtet wird. Unterschiedliche Perspektiven im Forschungsprozess konstruieren unterschiedliche Forschungsgegenstände und produzieren zwangsläufig unterschiedliche Ergebnisse (Fielding & Fielding, 1986, S. 30). „Es gibt nicht eine ‚richtige‘ Empirie, sondern verschiedene *Empirietypen*, die nach unterschiedlichen Regeln produziert werden und nicht aufeinander reduziert werden können“ (Bonß, 1982, S. 10). Welchen Erkenntnisgewinn bringt da noch die Triangulation? Es geht bei einer konstruktivistisch informierten Triangulation nicht darum, ‚einen‘ Gegenstand mittels einer subjektunabhängigen Naturempirie zu erfassen, sondern vielmehr soll ein tiefgreifendes, multiperspektivisches Verständnis über ein Phänomen entwickelt werden (Denzin, 1978; zit. n. Flick, 2011, S. 20). Die Reflexion über die Heterogenität eines Phänomens in Kombination mit methodologischer Offenlegung kann

verhindern, eine positivistische Vorstellung von einem singulären Forschungsobjekt zu entwickeln. Unter diesen Voraussetzungen kann Triangulation gelingen. Flick (2011, S. 25) bringt es auf den Punkt: „Triangulation liefert nicht übereinstimmende oder einander widersprechende Abbildungen des Gegenstandes, sondern zeigt unterschiedliche Konstruktionen eines Phänomens – etwa auf der Ebene des Alltagswissens und auf der Ebene des Handelns – auf.“

Denzin (1978, S. 295–304) unterscheidet zwischen vier Formen der Triangulation: die Kombination von Daten, Forschenden, Theorien oder Methoden. ‚Datentriangulation‘ meint die Einbeziehung verschiedener Datenquellen, die auf der gleichen Erhebungsmethode basieren; etwa die Beobachtung nicht nur eines Falles, sondern mehrerer (an verschiedenen Orten, zu verschiedenen Zeiten oder anhand verschiedener Personen). Die ‚Forschendentriangulation‘ („investigator triangulation“) bedeutet, dass mehrere Forschende sich an der Untersuchung eines Falls beteiligen. Damit sollen Verzerrungen durch die Person des Forschenden aufgedeckt werden (Flick, 2011, S. 14). Die ‚Theorientriangulation‘ wiederum bezeichnet den Versuch, sich an ein Phänomen aus unterschiedlichen theoretischen Richtungen anzunähern. Beispielsweise kann ein Interviewleitfaden so erstellt werden, dass er mehrere theoretische Stränge inkorporiert, oder Untersuchungsmaterial kann aus unterschiedlichen Perspektiven heraus analysiert werden. Auch hier geht es um Bias-Minimierung, wie Westie (1957, S. 154) verdeutlicht: Theorientriangulation verkleinert die Wahrscheinlichkeit, dass die Forschenden Irregularitäten übersehen oder ignorieren; sie erlaubt generalisiertere Aussagen; und schließlich schafft sie die Grundlage für die Entwicklung theoretischer Synthesen.

Typ vier, die ‚methodologische Triangulation‘, schließlich bezeichnet die Verknüpfung unterschiedlicher Untersuchungswerkzeuge. Es gibt hier die seltener eingesetzte Triangulation innerhalb einer Methode („within-method triangulation“), bei der etwa in der Analyse unterschiedliche Skalen angelegt werden, und die häufiger verwendete Triangulation zwischen mehreren Methoden („between-method triangulation“), bei der verschiedene Methoden miteinander kombiniert werden. Letztere Form kennt sowohl die Kombination von allein qualitativen (z.B. in der Ethnografie), ausschließlich quantitativen Ansätzen wie auch eine Mischung aus Werkzeugen beider Paradigmen. In einer Metaanalyse macht Bryman (2006) ein gutes Dutzend unterschiedlicher Begründungen für den Einsatz von Methodentriangulation aus – sie reichen von der Hoffnung auf Vollständigkeit bis hin zur Diversität der Ansichten. Kelle (2008) fasst die Vorteile der Methodentriangulation zusammen:

Methodenkombination kann die Möglichkeit zu einer wechselseitigen Methodenkritik bieten, d.h. mit Hilfe von Verfahren der einen Tradition können typische Validitätsprobleme und Fehlerquellen identifiziert werden, die sich mit der Anwendung von Methoden der anderen Tradition verbinden. Methodenkombination würde dann der Validierung von Daten, Methoden und Ergebnissen dienen. (...) Methodenkombination kann der wechselseitigen Ergänzung von Forschungsergebnissen dienen, d.h. mit Hilfe von Verfahren der einen Tradition können soziale Phänomene in den Blick genommen werden, die durch Methoden der anderen Tradition nicht oder nur ungenügend erfasst und beschrieben werden können, so dass die Ergebnisse qualitativer und quantitativer Forschung zusammen ein adäquates (oder auch nur umfassenderes) Bild des Untersuchungsgegenstandes ergeben. (S. 54–55)

Wenn, so wie in dieser Arbeit, qualitative und quantitative Methoden miteinander verschränkt werden, dann bieten sich verschiedene Abläufe der Kombination (Bryman, 1988/2004, S. 133–138; Flick, 2011, S. 80–84; Miles & Huberman, 1994, S. 42–43) an: Qualitative und quantitative Daten können fortlaufend parallel gesammelt und analysiert werden – in einem integrierenden Forschungsdesign; die (explorative) qualitative Forschung kann die Basis für die spätere quantitative Forschung bilden; die quantitative Forschung kann wiederum das Fundament für qualitative Erhebungen – etwa für die Fallauswahl oder zur Interpretation von Zusammenhängen von statistischen Analysen – liefern. Für alle Varianten gibt es unzählige Best-Practice-Beispiele (vgl. Bryman, 2006).

Ziel der in diesem Abschnitt sehr kompakt dargestellten Übersicht über die Triangulation war es, die Möglichkeiten bei der Entwicklung eines mehrstufigen Forschungsdesigns auszuloten und zu reflektieren, welche Konsequenzen sich daraus ergeben. In Kombination mit den zuvor dargelegten Punkten (der methodologischen Offenlegung nach Open-Science-Prinzipien, den Herausforderungen bei digitalen Methoden, den Einflussfaktoren auf ein Forschungsdesign sowie den Schematisierungen von Forschungsprozessen) bildeten diese Ausführungen die Grundlage für das Forschungsdesign, das im kommenden Abschnitt dargelegt wird.

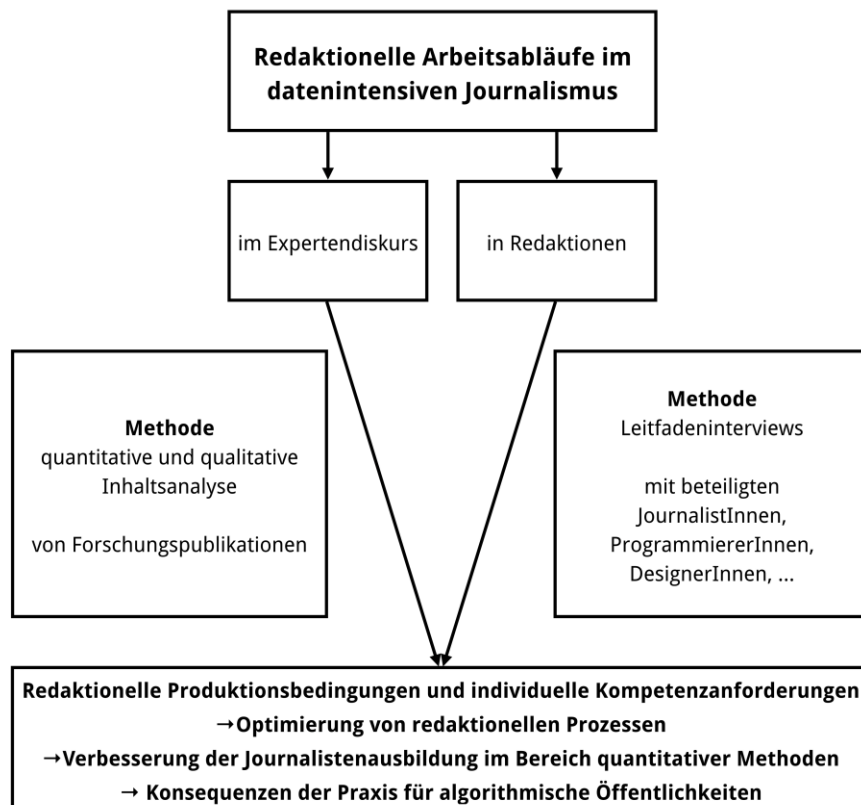
4.2 Das Forschungsdesign dieser Arbeit

Aufbauend auf den vorangegangenen wissenschaftstheoretischen Ausführungen und den Reflexionen zur methodologischen Position soll im Folgenden das Forschungsdesign des empirischen Teils dieser Arbeit erläutert werden. Diese Dissertation hatte sich als Ziele gesetzt, eine Forschungssynthese zu datenintensivem Journalismus zu entwickeln, die Arbeitsabläufe von datenintensivem Journalismus zu erläutern und, aufbauend auf diesen beiden empirischen

Arbeiten, einen theoretischen Beitrag zu leisten, der die Konsequenzen der Praxis des datenintensiven Journalismus für journalistische Datenkompetenz und die Konstruktion von vernetzten Öffentlichkeiten kontextualisiert.

Die Ziele wurden in den zentralen Fragestellungen operationalisiert. Um die Fragen zu beantworten und die Ziele der Arbeit zu erreichen, wurde ein mehrstufiges Verfahren angewandt, das in verschiedenen Phasen sowohl qualitative als auch quantitative Methoden beinhaltet. Abbildung 8 illustriert das Vorgehen.

Abbildung 8. Das Forschungsdesign dieser Arbeit



Adaptiert von Designing qualitative research (S. 47), von U. Flick, 2007, London: Sage.

Dieser gewählte Weg ist einer von vielen verschiedenen, die möglich gewesen wären, um die redaktionellen Produktionsbedingungen des datenintensiven Journalismus zu erfassen. In ihrer ursprünglichen Konzeption war die Arbeit als vernetzte Newsroom-Ethnografie konzipiert (P. N. Howard, 2002). Geplant war, die Praktiken des datenintensiven Journalismus in einer Redaktion in den Mittelpunkt der Untersuchung zu stellen und diese zu beobachten. Zusätzlich zu den Daten aus den (teilnehmenden) Beobachtungen und Interviews wären noch heterogene Dokumente in die Analyse eingeflossen. Damit hätte diese Arbeit an die Tradition von ‚klassischen‘ ethnografischen Arbeiten der Journalismusforschung wie etwa jene von

Tuchman (1978) und Gans (1979/2004) anschließen können (für einen Überblick s. Cottle, 2007), aber auch an neuere Arbeiten, die die Praktiken des Onlinejournalismus in einzelnen Redaktionen untersuchten (z.B. C. W. Anderson, 2013a; Domingo & Paterson, 2011; Steensen, 2010; Usher, 2014, 2016). Trotz mehrerer Anfragen bei verschiedenen Medien im deutschsprachigen Raum konnte in keiner Redaktion ein längerer ‚Feldaufenthalt‘ realisiert werden. Während die DatenjournalistInnen durchaus Verständnis und Unterstützung für das Vorhaben artikulierten, waren die Redaktionsleitungen der Idee gegenüber weniger aufgeschlossen. Als Gründe für die Ablehnung wurden etwa der vorherrschende Zeit- und Kostendruck des Mediums genannt. Auch wurde die Sorge geäußert, dass durch die Datenaufnahme in der Redaktion zu viel aus dem Unternehmensinneren preisgegeben werden würde, man die wissenschaftliche Veröffentlichung nicht kontrollieren könne, und man wirtschaftliche Risiken befürchte.

Nach den erfolglosen Versuchen, direkt den Zugang zu einer Redaktion zu erhalten, musste der Forschungsfokus leicht angepasst werden. Anstelle von Beobachtungen in einer Redaktion wurden vermehrt Interviews mit Beteiligten in verschiedenen Redaktionen geführt. Neben der systematischen Datenaufnahme für die strukturierte Literaturanalyse und die Interviewstudie wurden andere, vornehmlich öffentlich zugängliche Foren der Datenjournalismus-Community genutzt, um weitere Eindrücke zu erhalten und Daten zu erheben. Dazu zählten etwa die jahrelange Beobachtung verschiedener Onlinequellen und -communities – unter anderem auf Plattformen wie *Twitter*, *Slack*, *Facebook* und diversen Mailinglisten. Darüber hinaus nahm ich 2014 bzw. 2015 im Rahmen eines mehrmonatigen Forschungsaufenthalts an der *University of California, Berkeley* an zwei einschlägigen Lehrveranstaltungen der *Graduate School of Journalism* teil. In dieser Zeit besuchte ich auch zahlreiche Konferenzen (u.a. die NICAR-Konferenz 2015 in Atlanta, Georgia) und interviewte Personen, die sich mit datenintensivem Journalismus beschäftigten. Zwischen 2010 und 2016, im Rahmen meiner Anstellung als wissenschaftlicher Mitarbeiter am *Institut für Journalismus und Public Relations* der *FH JOANNEUM* in Graz setzte ich verschiedene Initiativen um, organisierte etwa ein Symposium zum Thema „Datenjournalismus und Infografik“³¹ und unterrichtete Studierende in Datenjournalismus in Theorie und Praxis. Auch diverse, von mir initiierte und kollaborativ durchgeführte Forschungs- und Entwicklungsprojekte beschäftigen

³¹ <http://datablog.fh-joanneum.at/projekte/symposium2013/>

sich in Graz mit dem Thema.³² Des Weiteren beteiligte ich mich an der Wiener Ortsgruppe der internationalen *Hacks/Hackers*-Bewegung und organisierte mehrere einschlägige Veranstaltungen.³³ Für das Österreich-,Chapter‘ der auf ‚open data‘ spezialisierten Organisation *Open Knowledge* war ich 2015 und 2016 an einem Projekt beteiligt, das Daten der öffentlichen Verwaltung so zur Verfügung stellte, dass sie JournalistInnen gut weiterverwenden konnten.³⁴ Zuletzt (2016–2017) arbeitete ich als assoziierter Forscher am *Alexander von Humboldt Institut für Internet und Gesellschaft* mit RedakteurInnen der Tageszeitung *Der Standard* an einem Projekt, in dem wir verschiedene Aspekte rund um die Onlinekommunikation von rechten Empörungsbewegungen in Europa ergründeten. Neben der inhaltlichen Arbeit stand in diesem Projekt insbesondere der Austausch über Methoden von Datenjournalismus und Kommunikationswissenschaft im Vordergrund.³⁵ Auch wenn sich die wenigsten dieser Aktivitäten direkt im empirischen Teil der Arbeit niederschlugen, so generierten sie doch essentielles Kontextwissen, das in viele Teile des Texts einfluss. Diese Tätigkeiten waren zentral, um einen tiefgründigeres Verständnis der Praxis des datenintensiven Journalismus zu erlangen. Sie trugen dazu bei, die Defizite, die sich durch die nicht realisierbare vernetzte Newsroom-Ethnografie ergaben, zu kompensieren.

Eine detaillierte Beschreibung des Vorgehens und der angewandten Methoden befindet sich am Beginn der jeweiligen Kapitel zur systematischen Literaturanalyse bzw. Interviewstudie.

³² Das Projekt *Visual Analytics in Data-driven Journalism (VALiD)* (<http://www.validproject.at/> bzw. <https://fh-joanneum.at/projekt/valid-visual-analytics-im-datenjournalismus/>) etwa zielte darauf ab, technische Lösungen zu entwickeln, mit denen DatenjournalistInnen leichter heterogene Daten explorieren können. Das Projekt *Steirische Vielfalt visualisiert* (<https://fh-joanneum.at/projekt/steirische-vielfalt-visualisiert/>) stellte steirische Diversitätsdaten in einer Form dar, die es RezipientInnen spielerisch ermöglichte, sich persönlich mit Vielfalt in der Steiermark in Beziehung zu setzen (<https://fh-joanneum.at/projekt/steirische-vielfalt-visualisiert/>, siehe auch Abbildung 4).

³³ <https://www.meetup.com/Hacks-Hackers-Vienna/>

³⁴ Das Projekt *Gute Taten für gute Daten* ist hier dokumentiert: <https://okfn.at/gutedaten/>. Der Begriff Open Data wurde bereits in einem eigenen Unterabschnitt weiter oben erläutert.

³⁵ <https://www.hiig.de/project/networks-of-outrage/>

5 Systematische Literaturanalyse: Datenintensive journalistische Arbeitsabläufe aus Forschungssicht³⁶

5.1 Methode: Structured Literature Reviews

Diese systematische Dokumentenanalyse folgt der Methode eines strukturierten Literaturvergleichs (Massaro, Dumay & Guthrie, 2016): Das grundsätzliche Ziel eines solchen ‚structured literature reviews‘ ist es, zu einem bestimmten Themengebiet Einsichten, kritische Reflexionen, zukünftige Forschungspfade und Forschungsfragen zu entwickeln; die Methode bildet auch das Fundament, auf dem weitere Forschung aufbauen kann. Darüber hinaus sind strukturierte Literaturanalysen die Grundlage für Forschungssynthesen (S. 795).

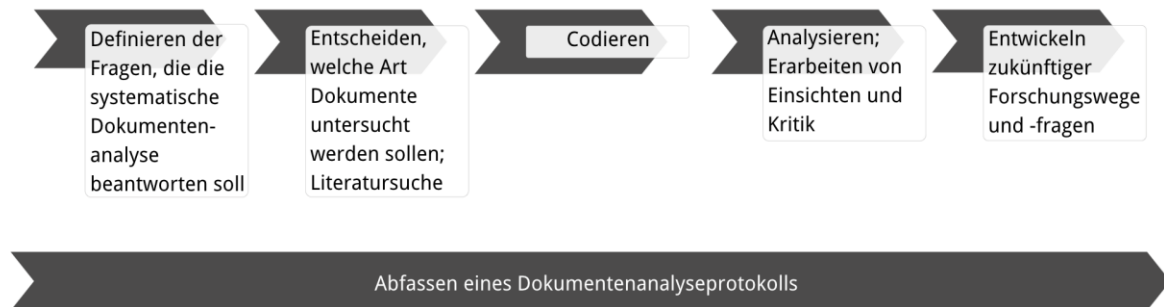
Strukturierte Literaturvergleiche basieren auf einem möglichst nachvollziehbaren und transparenten Prozess, der Bias zu minimieren versucht (Tranfield, Denyer & Smart, 2003, S. 209). Sie nutzen sowohl Techniken der quantitativen, form-orientierten Inhaltsanalyse als auch hermeneutische und interpretative Methoden – letztere insbesondere um Einsichten und Kritik zu entwickeln (Massaro u. a., 2016). Der Zweck von Kritik in diesem Kontext ist es, die dominanten Ziele, Ideen, Ideologien und Diskurse zu einem Phänomen zu hinterfragen (Alvesson & Deetz, 2000, S. 18).

Abbildung 9 illustriert die für diese Arbeit unternommenen Schritte (illustriert und adaptiert von Massaro u. a., 2015): Zunächst wurden die Fragen definiert, die die systematische Literaturanalyse beantworten sollte. Im Fall dieser Arbeit ergaben sich diese aus den übergeordneten Forschungsfragen. Anschließend wurde entschieden, welcher Typ von Dokumenten inkludiert werden sollte, und eine Literatursuche durchgeführt, die einem rigiden, reproduzierbaren Prozess folgt. Schließlich wurde das Material qualitativ codiert und einer Reihe von quantitativen Analysen unterzogen, bevor der Ergebnisteil abgefasst wurde. Parallel wurde ein Forschungsprotokoll abgefasst, in dem obligatorische Passagenpunkte und Auswahlentscheidungen dokumentiert wurden. Auch wenn der Prozess hier linear

³⁶ Teile dieses Kapitels basieren auf einem andernorts veröffentlichten Text (Ausserhofer u. a., 2017), der für diese Arbeit übersetzt, adaptiert und aktualisiert wurde.

schematisiert wurde – so darf man, wie weiter oben argumentiert, nicht vergessen, dass dennoch stets iterativ-reflexive Prinzipien angelegt wurden.

Abbildung 9. Schritte in einer systematischen Dokumentenanalyse



Eigene Visualisierung basierend auf „Public sector knowledge management: A structured literature review,“ von M. Massaro, J. Dumay und A. Garlatti, 2015, *Journal of Knowledge Management*, 19(3), S. 532–533.

5.2 Dokumentensuche und -auswahl

Der Prozess der Dokumentenerhebung startete mit einer Definition formaler Inklusions- und Exklusionskriterien für das Dokumentenkorpus. Diese Entscheidungen basierten auf der Evaluation anderer systematischer Literaturvergleiche (Fecher u. a., 2015; G. Guthrie & Murthy, 2009; Lecheler & Kruikemeier, 2016; Massaro u. a., 2015; Neuberger, Nuernbergk & Rischke, 2007). Obwohl zahlreiche hervorragende theoretische und konzeptionelle Publikationen existieren, sollten nur Publikationen mit empirischer Basis in die Analyse einbezogen werden, unabhängig davon, ob es sich um quantitative, qualitative oder Mixed-Methods-Ansätze handelte. Diese Entscheidung wurde einerseits getroffen, weil andere Metaanalysen einen ähnlichen Weg gewählt hatten, andererseits, weil die Operationalisierung der Forschungsfragen empirische Nähe verlangte. Der Fokus lag auf Publikationen aus den Sozialwissenschaften – und hier insbesondere aus der Kommunikationswissenschaft –, aber auch Publikationen aus anderen Disziplinen wurden inkludiert. In Hinblick auf die Art der Publikation wurden Zeitschriftenartikel, Buchkapitel, Konferenzbeiträge, Forschungsberichte und Dissertationen berücksichtigt. Aufgrund der hohen Zahl an Publikationen wurden Bachelor- und Masterarbeiten, Presseartikel sowie Blogposts ausgeschlossen. Lehrbücher wie etwa das *Data Journalism Handbook* (Gray u. a., 2012) oder *Precision Journalism* (P. Meyer, 1973/2002) wurden aufgrund ihrer Ziel-Leserschaft und ihres selbstberichtenden Stils ebenfalls nicht in Betracht gezogen. Alle zwischen 1996 und 2015 veröffentlichten Publikationen wurden inkludiert. Die vereinzelte Forschung, die vor 1996 veröffentlicht

wurde, schien eine ganz andere Praxis zu untersuchen. Zwei Journalbeiträge mit Erscheinungsdatum 2016, die online bereits 2015 verfügbar waren, wurden in die Analyse aufgenommen.

Der nächste Schritt war, eine Liste der relevanten Keywords zu erstellen, die als Suchbegriffe für die Dokumentrecherche aus wissenschaftlichen Datenbanken dienen. Dazu wurde eine vorläufige Suche auf *Google Scholar* mit den Begriffen ‚data-driven journalism‘ und ‚data journalism‘ durchgeführt und aus den Abstracts und Keyword-Abschnitten der Papiere einschlägige Begriffe extrahiert. Dies führte zu folgender Liste von Synonymen (Tabelle 1):

Tabelle 1. Synonyme Suchbegriffe für die Dokumentensuche

Synonym	
algorithmic journalism	datajournalism
data-driven reporting	data journalism
computational journalism	Datenjournalismus
database journalism	data-driven journalism
computer-assisted reporting	quantitative journalism

Während der deutsche Begriff ‚Datenjournalismus‘ inkludiert war, mussten die am meisten verbreiteten französischen und spanischen Ausdrücke (‚journalisme de données‘, ‚jornalismo de dados‘) aufgrund unzureichender Sprachkenntnisse ausgeschlossen werden. Weitere verwandte, allgemeinere Begriffe und Abkürzungen wie ‚accountability journalism‘, ‚crowdsourced journalism‘, ‚data visualization‘, ‚DDJ‘, ‚investigative journalism‘, ‚online journalism‘ und ‚open journalism‘ wurden zwar getestet, aber nicht in die endgültige Liste aufgenommen. Die Suchen, die diesen Begriffen zugrunde lagen, lieferten zu ungenaue Ergebnisse.

Basierend auf den synonymen Suchbegriffen aus Tabelle 1 wurde eine Literatursuche in 15 wissenschaftlichen Datenbanken durchgeführt und wurden insgesamt 772 Datensätze heruntergeladen. Die Suchbegriffe wurden mit dem Booleschen ODER-Operator verknüpft.

Tabelle 2. Durchsuchte wissenschaftliche Datenbanken

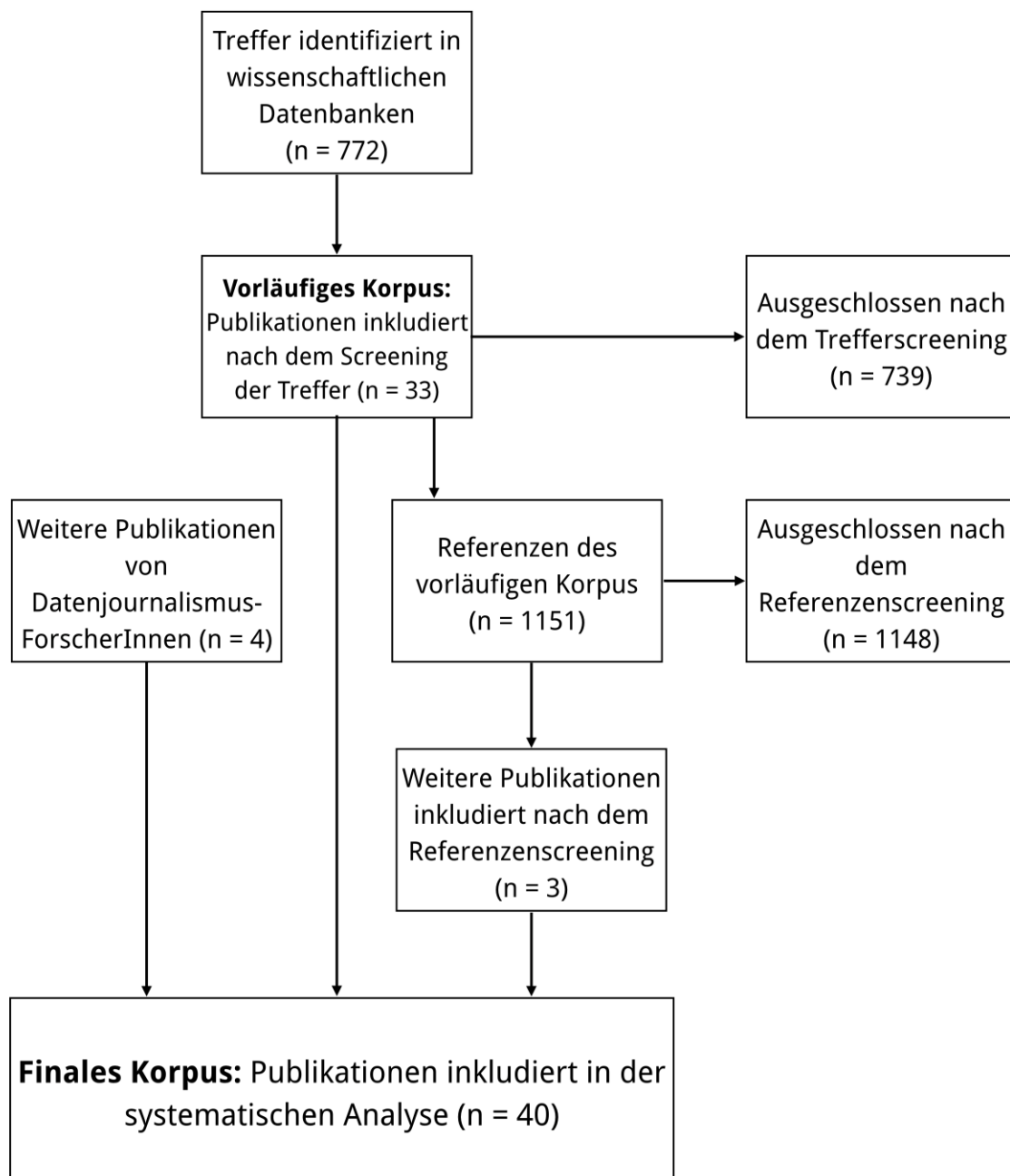
Datenbank	Treffer
ACM Digital	112 Sociological Abstracts 5
EBSCO	19 Sowiport 53
Google Scholar	400 Springer 29
IEEE	26 SpringerLink 135
JSTOR	33 Taylor & Francis Online 144
ProQuest	6 Web of Science 43
Science Direct	69 Wiley 73
Scopus	25
Summe	
772	

Google Scholar lieferte 3290 Treffer, jedoch wurden nur die ersten 400 importiert, die darauffolgenden Ergebnislisten lieferten keine relevanten Treffer mehr.

Alle Treffer wurden in eine Zotero-Datenbank (Roy Rosenzweig Center for History and New Media, 2015) importiert. Gemeinsam mit einem anderen Forscher, Robert Gutounig, bewertete ich alle Treffer unter Berücksichtigung von Titel, Zusammenfassung und Schlüsselwörtern. Unabhängig voneinander sortierten wir die Literatur nach den oben genannten formalen Kriterien und Forschungsschwerpunkten. Publikationen, die beide Forscher als ‚nicht relevant‘ markierten, wurden ausgeschieden. Diejenigen Veröffentlichungen, die beide Forscher als ‚relevant‘ markiert hatten, wurden in das vorläufige Korpus aufgenommen. Divergierende Beurteilungen wurden von uns diskutiert, bis eine Einigung erzielt wurde. Dieses Screening führte zu einem vorläufigen Korpus, den wir anschließend von Duplikaten reinigten.

Zwei weitere Maßnahmen wurden getroffen, um die Literatur für das Korpus auszuwählen. Erstens wurden drei Domain-ExpertInnen, alle ForscherInnen auf dem Gebiet des Datenjournalismus, gebeten, entsprechende Arbeiten aufgrund ihres impliziten Wissens hinzuzufügen. Zweitens wurden die Referenzen unserer Auswahl computergestützt und manuell überprüft: Einerseits extrahierte Michael Oppermann, Informatik-Student an der Universität Wien, die Referenzen aller ausgewählten Veröffentlichungen aus dem vorläufigen Korpus mittels eines bibliografischen Datenerkennungsalgorithmus (Lopez, 2009). Alle Referenzen, die mindestens von zwei Publikationen aus dem vorläufigen Korpus zitiert wurden, sowie alle Referenzen, die 2015 und 2016 veröffentlicht wurden, wurden erneut manuell von Robert Gutounig und mir getrennt auf ihre Eignung für die Dokumentenanalyse überprüft. Dieser Prozess führte zu einem Korpus von 40 Publikationen. Abbildung 10 veranschaulicht die Dokumentensuche und -auswahl in einem Flussdiagramm:

Abbildung 10. Ablauf der Dokumentensuche und -auswahl



Adaptiert von „What drives academic data sharing?“ von B. Fecher, S. Friesike und M. Hebing, 2015, *PLoS ONE*, 10(2), S. 4.

5.3 Quantitative und qualitative Inhaltsanalysen

Das finale Korpus von 40 Forschungspublikationen war der Ausgangspunkt für unterschiedliche quantitative und qualitative Inhaltsanalysen, in denen sowohl Struktur als auch Inhalt der Publikationen untersucht und interpretiert wurden.

Zunächst wurden verschiedene strukturelle Aspekte jeder Publikation extrahiert und miteinander verglichen. Erfasst wurden die Namen der AutorInnen, deren institutionelle Zugehörigkeit, das Veröffentlichungsjahr, der Publikationstyp (Zeitschriftenartikel, Konferenzbeitrag, Buchkapitel, Bericht, Dissertation), der Titel der Veröffentlichung, und – bei Zeitschriftenartikeln – der Titel des Journals. Zusätzlich wurde von Michael Oppermann mithilfe des *scholar.py*-Python-Moduls (Kreibich, 2016) aufgezeichnet, wie oft jede Publikation laut *Google Scholar* zitiert wurde. Außerdem wandte er den bibliografischen Datenerkennungsalgorithmus (Lopez, 2009), den er bereits zur Vervollständigung des vorläufigen Korpus angewandt hatte, auf den finalen Korpus an und extrahierte sämtliche Referenzen. Diese wurden gesäubert, vereinheitlicht, von Duplikaten befreit, mit *Google Scholar* referenziert und in einem Netzwerkanalyse-freundlichen Format analysiert, um einflussreiche Publikationen zu identifizieren.

Parallel zur rechnerischen Exploration des Korpus wurde eine softwaregestützte qualitative Inhaltsanalyse durchgeführt (Kaefer, Roper & Sinha, 2015; Schreier, 2012), beginnend mit der Entwicklung eines Codebuchs. Während die Kategorien des Codebuchs durch das Forschungsinteresse bzw. die Forschungsfragen definiert wurden, wurden die Unterkategorien manuell aus dem Material abgeleitet. Das Kategoriensystem wurde zusätzlich informiert durch das Material aus der Interviewstudie, die teils parallel zum systematischen Literaturüberblick durchgeführt und für deren Analysearbeit dasselbe Codebuch herangezogen wurde. Dieses induktive Codierungsverfahren trägt dazu bei, neuartige interpretative Verbindungen zu entdecken, die stärker auf dem Material basieren und weniger auf einem konzeptionellen Vorverständnis durch die Forschenden. In einem Pre-Test kodierte ich mit zwei ForschungskollegInnen, Robert Gutounig und Sarah Matiasek, zwei Publikationen. Dabei diskutierten und überarbeiteten wir laufend das Kategoriensystem. Anschließend wurden die Codierungen verglichen, Unterschiede erörtert und das Codebuch angepasst, bis ein Übereinkommen über die Codes erreicht wurde. Für das Codieren wurde die qualitative Datenanalyse-Software Nvivo (QSR International, 2016) eingesetzt.

Schließlich wurde die qualitative Inhaltsanalyse noch um eine automatisierte quantitative Inhaltsanalyse ergänzt, die auf eine Python-Implementierung des TF-IDF-Algorithmus zurückgriff, um die zentralen Begriffe jeder Publikation zu extrahieren. TF-IDF, der Kurzausdruck für ‚term frequency–inverse document frequency‘, ist ein Maß, das die Wichtigkeit eines Wortes innerhalb eines individuellen Dokuments im Verhältnis zu einem Korpus aus Dokumenten anzeigt (K. S. Jones, 1972). Im Unterschied zu einfachen ‚term

frequency'-Verfahren, in denen die Häufigkeit eines Ausdrucks innerhalb eines Dokuments gezählt und mit den Häufigkeiten des gleichen Ausdrucks in den anderen Korpusdokumenten verglichen wird, werden beim TF-IDF-Maß häufige Ausdrücke eines Dokuments stärker gewichtet, wenn sie seltener in den anderen Korpusdokumenten vorkommen.

Dieser Mixed-Methods-Ansatz lieferte unter anderem Einblicke in den Stand und die Entwicklung des Forschungsfelds sowie in die Produktionsstätten, Theorien, Forschungsdesigns und Forschungslücken der empirischen Datenjournalismusforschung. Auch konnte, aufbauend auf diesem Ansatz, ein systematischer Überblick über datenintensive Arbeitsabläufe in Redaktionen gegeben werden.

5.4 Ergebnisse der systematischen Dokumentenanalyse

Die Resultate, die im Folgenden vorgestellt werden, sind anhand der in der Einleitung formulierten Forschungsfragen strukturiert. Der erste Unterabschnitt beschreibt den aktuellen Stand und die Entwicklung der Datenjournalismusforschung, unter anderem durch eine Zitationsanalyse und die Berücksichtigung von Kollaborationsstrukturen. Die darauffolgenden drei Unterabschnitte explorieren die theoretischen und methodischen Schwerpunkte der bisherigen Forschung und zeigen Forschungslücken auf. Bei der daran anschließenden Erläuterung von datenintensiven Arbeitsabläufen werden insbesondere die Aspekte Themenfindung, Datensammlung, -säuberung, -verifikation, -analyse, Produktion und Visualisierung sowie Veröffentlichung und Evaluierung besprochen.

5.4.1 Korpus-Charakteristiken

Die erste Forschungsfrage dieser Arbeit lautete: Wie entwickelt sich die Forschungsliteratur zu datenintensivem Journalismus, unter anderem in Bezug auf Publikationsaktivität, Veröffentlichungen und Zitationen? Abbildung 11 zeigt die jährlich veröffentlichten Forschungspublikationen zu Datenjournalismus seit 1996. Obwohl ‚computer-assisted reporting‘ bereits seit den 1960ern vor allem im anglo-amerikanischen Raum praktiziert wird (M. Cox, 2000), begann die seriöse wissenschaftliche Auseinandersetzung mit dem Phänomen erst in den letzten Jahren. Vor 2010 beschäftigte sich nur eine kleine Anzahl isolierter Forschungspublikationen mit dem Thema Daten und Journalismus. Die jüngste Intensivierung der Forschungsaktivität muss in Verbindung mit der Zusammenführung oder (Um-)Benennung

von verschiedenen datenintensiven Methoden im Journalismus in die Oberbegriffe ‚Datenjournalismus‘ bzw. ‚data journalism‘ oder ‚data-driven journalism‘ in den späten 2000ern gesehen werden. Dieses Reframing schien das Interesse von PraktikerInnen wie auch der Journalismusforschung gleichermaßen geweckt zu haben.

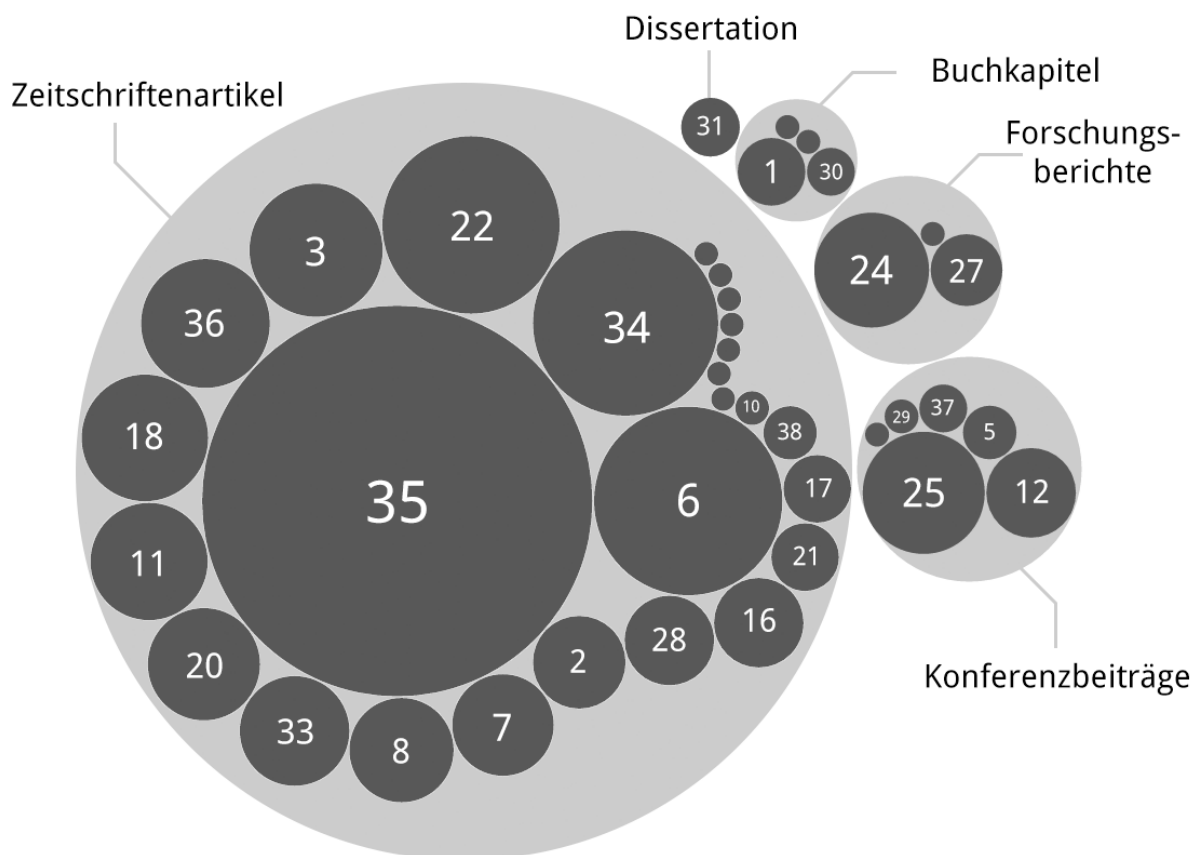
Abbildung 11. Entwicklung der Literatur über die Zeit



Jeder Block repräsentiert eine Publikation aus dem Korpus. Die zwei Publikationen aus 2016 wurden bereits 2015 – also innerhalb der Evaluierungsperiode – online veröffentlicht und konnten daher in der Analyse berücksichtigt werden.

Die größte Anzahl an Arbeiten, 26 der 40 Publikationen, wurde in wissenschaftlichen Fachzeitschriften veröffentlicht. Hier sind insbesondere die Zeitschriften der Journalismusforschung prominent vertreten: *Digital Journalism* (6 Publikationen), *Journalism Studies* (3), *Journalism* (2) und *Journalism Practice* (2) sind die populärsten Publikationsmedien. *Digital Journalism* schuldet seine Führungsposition der von Seth Lewis (2015) herausgegebenen Schwerpunktausgabe zum Thema „Journalism in the era of big data“. Zeitschriftenaufsätze sind nicht nur die populärsten Disseminationsformate, sie sind durchschnittlich auch signifikant öfter zitiert als Konferenzbeiträge (6), Buchkapitel (4) und andere Publikationstypen (4) (siehe Abbildung 12). Die meistzitierte Arbeit im Korpus (Segel & Heer, 2010) wurde in einem Informatik-Journal veröffentlicht.

Abbildung 12. Publikationen nach Typ und Anzahl an Zitationen



Jeder dunkle Kreis repräsentiert eine Publikation. Die Zahlen können in Tabelle 4 referenziert werden. Der Durchmesser des Kreises repräsentiert relational die Anzahl der Zitationen auf Google Scholar vom 12. April 2016.

5.4.1.1 Einflussreiche Publikationen

Was sind die einflussreichsten Publikationen im Bereich des Datenjournalismus bislang? Das Zählen der Referenzen kann eine erste Antwort darauf geben, wiewohl angemerkt werden muss, dass eine hohe Anzahl an Zitationen nicht notwendigerweise hohen Einfluss mit sich bringt; schließlich können auch wenig einflussreiche Publikationen oft zitiert worden sein – etwa Arbeiten von PraktikerInnen – bzw. auch Zitationskartelle vorliegen. Dennoch liefert die Anzahl der Nennungen Hinweise auf die Wichtigkeit einer Publikation. Nicht von ungefähr bauen die Systeme von *Google Scholar* und zahlreicher Metriken zur Messung von wissenschaftlichem Einfluss auf dem Zählen von Zitationen auf.

Phil Meyers (1973/2002) Buch *Precision Journalism*, das seit den 1970ern wichtige Narrative für die Computer-assisted-Reporting-Gemeinschaft in den USA lieferte, ist die meistzitierte Publikation im Korpus, gemeinsam mit Parasies und Dagirals (2013) Untersuchung der Epistemologien und Praktiken von DatenjournalistInnen in Chicago. Deren

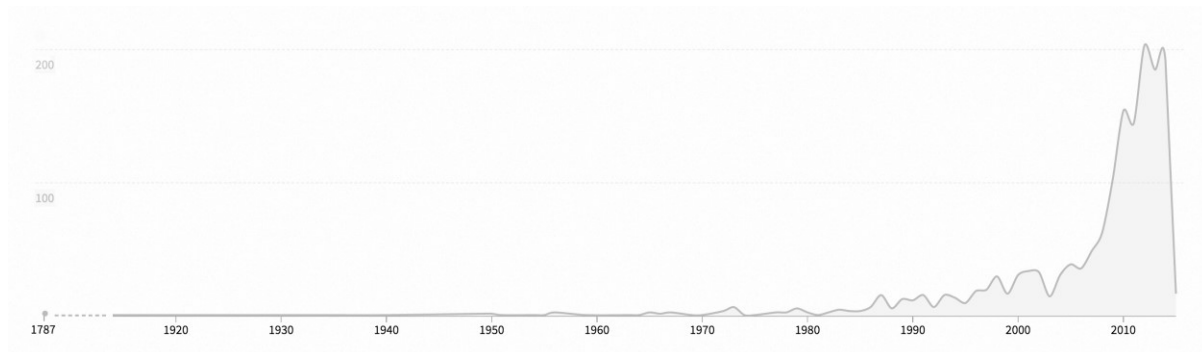
Beitrag kann als ein prototypisches Stück zeitgenössischer Datenjournalismusforschung angesehen werden, weil deren theoretisches Framework, Forschungsmethoden und Untersuchungsperspektive in vielen späteren Publikationen wieder auftauchen. Beide Publikationen, jene von Meyer sowie jene von Parasie und Dagiral, wurden von 15 der 40 Publikationen im Korpus zitiert. Andere oft zitierte und damit einflussreiche Publikationen, die den Diskurs um Datenjournalismus prägen, können teils als Advocacy-Beiträge klassifiziert werden (vgl. Gynnild, 2014, S. 714), die sich durch technologische Entwicklung die Verbesserung der journalistischen Praxis oder demokratischer Standards erhoffen (S. Cohen u. a., 2011a; S. Cohen, Li, Yang & Yu, 2011b; Flew u. a., 2012; Hamilton & Turner, 2009). Tabelle 3 listet die zehn meistzitierten Publikationen des Korpus auf:

Tabelle 3. Die meistzitierten Referenzen im Korpus

Publikation	Anzahl an Zitationen
P. Meyer (1973/2002)	15
Parasie & Dagiral (2013)	15
Gray u. a. (2012)	13
Flew u. a. (2012)	11
Hamilton & Turner (2009)	10
Royal (2012)	10
S. Cohen u. a. (2011a)	8
M. Cox (2000)	7
S. Cohen u. a. (2011b)	7
Lewis & Usher (2013)	6

Temporär verteilen sich die Referenzen über einen Zeitraum von 1787, der US-amerikanischen Unabhängigkeitserklärung, bis heute, wobei ab Mitte der 2000er die meisten Publikationen erschienen sind. Mit anderen Worten: Die meisten Publikationen, auf die sich Datenjournalismusforschung bezieht, wurden seit 2005 veröffentlicht. Abbildung 13 illustriert diese Entwicklung:

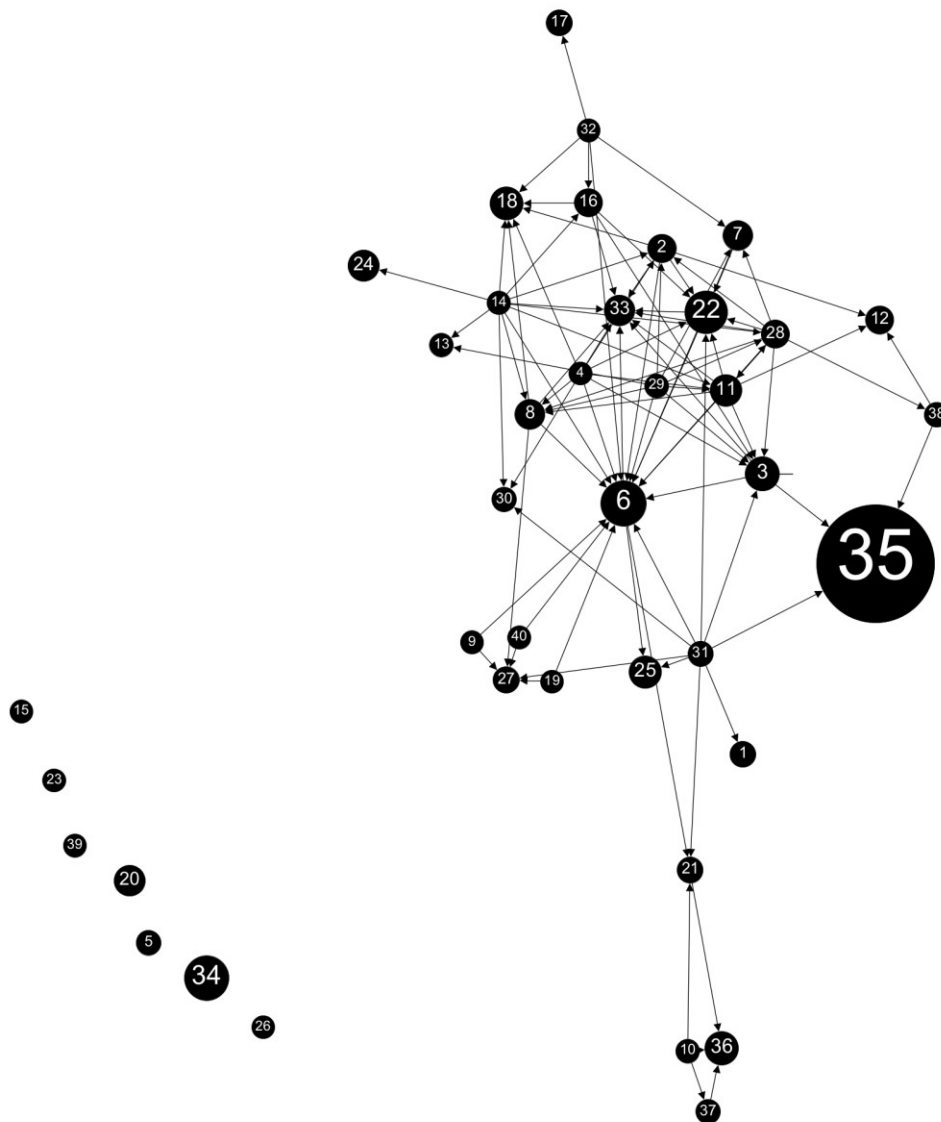
Abbildung 13. Referenzen nach Anzahl und Veröffentlichungsjahr



Eine weitere Perspektive auf den Diskurs über Datenjournalismus liefern die Abbildungen 14, 15 und 16. Hierfür wurden sämtliche Korpuspublikationen und deren Referenzen in Netzwerken visualisiert (für Details zur Methode siehe den Abschnitt Inhaltsanalysen). Schwarze Knoten repräsentieren eine Publikation aus dem Korpus, graue Knoten repräsentieren Referenzen. Die Zahlen in den Graphen können in Tabelle 4 referenziert werden. Der Durchmesser der schwarzen Knoten entspricht relational der Anzahl an Zitationen auf *Google Scholar*.³⁷

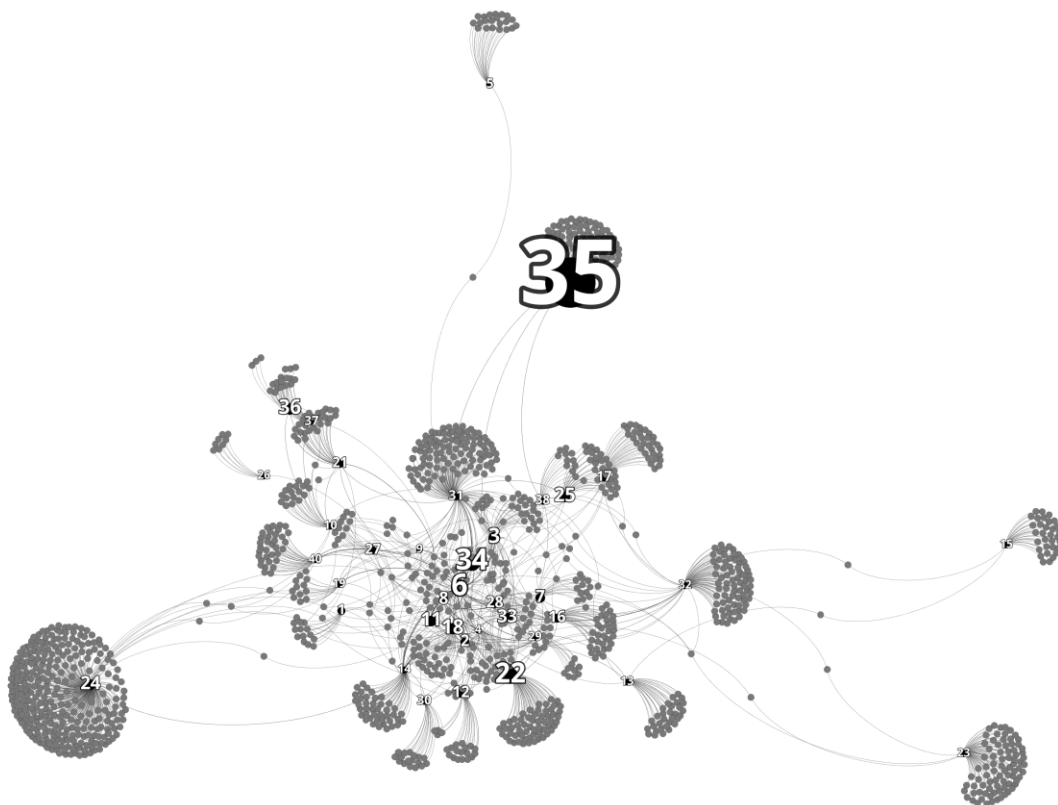
³⁷ Datum der Abfrage: 12. April 2016.

Abbildung 14. Zitationsnetzwerk



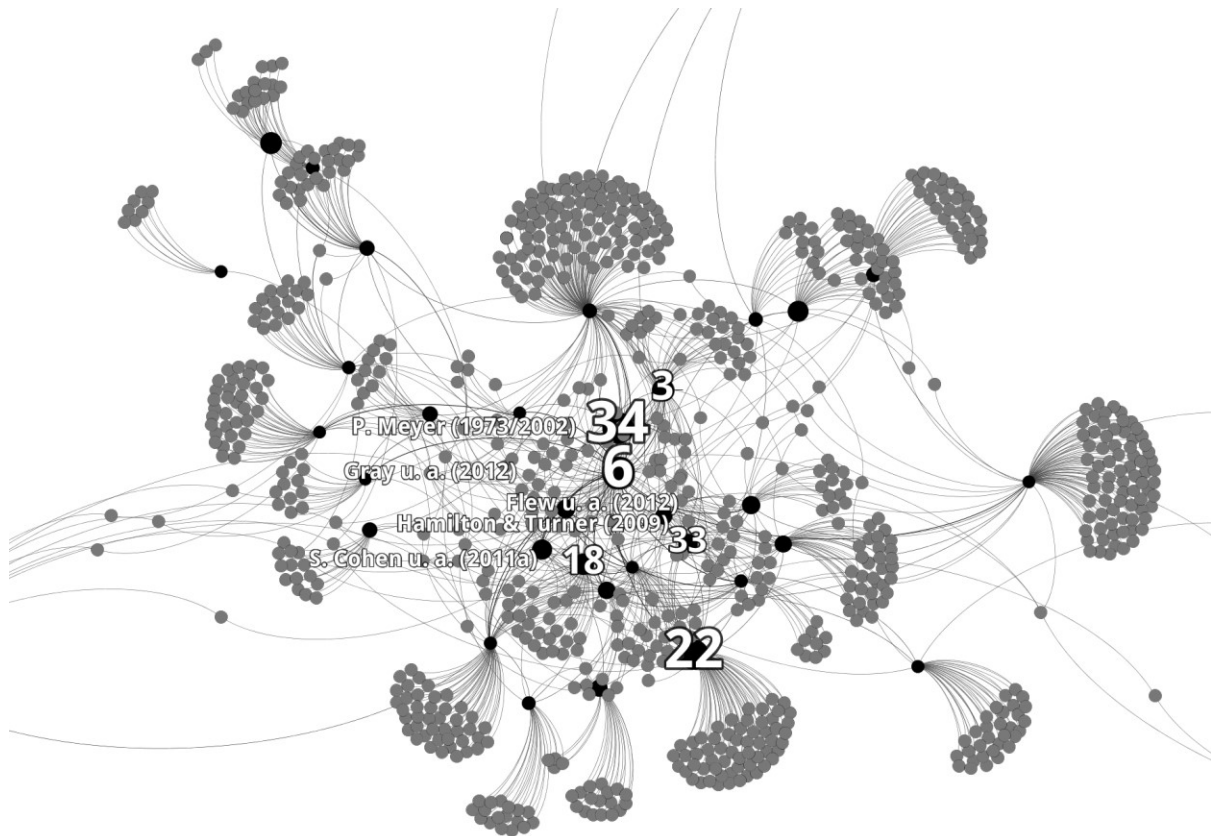
Das Netzwerk zeigt die Zitationen innerhalb des Korpus. Je zentraler ein Beitrag im Netzwerk, von desto mehr anderen Publikationen wurde er zitiert. Eine Zitation von vielen Publikationen ist ein Indikator für die Wichtigkeit eines Beitrags innerhalb des Korpus. Zu den Publikationen mit dem höchsten Einfluss (,eigenvector centrality' = Knotengröße) im Korpus zählen Royal (2012) (#33), Parasie & Dagiral (2013) (#6) sowie Lewis & Usher (2013) (#22). Knoten $n_n=40$, Kanten $n_e=95$. Netzwerklayout: Force Atlas.

Abbildung 15. Netzwerk von Korpus-Publikationen und Referenzen



Dieses Netzwerk stellt die Korpuspublikationen und ihre (geteilten) Referenzen dar. Durch diese Visualisierung wird etwa offenbar, dass A. B. Howard (2014) (#24) und Stavelin (2013) (#31) auf die meisten anderen Arbeiten verweisen. Das verwundert wenig, sind es doch auch die umfangreichsten Publikationen der Arbeit. Auch zeigt das Netzwerk, dass keine Korpuspublikation völlig isoliert ist. Jede Publikation teilt zumindest eine Referenz mit einer anderen Publikation. Im Kern des Netzwerks wiederum gibt es eine Reihe von Korpuspublikationen, die vielfach aufeinander verweisen (siehe dazu auch Abbildung 14) und eine Reihe von Referenzen teilen. Knoten $n_n=1367$, davon Korpuspublikationen (in schwarz) $n_{nb}=40$ und Referenzen (in grau) $n_{ng}=1327$; Kanten $n_e=1655$.
Netzwerklayout: Yifan Hu.

Abbildung 16. Netzwerk von Korpus-Publikationen und Referenzen (Ausschnitt)



Hierbei handelt es sich um einen Ausschnitt von Abbildung 15 mit anderer Beschriftung. Das Netzwerk nennt die von den Korpuspublikationen meistzitierten Publikationen – egal ob es sich dabei um einen Teil des Korpus oder um Referenzen handelt. Das Netzwerk kontextualisiert damit die Ergebnisse von Tabelle 3 (oben). Die Publikationen, die in diesem Ausschnitt vorkommen, bilden gleichsam den Kern der Datenjournalismusforschung.

Bei aller Einsicht, die die Abbildungen 14, 15 und 16 bieten, soll zugleich kritisch angemerkt werden, dass derartige Zitationsanalysen die Bedeutung von älteren Publikationen besser charakterisieren als jene von jüngeren Publikationen. Schließlich können neuere Publikationen nur ältere zitieren – und nicht umgekehrt. Ein Netzwerk, das sich hauptsächlich aus Publikationen jüngeren Datums zusammensetzt, kann deshalb immer nur ein vorläufiges sein, schließlich zeigt sich der Einfluss einer Publikation erst über längere Zeit.

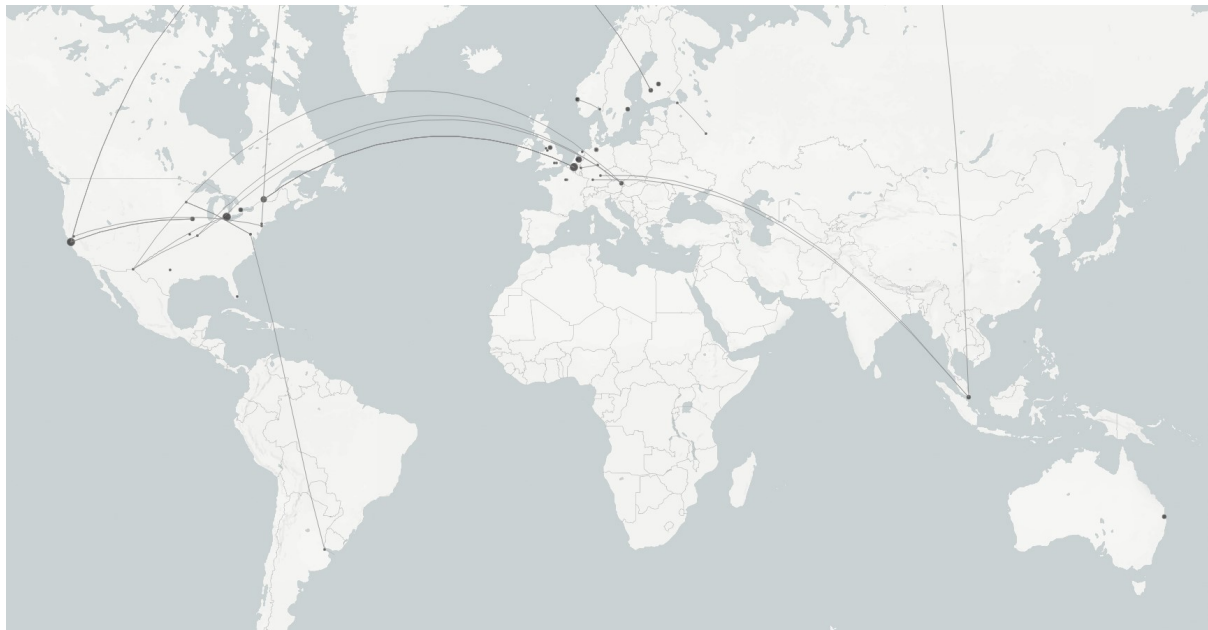
5.4.2 Produktionsstätten der Datenjournalismusforschung

Wo und wie wird zu Datenjournalismus geforscht? Um eine Antwort auf diese Frage zu erhalten, wurde die in jeder Publikation erstgenannte institutionelle Anbindung jedes Autors/jeder Autorin exzerpiert und geocodiert. Es zeigte sich, dass die meisten Forschenden zu Datenjournalismus in Institutionen in Europa (38) und Nordamerika (29) arbeiten. Von den 71 AutorInnen im Korpus waren 22 mit Institutionen in den Vereinigten Staaten verbunden.

Innerhalb der USA wurde ein starkes Ost-West-Gefälle offenbar: Die meisten Forschenden waren an Ostküsten-Institutionen tätig, während nur wenige an Westküsten-Universitäten tätig waren.

Forschung zu Datenjournalismus ist eine kollaborative Angelegenheit. Mehr als zwei Drittel (27) aller Datenjournalismus-Publikationen wurden von mehr als einer Person verfasst. Zum Vergleich: Im Jahr 2000 basierte etwa die Hälfte aller sozialwissenschaftlichen Arbeiten auf Kollaborationen (Wuchty, Jones & Uzzi, 2007). Sieben Publikationen basierten auf internationalen Kollaborationen. Abbildung 17 stellt die institutionelle Zugehörigkeit der AutorInnen auf einer Karte dar. Bemerkenswert ist die große Anzahl an transatlantischen Kollaborationen.

Abbildung 17. Institutionelle Anbindungen und Kooperationen in der Datenjournalismusforschung



Größere Punkte stehen für mehrere AutorInnen an einer Institution.

5.4.3 Theorien der empirischen Datenjournalismusforschung

VertreterInnen verschiedenster Fächer und Denkrichtungen betreiben Datenjournalismusforschung. Sowohl in den Geisteswissenschaften, den Sozialwissenschaften als auch den Technikwissenschaften gibt es Menschen, die sich mit dem Thema auseinandersetzen. Es ist deshalb wenig verwunderlich, dass sehr unterschiedliche Zugänge existieren, um das Phänomen zu rahmen und zu kontextualisieren.

Zugleich fällt auf, dass anteilig relativ wenige empirische Arbeiten überhaupt auf theoretische Ansätze rekurrieren. Das mag ein generelles Problem der Journalismusforschung sein (Scholl, 2011), doch in der empirischen Datenjournalismusforschung scheint die Theoriearmut besonders weit verbreitet zu sein. Das hat wohl damit zu tun, dass sich die angloamerikanische Journalismusforschung, die viel zur Datenjournalismusforschung beigetragen hat, seit jeher stärker an der Praxis, an empirischer Nähe orientiert als etwa die deutschsprachige Journalismusforschung (vgl. Wahl-Jorgensen & Hanitzsch, 2009, S. 5). Zugleich hängt diese Diagnose auch mit dem Auswahlverfahren (siehe oben) zusammen, das rein konzeptionelle Arbeiten für die Dokumentenanalyse ausschloss.

Viele Publikationen mit Theoriebezug sehen Datenjournalismus, ein Feld, das praktisch wie auch diskursiv von der Verschränkung von Journalismus und Technik lebt, entsprechend durch die Brille der Science and Technology Studies (STS) (Ausserhofer, 2015; Lewis & Usher, 2013, 2014; Stavelin, 2013). Aus den verschiedenen Ansätzen der STS sticht die Akteur-Netzwerk-Theorie (ANT) besonders hervor, die weiter oben im Kapitel zu wissenschaftstheoretischen Grundannahmen bereits diskutiert wurde. Eine große Anzahl an AutorInnen versucht, die ANT mit Datenjournalismus in Verbindung zu bringen (Ausserhofer, 2015; De Maeyer u. a., 2015; Parasie, 2015; Parasie & Dagiral, 2013; Stavelin, 2013). Die ANT ist die beliebteste Rahmentheorie des zeitgenössischen akademischen Datenjournalismusdiskurses.

Aber es finden sich auch einige Publikationen, die andere Wege eingeschlagen haben – meist, um die erforschten Aspekte rund um die datenintensive journalistische Praxis zu kontextualisieren:

Ebenfalls aus den STS – konkret der Wissenschaftsgeschichte und –philosophie – entlehnt, ist das Konzept der „trading zones“ (Galison, 1997), das zum besseren Verständnis der Foren der Datenjournalismuscommunity eingebracht wird (Lewis & Usher, 2014). ‚Trading zones‘ können virtuelle oder physische Orte des Austausches sein, in denen sich VertreterInnen verschiedener Disziplinen miteinander vernetzen.

Karlsen und Stavelin (2014) wiederum rekurrieren bei ihrer Rahmung der Praxis des datenintensiven Journalismus auf den aristotelischen Begriff der ‚techne‘ und sehen datenintensiven Journalismus als rhetorisches Handwerk, das im Unterschied zu ‚epistémé‘ (der Wissensdomäne) auf praktische Umsetzung ausgerichtet ist.

Weinacht und Spiller (2014) konzentrieren sich in ihrer Arbeit auf die Rollenbilder von DatenjournalistInnen in Deutschland. Diese Perspektive schärfen sie durch Verweise auf die struktur-funktionalistische Rollentheorie, unter anderem entwickelt von Linton, Merton und Parsons. Diese Theorie beschreibt die kontextabhängige Ausprägung von Rollen in verschiedenen Situationen.

5.4.4 Forschungsdesigns der Datenjournalismusforschung

Nachdem im vorangegangenen Abschnitt die theoretischen Kontexte der empirischen Datenjournalismusforschung diskutiert wurden, stehen hier deren konkrete Forschungsdesigns im Zentrum. Tabelle 4 fasst die gewählten Methoden und geografischen Forschungsschwerpunkte der einzelnen Publikationen zusammen. Zahlreichen qualitativen, explorativen Verfahren stehen einige wenige quantitative Forschungsdesigns gegenüber. Mit Ausnahme der Arbeit von Appelgren und Nygren (2014a, 2014b) gibt es keine jüngere Datenjournalismusforschung, die auf Umfragen aufbaut. Dieser starke Überhang qualitativer Verfahren ist nichts Ungewöhnliches, schließlich müssen die Charakteristiken der doch relativ neuen Praxis erforscht und definiert werden, bevor ein quantitativ orientiertes Forschungsdesign auf den Ideen aufbauen kann.

5.4.4.1 Qualitative Interviews

Qualitative Interviews waren die am häufigsten angewandte Methode innerhalb des untersuchten Literaturkorpus. 25 von 40 Publikationen griffen auf Interviews zurück. Die meisten Forschenden führten die Interviews nach einem semi-strukturierten Leitfaden durch. Befragt wurden sowohl PraktikerInnen wie auch ExpertInnen. Die Anzahl der für jede Publikation Befragten variierte zwischen fünf und 35. Howard (A. B. Howard, 2014) sprach sogar mit über 100 Personen, die sich mit datenintensivem Journalismus beschäftigten. Die meisten Interviews wurden persönlich durchgeführt, gefolgt von Telefoninterviews und Gesprächen über Skype. In vielen Fällen wurden Interviews mit anderen Methoden, insbesondere Verfahren der Inhaltsanalyse, kombiniert. Maeyer et al. (2015) etwa wählten ein solches Forschungsdesign, das durchaus als prototypisch angesehen werden kann: Sie führten 20 semi-strukturierte Interviews mit Personen durch, die an Datenjournalismus in Belgien beteiligt waren, und kombinierten diese Daten mit der Analyse von Dokumenten und Artefakten.

5.4.4.2 Inhaltsanalysen

Die zweithäufigste Methode der Datenerhebung war die Inhaltsanalyse. Insgesamt konnten 21 Publikationen identifiziert werden, die systematisch Texte, Bilder oder andere Inhalte analysierten. Auffällig ist, dass nur sehr wenige AutorInnen auch tatsächlich datenjournalistische Beiträge untersuchten. Die meisten beschäftigen sich mit Meta-Inhalten, die indirekt Rückschlüsse auf die datenjournalistische Praxis zuließen, wie etwa Stellenanzeigen (Bakker, 2014), Nutzerkommentare unter Visualisierungen (Hullman u. a., 2015) oder einschlägige Mailinglisten und Blogs, die über Journalismusinnovation berichten (Gynnild, 2014).

Als ‚klassische‘ Inhaltsanalyse von datenintensiven Medienprodukten kann etwa das Paper von Lugo-Ocando und Brandão (2016) klassifiziert werden, das evaluierte, inwieweit Statistiken in britischen Boulevard- und Qualitätsmedien richtig interpretiert wurden. In die Kategorie der ‚klassischen‘ Inhaltsanalyse fällt auch die Publikation von Tandoc und Oh (2015), die 260 Beiträge des *Guardian Datablog* hinsichtlich ihrer Nachrichtenwerte, Quellen und Themen analysierten. Loosen, Reimer und Schmidt (2015) beschäftigten sich mit jenen 120 journalistischen Produkten, die 2013 und 2014 für die internationalen *Data Journalism Awards* nominiert waren. Hier wurden unter anderem die Themen, Datenquellen, Visualisierungsstrategien sowie Hintergründe zum veröffentlichenden Medium erfasst. Segel und Heer (2010) klassifizierten 58 Visualisierungsbeispiele aus Online-Medien.

5.4.4.3 Andere Methoden

Wenig verbreitet in der Datenjournalismusforschung sind Umfragen mittels (Online-)Fragebogen. Eingesetzt wurden sie in den letzten Jahren nur von Appelgren und Nygren (2014a, 2014b) – in Kombination mit qualitativen Interviews. Beobachtungen hingegen werden häufiger angewendet. Hervorstechend ist die Publikation von Royal (2012), die Mitglieder des *Interactive News Technology Department* der *New York Times* für mehrere Tage bei ihrer Arbeit beobachtete. Die Arbeit wurde aufgrund der längeren Beobachtungsdauer als einzige als Newsroom-Ethnografie gewertet, die durchaus als ein eigenes Genre innerhalb der Journalismusforschung angesehen werden kann (Domingo & Paterson, 2011; z.B. Eliasoph, 1988; Gans, 1979/2004; Tuchman, 1978; Usher, 2014, 2016). Smit, Haan und Buys (2014) nahmen an Redaktionstreffen und Brainstormingsitzungen einer führenden

Sendeanstalt in den Niederlanden teil, während Dick (2014) acht Stunden lang das *BBC News Online Specials Team* beobachtete.

Tabelle 4. Forschungsmethoden, geografischer Fokus und Themen der Datenjournalismusforschung

Id	Publikation	QI	IA	Andere Methode	Örtlicher Fokus (Land)	Thema
1	Baack (2013)			Fallstudie	n. a.	Fallstudie: <i>Wikileaks</i>
2	Lewis & Usher (2014)	X		Beobachtung (kurz)	International	Fallstudie: <i>Hacks/Hackers</i>
3	Karlsen & Stavelin (2014)	X			Norwegen	DIJ in Norwegen
4	Hannaford (2015)	X	X		England	DIJ bei der <i>BBC</i> und der <i>Financial Times</i>
5	Hullman u. a. (2015)		X		England	Nutzerkommentare unter dem Graphic Detail Blog des <i>Economist</i>
6	Parasie & Dagiral (2013)	X	X		USA	DIJ bei der <i>Chicago Tribune</i>
7	Parasie (2015)	X	X		USA	Epistemologie von DIJ
8	Appelgren & Nygren (2014b)	X		Umfrage	Schweden	DIJ bei schwedischen Nachrichtenorganisationen
9	Appelgren & Nygren (2014a)	X	X	Umfrage	Schweden	DIJ bei schwedischen Nachrichtenorganisationen
10	Knight (2015)		X		England	DIJ bei Tages- und Sonntagsblättern
11	Fink & Anderson (2015)	X			USA	DIJ bei kleinen, mittelgroßen und großen Zeitungen
12	Weber & Rall (2012)	X			USA, Deutschland, Schweiz	DIJ bei verschiedenen Onlinetiteln
13	Weinacht & Spiller (2014)	X			Deutschland	Selbstwahrnehmung von deutschen DatenjournalistInnen
14	Ausserhofer (2015)	X			International	Arbeitsabläufe von DIJ
15	Felle (2016)	X			International	DIJ und die vierte Säule der Demokratie
16	Young & Hermida (2015)	X	X	Beobachtung (kurz)	USA	Chronikjournalismus bei der <i>LA Times</i>
17	Dick (2014)	X		Beobachtung (kurz)	England	Infografiken bei Onlinezeitungen
18	Gynnild (2014)		X		England, USA	Innovation(rhetorik) im Journalismus
19	Uskali & Kuutti (2015)	X	X		Finnland, England, USA	DIJ-Praktiken in unterschiedlichen Newsrooms
20	Bakker (2014)		X		Niederlande	Rollenwandel von JournalistInnen
21	Garrison (1999)			Umfrage	USA	DIJ-Ressourcen von Tageszeitungen
22	Lewis & Usher (2013)			Close reading	n. a.	Open Source im Journalismus

Id	Publikation	QI	IA	Andere Methode	Örtlicher Fokus (Land)	Thema
23	Lugo-Ocando & Brandão (2016)	X	X		England	Kriminalberichterstattung und Statistik
24	A. B. Howard (2014)	X			International	Genereller Status von DIJ
25	Flew u. a. (2010)		X		England	Die Berichterstattung des <i>Guardian</i> über einen Spesenskandal
26	Radchenko & Sakoyan (2014)		X		Russland	Offene Daten und DIJ in Russland
27	Aitamurto u. a. (2011)	X			England, USA, Argentinien	Genereller Status von DIJ
28	De Maeyer u. a. (2015)	X	X		Belgien	Regional DIJ und sein Diskurs
29	Loosen u. a. (2015)		X		International	Für die International Data Journalism Awards nominierte Beiträge
30	Weber & Rall (2013)	X			USA, Deutschland, Schweiz	DIJ bei verschiedenen Onlinemedien
31	Stavelin (2013)	X	X		Norwegen	DIJ in Norwegen
32	Tandoc & Oh (2015)		X		England	Nachrichtenergebnisse, Normen und Routinen beim <i>Guardian Datablog</i>
33	Royal (2012)			Newsroom-Ethnografie	USA	DIJ bei der <i>New York Times</i>
34	S. Cohen u. a. (2011a)	X			n. a.	Einfluss von DIJ auf investigativen Journalismus
35	Segel & Heer (2010)		X		England, USA	Narrative Visualisierungen im Onlinejournalismus und verwandten Branchen
36	L. D. Davenport, Fico & Weinstock (1996)			Umfrage	USA	Computereinsatz in Michigans Redaktionen
37	L. D. Davenport u. a. (2000)			Umfrage	USA	DIJ in Michigan Newsrooms
38	Smit u. a. (2014)	X	X	Beobachtung (kurz)	Niederlande	Produktion von Nachrichtervisualisierung
39	Zanchelli & Crucianelli (2012)	X	X		England, USA, Brasilien	DIJ-Prozesse in Newsrooms
40	Tabary u. a. (2016)	X	X		Kanada	DIJ in Quebec

QI = Qualitative Interviews, IA = Inhaltsanalyse, DIJ = Datenintensiver Journalismus.

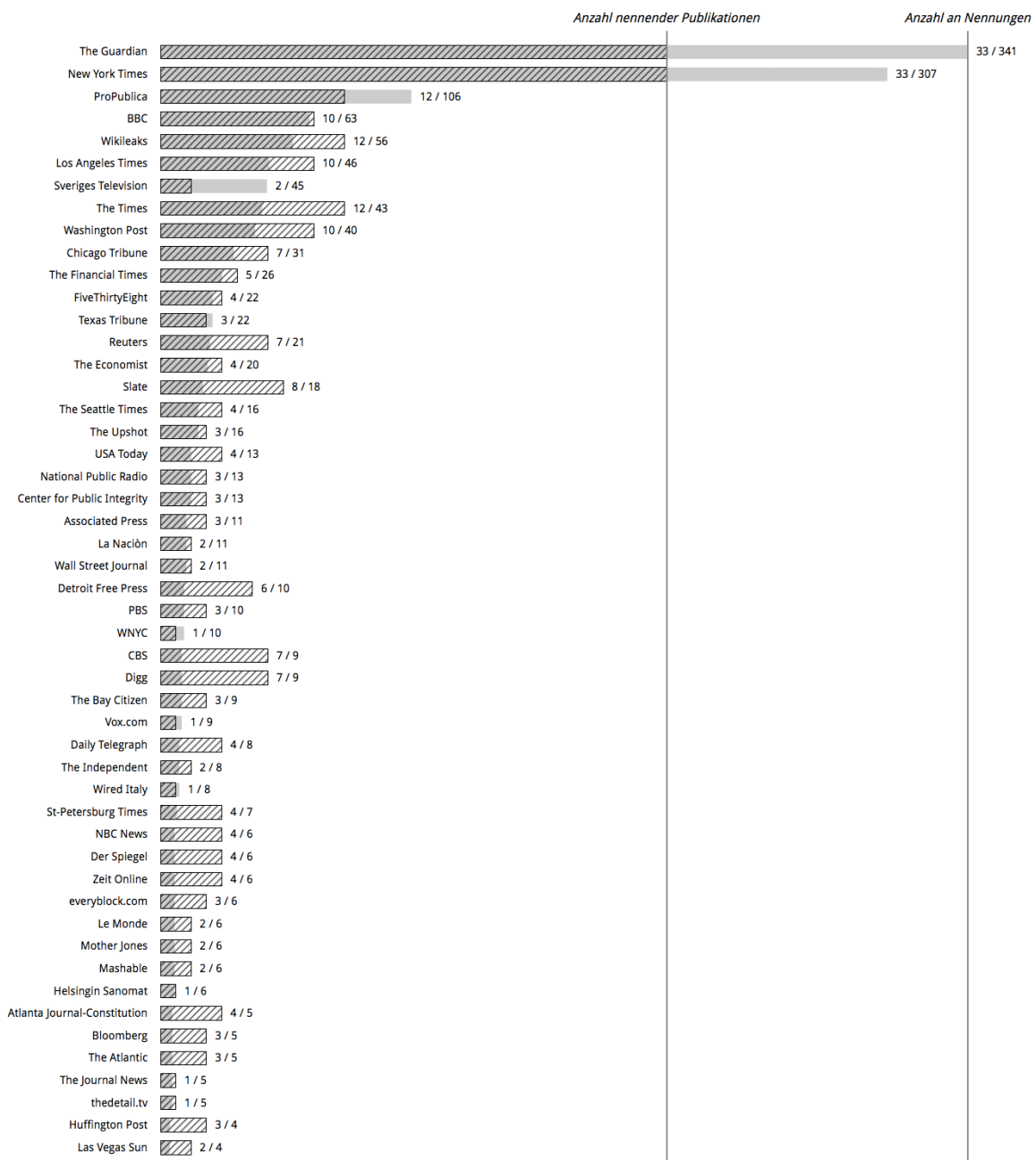
5.4.4.4 Örtlicher Fokus der Untersuchungen

Der geografische Fokus der Studien korrelierte weitgehend mit dem Land, in dem die AutorInnen institutionell angebunden waren. Die untersuchten Redaktionen lagen überwiegend in den Vereinigten Staaten und Großbritannien. Auch Praktiken aus anderen westlichen Ländern wie Deutschland oder Skandinavien standen im Fokus der

Untersuchungen. Was darüber hinaus auffällt: Es existiert fast keine englischsprachige Forschung über datenintensiven Journalismus außerhalb von Europa und Nordamerika. Einige Autorenkollektive konzentrierten sich in ihren Arbeiten auf die Szene eines ganzen Landes – etwa Appelgren und Nygren (2014a, 2014b) auf Schweden, Stavelin allein (2013) beziehungsweise mit Karlsen (2014) auf Norwegen, De Maeyer und KollegInnen (2015) auf Belgien, Weinacht und Spiller (2014) auf Deutschland oder Radchenko und Saroyan (2014) auf Russland. Seltener sind vergleichende Ansätze zu finden, in denen die Praxis in verschiedenen Ländern miteinander in Beziehung gesetzt wurde (Uskali & Kuutti, 2015; Weber & Rall, 2012, 2013; Zanchelli & Crucianelli, 2012).

Welche Medienorganisationen haben die Forschenden mit datenintensivem Journalismus assoziiert und untersucht? Die Liste in Abbildung 18 wurde durch manuelle Codierung in Kombination mit automatischer Termextraktion (TF-IDF) erstellt. Die Visualisierung zeigt die Häufigkeit und die Gesamtzahl an Nennungen einer bestimmten Medienorganisation sowie die Anzahl der Publikationen, in denen diese erwähnt wird. Der *Guardian*, die *New York Times* und *ProPublica* traten am häufigsten und in den meisten Dokumenten auf, aber Abbildung 18 hilft auch, kleinere Nachrichtenorganisationen zu identifizieren, die nur selten erwähnt wurden, sich aber für künftige Forschung eignen.

Abbildung 18. Mit datenintensivem Journalismus assoziierte Medienorganisationen



5.4.5 Forschungslücken der Datenjournalismusforschung

Angesichts des Umstands, dass die Erforschung von datenintensivem Journalismus seit den letzten Jahren intensiver betrieben wird, ist es relevant, was diejenigen, die sich bereits mit dem Phänomen beschäftigt haben, als offene Forschungslücken identifizieren. Zu den Empfehlungen für die künftige Forschung zählen unter anderem transnationale Untersuchungen sowie ethnografische Studien. Parasie und Dagiral (2013) etwa regen an, die Praktiken in verschiedenen Ländern zu untersuchen, wobei die lokal gewachsenen Kulturen

von JournalistInnen und HackerInnen berücksichtigt werden sollten. Die Unterschiede hätten die Art und Weise, wie datenintensiver Journalismus in verschiedenen Ländern praktiziert wird, beeinflusst. Sie schlagen auch vor, dass ethnografische Studien von Newsrooms klären könnten, wie die VertreterInnen von datenintensivem Journalismus tatsächlich in verschiedenen Organisationen integriert würden. Auch Appelgren und Nygren (Appelgren & Nygren, 2014b) empfehlen, vergleichende internationale Studien durchzuführen, die nationale Gesetzgebungen und andere Regulatorien berücksichtigen.

Die meisten Publikationen zu datenintensivem Journalismus konzentrierten sich auf eine relativ kurze Zeitspanne. Auf diesen Mangel weist Knight (2015) hin und schlägt Langzeitstudien vor. Bereits 15 Jahre zuvor hatten Davenport et al. (2000) das Gleiche empfohlen. Vor dem Beginn der wirtschaftlichen Krise vieler Medienhäuser mit Anfang der 2000er Jahre sprach sich Garrison (1999) dafür aus, die Ressourcen von Zeitungen stärker in den Mittelpunkt der Aufmerksamkeit zu rücken: Inwieweit würden Ressourcenunterschiede von Zeitungen unterschiedlicher Größe unterschiedliche Geschichten (Umfang, Tiefe, Größe der Datenbanken) nach sich ziehen? Lewis und Usher (2013) schlagen außerdem vor, den Einfluss von Technologien auf die Nachrichtenproduktion zu untersuchen. In ähnlicher Weise empfiehlt Stavelin (2013) zu erforschen, wie sich das Design und die Verwendung von Software auf die datenintensive Nachrichtenproduktion auswirken. Segel und Heer (2010) schlagen ihrerseits Rezeptionsforschung vor, um zu klären, wie die RezipientInnen von datenintensivem, interaktivem Journalismus damit umgehen. Darauf aufbauend ließe sich das Design verbessern. Trotz dieser Vielzahl an unterschiedlichen Anregungen – die Mehrheit der Publikationen im Korpus diskutierten keine Forschungslücken oder schlugen Wege für die weitere Forschung vor.

5.4.6 Datenintensive Arbeitsabläufe in Redaktionen³⁸

Nachdem auf den vorangegangenen Seiten die Forschung zu datenintensivem Journalismus multiperspektivisch dargestellt wurde, steht im Folgenden die konkrete Praxis im Mittelpunkt. Beschrieben werden redaktionelle Arbeitsabläufe anhand der verschiedenen Schritte, die ein datenintensiver journalistischer Beitrag im Rahmen seiner Produktion durchläuft. Konkret wird dabei auf folgende Bereiche eingegangen: Themenfindung, Auftragserteilung und

³⁸ Teile dieses Unterabschnitts basieren auf einem andernorts veröffentlichten Text (Ausserhofer, Goldgruber, Gutounig & Matiassek, 2017), der für diese Arbeit übersetzt, adaptiert, aktualisiert und erweitert wurde.

Datensammlung, -säuberung, -verifikation, -analyse, Produktion und Visualisierung sowie Veröffentlichung und Evaluierung. Es sei in Erinnerung gerufen, dass dieser Abschnitt auf der systematischen manuellen Verdichtung der codierten Stellen auf Grundlage des Codebuchs basiert. Er gibt die Aussagen anderer Studien wieder, teilweise ergänzt durch Zusatzinformationen. Auch werden nur Aspekte erläutert, soweit sie sich von gewöhnlichen redaktionellen Arbeitsabläufen unterscheiden. Der Text versucht, einen möglichst detaillierten Überblick zu generieren, ohne Wertungen oder Erkenntnisse aus der eigenen Forschung vorwegzunehmen. Diese sind im Abschnitt zur Diskussion der Ergebnisse der systematischen Literaturanalyse bzw. in den darauffolgenden Kapiteln (Interviewstudie, Conclusio) zu finden.

5.4.6.1 Themenfindung, Auftragserteilung und Datensammlung

Nicht jedes Ressort und jedes Thema in einer Redaktion eignen sich ähnlich gut für eine datengestützte Aufbereitung. In der Vergangenheit haben sich bestimmte thematische Bereiche als besonders affin für datenintensiven Journalismus erwiesen. Abseits von individuellen Schwerpunkten und Interessen einzelner MitarbeiterInnen in den Redaktionen hängt die Themenaffinität des datenintensiven Journalismus mit der Verfügbarkeit von (einfach) weiterzuverarbeitenden Daten sowie nachrichtenrelevanten Fragestellungen zusammen. Zu typischen Themen(gebieten), die sich für eine datenjournalistische Aufbereitung eignen, zählen Demografie, Kriminalität, Politik, Gesundheit, Bildung, Umwelt, Wirtschaft, Arbeitsbedingungen, städtische Dienstleistungen und öffentlicher Transport, Kultur sowie Sport (Appelgren & Nygren, 2014a; Fink & Anderson, 2015; Parasie & Dagiral, 2013). Auch in der Wissenschaftsberichterstattung spielen Daten eine große Rolle; interaktive Datenelemente sind dort substantielle Teile von vielen Beiträgen (Felle, 2016; Knight, 2015).

Die Literatur unterscheidet grob zwischen zwei unterschiedlichen Ansätzen, einen datenintensiven Beitrag zu beginnen: entweder mit einem Thema oder mit einem Datensatz (Aitamurto u. a., 2011; Zanchelli & Crucianelli, 2012). Bei ersterem, einem themeninduzierten Start, stößt ein Redaktionsmitglied bei der Recherche zu einem bestimmten Thema auf einen Datensatz, der sich für die weitere Aufbereitung eignet. Nach einer Durchsicht der Daten wird entschieden, die Berichterstattung auch auf Grundlage von Datenanalysen durchzuführen. Das Thema wird meist von der Tagesagenda bestimmt; in vielen Fällen fungieren die datenintensiven Beiträge als Ergänzungen zur gewöhnlichen Berichterstattung (Dick, 2014, S. 498). Beim anderen Ansatz, einem dateninduzierten Start, wird der Redaktion ein Datensatz zugetragen. Sie unterzieht den Datensatz einer Erstanalyse und entscheidet sich dafür, einen

Beitrag zu publizieren, der auf diesem Datensatz (und weiterer Recherche) basiert. Dieser dateninduzierte Ansatz ist für den Journalismus untypisch und eher bekannt aus den Computerwissenschaften. Cohen et al. (2011b, S. 150) fassen diesen Perspektivwechsel mit „from finding answers to finding questions“ zusammen (siehe auch dazu den Unterabschnitt zum Datenbegriff im Einleitungskapitel sowie Gynnild, 2014, S. 722).

In Redaktionen kommt es auch häufig vor, dass RedakteurInnen mit Datenkompetenz von ihren KollegInnen gebeten werden, sie bei der Datenanalyse zu unterstützen (Royal, 2012). Dick (2014, S. 498) erläutert das am Beispiel der *BBC*:

Commissioning in the BBC Specials team is approximately 40 per cent internal (where a team member is commissioned to create infographics by the Specials editor) and 60 per cent external (where other teams and departments engage the Specials editor with work).

Für die Auftragsvergabe hat sich vielerorts der Begriff des „data desk“ etabliert (Uskali & Kuutti, 2015, S. 79), ein Tisch im Newsroom, an dem Ansprechpartner für Datenbelange sitzen. Im Idealfall, so sagen mehrere Interviewte unisono bei Zanchelli und Crucianelli (2012), ist dieser Tisch möglichst zentral im Newsroom positioniert, um viel vom Geschehen mitzubekommen. Eine Ausnahme bilden freilich Produktionen, die mehrere Wochen oder Monate benötigen.

Betont wird auch in verschiedenen Publikationen die wichtige Rolle von Abstimmungsbesprechungen, an denen Personen mit unterschiedlichen Hintergründen teilnehmen (Zanchelli & Crucianelli, 2012, S. 2): „Because developers and reporters often have specialized skill sets, it’s important to bring them together to brainstorm story ideas. This increases the likelihood of identifying unique and significant angles for data-driven stories.“

Die Daten erreichen die Redaktionen über verschiedene Quellen und Zugänge³⁹:

³⁹ Einige Elemente in dieser Liste sind nicht völlig trennscharf von anderen Elementen zu unterscheiden. Würde man diese Liste als Ausgangspunkt für eine quantitative Inhaltsanalyse nehmen, müssten weitere Verfeinerungen und Verallgemeinerungen getroffen werden. Vollkommen kann ein Kategoriensystem ohnehin nie sein und Standards werden idealisiert (Bowker & Star, 1999b). An diesem Punkt sei an die Tier-Taxonomie aus einer „gewissen chinesischen Enzyklopädie“ erinnert, die Jorge Luis Borges (1966) in einem seiner Essays zitiert und die durch Foucaults (1966/2003) Vorwort zur *Ordnung der Dinge* Berühmtheit erlangte: Auf den uralten Blättern der Enzyklopädie „steht geschrieben, daß die Tiere sich wie folgt unterteilen: a) dem Kaiser gehörige, b) einbalsamierte, c) gezähmte, d) Milchschweine, e) Sirenen, f) Fabeltiere, g) streunende Hunde, h) in diese Einteilung aufgenommene, i) die sich wie toll gebärden, j) unzählbare, k) mit feinstem Kamelhaarpinsel

- Öffentliche Stellen (06, 07, 08, 10, 13, 19, 27, 28, 29, 39, 40)⁴⁰
 - hier insbesondere über Auskunftsbeglehen, Anträge auf Aktenauskunft, ‚freedom of information requests‘, ‚access to information requests‘, ‚personal information requests‘, etc. (03, 07, 10, 18, 27, 37, 40)
- Eigenerhebungen und -umfragen (03, 06, 10, 13, 27, 29, 39, 32, 40)
- Crowdsourcing (03, 08, 13, 15, 17, 32)
- Leaks (10, 13, 27, 19, 30)
- Computergestützte Techniken wie eigene Sensoren, ‚scraping‘, ‚crawling‘⁴¹ oder ‚data mining‘ (13, 14, 18, 27, 29)
- Nicht-profitorientierte Organisationen wie Universitäten, Forschungsinstitutionen, NGOs, *Wikileaks* (10, 27, 29)
- Presseaussendungen (10, 13, 34)
- Sonstige öffentlich zugängliche digitale Datenbanken (03, 36)
- Unternehmen (10)
- Transnationale Organisationen (z.B. die Vereinten Nationen) (10)
- Social-Media-Daten (14, 32)
- Eigene Nachrichtenarchive (14)

Die Forschungsliteratur nennt am häufigsten öffentliche Stellen als Quellen bzw. Zugangswege für Daten im datenintensiven Journalismus. Angesichts der Themen, die datenintensiver Journalismus behandelt, dessen Geschichte und der engen Verbindung zum investigativen Journalismus, sowie zu Open-Government-Communities, ist das durchaus naheliegend. Die Qualität des Zugangs zu Daten von öffentlichen Stellen variiert von Land zu Land. In Norwegen beispielsweise hatten interviewte JournalistInnen laut eigener Aussage geringe Probleme, an Daten zu gelangen; eine Herausforderung sei jedoch, dass zunehmend auch für Daten der öffentlichen Hand bezahlt werden müsse (Karlsen & Stavelin, 2014). In Belgien wiederum gibt es Kritik an den DatenbereiterInnen, insbesondere der nationalen

gezeichnete, l) und so weiter, m) die den Wasserkrug zerbrochen haben, n) die von weitem wie Fliegen aussehen.“ (Borges, 1966, S. 212).

⁴⁰ Aus Gründen der Lesbarkeit wurde in dieser Liste von der Standard-Zitierweise abgesehen. Die Zahlen in Klammer beziehen sich auf die „ID“-Spalte von Tabelle 4 (oben).

⁴¹ Crawler und Scraper sind Skripte, die ganze Webseiten bzw. Datenbanken oder Teile davon abgreifen, auf eine lokale Repräsentanz kopieren und für die weitere Analyse aufbereiten.

und EU-Statistikbehörde; daran scheinen die Open-Data-Initiativen von öffentlichen Stellen kaum etwas geändert zu haben (De Maeyer u. a., 2015, S. 438). In den USA wiederum gibt es eine lange Tradition, Daten von öffentlichen Stellen über „freedom of information requests“ zu erfragen (Fink & Anderson, 2015; Parasio, 2015; Parasio & Dagiral, 2013). Im Vergleich zu anderen Ländern empfinden DatenjournalistInnen in Deutschland das Datenangebot als deutlich restriktiver; auch sei dort die Skepsis gegenüber „Datensammeln“ ausgeprägter (Weinacht & Spiller, 2014, S. 429).

Eine Rolle in der Literatur spielen auch selbst erhobene Daten – etwa in Form von Umfragen oder Crowdsourcing. Auch wenn dies häufig als zentraler Aspekt von Datenjournalismus beschrieben wird (z.B. bei Appelgren & Nygren, 2014b), spielt es in der alltäglichen redaktionellen Praxis eine eher untergeordnete Rolle (Tabary u. a., 2016). Ähnlich verhält es sich mit Leaks, also der Veröffentlichung bzw. dem Zuspielen von vertraulichen Daten, deren Enthüllung nicht autorisiert wurde: Zwar hat der datenintensive Journalismus durch die Aufbereitung von Leaks größere Bekanntheit erfahren und zahlreiche Referenzprojekte hervorgebracht, im datenjournalistischen Alltag sind Leaks dennoch nur gelegentlich Thema (Knight, 2015, S. 69; Loosen u. a., 2015, S. 13).

Des Weiteren gibt es Indizien für ‚datenbasierte Presseaussendungen‘ als Ausgangspunkt für einschlägige Berichterstattung: „A substantial proportion of the stories showed evidence of a body of data being released wholesale by research institutes and other vested bodies. Particularly in social issues and health, most of the data presented were acquired in this way“ (Knight, 2015, S. 69). Mit anderen Worten: Nicht anders als im traditionellen Textjournalismus wird die Arbeit des datenintensiven Journalismus massiv von Öffentlichkeitsarbeit bestimmt. Damit eng zusammenhängend gibt es das Problem, dass aus Ressourcengründen für tagesaktuelle Beiträge oftmals nur ein oder zwei Quellen herangezogen werden können (Lugo-Ocando & Brandão, 2016, S. 722). Ideal, so werden Beteiligte in der Forschungsliteratur zitiert, wäre es, verschiedene Daten zu einem Sachverhalt in einer Analyse kombinieren zu können, um daraus einen Mehrwert für die LeserInnen zu schaffen (z.B. Parasio & Dagiral, 2013, S. 862–863).

5.4.6.2 Datensäuberung, -verifikation, -analyse

Nachdem die ersten Datensätze für einen Beitrag vorliegen, werden sie begutachtet, ausgewählt, gefiltert und gereinigt (Karlsen & Stavelin, 2014, S. 41). Dieser Prozess ist in den

meisten Fällen der zeitintensivste in der gesamten Produktion (A. B. Howard, 2014, S. 58). Ein erster notwendiger Schritt ist in vielen Fällen die Digitalisierung von Daten bzw. die Vorbereitung von Datensätzen für die Begutachtung. Dann wird versucht, einen ersten Eindruck über den Datensatz zu erhalten. Computergenerierte Zusammenfassungen über das Datenmaterial geben einen Überblick – oft auch in Form erster Visualisierungen. Diese Begutachtung hat zwei Ziele: Einerseits sollen Fehler im Datenbestand entdeckt werden (1), andererseits gilt es, ein Narrativ zu entwickeln (2).

(1) Wenn die Daten nicht durch die RedakteurInnen selbst erhoben wurden, ist es oftmals für die mit Datenjournalismus befassten Personen schwierig, diese mit ihren physischen Repräsentationen zu verbinden (Parasie, 2015, S. 372). Es setzen ethische Überlegungen ein (Parasie, 2015, S. 372): Wie belastbar sind die Daten? Für die Verifikation und Plausibilitätsprüfung von Daten werden häufig interne und externe ExpertInnen hinzugezogen, etwa mittels Telefonanrufen oder Treffen, im Rahmen derer gemeinsam die Daten geprüft werden (Lugo-Ocando & Brandão, 2016, S. 722). Auch viele Daten, die von öffentlichen Stellen zur Verfügung gestellt werden, sind fehlerhaft (Parasie, 2015, S. 371). ReporterInnen begegnen Daten der offiziellen Verwaltung mit einer gewissen Grundskepsis, da sie Manipulationen an den Daten befürchten (Parasie & Dagiral, 2013, S. 868). Wie bei anderem Quellenmaterial kommt es häufig vor, dass Datensätze sich aus verschiedenen Gründen nicht für eine Veröffentlichung eignen. Sie werden dann als interne Ressourcen verwendet und nicht für online zu publizierende Beiträge (Karlsen & Stavelin, 2014, S. 41).

(2) Neben dem Auffinden von Fehlern wird diese Phase dazu genutzt, um Fragen und Narrativ(e) zu entwickeln. Das passiert, indem etwa verschiedene Werte miteinander verglichen werden oder Ausreißer hinterfragt werden – zum Beispiel Unterschiede zwischen verschiedenen Entitäten und/oder über die Zeit (Loosen u. a., 2015, S. 16). Verkürzt ließe sich das so formulieren: „Journalists ask questions; data journalists ask the data“ (Appelgren & Nygren, 2014a, S. 5). Dabei unterscheidet sich der Zugang im Endeffekt wenig vom Zugang im ‚klassischen Textjournalismus‘:

A crucial part of what’s needed to practice data journalism, however, has little to do with tools and technology and everything to do with perspective and critical thinking. „You need a mindset which is about putting this in the context of the story and spotting stories, as well as having creative and interesting ideas about how you can actually collect this material for your own stories,“ said Emily Bell. „It’s not a passive kind of processing function if you’re a data journalist: It’s an active speaking, inquiring, and discovery process. (...)“ (A. B. Howard, 2014, S. 58)

In dieser Phase gibt es gewöhnlicherweise noch viele Probleme mit Daten. Zum Beispiel liegen sie in einem falschen Dateiformat vor, sind nicht lesbar oder nicht miteinander kompatibel. Oder die gleichen Entitäten sind unterschiedlich benannt. Ein Interview mit einem Praktiker fasst das Problem gut zusammen (Parasie, 2015, S. 371): „No information is ever clean. No data is ever perfect. I’m willing to accept that. But given that limitation, it’s really important and really necessary and really hard to figure out what we can responsibly say with this information.“

Darüber hinaus kann diese erste Analyse eine Grundlage für Follow-Up-Interviews und weitere Datenrecherchen liefern (Appelgren & Nygren, 2014a, S. 5).

5.4.6.3 Produktion und Visualisierung

Nachdem sich in der Analyse erste Trends abzeichnen oder sich ein Narrativ herausgebildet hat, stellt eine sinnvolle Form der Präsentation die nächste Herausforderung dar (Karlsen & Stavelin, 2014, S. 45). Die Daten unterstützen dabei die Konstruktion des Narrativs; sie helfen, ein Argument zu untermauern (Parasie, 2015, S. 372). Darüber hinaus helfen sie den Lesenden bei der visuellen Orientierung im Text (Segel & Heer, 2010).

How data are joined, analysed and presented is the direct result of journalistic sense-making combined with proficiency in programming. This is often an inductive process of trial and error, not guided by reasoning alone. The programming becomes inseparable from journalism and vice versa. (Karlsen & Stavelin, 2014, S. 43)

Zuweilen wird in der Produktion zugunsten eines linearen Narrativs auf eine Visualisierung verzichtet (Karlsen & Stavelin, 2014, S. 42; Loosen u. a., 2015, S. 18; Tabary u. a., 2016). Für die richtige Wort- und Gestaltungswahl (z.B. Farben und Symbole) gibt es sogenannte ‚styleguides‘, i.e. Handbücher, die derartige Aspekte festlegen (Dick, 2014, S. 501). Einige dieser Style Guides sind auch öffentlich einsehbar, z.B. jener der *BBC* (2017), des *Guardian* (Marsh & Hodsdon, 2017) oder der *Dallas Morning News* (2005). Teil dieser Style Guides sind vielfach auch Artikelvorlagen, denen Publikums- und Marktforschung zugrunde liegt (Dick, 2014). In vielen Redaktionen scheinen derartige Leitlinien aber auch zu fehlen, was eine Herausforderung für die DatenjournalistInnen darstellt (Smit u. a., 2014, S. 348).

Je nach Größe der Redaktion und Komplexität des Projekts werden Daten vom/von der DatenredakteurIn unter Zuhilfenahme einfacher Tools selbst visualisiert oder einE DesignerIn nimmt sich der Visualisierung an (Baack, 2013). Bei aufwändigeren Projekten wird datenintensiver Journalismus im Team organisiert (Bakker, 2014; Weinacht & Spiller, 2014,

S. 425). Nicht selten dauern Projekte dabei mehrere Wochen, Monate oder sogar Jahre (Uskali & Kuutti, 2015, S. 85). Beteiligte JournalistInnen müssen bei dieser Zusammenarbeit kein „full repertoire of computational thinking and skills“ vorweisen (Gynnild, 2014, S. 719). Sie sind für gewöhnlich koordinierend tätig, und ermöglichen es, dass Menschen mit komplementären Fertigkeiten, Einstellungen und Werten gemeinsam Ideen implementieren (Gynnild, 2014, S. 725). Der Rest des Teams sind Menschen (mit oder ohne journalistischem Hintergrund) mit Fertigkeiten in statistischer Analyse, Datenverarbeitung, Visualisierung und Infografik (Aitamurto u. a., 2011, S. 11). Die DesignerInnen haben dabei häufig Schwierigkeiten, journalistische Prinzipien auf ihre Visualisierungen anzuwenden, während den JournalistInnen wiederum die Erfahrung in der Anpassung von Designregeln an ihre Geschichten fehlt (Smit u. a., 2014, S. 348). Die Kompetenzen, die Datenjournalismus-Akteure beherrschen, sind anderen in der Redaktion kaum bekannt (Karlsen & Stavelin, 2014, S. 44). Schließlich sind auch ProgrammiererInnen beteiligt; die Rollen und Verantwortlichkeiten stellen sich dabei folgendermaßen dar (Zanchelli & Crucianelli, 2012, S. 112):

Generalizing, developers or data miners tend to specialize in analyzing, interpreting and visualizing data. These skills give developers a knack for understanding how to extract the numbers, see patterns and trends and interpret what those numbers may mean. Reporters tend to bring an understanding of the history and significance of the numbers. They can put the numbers into a broader political, social and economic context and see trends developers might not immediately connect to issues that matter to citizens. Reporters also have storytelling skills.

Obwohl viele TechnikerInnen und DesignerInnen keine journalistische Ausbildung haben, sehen sie sich selbst als JournalistInnen und nicht als DienstleisterInnen für ReporterInnen oder RedakteurInnen (Weber & Rall, 2013). In deutschen und Schweizer Redaktionen scheint der Literatur zufolge jedoch noch ein größeres Gefälle vorzuherrschen, auch wenn Anzeichen existieren, dass dieses langsam kleiner wird (Weber & Rall, 2012).

Die Ideen für Visualisierungen und Newsapps werden oft gemeinsam entwickelt (Zanchelli & Crucianelli, 2012, S. 9). Feedback-Sessions und Diskussionen im Team benötigen dabei viel Zeit (Smit u. a., 2014, S. 350). In manchen Redaktionen sind diese Kollaborationen in Form von fest zusammenarbeitenden Gruppen institutionalisiert, in anderen basiert die Zusammenarbeit auf einer informellen Basis (Karlsen & Stavelin, 2014, S. 39). Vielerorts, wo sich erste Teamstrukturen gebildet haben, haben sich keine festen Routinen etabliert; ständig wird versucht, die Prozesse zur Umsetzung datenintensiver Beiträge zu verbessern (Karlsen & Stavelin, 2014, S. 44). Mit den klassischen IT-Abteilungen gibt es oftmals Differenzen, weil

die mit Datenjournalismus befassten Personen mehr Ressourcen als klassische TextjournalistInnen benötigen (Software, Server, Datenbanken, etc.), was wiederum mit höheren Kosten und Sicherheitsproblemen verbunden ist (Karlsen & Stavelin, 2014, S. 39).

Bei aufwändigeren interaktiven Projekten werden in der Produktion auch sogenannte Storyboards eingesetzt. Sie dienen einerseits der Materialisierung der Gedanken der Haupt-Visualisierenden, andererseits helfen sie bei der Abstimmung mit anderen Teammitgliedern (Royal, 2012, S. 16; Weber & Rall, 2012, S. 352, 2013, S. 162–163). Ein Storyboard können einfache handgezeichnete ‚Szenen‘ sein, wie die Visualisierung nach bestimmten Interaktionen durch die UserInnen aussehen soll, oder aber computergenerierte Vorvisualisierungen (Weber & Rall, 2012, S. 352). Das Ziel ist dabei weniger ein ‚polierter Look‘ als vielmehr ein klarer visueller Plan (Weber & Rall, 2012, S. 352–353). Ihren Ursprung hat die Technik in der Fernseh- und Filmproduktion (Weber & Rall, 2013, S. 169).

Für gewöhnlich werden datenintensive Beiträge mit einer oder mehr der folgenden Visualisierungsformen dargestellt (Fink & Anderson, 2015; Karlsen & Stavelin, 2014; Knight, 2015; Loosen u. a., 2015; Parasie & Dagiral, 2013; Royal, 2012; Zanchelli & Crucianelli, 2012):

- Bilder und Diagramme
- Karten
- Tabellen
- Infografiken
- Zeitlinien
- Animationen
- Listen
- Textvisualisierungen (z.B. Wordclouds)
- Töne und Videos

Diese Visualisierungen können dabei sowohl statisch als auch interaktiv sein. Je mehr sich eine Redaktion auf datenintensiven Journalismus einlässt, desto komplexer entwickeln sich die Visualisierungen (Weber & Rall, 2012, S. 352). Auch wenn im Datenjournalismus technisch viel möglich ist, gibt es noch zahlreiche Fälle, in denen bereits zusammengesetzte Datensätze einfach mit automatisierten Programmen ohne weitere Analyse visualisiert werden (Tabary u. a., 2016). Auch völlig statische Visualisierungen sind keine Seltenheit. Das Resultat der

journalistischen Bearbeitung von Daten kann eine traditionelle Geschichte sein, eine Karte, eine Grafik oder auch ein Werkzeug (Zanchelli & Crucianelli, 2012, S. 8).

5.4.6.4 Veröffentlichung und Evaluierung

In dieser Prozessphase sind die Rahmenbedingungen rund um die Veröffentlichung geklärt. Die Forschung hat bislang weniger darüber geschrieben, wie die Publikation von datenintensiven Beiträgen aus der Perspektive redaktioneller Prozesse ganz konkret funktioniert. Es wird dies im nächsten Kapitel erläutert. Diskutiert wird in Forschungsliteratur zu Datenjournalismus vor allem die Veröffentlichung von Rohdaten (1) gemeinsam mit dem Beitrag sowie die Einbindung des Publikums (2).

(1) Die Veröffentlichung der Daten, die dem Beitrag zugrunde liegen, sehen befragte DatenjournalistInnen als integralen Bestandteil der Produktion (Aitamurto u. a., 2011, S. 11). Die Daten werden aus Gründen der Reproduzierbarkeit und Transparenz veröffentlicht. Das soll die Glaubwürdigkeit des Mediums erhöhen (Felle, 2016, S. 92). Zugleich sehen die Akteure (bzw. die Forschungsliteratur) die Veröffentlichung als Einladung an das Publikum, sich selbst mit den Daten auseinanderzusetzen und eigene Analysen anzustellen (Aitamurto u. a., 2011, S. 11). Die Daten werden unter anderem auf der Website des Mediums, in *Google Spreadsheets*⁴², auf *GitHub*⁴³ oder auch in hauseigenen, sogenannten ‚Datastores‘⁴⁴ veröffentlicht (Aitamurto u. a., 2011; A. B. Howard, 2014, S. 20). Für die Veröffentlichung von unstrukturierten Dokumenten wie etwa PDFs wird häufig die Plattform *DocumentCloud*⁴⁵ eingesetzt, ein Open-Source-Projekt, das die Analyse, Annotation und Publikation von Dokumenten ermöglicht (Lewis & Usher, 2013, S. 610).

Die kombinierte Veröffentlichung von Beitrag und Daten wurde zwar als publizistischer Idealstandard häufig in der Forschung genannt, in der Praxis kann dieser

⁴² <https://docs.google.com/spreadsheets/>

⁴³ *GitHub* ist ein webbasierter Dienst, der unter anderem Programmiercode und Daten zugänglich macht. Grundlage ist das namensgebende *Git*-System zur verteilten Versionsverwaltung, mithilfe dessen kollaborativ Texte, Daten und – vor allem – Programmcode entwickelt werden. Eine Liste von Redaktionen, die auf *GitHub* publizieren, findet sich unter: <https://github.com/OpenNewsLabs/news-graphics-team>.

⁴⁴ Bekannte redaktionelle ‚One-Stop-Shops‘ für Daten sind etwa der *ProPublica Data Store* (<https://www.propublica.org/datastore/>), oder die *NICAR Data Library* (<https://www.ire.org/nicar/database-library/>). Der Datastore des *Guardian* wurde bald nach seiner Einführung wieder eingestellt.

⁴⁵ <https://www.documentcloud.org/>

vielfach nicht eingehalten werden, wie Inhaltsanalysen belegen: In Knights (2015, S. 65) Untersuchung datenintensiver Beiträge von Onlineportalen englischer Zeitungen hatten 40 % der Beiträge nicht einmal Angaben über die Quelle der Daten – von der Veröffentlichung des Datensatzes ganz zu schweigen. Dieses Ergebnis wird von Loosen et al. (2016) in der Untersuchung des ‚Gold Standard‘ des datenintensiven Journalismus, der Nominierten für die *International Data Journalism Awards*, untermauert: Hier hatten 6,1 % der Projekte keine Quellen angegeben.

(2) Wer die RezipientInnen von datenintensivem Journalismus sind, und ob bzw. wie sie sich von anderen Rezipientengruppen unterscheiden, ist den DatenjournalistInnen kaum bekannt (Fink & Anderson, 2015, S. 474; Smit u. a., 2014, S. 352). Nur ein kleiner Teil der Leserschaft nutzt die Möglichkeit der Interaktion bei datenintensiven Beiträgen – etwa die Daten zu filtern oder zu explorieren (Karlsen & Stavelin, 2014, S. 40). Viele ProduzentInnen der Beiträge frustriert das – viel Aufwand, aber wenig Interesse seitens des Publikums; ein Informant dazu (zit. n. Karlsen & Stavelin, 2014, S. 40): „When we have a large dataset we often give the audience the whole package with lots of buttons and analytic tools, etc. Then people aren’t really interested.“ Zugleich gibt es Aussagen darüber, dass datenintensive Beiträge dazu tendieren, nachhaltiger zu sein als Textjournalismus im Netz – denn die Beiträge würden selbst Wochen und Monate nach der Veröffentlichung wiederholt aufgerufen werden (Felle, 2016, S. 92; Usher, 2016). Die Besucherstatistiken der Content-Management-Systeme und Social-Media-Verteilungsmetriken sind meist die einzigen Indikatoren für den Erfolg eines Beitrags bei den RezipientInnen (Fink & Anderson, 2015, S. 474). Sie stellen eine Entscheidungsgrundlage in der Redaktion dar, ob ein Projekt mit neuen Daten aktualisiert oder das Thema weiter behandelt wird (Fink & Anderson, 2015, S. 474).

Visualisierungen ziehen UserInnen an, verändern aber auch die Leseerfahrung (Tandoc & Oh, 2015, S. 5). Aspekte wie Farben, Kontrast, Platzierung, aber auch kulturelle Faktoren wie etwa Leserichtung spielen dabei eine besondere Rolle (Segel & Heer, 2010, S. 2). Ein Teil des Publikums scheint auch von Visualisierungen eingeschüchtert zu sein, weil es unliebsame Erinnerungen an Mathematik in der Schule aufleben lässt; ebenso wirken gewisse Visualisierungsformen wie etwa Kreise ansprechender als andere (Dick, 2014, S. 502). Nur wenige Medienhäuser evaluieren systematisch, wie die interaktiven Visualisierungen vom Publikum angenommen werden (Smit u. a., 2014, S. 348).

In manchen Redaktionen wird versucht, das Publikum möglichst intensiv einzubinden. Ein Beispiel ist der *Guardian* (Gynnild, 2014, S. 719): „The Guardian’s readers are encouraged to help analyze data sets, provide feedback or additional data, formulate research questions, and submit applications and visualizations that they have created from accessible data in the Data Store.“ Auch sollen die LeserInnen die Beiträge in sozialen Medien teilen (Aitamurto u. a., 2011, S. 11). Befragten JournalistInnen sind derartige Interaktionen mit dem Publikum sehr wichtig, wiewohl die Häufigkeit und Intensität der Einbindung je nach Redaktion und Land variiert (Felle, 2016, S. 91). In Schweden etwa gab bei einer Umfrage ein Drittel der befragten DatenjournalistInnen (N = 164) an, Crowdsourcing für Datensammlung oder Publikumseinbindung eingesetzt zu haben (besonders *Aftonbladet* und *Svenska Dagbladet*) (Appelgren & Nygren, 2014b, S. 399). Während es in den USA ebenfalls größeren Enthusiasmus für die Einbindung des Publikums zu geben scheint, gaben RedakteurInnen aus Europa, Australien und Asien zu Protokoll, gemischte Erfahrungen mit Publikumseinbindung gemacht zu haben (Felle, 2016, S. 91). Insbesondere bei politischen Themen schien das Interesse des Publikums verhalten zu sein, während Themen, die persönliche Lebensbereiche wie etwa Kriminalität, Gesundheit, Bildung, private Finanzen und Steuern betreffen, höheres Engagement durch die RezipientInnen hervorriefen (Felle, 2016, S. 91–92).

5.5 Diskussion der Ergebnisse der systematischen Literaturanalyse

5.5.1 Agenda zur weiteren Erforschung des datenintensiven Journalismus

In einer Literaturanalyse sondierte Diakopoulos (2012) das „Potenzial für technische Innovationen im Journalismus“ (S. 2): Durch die systematische Analyse von Informatik-Publikationen, die sich mit Journalismus beschäftigten, konnte er Bereiche identifizieren, die die Informatik-Community bislang vernachlässigt hatte, deren Weiterentwicklung sich jedoch lohnen würde. Die Dokumentenanalyse dieses Abschnitts kann als Komplementärarbeit zu Diakopoulos’ Analyse gesehen werden. Während er das Potenzial diverser Technologien – von „machine learning“ über „natural language processing“ bis hin zu Modellierung – für den Journalismus auslotete, wurde hier die sich mit datenintensivem Journalismus beschäftigende Forschung aus den Sozialwissenschaften und benachbarten Disziplinen analysiert. In Kombination bieten beide Arbeiten einen funktionalen Fahrplan für die weitere Entwicklung und Erforschung des Felds des datenintensiven Journalismus.

In diesem Abschnitt wurden die Struktur und die Themen jener Forschungsliteratur analysiert, die sich in den letzten 20 Jahren mit datenintensivem Journalismus auseinandergesetzt hat. Der rigiden Methode einer systematischen Literaturanalyse folgend wurde ein Korpus von 40 Publikationen ausgewählt und mithilfe von computergestützten Methoden und einer qualitativen softwaregestützten Inhaltsanalyse untersucht.

Sowohl datenintensiver Journalismus als auch seine begleitende Erforschung haben sich in den letzten zwei Jahrzehnten rasant entwickelt. Insbesondere seit 2010 ist eine starke Zunahme der Forschung zum Thema zu verzeichnen. Dieses Wachstum hat zu Qualitätsverbesserungen geführt und dazu beigetragen, eine solide Grundlage für das Feld zu schaffen. Ein Indikator für die Qualitätsverbesserung sind Kanäle der Referenzen: Statt auf Publikationen zu verweisen, die von DatenjournalistInnen selbst geschrieben wurden, wird zunehmend auf wissenschaftliche Publikationen verwiesen, die von Forschenden nach konsistenten Methoden erstellt und in Zeitschriften mit Peer-Review-Verfahren veröffentlicht wurden. Ein weiterer Indikator für die Qualitätsverbesserung ist der hohe Anteil an Kooperationen innerhalb der Forschungsgemeinschaft. Dies stimmt mit dem „Kooperationsimperativ“ der modernen Forschung überein, der auch positive Auswirkungen auf die Schaffung und Verbreitung von Wissen hat (Bozeman & Boardman, 2014). Die häufige Zusammenarbeit über Ländergrenzen hinweg kann durchaus als eine Tendenz zur Internationalisierung und Professionalisierung des Feldes gelesen werden.

Gleichzeitig finden sich auch Probleme bei Teilen der aktuellen Literatur. Zum Beispiel bezieht sich nur eine Minderheit der empirischen Datenjournalismusforschung auf theoretische Konzepte. Viele Publikationen berichten nur, was untersucht wurde, ohne theoretischen oder methodologischen Fragen (viel) Platz einzuräumen. Natürlich ist deskriptive Forschung sehr wichtig – vor allem bei einem neuen Phänomen wie es datenintensiver Journalismus ist –, aber es scheint doch, dass einige Forschende die objektivistische Epistemologie vieler JournalistInnen angenommen haben, die mit quantitativen Daten arbeiten (Godler & Reich, 2013). Forschungsarbeiten, die sich aus neuen Theorietraditionen speisen, oder Arbeiten, die theoretische Ansätze weiterentwickeln, wären in Zukunft wünschenswert. Dabei ließe sich etwa auf zahlreichen Arbeiten aufbauen, die sich bereits mit der Rolle von Daten und Journalismus in der Gesellschaft auseinandersetzen (z.B. C. W. Anderson, 2015; Bunz, 2011; N. S. Cohen, 2015; Fairfield & Shtein, 2014; Lewis & Westlund, 2015; Peters, 2001; Schudson, 2001, 2010). Ein bereichernde Perspektive könnte etwa die Rahmung von Konzepten des datenintensiven Journalismus durch eine ideengeschichtliche Linse sein.

Ein weiteres Problem ist die schwach ausgeprägte Methodenvielfalt. Eine Mehrzahl der Publikationen baut ihre Ergebnisse auf semistrukturierten Interviews auf. Die Tendenz zu erwünschten Antworten, die bei rekonstruktiven Verfahren ohnehin ausgeprägt ist, wird in diesem Fall durch zwei Aspekte besonders befördert: Erstens sind die DatenjournalistInnen in den meisten Publikationen persönlich identifizierbar – in vielen Ländern wäre eine Anonymisierung aufgrund der kleinen Grundgesamtheit wohl nur bedingt möglich. Auch nach den Interviews sind sie ihrem Umfeld, insbesondere der Kollegenschaft und ihren ArbeitgeberInnen, Rechenschaft schuldig. Es ist deshalb davon auszugehen, dass sich die Befragten bestmöglich darstellen wollen und ihre Antworten in den semi-strukturierten Interviews idealisieren. Zweitens sind DatenjournalistInnen untereinander stark vernetzt, vielfach auch international (Lewis & Usher, 2014; Usher, 2016). Ganz so wie bei vielen anderen „communities of practice“ (Wenger, 1998; Wenger, McDermott & Snyder, 2002) gibt es einen transnationalen Diskurs, der über diverse Onlinekanäle und physische Treffen verhandelt wird und der sich normativ auf die Einstellungen, die Praxis und letztendlich auch auf die Antworten der DatenjournalistInnen auswirkt. Bei aller Kritik an der Dominanz der Erhebungsform ‚Interview‘ ist aber natürlich nicht zu vergessen, dass viele Publikationen verschiedene Qualitätssicherungsmaßnahmen getroffen haben, etwa Interviews mit anderen Datenerhebungsmethoden triangulieren.

Was in Bezug auf die Methodenvielfalt darüber hinaus noch auffällt ist, dass es fast keine Forschung auf der Grundlage quantitativer Methoden gibt. Während qualitative, explorative Studien typisch für im Entstehen begriffene Forschungsgebiete sind, scheint es angebracht, einige bestehende Erkenntnisse mit größeren Samples zu testen. Ebenso wenig gibt es Arbeiten, die auf digitalen Methoden beruhen (zum Begriff ‚digitaler Methoden‘ siehe R. Rogers, 2013 sowie Schumann u. a. (2015) und den Abschnitt „Methodologische Herausforderungen digitaler Methoden“ weiter oben). Etwa ließe sich das Feld durch die Analyse von Onlinespuren oder die Beobachtung von Interaktionen auf den größeren Plattformen, auf denen der Datenjournalismusdiskurs stattfindet (z.B. *GitHub*, *Slack*, *Twitter*, *Facebook* oder *Meetup*), aus anderer Perspektive begreifen.

Zusätzlich zu den (oben beschriebenen) Forschungslücken hat die Forschung zu datenintensivem Journalismus derzeit wenig zu Fragen rund um Gender zu sagen. Frauen scheinen eine Minderheit unter den DatenjournalistInnen. Was sind ihre speziellen Herausforderungen? Wie nehmen sie Rollenbilder wahr? Wie wirken sich Aspekte wie Sprache, Bilder und Software im datenjournalistischen Redaktionsalltag auf

Genderstereotypen aus? Forschung, die diese und ähnliche Fragen adressiert, bleibt ein Desiderat.

Schließlich ist sehr wenig über Datenjournalismus außerhalb der Redaktionen berühmter Nachrichtenorganisationen bekannt. Fruchtbar könnten Forschungsarbeiten sein, die sich mit lokalem oder mobilem Datenjournalismus beschäftigen, mit kleinen Medienhäusern, die an datenintensiven Projekten arbeiten, und mit den vielen Freischaffenden, die Dienstleistungen für Nachrichtenorganisationen anbieten. Hier wäre eine ökonomische Perspektive ein besonders wertvoller Beitrag für das Feld.

In der Journalismusforschung gibt es eine starke Verbindung zwischen dem Forschungsinteresse, der gewählten Theorie, den Methoden der Datenerfassung und -analyse sowie der Ergebnisdarstellung (Scholl, 2011). Manche Arbeiten kombinieren diese Aspekte systematisch und sinnvoll, während andere neue Perspektiven aus der Integration neuer oder bewährter Perspektiven, Theorien und Methoden in neuartige Rahmenbedingungen entwickeln (Markham, 2013). Unabhängig davon, welcher Weg für künftige Untersuchungen zum Datenjournalismus gewählt wird – ein traditioneller oder ein Bricolage-Ansatz – dieses Kapitel lieferte die Grundlage für beide Ansätze. Durch die kritische Vermessung wesentlicher Elemente der Publikationen und die Bereitstellung zusätzlicher Forschungsvorschläge wird es künftigen Forschenden ermöglicht, ihre Forschungsinteressen, ihre theoretischen Konzepte und Methoden in Hinblick auf die Kontinuität und Innovation auf dem Gebiet der Datenjournalismusforschung zu wählen.

5.5.2 Datenintensive Arbeitsabläufe im Überblick

Neben dem Literaturüberblick und den daraus abgeleiteten Vorschlägen für die weitere Erforschung der Subdisziplin des datenintensiven Journalismus bot der vorangegangene Abschnitt einen komprimierten Überblick über die bisherigen Erkenntnisse zu datenintensiven Arbeitsabläufen in Redaktionen. Es konnte gezeigt werden, dass bestimmte Themen und Ressorts ‚datenjournalismusaffiner‘ sind als andere. Auch wurden unterschiedliche Herangehensweisen an einen datenintensiven Beitrag beschrieben.

Die Inhaltsanalysen von datenjournalistischen Beiträgen (u.a. Knight, 2015; Loosen u. a., 2016, 2015) liefern wichtige Rückschlüsse auf die Praxis und rücken das in vielen Interviews durchaus idealisierte Bild des datenintensiven Journalismus als Hüter der Demokratie zurecht. Denn in Bezug auf den Datenzugang ist datenintensiver Journalismus

massiv vom Entgegenkommen von öffentlichen Stellen sowie von Öffentlichkeitsarbeit, zum Beispiel in Form von institutionellen und kommunalen Open-Data-Initiativen, abhängig. Nur wenige Beiträge basieren auf ‚geleakten‘ Daten oder ‚freedom of information requests‘. Zwischen dem, was viele Texte für den datenintensiven Journalismus reklamieren – etwa „digitale Wachhunde“ in einer modernen Demokratie zu sein (Felle, 2016) –, und der alltäglichen Praxis klafft eine beträchtliche Lücke. Deshalb ist der datenintensive Journalismus aus einer Gesamtsicht nicht als investigativer Journalismus zu klassifizieren – auch wenn in einzelnen Regionen und Medien doch investigativer Datenjournalismus betrieben wird.

Weiters lieferte die Literatursynthese erste Befunde zur konkreten Arbeit an den Daten. Es zeigte sich, dass auch Datenarbeit im Journalismus immer mit Experimentieren verbunden ist. Die Prozesse, die für die Entwicklung eines Narrativs angestrengt werden – Digitalisierung, Reinigung, Filterung, Plausibilitätsprüfung, etc. –, sind im höchsten Maße subjektiv. Das an diesen Abschnitt anschließende Kapitel führt insbesondere diesen Aspekt noch genauer aus.

Ähnlich verhält es sich in der Phase der Produktion und Visualisierung – auch hier wird an vielen Stellen in der Produktion ausgesondert, interpretiert und zurechtgerückt, um Aspekte wie Thema, Daten, Narrativ und redaktionelle Linie in Einklang zu bringen. Die Arbeit an Daten und der Visualisierung amalgamiert mit der Arbeit am Text – manchmal durch ein Individuum, immer öfter durch ein Team, in dem jede Person ihre spezifischen Fähigkeiten einbringen kann. Wenig ist aus der Literatur über verwendete Werkzeuge, technische Standards und Infrastrukturen zu erfahren, wiewohl diese Akteure wichtige Determinanten des datenintensiven Journalismus darstellen.

Über die Besonderheiten beim Veröffentlichungsprozess von datenintensiven Beiträgen wusste die Forschung bislang wenig zu berichten. Es ist anzunehmen, dass sich kaum etwas vom klassischen Textjournalismus im Onlinebereich unterscheidet – mit Ausnahme der Freigabe von Visualisierungen, für die es wohl eigener redaktioneller Prozesse Bedarf. Eine weitere Erkenntnis aus der Literatursynthese ist, dass den in einer Redaktion mit Datenjournalismus befassten Personen wenig bis gar nicht klar ist, wie die datenintensiven Beiträge beim Publikum ankommen. Systematisch wird das nicht erhoben; allein über Metriken aus Content-Management-Systemen und Social-Media-Plattformen sowie über direktes Feedback wie Leserbriefe lassen sich Rückschlüsse ziehen; diese sollten daher auch verstärkt zur Analyse eingesetzt werden.

Eine letzte Erkenntnis, die sich aus der Inhaltsanalyse ergab, ist, dass datenintensiver Journalismus die Interaktion mit RezipientInnen auf zumindest zwei Ebenen umdefiniert: Einerseits wird das Publikum in manchen Fällen weit stärker als im traditionellen Textjournalismus eingebunden, indem es etwa aufgefordert wird, selbst Daten beizutragen, zu analysieren oder zu visualisieren – Stichwort Crowdsourcing und Co-Produktion. Andererseits lassen sich datenintensive Beiträge weit besser auf die einzelne rezipierende Person anpassen – entweder durch diese selbst in Form interaktiver Features, andererseits automatisiert aufgrund von Daten, die über UserInnen bereits vorliegen – Stichwort Personalisierung.

6 Interviewstudie: Die Praxis datenintensiver journalistischer Arbeitsabläufe⁴⁶

Das vorangegangene Kapitel analysierte einerseits aus einer Metaperspektive die Forschung der letzten zwei Jahrzehnte zu datenintensivem Journalismus, andererseits fasste es zusammen, was bislang zu Arbeitsabläufen im Datenjournalismus festgestellt worden war. Dieses Kapitel exploriert letzteren Aspekt anhand einer konkreten, lokalen Forschungsarbeit, einer empirischen Interviewstudie, die insbesondere die soziotechnische Co-Kreation in den Mittelpunkt rückt.

Die epistemologische Basis für dieses Kapitel bilden wiederum die im Kapitel zu den wissenschaftstheoretischen Überlegungen formulierten Annahmen, dass die Realitätskonstruktion von historischen, sozialen, kommunikativen und technischen Faktoren abhängt. Wenn wissenschaftliche Erkenntnis sich derart konstruiert, dann auch die Produktion eines datenintensiven Beitrags. Bauer (2005) beschreibt diese Perspektive am Beispiel von Nachrichtenagenturen:

News agencies are more than just input-output organisations for the flux of information. They are socially, politically and/or economically organized networks of sharing societal discourses. News agencies socialize information and deal with fragile constructs of reality rather than with real new events. In that sense, news agencies are making news. And news are making views.

Entsprechend gilt es, diese Aspekte in der Interviewstudie zu berücksichtigen. Um die Forschungsfragen 2.1–2.3 zu beantworten, wurde eine Interviewstudie (1) durchgeführt, die mit Daten aus teilnehmenden Beobachtungen (2) kombiniert wurde.

⁴⁶ Teile dieses Kapitels basieren auf einem andernorts veröffentlichten Text (Ausserhofer, 2015), der für diese Arbeit adaptiert und aktualisiert wurde.

6.1 Methode: Semistrukturierte Interviews und teilnehmende Beobachtungen

Insgesamt wurden semistrukturierte Interviews mit 15 Personen durchgeführt, die hauptberuflich datenintensiven Journalismus betreiben.⁴⁷ Ihre Beschäftigung übten sie entweder fest angestellt in Redaktionen aus (11) oder als externe DienstleisterInnen für Medienhäuser (4). Zum Zeitpunkt des Interviews waren die GesprächspartnerInnen in den folgenden elf Organisationen tätig: *New York Times*, *Guardian*, *ProPublica*, *Twitter* (ehemals *Guardian*), *Der Spiegel*, *Die Zeit*, *Deutsche Welle*, *Berliner Morgenpost*, *Der Standard* sowie den Datenjournalismus-Agenturen *Open Data City* und *Journalism++*.⁴⁸ Die Mehrheit der Interviewten arbeitete täglich direkt mit Daten, ein kleinerer Teil hatte eine redaktionelle Leitungsfunktion inne und koordinierte datenintensive Projekte bzw. damit befasste Teams. Elf der Interviewten arbeiteten in Deutschland oder Österreich; dementsprechend betrifft die Ergebnisdarstellung stärker diesen Raum. Die vier anderen Interviews (drei aus den USA, eines aus Großbritannien) werden gemeinsam mit der lokal geerdeten Forschungsliteratur als Kontrolldaten verwendet, um mögliche regionale Unterschiede auszumachen. Die Interviews wurden zwischen Ende August 2013 und Juni 2015 geführt. Weiter unten wird der methodologische Entscheidungsprozess erläutert.

Ergänzt werden die Interviews mit einer mehrjährigen teilnehmenden Beobachtung des Felds des datenintensiven Journalismus. Wie im Abschnitt über das Forschungsdesign der Arbeit beschrieben, beobachtete ich die Subdisziplin des datenintensiven Journalismus sowohl on- als auch offline, führte dutzende Gespräche, unterrichtete Datenjournalismus und arbeitete in einschlägigen kooperativen Forschungsprojekten. Insbesondere eine 15-monatige Zusammenarbeit mit der Tageszeitung *Der Standard* im Rahmen eines gemeinsamen Projekts zu rechten Empörungsbewegungen in sozialen Medien half mir dabei, die redaktionellen Arbeitsabläufe im datenintensiven Journalismus noch besser zu verstehen. Die ‚Daten‘ aus dieser Beobachtung existieren zum Teil in Form von implizitem Wissen, zum Teil wurden sie

⁴⁷ Die Befragten waren (in alphabetischer Reihenfolge): Michael Bauer, Mona Chalabi, Sarah Cohen, Christina Elmer, Florian Gossy, Markus Hametner, Thomas Jöchler, Nicolas Kayser-Bril, Mirko Lorenz, Lorenz Matzat, T. Christian Miller, Simon Rogers, Maximilian Schäfer, Julius Tröger, Sascha Venohr.

⁴⁸ Manche der Befragten waren (auch) für die Onlinepublikationen dieser Organisationen tätig.

systematisch in einem digitalen Forschungstagebuch dokumentiert. Sie werden insbesondere für die Fundierung und Kontextualisierung der Interviewdaten herangezogen.

6.1.1 Sampling

Sampling in qualitativer Forschung kann unterschiedlichen Logiken folgen (Flick, 2007, S. 26). Aufbauend auf den Arbeiten von Kuzel (1992) und Patton (1990) haben Miles und Huberman (1994, S. 28) eine umfassende Typologie vorgelegt, die verschiedene Samplingstrategien erläutert. Diese stellte auch die Grundlage für die im Zuge meiner Arbeit gefällten Samplingentscheidungen dar. Von Anfang an war klar, dass aus forschungsökonomischen Gründen keine Gesamterhebung, das heißt Interviews mit allen Personen, die sich als DatenjournalistInnen identifizieren, durchgeführt hätte werden können. Es musste also eine Auswahl getroffen werden. Vor dem Hintergrund des Forschungsinteresses (Arbeitsabläufe und Produktionsbedingungen im datenintensiven Journalismus mit einem Schwerpunkt im deutschsprachigen Raum und einem besonderen Blick auf nicht-menschliche Akteure) war klar, dass „größtmögliche Variation“ (Miles & Huberman, 1994, S. 28) bei der Auswahl der InterviewpartnerInnen eine Rolle spielen musste. Wenn verschiedene Akteure, die in verschiedenen Medienorganisationen arbeiten, Ähnliches berichten, sollte es möglich sein, prototypische Beschreibungen von Arbeitsabläufen zu formulieren. Eine weitere wichtige Samplingstrategie, die der Forschung zugrunde gelegt wurde, war die „Homogenität“ (Miles & Huberman, 1994, S. 28) von Fällen. Dies steht dem Prinzip der größtmöglichen Variation nicht entgegen, schließlich brauchte es auch verbindende Kriterien, die Aussagen ermöglichen, die über den einzelnen Fall hinausgehen: Im Zentrum sollten also Redaktionen von Medien bzw. die dort tätigen Personen und Infrastrukturen stehen, die bereits datenintensiven Journalismus betreiben, beziehungsweise im Optimalfall sogar Routinen in ihren datenintensiven Arbeitsabläufen etabliert haben. Eine weitere Strategie war, bei einzelnen Fällen eine gewisse „Intensität“ (Miles & Huberman, 1994, S. 28) zu erreichen, also informationsreiche Fälle zu finden. Dieses Prinzip sollte befolgt werden, indem, soweit möglich, mehrere Personen pro Medium für ein Interview angefragt werden sollten. Zusätzliche Intensität sollte durch Unmittelbarkeit erreicht werden: Nach Möglichkeit sollten die Interviews face-to-face und in den jeweiligen Arbeitsumgebungen geführt werden, sodass neben den Interviewfragen auch anwesende Personen und Objekte ein Auslöser für die Schilderung weiterer Aspekte sein könnten. Als letzte Leitlinie schließlich sollte das

„Schneeball“-Prinzip zum Tragen kommen: Die Interviewten sollten am Ende des Gesprächs nach weiteren relevanten Fällen und potenziellen InterviewpartnerInnen gefragt werden.

Diesen Prinzipien entsprechend wurden zunächst Personen für ein Interview kontaktiert, weil sie in der einschlägigen Berichterstattung über Datenjournalismus wiederholt genannt worden waren oder weil sie sich auf *Twitter* unter dem Hashtag #ddj geäußert hatten und sich in ihrem dortigen Profil entsprechend deklarierten. Darüber hinaus spielten auch die im Abschnitt über das Forschungsdesign erläuterte eigene Einbindung in das Feld sowie Forschungsaufenthalte in den USA (Berkeley) und Berlin eine Rolle. In den ersten Interviews sowie in Gesprächen auf Veranstaltungen wurden weitere Personen genannt, die sich für ein Interview eignen würden. Im Laufe der Forschung wurde zum Zweck der besseren Fundierung durch Iteration auch die Samplingstrategie angepasst.⁴⁹

Schließlich konnten die folgenden Personen für ein Interview gewonnen werden:

Tabelle 5. Übersicht über die geführten Interviews

Name	m/w	Medium	fest/frei	Land	Ort	Sprache	fzf	AU
Julius Tröger	m	Berliner Morgenpost	fest	Deutschland	Berlin	dt.	X	
Christina Elmer	w	Der Spiegel	fest	Deutschland	Hamburg	dt.	X	X
Maximilian Schäfer	m	Der Spiegel	fest	Deutschland	Hamburg	dt.	X	X
Michael Bauer	m	Der Standard	fest	Österreich	Wien	dt.	X	X
Florian Gossy	m	Der Standard	fest	Österreich	Wien	dt.	X	X
Markus Hametner	m	Der Standard	fest	Österreich	Wien	dt.	X	X
Mirko Lorenz	m	Deutsche Welle / Data Wrapper	n/a	Deutschland	Skype	dt.		
Thomas Jöchler	m	Die Zeit	fest	Deutschland	Berlin	dt.	X	X
Sascha Venohr	m	Die Zeit	fest	Deutschland	Berlin	dt.	X	X
Mona Chalabi	w	Guardian	fest	Großbritannien	Graz	en.	X	
Nicolas Kayser-Bril	m	Journalism++	frei	Deutschland	Berlin	en.	X	X
Sarah Cohen	w	New York Times	fest	USA	Skype	en.		
Lorenz Matzat	m	Open Data City	frei	Deutschland	Berlin	en.	X	X

⁴⁹ Zur Rolle von Iteration in der Forschung siehe auch den Abschnitt Forschungsschemata im Kapitel methodologische Verortung).

Name	m/w	Medium	fest/frei	Land	Ort	Sprache	f2f	AU
T. Christian Miller	m	Pro Publica	fest	USA	Berkeley, CA	en.	X	
Simon Rogers	m	Twitter	n/a	USA	Skype	en.		

Sämtliche Angaben stellen die Situation zum Zeitpunkt des Interviews dar. m/w = männlich/weiblich; fest = Anstellung in einer Redaktion; frei = externe(r) DienstleisterIn; Land = Hauptsitz des Mediums; dt. = deutsch; en. = englisch; f2f = face-to-face Interview; AU = Interview in der Arbeitsumgebung.

Dieses Sample besitzt einerseits eine relativ große ‚Variation‘ – mit DatenjournalistInnen, die in verschiedenen Ländern für verschiedene Redaktionen arbeiteten, teils freischaffend, teils fest angestellt. Zugleich durchzieht das Sample eine gewisse ‚Homogenität‘ – alle Interviewten waren für überregional bekannte Qualitätsmedien tätig, ein Großteil für Medien in Deutschland und Österreich. Auch dem Sample-Prinzip der ‚Intensität‘ konnte nachgekommen werden: Bei den deutschsprachigen Redaktionen konnten mit drei interviewten Personen bei der Tageszeitung *Der Standard*⁵⁰ und jeweils zwei Personen bei den Wochenzeitungen *Die Zeit* und *Der Spiegel* mehrere Perspektiven berücksichtigt werden. Zwölf der Interviews fanden face-to-face statt, davon neun im jeweiligen Arbeitsumfeld (in der Regel in der Redaktion).⁵¹ Im Anschluss an die Interviews wurden noch Beobachtungs- bzw. Gedankenprotokolle angefertigt, die zusätzlich zum Prinzip der ‚Intensität‘ beitragen.

6.1.2 Anbahnung und Ablauf der Datenerhebung

Der ersten Anfrage für ein Interview gingen umfassende Hintergrundrecherchen zur Person voraus. Die erste Kontaktaufnahme erfolgte in der Regel per E-Mail. Es wurde der Zweck der Forschung mitgeteilt und um Teilnahme an der Studie gebeten.

Ausgehend von den Forschungsinteressen und der methodologischen Fundierung wurde ein schwach-strukturierter Leitfaden entwickelt, der sich im Anhang dieser Arbeit befindet. Beim angehängten Dokument handelt es sich um einen Grundstock an Fragen. Den Empfehlungen der Methodenliteratur folgend, dass Forschende jedem Gegenüber als Individuum begegnen sollten (Krotz, 2005, S. 139), wurde der Fragenkatalog für jedes Gespräch auf den/die BefragteN abgestimmt. Vor dem ersten ‚Praxiseinsatz‘ wurde der

⁵⁰ Zwei der Interviews bei der Tageszeitung *Der Standard* wurden von Sarah Matiasek geführt, zum damaligen Zeitpunkt Studentin und später studentische Mitarbeiterin am Studiengang für Journalismus und PR der FH Joanneum – meiner ehemaligen Arbeitsstätte. Der von mir entwickelte Leitfaden wurde ihr dafür zur Verfügung gestellt. Sie verwertete die Gespräche in ihrer Bachelorarbeit (Matiasek, 2015).

⁵¹ Face-to-face-Interviews liefern etwas mehr Tiefe als Interviews, die mittels Telefon oder Skype geführt wurden (Irvine, 2011), auch wenn die Unterschiede nicht signifikant sind (Sturges & Hanrahan, 2004).

Leitfaden mit KollegInnen im Rahmen eines Pre-Tests erprobt. Ein Ziel war es, die Lebenswelt der Interviewten zu beschreiben (Rubin & Rubin, 2012), deshalb begann jedes Interview nach der Eröffnung mit Fragen zur Person, zu deren Motiven und Hintergründen. Dabei interessierte mich etwa die Ausbildung, der berufliche Werdegang und der Einstieg in den Datenjournalismus. Anschließend wurden die redaktionellen Arbeitsabläufe bei der Produktion von datenintensiven Beiträgen besprochen. Dabei wurde im Speziellen auf die Ebene der Kollegenschaft, die Organisation, sowie die konkrete Praxis von Methoden und Technik eingegangen. Auch rückten dabei die menschlichen und nicht-menschlichen Akteure in den Mittelpunkt des Gesprächs. Der gesamte datenintensive redaktionelle Prozess, von der Themenfindung über die Produktionsbedingungen bis hin zur Dissemination datenjournalistischer Projekte, wurde im Detail durchgegangen. Den Abschluss bildeten Fragen zu digitalen Methoden und zur allgemeinen Einschätzung datenintensiver Verfahren in Redaktionen. Auch wurde versucht, mehr über Reflexionen zur eigenen Praxis zu erfahren. Die Fragen waren dabei immer so offen formuliert, dass den Befragten der Raum gelassen wurde, die eigene Sicht der Dinge deutlich zu machen (Krotz, 2005, S. 138).

Die Durchführung sämtlicher Interviews verlief friktionsfrei. Die Gespräche wurden allein gelegentlich durch Telefonanrufe oder KollegInnen unterbrochen, die jedoch nur kurzzeitige Ablenkungen darstellten. Die durchschnittliche Gesprächsdauer betrug genau eine Stunde (Median: 55 Minuten). Bei jenen Interviews, die vor Ort in Redaktionen geführt wurden, fand darüber hinaus eine gemeinsame Begehung der Redaktionsräumlichkeiten statt. Im Anschluss an die Interviews wurden Beobachtungsprotokolle angefertigt, in denen Eindrücke niedergeschrieben wurden, die nicht mit dem Aufnahmegerät festgehalten werden konnten – Aspekte wie räumliche Ordnung, Gesten und Materialien in interpersonellen Interaktionen.

Die Interviews wurden zeitnah nach dem Interviewtermin transkribiert. Dabei wurde weitestgehend dem Regelwerk von Dresing und Pehl (2015) gefolgt: Das bedeutet, dass Interviews wörtlich transkribiert wurden, nicht lautsprachlich oder zusammenfassend. Dialekte wurden möglichst wortgetreu der Schriftsprache angepasst, inhaltlich jedoch nicht verändert. Die auf Englisch geführten Interviews wurden auch in dieser Sprache belassen und nur einzelne Passagen für die Ergebnisdarstellung übersetzt. Wortverschleifungen wurden an das Schriftdeutsch angenähert.

6.1.3 Auswertung

Für die Analyse des Materials wurde dasselbe Codebuch herangezogen, das bereits für die systematische Analyse der Forschungsliteratur verwendet worden war und im Anhang der Arbeit dargestellt ist. Während sich die Überkategorien aus dem Forschungsinteresse abgeleitet hatten, wurden die Codes auf Basis der Transkripte und der empirischen Forschungsliteratur entwickelt. In einem iterativen Prozess wurden die Codes auf die Gesamtheit des Materials (Transkripte, Beobachtungsprotokolle) angewendet bis ein zufriedenstellendes Kategoriensystem bzw., um mit Glaser und Strauss (1967) bzw. Strauss und Corbin (1990/1996) zu sprechen, theoretische Sättigung im Kategoriensystem erreicht war. Das Material wurde anschließend in ein Programm zur computerunterstützten qualitativen Auswertung, den *TAMS Analyzer* (Weinstein, 2006), importiert und codiert. Die verdichtete Zusammenfassung dieser Analyse wird im folgenden Ergebnisteil wiedergegeben. Wie von Kvale (2007) empfohlen, wurden zum Zweck der kommunikativen Validation die Zitate vor der Veröffentlichung nochmals den Interviewten vorgelegt. Auch wurden die Namen der Interviewten durch Nummern ersetzt.

Das Material wurde in der Analyse als Beitrag zur Theoriengenerierung gesehen (Bogner, Littig & Menz, 2014, S. 76–83). Vor dem Hintergrund, dass sich eine intersubjektive Realität nicht erschließen lässt (vgl. auch dazu das Kapitel zu den wissenschaftstheoretischen Grundlagen), hatte die Auswertung zum Ziel, das Deutungswissen jeder mit Datenjournalismus befassten Person hinsichtlich ihrer Arbeitsabläufe zu erschließen (Bogner u. a., 2014, S. 76). Im Kern ging es darum, das „Überindividuell-Gemeinsame“ (Meuser & Nagel, 2002, S. 80) herauszuarbeiten; mit anderen Worten: „gemeinsam geteilte, gewissermaßen typische Wissensbestände, Relevanzstrukturen und Deutungsmuster zu rekonstruieren“ (Bogner u. a., 2014, S. 78). Aus diesem Grund wurde bei der Entwicklung der Codes und insbesondere beim Codieren selbst auf diese Aspekte Rücksicht genommen. Nach dem Codieren erfolgte die Konzeptualisierung und theoretische Generalisierung (Bogner u. a., 2014, S. 78), deren Ergebnisse im folgenden Unterabschnitt dargestellt werden.

6.2 Ergebnisse der Interviewstudie

6.2.1 Interdisziplinäre Kollaborationen

Die Produktion eines datenjournalistischen Stücks ist in der Regel ein kollektives Unterfangen, erklärte die Mehrzahl der Interviewten. Viele unterschiedliche Akteure helfen mit, bis ein Projekt erscheint. ‚Den Datenjournalisten‘ oder ‚die Datenjournalistin‘ gibt es in den wenigsten Organisationen. Gewöhnlich werden datenjournalistische Stücke arbeitsteilig von SpezialistInnen produziert. Ihre Expertise liegt in Programmierung, Statistik, Software-Projektmanagement, interaktiver Grafik, nutzerzentriertem Design, Kartografie, Fact-Checking oder in einem spezifischen Themengebiet. Viele der Interviewten sind ExpertInnen in mehreren Domänen. Mit geringer externer Unterstützung können sie allein datenjournalistische Stücke veröffentlichen. Zentral dafür sind das Beherrschen dreier Kompetenzen: Datenanalyse bzw. Programmierung, Design sowie Formulierungsstärke. Mängel in Visualisierungsfähigkeiten können in vielen Fällen durch den Einsatz webbasierter Tools kompensiert werden.

Die finanziellen Probleme, von denen viele Medienhäuser betroffen sind, beeinträchtigen auch die Personalpolitik im Datenjournalismus. So wie im Textjournalismus ist es im Datenjournalismus nichts Ungewöhnliches, für bestimmte Aufgaben externe Freischaffende zu beauftragen. Nur ressourcenstärkere Nachrichtenorganisationen leisten es sich, Vollzeitkräfte oder sogar Teams zu beschäftigen, die sich allein um datenjournalistische Projekte kümmern. Viele, die sich in Redaktionen an journalistischen Datenprojekten beteiligen, interessieren sich privat für Informationsvisualisierung oder Datenbanken. Sie sind ReporterInnen oder RedakteurInnen, häufig auch MitarbeiterInnen der Grafik-, Archiv- oder Entwicklungsabteilung, die sich innerhalb der Organisation ein neues Aufgabenfeld suchen und einen Teil ihrer Arbeitszeit auf einschlägige Projekte verwenden. Je nach Ressourcen und Projektanforderungen beteiligen sie sich in interdisziplinären Teams. Die Leitung in diesen Teams übernimmt meist eine Person mit journalistischem Profil, die die Arbeit dann zwischen sich selbst und den Verantwortlichen für Datenanalyse, Programmierung, Visualisierung und dem Fachressort, dem die Geschichte thematisch zugeordnet ist, verteilt.

Insofern nehmen Personen, die überwiegend einer mit Datenjournalismus assoziierten Tätigkeit nachgehen, in der Aufgabenverteilung in ihren Organisationen eine selten klar definierte Funktion ein: Einerseits sind sie selbstständige ReporterInnen und RedakteurInnen,

die selbst recherchierte Projekte publizieren, andererseits unterstützen sie RedakteurInnen, die keine ausgeprägte Datenkompetenz haben, bei ihren Veröffentlichungen. Sie sehen sich sowohl als Teil der Redaktion als auch als Teil der Supportdienste. RedakteurInnen „suchen regelmäßig die fachliche Beratung [bei uns]“, erzählt ein Datenjournalist (I7). Und weiter: „Es ist mittlerweile so, dass wir in der Redaktion etabliert sind, dass wir uns als geschätzten Partner sehen, der Arbeit nicht wegmoderiert. (...) [Die Fachredakteure] wissen: Wenn wir sagen, etwas geht nicht, dann geht es wirklich nicht.“ Zu dieser Rolle zwischen Redaktionsmitglied und DienstleisterIn passt auch die Tatsache, dass DatenjournalistInnen interne Weiterbildungen durchführen, in denen sie die technischen Kompetenzen von FachredakteurInnen verbessern und ihnen Tools vermitteln, mit Hilfe derer sie ohne externe Hilfe Datenvisualisierungen anfertigen können.

Interessierte treffen sich in internen Fortbildungen genauso wie auf einschlägigen, klassisch abgehaltenen Konferenzen. Darüber hinaus werden auch informelle Treffen mit Gleichgesinnten organisiert – sowohl hausintern als auch extern. Eine besondere Rolle spielen Formate wie die Veranstaltungen der internationalen Graswurzelbewegung *Hacks/Hackers*, bei der ‚Hacks‘⁵², also JournalistInnen, mit ‚Hackern‘⁵³ zusammenkommen, um sich auszutauschen und zusammenzuarbeiten (detaillierter beschrieben bei Lewis & Usher, 2014). Ein Reporter (I8), der sich hauptsächlich mit Datenprojekten beschäftigt und auf einem solchen Treffen ein Projekt vorgestellt hat, berichtet von seinen Erfahrungen im Anschluss an sein Referat: „Dort kommen die entsprechenden ‚Nerdfragen‘, die du sonst nicht bekommst. Und ich erzähle dann genau, wie wir das gemacht haben, helfe dem einen oder anderen auch damit.“ Die gegenseitige Unterstützung ist im Datenjournalismus laut den Interviewten stark ausgeprägt – auch über Medienunternehmen hinweg. Selbst MitarbeiterInnen von Unternehmen, die miteinander in Konkurrenz stehen, würden sich gegenseitig helfen. Das gute Klima der Kollaboration betonen alle Befragten in ihren Aussagen. Sie unterstützen einander etwa bei technischen Problemen oder inhaltlichen Fragen. „In dieser Szene geht das sehr schnell. Man bringt eine Frage auf Twitter ein und dann antworten einem drei Koryphäen aus den USA oder von wo auch immer“, erklärt ein Interviewter (I7). *Twitter* wird

⁵² ‚Hacks‘ ist ein umgangssprachlicher englischer Ausdruck für (Print-)JournalistInnen, meist humoristisch oder abschätzig verwendet („Hack. Def. 4b.“, 2016).

⁵³ ‚Hacker‘ in diesem Kontext bezeichnet nicht Personen, die Sicherheitslücken in Computersystemen ausnützen, sondern technologieaffine Personen, die mit ihren Programmierkenntnissen spielerisch gesellschaftliche Probleme lösen wollen (Coleman, 2013).

übereinstimmend als ein zentraler Disseminations- und Feedbackkanal genannt. Die Beteiligten organisieren ihre Kommunikation unter anderem durch die Verwendung des Hashtags #ddj. Auch kommunizieren viele Akteure miteinander per Telefon, Skype oder E-Mail. Praktiken wie etwa das Einbetten fremder Inhalte oder die Übernahme von Visualisierungsmethoden oder von Stories mit lokalisiertem Inhalt sind üblich.

6.2.2 Routinen und Arbeitsabläufe

Als große Herausforderung nannte ein Großteil der Interviewten die Einführung von Routinen in die Produktion datenjournalistischer Stücke. Viele der Projekte haben experimentellen Charakter – selbst in Organisationen, die bereits seit mehreren Jahren Datenjournalismus betreiben. Die Installation von Routinen wird nicht zuletzt durch die Heterogenität der Projekte erschwert. Diese zeigt sich an drei Aspekten: Erstens sind Datenjournalismus-Verantwortliche in vielen Organisationen sowohl für tagesaktuelle als auch für langfristig angelegte Projekte zuständig. Da diese Projekte je nach Realisierungshorizont unterschiedliche Herangehensweisen erfordern, ist es schwer, Routinen zu etablieren. Zweitens macht das Themenspektrum, das von einer einzelnen Person oder einem Team behandelt wird, nicht vor Ressortgrenzen halt: Mit Datenjournalismus betraute Personen arbeiten gewöhnlich an Fragestellungen und heterogenen Daten aus allen Themenbereichen einer Redaktion – von Außenpolitik über Sport bis zu Chronik. Drittens setzen viele Projekte den Einsatz neuer Tools, Software oder Visualisierungsbibliotheken voraus. „Man hat keine Strategie für die nächsten Jahre (...), man kann online nur die nächsten zwei Wochen planen und ‚wurstet sich durch‘ [arbeitet durchgehend behelfsmäßig]“, resümiert ein Journalist (I8).

Auch wenn kaum datenjournalistische Routinen etabliert werden, so können trotzdem abgeschlossene Praktiken identifiziert werden, die miteinander in Beziehung stehen: Initiiert werden datenjournalistische Projekte durch selbstständige Ideen und Recherchen sowie durch redaktionsinterne oder externe Anregungen und Einflüsse. Mehrere der Interviewten nennen die Veröffentlichung neuer Daten durch statistische Dienste als zentrales Ausgangsmoment ihrer Projekte. Die Datenerfassung geschieht in komplexen Projekten oftmals automatisiert mit Hilfe von Scrapern oder Crawlern. Auch werden Daten auf ‚Vorrat‘ gesammelt, obwohl zum Zeitpunkt der Akquise noch nicht klar ist, inwieweit diese sich für eine Projektrealisierung eignen. Ein freischaffender Datenjournalist (I1) erläutert den Hintergrund:

Am besten sammelst du einmal alles ein, was es gibt. Du kannst später immer noch filtern. Der Scraper läuft dann, du lässt ihn, wenn man so will, ‚reifen‘. Nach einem Monat oder zwei guckst du in die Daten und schaust, ob du Muster findest.

Das Finden von Mustern ist Teil einer „explorative[n] Phase, in der man erst einmal schaut: Was sind das überhaupt für Daten? In welchem Zustand sind sie? Was für ein Aufwand ist es, sich damit zu beschäftigen? Welche Form kann das am Ende annehmen?“, erklärt ein anderer Interviewter (I3). Weitere Praktiken, die ebenso am Beginn eines Projekts stehen, sind die Erstreinigung der Daten, die Formulierung von Hypothesen und das Herausbilden eines Narrativs. Nicht jedes Projekt enthält zwangsläufig jede Praxis – und auch die Reihenfolge unterscheidet sich von Mal zu Mal. In einigen Projekten etwa verzichten die ProduzentInnen auf das Formulieren von Hypothesen oder das Reinigen der Daten.

Die Explorationsphase ist geprägt durch Experimentieren und Sondieren. Daten werden geprüft und miteinander verglichen. Filter werden hinzugezogen und angepasst. Fehler in den Daten werden ausgebessert. Pivottabellen oder Visualisierungen in Form von Achsendiagrammen (z.B. Säulen-/Balken-, Streu-, oder Liniendiagramme), Netzwerkgraphen oder Karten liefern erste Eindrücke. Zitat eines Datenjournalisten (I8): „Es passiert ständig: Wir haben irgendwelche Ideen, werfen Datensätze rein und dabei kommt gar nichts raus.“ „Es geht viel um herumspielen“, charakterisiert ein anderer (I10) diese Phase. Es werden Fragen an die Daten gerichtet und probiert, diese mithilfe der zur Verfügung stehenden Ressourcen zu beantworten. Diese Fragen bilden dann die Grundlage für die Formulierung des Narrativs. Damit wiederum geht für gewöhnlich die Entwicklung der Datenvisualisierung einher. Dabei werden Skizzen per Hand aufs Papier gebracht. Bei komplexeren Projekten werden auch sogenannte Wireframes, also am Computer gezeichnete Entwürfe, angefertigt. Diese sollen dabei helfen, die Sicht der EndanwenderInnen von Anfang an in die Konzeption zu integrieren. Sofern mehr als eine Person an einem Projekt arbeitet, ist in diesem Produktionsabschnitt die Abstimmung zwischen den einzelnen Teammitgliedern besonders intensiv. Allerdings gibt es kaum formalisierte Besprechungen, dafür aber starken informellen Austausch: Indem die Teammitglieder auf Abbildungen deuten, Skizzen anfertigen oder unter Zuhilfenahme von Gestik ihre Ideen artikulieren, aktualisieren sie wechselseitig ihre Vorstellungen über das Projekt. Unterstützend kommen in den Organisationen verschiedene physische und elektronische Projektkoordinationstools wie Pinnwände, die den Projektstatus abbilden, gemeinsame Chatkanäle in *IRC* oder *Slack*, *Skype*, Wikis, oder Mailinglisten zum Einsatz.

So werden datenjournalistische Projekte in Arbeitsteilung und intensiver gegenseitiger Abstimmung weiterentwickelt, bis sie kurz vor der Finalisierung stehen. Es folgen interne Tests, Feedbackschleifen mit den Ideen- bzw. den AuftraggeberInnen. Ein wichtiges Element, das enorme Ressourcen benötigt, ist die Adaption der Visualisierungen für verschiedene Plattformen, Betriebssysteme und Browser. Gerade in Projekten, bei denen NutzerInnen mit den Inhalten interagieren können, sie also die Informationenerst durch weitere Eingaben erfassen können, stellen unterschiedliche Darstellungs- und Interaktionstechnologien die EntwicklerInnen vor große Herausforderungen. Insbesondere die Adaption für mobile Endgeräte mit ihren kleineren Bildschirmen und anderen Interaktionsmöglichkeiten (,Gesten' statt ,Clicks') machen dabei große Probleme. Schließlich spielt der Veröffentlichungszeitpunkt gerade bei aufwändigeren Projekten eine wichtige Rolle, geben die Interviewten relativ übereinstimmend an. Neben der prominenten Platzierung auf der Startseite eines Mediums setzen sie im ,Rollout' intensiv auf die Distributionseffekte sozialer Medien.

In der Evaluierungsphase wird die Qualität von Unternehmungen von den PraktikerInnen nicht zuletzt danach beurteilt, inwieweit sie die Grenzen des technisch Machbaren ausloten. Mit anderen Worten: Die Qualität eines Projekts bemisst sich nicht allein am Inhalt bzw. dessen erfolgreicher Vermittlung an das Publikum, sondern auch am Arbeitsaufwand, am Komplexitätsgrad der eingesetzten Daten und Technologien sowie an der Weiterentwicklung von Visualisierungsansätzen.

Was beeinträchtigt die Güte eines Projekts? So gut wie alle Befragten nennen Ressourcenknappheit, insbesondere den Mangel an Zeit und Geld, an erster Stelle. Weiters fehlt es in sehr spezialisierten Projekten zuweilen an technischer Expertise. Exemplarisch hierfür die Aussage einer Redakteurin (I2):

Das ist so eine Zwickmühle im tagesaktuellen Online-Journalismus –, dass das, was wir datenjournalistisch machen, in einer perfekten Welt natürlich viel länger und noch reflektierter und optisch opulenter daherkommen könnte, wir uns das aber oft einfach nicht leisten können.

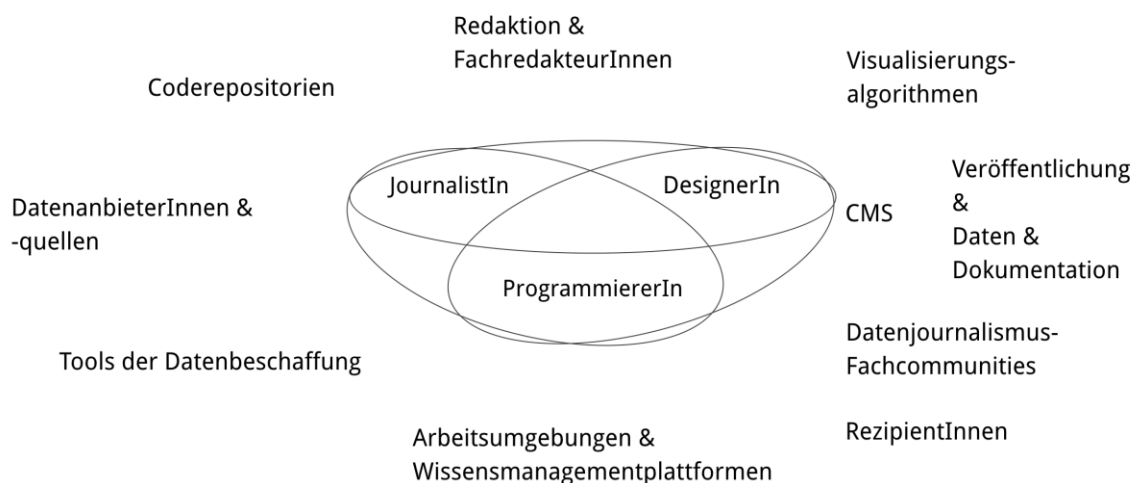
Die Knappheit an Ressourcen führt unter anderem dazu, dass die Nachbereitung der Projekte in vielen Fällen nicht so abläuft, wie es sich die PraktikerInnen wünschen würden. Insbesondere betrifft das die Veröffentlichung der Daten, der selbst entwickelten Programme und Werkstattberichte, die von allen Interviewten als wichtige Elemente am Ende des datenjournalistischen Workflows genannt werden. Dadurch soll die eigene Arbeit transparent und nachvollziehbar gemacht werden sowie mögliche Fehler nach der Veröffentlichung

korrigiert werden. Auch wenn DatenjournalistInnen diese selbst gesteckten Ziele nicht immer erreichen, halten sie die Idee sehr hoch, möglichst viel Transparenz über das eigene Schaffen herzustellen. „Wir machen alles open source, weil wir sehen nicht, warum es anders sein sollte“, erklärt ein Gründer einer Datenjournalismusagentur (I4) seine Motive.

6.2.3 Akteure digitaler datenjournalistischer Routinen

In den Beschreibungen der Arbeitsabläufe im vorhergehenden Unterabschnitt wurden bereits zahlreiche Akteure erwähnt, die an datenjournalistischen Methoden beteiligt sind. Eine Auswahl an Akteuren, die mehrere Interviewten nannten, ist in Abbildung 19 illustriert. Die Mehrheit der Interviewten schildert, dass Software-Komponenten im gesamten Prozess zentrale Rollen einnehmen. In der Phase der Datenakquise sind es etwa sogenannte ‚digitale Briefkästen‘. Dabei handelt es sich um Webformulare, über die eine Organisation verschlüsselt und anonymisiert Daten empfangen kann. Auch Crawler und Scraper kommen zum Einsatz. Nicht zuletzt sind auch APIs, also Schnittstellen zu Datenbanken, zentrale Agenten der datenjournalistischen Methode. Sie kontrollieren nicht nur den Zugang zu strukturierten Informationen, sondern formatieren gemeinsam mit den Datenbanken und deren Nutzungsbedingungen auch die Weiterverarbeitung der Daten vor (ausführlicher erläutert im Abschnitt zu den Herausforderungen digitaler Methoden dieser Arbeit sowie bei Ausserhofer, in Druck; Bodle, 2011; T. Bucher, 2013; Puschmann & Ausserhofer, 2017).

Abbildung 19. Akteure einer datenjournalistischen Produktion



Weitere Hard- und Software-Agenten beeinflussen die lokale Datenaufbereitung und -visualisierung: In den Interviews wurden unter anderem Tabellenkalkulationsprogramme und Visualisierungswerkzeuge genannt. Die DatenjournalistInnen würden die eigenen Vorstellungen ständig anpassen, weil die Software-Anwendungen nur bestimmte Operationen zuließen. Durch Programmieren könne man diese Restriktionen jedoch weitestgehend überwinden. Ein Datenjournalist (I10) erläutert: „Die Software gibt vor, wie Geschichten auszusehen haben. Zugleich ist es ein großer Vorteil, dass wir im Newsroom einen Entwickler haben, der mit uns eigene Lösungen baut, damit wir uns nach den Daten richten können und nicht nach der Software.“

Daten werden auf lokalen Servern oder in der ‚Cloud‘ abgelegt. Darüber hinaus dienen die Server auch als ‚Hosts‘ für Datenscraper und -crawler sowie News-Apps. Äußerst wesentlich erscheinen Hosting-Dienste wie *GitHub*, mit Hilfe derer einerseits kollaborativ Programmcode entwickelt wird, andererseits Daten veröffentlicht werden. Immer wieder stören lokale Sicherheitsvorkehrungen und Kompatibilitätsprobleme die Arbeit an Projekten, zum Beispiel wenn der Laptop, auf dem die Visualisierung entwickelt wird, bestimmte Programme nicht ausführen kann oder wegen Sicherheitsbedenken von den lokalen Servern ausgesperrt wird – alltägliche Situationen, wie sie mehrfach geschildert wurden.

Eine nicht unwesentliche Rolle spielt auch die räumliche Position der DatenjournalistInnen in der Organisation. Aus den Aussagen der Interviewten und den Beobachtungen während der Redaktionsbesuche lässt sich keine eindeutige Tendenz ablesen – auch weil die räumliche Situation in jeder Organisation unterschiedlich ist. Manche Akteure waren mitten im Newsroom stationiert, andere in der Entwicklungsabteilung, dritte waren ohne zugeteilten Arbeitsplatz, saßen mit ihrem Laptop an freien Tischen oder besetzten eigens für Recherche vorgesehene Räume. In den Agenturen dominierten Gemeinschaftsbüros.

Zunehmend werden traditionelle Web-Aufmerksamkeitsmetriken wie Verweildauer durch andere quantitative Indikatoren aus sozialen Medien ergänzt. Im Datenjournalismus sind Social-Web-Währungen (Gerlitz & Helmond, 2013) wie ‚Shares‘, ‚Likes‘, ‚Fans‘ oder ‚Follower‘ doppelt relevant: einerseits bei der Veröffentlichung fertiger Projekte als Maßstab für deren Erfolg, andererseits bei Projekten, die sich inhaltlich mit dem Web beschäftigten, als Datenquelle für Analysen und Visualisierungen. Mit letzterem beschäftigt sich der folgende Unterabschnitt: mit datenjournalistischen Projekten, die ähnlichen Fragen wie kommunikationswissenschaftliche Internetforschung nachgehen.

6.2.4 Digitale datenjournalistische Methoden

Automatisierungen, Fokus auf digitale Objekte und die Analysen von Web-Plattformen: Das sind einige zentrale Merkmale von digitalen Methoden, wie sie weiter oben in einem Unterabschnitt im Methodologie-Kapitel besprochen wurden. Diese Prinzipien sind auch vielen datenjournalistischen Projekten inhärent. So gibt es zahlreiche journalistische Analysen, die die Kommunikationsinhalte, das Verhalten oder die Netzwerke von NutzerInnen untersuchen. Das Themenspektrum ist dabei ähnlich breit wie in der kommunikationswissenschaftlichen Internetforschung und reicht von Mappings von NutzerInnenhalten bei Naturkatastrophen bis hin zur Analyse politischer und persönlicher Öffentlichkeiten. Aufgrund des relativ einfachen Datenzugangs greifen viele Analysen auf Social-Media-Dienste und hier insbesondere auf *Twitter* zurück.

Bei größeren Nachrichtenereignissen werden datenjournalistische Digital-Methods-Untersuchungen immer wieder herangezogen, um die Kommunikationsintensität oder Meinungen von Menschen zu erheben. Ein Datenjournalist (I7) erklärt, warum er Social-Media-Daten für seine Berichterstattung heranzieht: „Jetzt haben wir die Möglichkeit, zumindest ein Gefühl dafür zu bekommen, wie tausende Menschen zu einem Thema denken.“ Beispielhaft für diese Herangehensweise ist das Projekt *Die Freunde von Pegida* (Biermann, Blickle & Venohr, 2015), bei dem die Anhängerbasis der Facebook-Seite der ausländerfeindlichen Dresdner Organisation *Pegida* analysiert wurde. Nachdem der Ertrag klassischer soziologischer Erhebungen bislang ‚dünn‘ sei, wandten sich die AutorInnen der Datenbasis der Fanseite zu, untersuchten etwa deren Wachstum, deren angegebenen demografischen Besonderheiten und Vorlieben. Dass diese Social-Media-Öffentlichkeiten nicht repräsentativ sind und mit ihrer Konstruktion auch Politik gemacht wird, ist den befragten JournalistInnen dabei durchaus bewusst. Die Interviews zeigten, dass insbesondere aufgrund des häufig vorherrschenden Ressourcenmangels Hoffnungen in Automatisierungen gesetzt werden: Einige der Interviewten würden gerne Scripts und Scraper so konfigurieren, dass sie manuelle Arbeit verringern, und die Veröffentlichung schneller bzw. annähernd in Echtzeit (etwa bei Social-Media-Daten) passieren kann.

Viele journalistische Digital-Methods-Projekte beschäftigen sich auch mit den persönlichen Öffentlichkeiten von Personen, die im Medieninteresse stehen. Exemplarisch seien die Netzwerkanalysen über den mutmaßlichen Boston-Marathon-Attentäter (Lotan, 2013) sowie die Analysen über die Twitteraktivität von Berliner (Tröger, 2012) und Schweizer

(Thelitz u. a., 2014) Abgeordneten angeführt. Ein gut dokumentiertes Beispiel für digitale Methoden im Datenjournalismus ist eine Untersuchung über die Verbreitung von Falschmeldungen auf *Twitter* während der Ausschreitungen in London im Sommer 2011 (Dant & Richards, 2011; Guardian Interactive Team, Procter, Vis & Voss, 2011). Für dieses Projekt arbeitete der *Guardian* mit drei britischen Universitäten zusammen. Insgesamt wurden 2,6 Millionen Tweets mit spezifischen Hashtags gesammelt, gefiltert und ein Teilkorpus davon manuell codiert. Mit Hilfe einer zeitbasierten interaktiven Visualisierung konnten die AutorInnen Momente aufzeigen, in denen Gerüchte auf *Twitter* verbreitet und entkräftet wurden (Procter u. a., 2013).

6.3 Diskussion der Ergebnisse der Interviewstudie

Dieses Kapitel beleuchtete datenintensive journalistische Arbeitsabläufe mit einem besonderen Augenmerk auf die darin involvierten menschlichen und nichtmenschlichen Akteure. Aufbauend auf qualitativen Analysen von 15 semistrukturierten Interviews und Beobachtungsprotokollen aus Redaktionsbesuchen, die überwiegend im deutschsprachigen Raum stattfanden, konnten neue Perspektiven auf Methoden und Routinen im Datenjournalismus gegeben werden. Es wurde herausgearbeitet, welche Kompetenzen die in Datenjournalismus involvierten Individuen haben. Es zeigte sich, dass nicht allein Datenkompetenz, ‚data literacy‘, in Form von Programmierkenntnissen in der Produktion von datenintensiven Beiträgen gefragt ist. Neben Visualisierungsfähigkeiten benötigt es insbesondere auch Fertigkeiten, die traditionell zum journalistischen Kernkompetenzrepertoire gehören – etwa die Fähigkeit zur Entwicklung eines Narrativs aus heterogenem Quellenmaterial. Wie die Analyse jedoch aufzeigen konnte, wird das Narrativ im datenintensiven Journalismus oftmals aus einer großen Menge strukturierter Daten gewonnen. Die Techniken, die dabei angewendet werden – (Vor-)Visualisierungen, Exploration von Ausreißern, simultane Vergleiche vieler Werte, etc. – ähneln nur bedingt den Techniken des klassischen Textjournalismus. Auch wenn die Fragen, die im datenintensiven Journalismus an das Material gestellt werden, ähnliche bleiben und sich bei den Nachrichtenwerten wenig ändert (Dick, 2014), so sind die Techniken, die für die Herausbildung eines Narrativs angewendet werden, sehr verschieden.

Neben der Entwicklung von Narrativen aus komplexen Datenstrukturen verlangt der datenintensive Journalismus immer öfter auch nach Individuen mit der Fähigkeit zum

Projektmanagement. Die Abläufe, die dabei vor allem in größeren Redaktionen eingeführt wurden, wirken wie eine redaktionelle Adaption der Prozesse agiler Software-Entwicklung (ethnografisch beschrieben von R. Schmidt, 2012). Angesichts der Tatsache, dass viele Akteure des datenintensiven Journalismus ihren Hintergrund in diesem Bereich haben, ist das wenig verwunderlich. Ähnlich wie beim Scrum-Vorgehensmodell werden in Redaktionen sogenannte ‚Sprints‘ veranstaltet, in denen ein gewisser Aspekt einer Funktionalität eines datenintensiven Beitrags implementiert wird. Anstatt schon zu Beginn einen großen Plan inklusive aller Details für das Endprodukt zu entwickeln, gelangt ein Beitrag in Iteration bis zur Publikationsreife. Auch die regelmäßigen Informationsaustauschtreffen (‚daily scrum‘) erinnern an diese Methode. Die Entwicklung „wird (...) ein kollektives kreatives Geschehen, als ein Zusammenspiel zwischen ständig kommunizierenden, körperlich ko-präsenten Teilnehmerinnen, Code, Flipcharts, Haftetiketten und Tafelskizzen neu in Szene gesetzt“ (R. Schmidt, 2012, S. 191).

Die Journalistenausbildung ist gerade erst dabei, Kompetenzen wie Narrativentwicklung aus großen Datenbeständen oder Scrum-Projektmanagement in ihre Curricula zu implementieren. Dazu gibt es seit Jahren eine große Diskussion darüber, ob (zukünftige) JournalistInnen nun Programmieren beherrschen sollten (Creech & Mendelson, 2015; Spinner, 2014). Die Analyse der Praxis des datenintensiven Journalismus in diesem Kapitel zeigte, dass derartige Kompetenzen nicht unbedingt notwendig sind, aber in der Produktion doch von großem Nutzen sind. Zahlreiche renommierte Akteure des Datenjournalismus beherrschen zwei komplementäre Kompetenzen – etwa Programmierung und professionelles Schreiben; oder sie sind in der Lage, selbst Daten zu erheben und zu visualisieren. So können sie datenintensive Beiträge oft selbstständig bis zur Veröffentlichung bringen. In vielen Fällen ist allein das Beherrschen einer Kompetenz und das Verständnis für die anderen Bereiche bereits von Vorteil – etwa, um die Beauftragung externer DienstleisterInnen besser handhaben zu können.

Wie die Interviewstudie auch demonstrierte, ist die Rolle von vielen, die sich der jungen Subdisziplin des datenintensiven Journalismus zugehörig fühlen, in ihren Organisationen nicht völlig klar definiert. Als MethodenexpertInnen (im Unterschied zu den DomänenexpertInnen, den FachredakteurInnen) wurden viele PraktikerInnen von ihren KollegInnen als interne DienstleisterInnen, als eine Art ‚Helpdesk‘, gesehen. Bei den DatenjournalistInnen löste das nicht unbedingt Zufriedenheit aus, zeichnen sie doch häufig für Leuchtturmprojekte verantwortlich, die weiter strahlen als gewöhnliche Texte. Nicht zuletzt deshalb sollten

Redaktionsleitungen die Rolle dieser relativ jungen journalistischen Subdisziplin klar definieren. Vorbild könnte etwa der Umgang mit FotojournalistInnen sein.

Die verschiedenen Datenjournalismus-Communities, in die PraktikerInnen on- oder offline involviert sind, scheinen sich auch beträchtlich auf die Praxis auszuwirken. Das wird einerseits in Kollaborationen manifest, die mehr oder weniger offiziell über die Grenzen von Medienhäusern hinweg durchgeführt werden und Einfluss auf Themenwahl und Produktion haben. Andererseits zeigt sich das am Verve, mit dem datenintensive Beiträge diskutiert werden. Ein Erklärungsansatz für diese wichtige Rolle der extraredaktionellen Gemeinschaften könnten die in vielen Redaktionen fehlenden Austausch- und Feedbackmöglichkeiten sein: Beispielsweise sind DatenjournalistInnen vielfach die einzigen, die eine gewisse Erhebungs- oder Analysemethoden beherrschen. Aufgrund der Komplexität können sie sich intraredaktionell nur bedingt über diese Methoden austauschen. Die technische Finesse ist von Nicht-TechnikerInnen nur schwer zu beurteilen. Außerhalb der Redaktionen gibt es wiederum viele Interessierte, die fundierter Beiträge kritisieren und Verbesserungsansätze liefern können. Hinzu kommt, dass viele DatenjournalistInnen datenintensive Beiträge nicht allein an inhaltlichen Kriterien wie etwa dem Nachrichtenwert beurteilen, sondern auch daran, wie aufwändig die Produktion war und wie technisch komplex die Lösung dahinter ist.

Die Untersuchung der Rolle nichtmenschlicher Akteure förderte einige Technologien und Infrastrukturen zutage, deren Beitrag zum datenintensiven Journalismus bislang nicht oder kaum besprochen wurde. Deren Einfluss, der sich insbesondere in diversen Formatierungen niederschlägt, ist kaum zu unterschätzen und geschieht in allen Phasen des Produktionsprozesses. Das beginnt etwa bei DatenbereitstellerInnen, die Daten nur in bestimmten Formaten oder nur mit gewissen Ausformungen anbieten (vgl. den Unterabschnitt Methodologische Herausforderungen digitaler Methoden) und endet bei Metriken der Aufmerksamkeitsmessung in Content-Management-Systemen und in sozialen Medien. Beträchtlich ist auch die ‚Agency‘ von Visualisierungssoftware und -bibliotheken.⁵⁴ Häufig eingesetzte Tools wie *Datawrapper* oder *Tableau* und Visualisierungsbibliotheken wie *D3.js* prägen die Ästhetik des datenintensiven Journalismus über viele Medien und Webseiten hinweg. Dieses Kapitel konnte den Beitrag ebensolcher Akteure nur anreißen. Zukünftige

⁵⁴ Eine beeindruckende Demonstration der Handlungsmacht von Visualisierungssoftware lieferte etwa Rost (2016), die dieselben Daten und denselben Diagrammtyp mit 15 verschiedenen Visualisierungsbibliotheken darstellte und teils völlig unterschiedliche Ergebnisse erzielte.

Forschungsarbeiten könnten etwa die Produktion eines datenintensiven Beitrags aus Sicht eines solchen Akteurs (z.B. einer Visualisierungsbibliothek) detailliert beschreiben. Durch den damit verbundenen Perspektivenwechsel ließe sich die Handlungsmacht eines solchen Akteurs bei der Produktion datenjournalistischer Stücke wohl noch besser fassen.

Schließlich ist die in den Interviews vielfach noch geäußerte Hoffnung auf Stabilisierung von Prozessen durch Automatisierung anzusprechen. Als mögliche Quelle für derartige Anwendungen wurden häufig Daten von Social-Media-Plattformen genannt – aber nicht ausschließlich. Automatisierter Journalismus (oder algorithmischer Journalismus) wird aktuell jedoch vielfach getrennt vom Datenjournalismus diskutiert (für einen Überblick siehe z.B. Dörr, 2016; Graefe, 2016). Der Ausdruck bezeichnet die massenhafte maschinelle Generierung von singulären Texten ohne (großes) menschliches Zutun. Im Gegensatz dazu ist der in dieser Arbeit charakterisierte datenintensive Journalismus derzeit noch viel mehr Handarbeit (vgl. die Definition in Kapitel 2). Es ist jedoch anzunehmen, dass diese beiden Bereiche stärker verschmelzen werden. In sämtlichen Prozessen der Produktion von datenintensiven Beiträgen gibt es dafür Ansätze – etwa Assistenzsysteme, die Muster bei der Datenerhebung selbstständig erkennen – dabei unter Umständen auch lernen – oder mittels ‚intelligenter‘ Erstanalyse Visualisierungen oder Formulierungsvorschläge liefern. Die vielfach in den Massenmedien diskutierte Frage, ob menschliche JournalistInnen einmal durch ‚Roboter‘ ersetzt werden (z.B. bei Narodslawsky, 2016) ist also durch eine Frage nach der soziotechnischen Natur der Kollaboration abzulösen: Wie wird sich die Zusammenarbeit von Menschen und Technologien in der Nachrichtenproduktion gestalten?

Dieses Kapitel beleuchtete datenintensive Prozesse in Redaktionen aus verschiedenen Perspektiven. Schwerpunktmäßig befasste es sich dabei mit redaktionellen Produktionsbedingungen und den individuellen Kompetenzanforderungen, mit den in einer Produktion vermittelnden technischen Artefakten sowie mit Stabilisierungsmechanismen durch Automatisierungen. Gemeinsam mit dem Vorkapitel, der systematischen Literaturanalyse, bildete dieses Kapitel den empirischen Kern dieser Dissertation. In Kombination geben sie umfassend Auskunft über datenintensive Arbeitsabläufe in der journalistischen Produktion. Das nächste Kapitel versucht zum Abschluss, diese empirienahen Texten zusammenzufassen und theoretisch zu fundieren. Was befördert und behindert den datenintensiven Journalismus und welche Konsequenzen hat dessen Praxis für zunehmend datafizierte Gesellschaften und deren Medienkompetenz sowie menschlich-algorithmisch co-konstruierte Öffentlichkeiten?

7 Resümee und Ausblick

Das abschließende Kapitel dieser Arbeit dient der zusammenfassenden Beantwortung der in Kapitel 1 entwickelten Forschungsfragen.

7.1 Forschungssynthese

Das erste Ziel der Arbeit, eine Forschungssynthese zu datenintensivem Journalismus mit einem Schwerpunkt auf Arbeitsabläufen zu entwickeln, wurde in den Forschungsfragen 1.1-1.5 operationalisiert und in Kapitel 5 ausgeführt.

- FF1.1: Wie entwickelt sich die Forschungsliteratur zu datenintensivem Journalismus, unter anderem in Bezug auf Publikationsaktivität, Veröffentlichungen und Zitationen?

Datenintensiver Journalismus wurde und wird bis heute in unterschiedlichen Disziplinen diskutiert. Neben der Journalismusforschung sind es vor allem technische Disziplinen wie die Informatik, die Statistik und die Visualisierungsforschung, die sich des Themas aus verschiedenen Perspektiven annehmen. Während die Bewegung des ‚computer-assisted reporting‘, die sich vor allem in den USA seit der zweiten Hälfte des 20. Jahrhunderts stetig stärker etablierte⁵⁵, von der Forschung kaum beachtet wurde, ist dies seit dem Ende der 2000er Jahre mit dem Reframing datenintensiver journalistischer Praktiken als ‚Datenjournalismus‘ bzw. ‚data-driven journalism‘ anders. Seit 2010 ist eine große Zunahme der Publikationsaktivität zu verzeichnen – zunächst zögerlich und in den letzten Jahren immer stärker. Langsam scheint sich auch ein kritisches Forschungsfeld der ‚data journalism studies‘ herauszuschälen, das sich von der positivistischen Praxisliteratur emanzipiert. Meyers (1973/2002) *Precision Journalism* ist jedoch nach wie vor einer der wichtigsten Bezugspunkte.

- FF1.2: Was sind die Theorien der Datenjournalismusforschung?

Der empirischen Datenjournalismusforschung muss eine gewisse Theoriearmut attestiert werden. Wenn Arbeiten überhaupt mit Theorien gerahmt wurden, dann vor allem durch die Perspektive der Wissenschafts- und Technikforschung – und hier insbesondere durch jene der Akteur-Netzwerk-Theorie (ANT). Diese Wahl ist sinnvoll und nachvollziehbar, gibt es

⁵⁵ Die Wurzeln des quantitativen Journalismus und ‚computational journalism‘ liegen viel weiter zurück (vgl. dazu C. W. Anderson, 2015; M. Cox, 2000).

zwischen der redaktionellen Praxis des datenintensiven Journalismus und jener in wissenschaftlichen Laboren oder der Softwareentwicklung doch zahlreiche Parallelen. Gerade in Situationen, in denen diverse Technologien und Infrastrukturen eng mit RedakteurInnen kooperieren und großen Anteil am Zustandekommen eines datenintensiven Beitrags haben, ist die Perspektive der ANT sehr fruchtbar. Aus diesem Grund basierte die empirische Vor-Ort-Forschung dieser Arbeit unter anderem auch auf der ANT. Die empirische Datenjournalismusforschung kennt jenseits der Science and Technology Studies natürlich noch weitere Theorien. Diese wurden meist eingesetzt, um die erforschten Aspekte rund um die datenintensive journalistische Praxis zu kontextualisieren.

- FF1.3: Was sind die Methoden der Datenjournalismusforschung?

Die Reflexion des eigenen forschenden Handelns im Rahmen einer methodologischen Diskussion findet in der Literatur kaum statt. Die Methodenbeschreibungen der akademischen Datenjournalismusforschung bestehen in der Regel aus Berichten darüber, welche Maßnahmen im Rahmen der Forschung gesetzt wurden, um die Forschungsfrage(n) zu beantworten. Bevorzugt wurden dafür qualitative Methoden (insbesondere Interviews mit PraktikerInnen sowie Inhaltsanalysen datenintensiver Beiträge sowie Beobachtungen) eingesetzt. Quantitative oder digitale Methoden zur Erforschung des datenintensiven Journalismus fanden bislang kaum Berücksichtigung. Der örtliche Schwerpunkt der Datenjournalismusforschung korreliert stark mit jenen westlichen, vor allem englischsprachigen Ländern, in denen die ForscherInnen arbeiten. Neben der englisch- und deutschsprachigen Forschung existieren vereinzelte Veröffentlichungen zu datenintensivem Journalismus unter anderem in Französisch und Spanisch. Die Literatur verbindet bekannte (Medien-)Organisationen wie den *Guardian*, die *New York Times*, *ProPublica* und *Wikileaks* mit datenintensivem Journalismus. Darüber hinaus konnte die Analyse einige lokale Medienhäuser identifizieren, die bislang in der öffentlichen Debatte kaum oder nicht präsent waren und sich für zukünftige Fallstudien eignen könnten.

- FF1.4: Welche Forschungslücken haben Datenjournalismus-Forschende identifiziert und was sind ihre Vorschläge für zukünftige Forschung?

Die Inhaltsanalyse erlaubte es, die wahrgenommenen Forschungslücken zu erheben und zu kontextualisieren. Mitglieder der Forschungsgemeinschaft würden vergleichende Forschung begrüßen, die nationale Besonderheiten berücksichtigt. Auch gibt es weiteren Bedarf nach ethnografischen Studien, wiewohl mittlerweile Usher (2016) mit ihrer „hybriden Ethnographie“ dazu beigetragen hat, diese Forschungslücke zumindest in Bezug auf die

Leuchttürme des datenintensiven Journalismus in USA zu schließen. Auch gibt es die Meinung, dass man bislang zu wenig über die Rolle von Software in den Produktionsprozessen weiß.

Während die redaktionellen Praktiken mittlerweile relativ gut erforscht sind – auch diese Arbeit liefert einen Beitrag dazu –, ist wenig über die NutzerInnen von datenintensivem Journalismus bekannt. Auch zu Rezeptionserfahrungen von datenintensivem Journalismus gibt es keine mir bekannte Forschung. Daher schlage ich aufbauend auf der Inhaltsanalyse vor, die Onlinespuren jener Plattformen zu analysieren, die im Rahmen des datenintensiven Journalismus genutzt werden. Zuletzt fällt auf, dass Gender- und Minderheitenfragen sowie die (ökonomische) Lage von freischaffenden DienstleisterInnen bislang kaum thematisiert wurden.

7.2 Datenintensive journalistische Arbeitsabläufe in der Forschungsliteratur

- FF1.5: Was hat die Datenjournalismusforschung bislang zu datenjournalistischen Arbeitsabläufen festgestellt?

Im Sinne einer Zusammenfassung des Deutungswissens verfolgte die Inhaltsanalyse hinsichtlich dieser Forschungsfrage das Ziel, auch das „Überindividuell-Gemeinsame“ (Meuser & Nagel, 2002, S. 80) hinsichtlich datenintensiver journalistischer Routinen herauszuarbeiten. Dabei wurden verschiedene redaktionelle Praktiken identifiziert und anhand der verschiedenen Produktionsschritte beschrieben. Die detaillierte Erläuterung einzelner Routinen würde an dieser Stelle zu weit führen und kann im entsprechenden Unterabschnitt von Kapitel 5 nachgelesen werden. Stattdessen seien hier die zentralen Ergebnisse zusammengefasst:

Verschiedene Themen und Ressorts eignen sich unterschiedlich gut für datenintensiven Journalismus. Der Ausgangspunkt eines Beitrags kann sowohl daten- als auch themeninduziert sein. Manche DatenjournalistInnen sind in ihren Redaktionen interne DienstleisterInnen, deren Aufgabe sich mit jener von FotojournalistInnen vergleichen lässt. Andere wiederum sind ein Aushängeschild für ihre Redaktion. Öffentliche Stellen werden von der Forschungsliteratur am häufigsten als Quellen für Daten im datenintensiven Journalismus genannt. Open-Government-Data-Initiativen werden von vielen mit Datenjournalismus befassten Personen kritisch gesehen

und vor allem mit ‚Schönfärberei‘ in Verbindung gebracht. Selbst erhobene Daten werden zwar als Ideal gesehen, spielen in der alltäglichen Arbeit aber eine untergeordnete Rolle. Der datenintensive Journalismus ist, dem Tenor der Forschungsliteratur zufolge, stark von externen Datenquellen und zunehmend von datenbasierter PR abhängig.

Datensäuberung, -verifikation und -analyse sind für gewöhnlich die zeitaufwändigsten Prozesse in der gesamten Produktion. In dieser Phase gibt es häufig Probleme mit der Datenqualität. Die DatenjournalistInnen stellen Plausibilitätsprüfungen und ethische Überlegungen an. Auch wird diese Phase dazu genutzt, ein Narrativ für den datenintensiven Beitrag zu entwickeln. Der Kernprozess der Fertigung ist ein iterativer: Wiederholt werden Texte, Daten und Visualisierungen aufeinander abgestimmt und getestet. Kleinere datenintensive Projekte werden allein von einer/einem einschlägigen RedakteurIn produziert; bei größeren kollaborieren Personen mit verschiedenen Schwerpunkten.

Über den eigentlichen Veröffentlichungsprozess weiß die bisherige Forschung zu datenintensivem Journalismus wenig zu berichten. Es wird weitgehend versucht, die Daten aus Gründen der Reproduzierbarkeit und Transparenz auch zu veröffentlichen. Jedoch kann diesem Ideal oft nicht entsprochen werden. Nur ein kleiner Teil der RezipientInnen nutzt tatsächlich die interaktiven Funktionen, die viele datenintensive Beiträge ausmachen. Datenjournalistische Beiträge scheinen im Vergleich zum Durchschnitt des ‚Textjournalismus‘ höhere Zugriffsraten und länger anhaltendes Interesse verzeichnen zu können. Ein weiteres (neben der Veröffentlichung von Quelldaten oftmals) nur schwer erreichbares Ideal ist die Interaktion mit dem Publikum während der Produktion und nach der Veröffentlichung. Zwar scheint dies als Königsweg zu gelten, aber vielen Organisationen fehlen für derartige Prozesse die personellen, zeitlichen und finanziellen Ressourcen.

7.3 Datenintensive journalistische Arbeitsabläufe in der Praxis

Das zweite Ziel der Arbeit war, die Praxis datenintensiver redaktioneller Arbeitsabläufe mit einem besonderen Augenmerk auf den deutschsprachigen Raum zu beschreiben. Dieses Ziel wurde in den Forschungsfragen 2.1-2.3 operationalisiert, die in Kapitel 6 beantwortet wurden.

- FF2.1: Wie stellen sich die redaktionellen Produktionsbedingungen für mit datenintensivem Journalismus befasste JournalistInnen im deutschsprachigen Raum dar?

DatenjournalistInnen im deutschsprachigen Raum sind eher Methoden- als ThemenspezialistInnen. Als Teil einer durch Technologie spezifizierten journalistischen Spezialdisziplin⁵⁶ beherrschen sie Methoden, die anderen Personen in der Redaktion für gewöhnlich nicht oder kaum geläufig sind: Programmierung, Statistik, Projektmanagement, interaktive Grafik, nutzerzentriertes Design, Kartografie und anderes mehr. In der Redaktion agieren sie als eine Art Stabsstelle oder Service Desk, kooperieren mit FachredakteurInnen zu einem bestimmten Thema oder helfen dabei, Daten aufzubereiten.

Größere Medienunternehmen leisten es sich, mehrere SpezialistInnen für datenintensive journalistische Aufgaben zu beschäftigen. Für gewöhnlich beteiligen sich dann Personen mit unterschiedlichen professionellen Hintergründen an solchen größeren Produktionen. Grob gesprochen behält ein Part sämtliche Aspekte im Auge, die mit Journalismus und Storytelling zu tun haben, ein anderer Part kümmert sich um die Datenaufbereitung, und ein dritter Teil um User Experience und grafische Elemente. Abseits dieser Teams sind auch ‚EinzelkämpferInnen‘ keine Seltenheit, also Personen, die die gesamte Produktion selbst durchführen – von der Themenfindung über die Datenaufbereitung und Visualisierung bis hin zur Veröffentlichung und Evaluierung.

- FF2.2: Wie stellen sich redaktionelle Routinen und Arbeitsabläufe dar?

Die Etablierung von Routinen ist eine große Herausforderung, da beim Umgang mit Daten häufig Probleme auftreten, deren Ausräumung viel Zeit in Anspruch nehmen kann. Erschwerend kommt hinzu, dass Personen, die mit datenintensivem Journalismus befasst sind, für gewöhnlich an mehreren Projekten parallel arbeiten. Bei kollaborativen Produktionen greifen manche Redaktionen auf Methoden aus der agilen Softwareentwicklung zurück, die sich etwa durch morgendliche Kurzbesprechungen im Stehen, ständig präsente Übersichtspläne und kollektive ‚Sprints‘ auszeichnet.

Ein datenintensiver Beitrag durchläuft in der redaktionellen Produktion ähnliche Phasen wie ein gewöhnlicher Textbeitrag, jedoch gestalten sich die einzelnen Teilprozesse durchaus unterschiedlich. Sind die Daten beschafft, werden sie mit verschiedenen Applikationen exploriert und gereinigt. Dabei wird viel mit unterschiedlichen Darstellungsmöglichkeiten experimentiert. Auch sind die Daten entgegen ihrer

⁵⁶ Powers (2012) spricht von „technologically specific work“, Usher (2016) von einer „subspeciality“.

etymologischen Bedeutung nichts ‚Gegebenes‘⁵⁷, sondern werden ständig modifiziert und in andere Formate überführt. Aus diesem Experimentierprozess schälen sich langsam ein Narrativ und mögliche Visualisierungsformen heraus.

Die Veröffentlichung eines Beitrags ist ein ebenso komplexer Prozess wie die Datenexploration und Visualisierungsproduktion. Etwa haben viele gängige Content-Management-Systeme Schwierigkeiten mit der Einbindung von interaktiven Visualisierungen. Ähnlich gelagert ist das Problem, dass viele Beiträge schwer auf mobilen Endgeräten dargestellt werden können. Grund dafür sind die Interaktionsmöglichkeiten, die auf Geräten mit Touchscreen anders funktionieren müssen. Verschärft wird dieses Problem dadurch, dass so gut wie jede Produktion versucht, die technische Lösung zu optimieren und etwa neue Visualisierungstechnologien und -methoden einzusetzen.

- FF2.3: Welche menschlichen Akteure und nicht-menschlichen Entitäten beteiligen sich an der Entstehung datenjournalistischer Produkte?

Die Forschung konnte verschiedene menschliche und nichtmenschliche Akteure identifizieren, die an der Produktion eines datenintensiven journalistischen Beitrags beteiligt sind. Da der Beitrag menschlicher Akteure bereits hinreichend dargelegt wurde, konzentriert sich der folgende Teil der Antwort auf ihre nichtmenschlichen Assoziierten. Einige davon wurden in Abbildung 19 dargestellt. An erster Stelle sind natürlich Datensätze zu nennen, Quellmaterial und ideologisches Artefakt des datenintensiven Journalismus. Darüber hinaus haben verschiedene Software-Agenten großen Einfluss auf ein datenjournalistisches Produkt, zum Beispiel Visualisierungsbibliotheken und -tools, Datenbankapplikationen, Coderepositorien, Programme zur Datenbeschaffung und Content-Management-Systeme. Nicht zuletzt sind interne Logik und Funktionen von Tabellenkalkulationsprogrammen wie *Excel* Determinanten eines datenjournalistischen Produkts. In diesem Kontext ist erwähnenswert, dass Programmiersprachen von den Interviewten als Möglichkeit gesehen wurden, die Einschränkungen, die diverse Applikationen mit sich bringen, zu überwinden. Nun ermöglichen Programmiersprachen in der Regel mehr Gestaltungsfreiheit als Software für EndnutzerInnen. Doch auch ihnen ist eine gewisse Logik inhärent, die eine datenintensive journalistische Produktion nachhaltig beeinflusst.

⁵⁷ Zur Etymologie und zur Bedeutungsverschiebung des Datenbegriffs, siehe auch entsprechenden Abschnitt in Kapitel 2.

Nicht zu unterschätzen ist darüber hinaus der Beitrag von externen institutionellen Infrastrukturen: Förderinstitutionen und insbesondere Fachorganisationen, Online-Netzwerke und geteilte Hashtags, die so etwas wie ein Feld des datenintensiven Journalismus bilden und zusammenhalten. Der Austausch zwischen DatenjournalistInnen ist über die Grenzen von Medienhäusern hinweg weit intensiver als in anderen journalistischen Spezialdisziplinen.

Weitere wichtige Akteure, die die Produktion von datenjournalistischen Beiträgen begleiten, sind schließlich Statistiken und Metriken von Content-Management-Systemen und Social-Media-Plattformen. Ihre Werte haben großes Gewicht für die Rechtfertigung des datenintensiven Journalismus innerhalb der Redaktionen, aber auch in der Öffentlichkeit.

7.4 Die Konsequenzen des datenintensiven Journalismus für mediale Vertrauensbildung und Medienkompetenz

Als drittes Ziel der Arbeit wurde eine Kontextualisierung der Praxis des datenintensiven Journalismus angestellt – und zwar in Gesellschaften, die zunehmend von ‚data literacy‘ als Kulturtechnik und algorithmisch-menschlich co-konstruierten Öffentlichkeiten bestimmt sind. Die Forschungsfrage lautete dabei wie folgt:

- FF3: Welche Konsequenzen hat die Praxis des Datenjournalismus für zunehmend von Datenkompetenz abhängige Gesellschaften und deren menschlich-algorithmisch co-konstruierte Öffentlichkeiten?

Die Beantwortung dieser Frage erfolgt theoretisch und nimmt dabei zugleich auf die Forschungsergebnisse der Arbeit Rücksicht. Zunächst wird auf die Faktizität des datenintensiven Journalismus eingegangen, anschließend auf die Rolle von menschlich-algorithmisch co-konstruierten Öffentlichkeiten, bevor schließlich diese Phänomene im Kontext von journalistischer und gesellschaftlicher Vertrauensbildung und ‚data literacy‘ thematisiert werden.

7.4.1 Zur Faktizität des datenintensiven Journalismus

Datenintensiver Journalismus wird heute mit sehr viel Aufwand arbeitsteilig in verschiedensten Redaktionen auf der ganzen Welt produziert. Das haben die vorangegangenen zwei Kapitel umfassend dargestellt. Egal ob Themenfindung, Datenerhebung, -verifikation und

-analyse, Produktion und Visualisierung oder Veröffentlichung und Evaluierung eines Beitrags: Sämtliche Arbeitsabläufe im Datenjournalismus gestalten sich als sehr zeit- und geldintensiv. Die Etablierung eines tagesaktuellen datenintensiven Journalismus ist in vielen Organisationen nicht möglich, in manchen nur mit sehr großem personellen Einsatz und mit Unterstützung spezieller Softwareanwendungen.⁵⁸

Die in dieser Arbeit beschriebene Praxis ist über weite Strecken ein Phänomen des Qualitätsjournalismus.⁵⁹ Betrieben wird der datenintensive Journalismus hauptsächlich von großen, überregionalen Medienhäusern, die die finanziellen Mittel haben, RedakteurInnen für entsprechende Tätigkeiten zu beschäftigen. Das zeigte sich auch bei der Analyse des örtlichen Fokus der Datenjournalismusforschung in Kapitel 5: In Abbildung 18 finden sich kaum Redaktionen, die in ihren jeweiligen Ländern nicht dem Qualitätssegment zugeordnet werden können.⁶⁰ Die wenigen Versuche, ‚tabloid data journalism‘ zu etablieren, waren wenig erfolgreich; etwa wurde das Projekt *Ampp3d* des britischen Boulevardblatts *Trinity Mirror* nach weniger als zwei Jahren wieder eingestellt (Jackson & Sweney, 2015; Warnes, 2017).

Traditionell versuchen besonders Qualitätsmedien ihre Arbeit an publizistischen Qualitätsstandards zu orientieren (Singer, 2007). Nicht zuletzt ist diese Orientierung bei vielen dieser Medien auch integraler Bestandteil der Außenkommunikation. Der datenintensive Journalismus inszeniert sich dabei häufig als Avantgarde dieser Bemühungen um Qualität: Er formuliert den Anspruch, präziser als gewöhnlicher „Narrativjournalismus“ zu sein (P. Meyer, 1973/2002, 2012); Fakten seien ihm heilig (S. Rogers, 2013); und mit ihm ließen sich

⁵⁸ Dass tagesaktueller Datenjournalismus eine große Herausforderung darstellt, zeigt sich etwa auch daran, dass bei den *Data Journalism Awards* eine eigene Kategorie eingeführt wurde, die „the best use of data in a breaking news story“ auszeichnet (Data Journalism Awards, 2017). Der in den Kriterien definierte Zeitrahmen zwischen Ereignis und Veröffentlichungszeitpunkt beträgt 36 Stunden, was im klassischen Textjournalismus wohl nicht mehr unbedingt als ‚zeitnah‘ gelten würde.

⁵⁹ Es soll hier keine Qualitätsdebatte eröffnet werden, die ohnehin „fast so alt [ist] wie die periodische Presse selbst“ (Wilke, 2003, S. 40) und wohl niemals endgültig abgeschlossen werden kann (H.-J. Bucher, 2003). Als Kriterium für Qualitätsjournalismus sei auf die Forschungssynthese von Engesser (2013, S. 190–197) verwiesen, der für den partizipativen Journalismus im Web insgesamt 36 Kriterien herausarbeitet, die größtenteils auch für die Qualität im datenintensiven Journalismus gültig sind.

⁶⁰ JournalismusforscherInnen beschäftigen sich generell häufiger mit Qualitätsmedien als mit Boulevardmedien. Es ist jedoch nur sehr unwahrscheinlich, dass Boulevardmedien allein aufgrund dieses Bias bislang fast völlig ausgeblendet wurden und deshalb nicht in der Liste der Organisationen aufscheinen, die mit Datenjournalismus assoziiert werden (Abbildung 18). Weit wahrscheinlicher ist, dass es sich bei datenintensivem Journalismus tatsächlich vor allem um ein Qualitätsphänomen handelt.

schließlich die Nachrichten verbessern (Gray u. a., 2012). Auch wenn diese Ideologie, diese normative Epistemologie vielfach kritisiert wurde (z.B. von Haas, 1999, S. 108–112; aber auch von PraktikerInnen selbst, z.B. N. Silver, 2014), haftet sie dem datenintensiven Journalismus weiterhin an. Einen nicht unerheblichen Beitrag für diese Faktizität haben die Datenvisualisierungen, die elementarer Teil des Datenjournalismus sind. Ihnen wird eine höhere Glaubwürdigkeit attestiert (Carney & Levin, 2002).

Nun wächst der datenintensive Journalismus in einer Zeit, in der ‚Fakten‘⁶¹ von vielen Seiten Dekonstruktion erfahren. Diese Abundanz der Dekonstruktion zeigt sich etwa daran, dass „postfaktisch“ bzw. „post-truth“ als Wort des Jahres 2016 gewählt wurde (Gesellschaft für deutsche Sprache, 2016; Oxford Dictionaries, 2016); oder, dass Formulierungen wie ‚alternative facts‘ bzw. ‚alternative Fakten‘, ‚Lie press‘ bzw. ‚Lügenpresse‘ mittlerweile (wieder) zu geflügelten Worten avanciert sind; oder, dass der Ausdruck ‚fake news‘ zunehmend auch im deutschsprachigen Raum um sich greift.

Datenintensiver Journalismus entwickelt sich in einer (europäischen) Gesellschaft, in der eine Mehrheit der Bevölkerung wenig Vertrauen in Medien hat und an deren Unabhängigkeit zweifelt; gerade einmal etwas mehr als die Hälfte der Bevölkerung ist davon überzeugt, dass traditionelle Medien glaubhafte Informationen verbreiten (Europäische Kommission, 2016).

7.4.2 Menschlich-algorithmisch-co-konstruierte Öffentlichkeiten

Zum Vertrauensverlust kommt hinzu, dass der Journalismus schon seit einiger Zeit das Quasi-Monopol der öffentlichen Berichterstattung und Interpretation von Nachrichteneignissen abgegeben hat. Aus SchleusenwärterInnen sind SchleusenbeobachterInnen geworden (Bruns, 2005). Zugleich scheinen sich Ideen einer Demokratisierung der Öffentlichkeit (z.B. bei Rheingold, 1993), einer „multitude of ‚virtual coffeehouses““ (Geiger, 2009, S. 11), dabei ebenso wenig realisiert zu haben wie die Dystopien der gesellschaftlichen Fragmentierung und Isolierung durch das Internet (Habermas, 2006), algorithmisch konstruierte „filter bubbles“ (Pariser, 2011) oder Echokammern (Sunstein, 2007), in denen ideologische Segregation

⁶¹ Der Ausdruck ist hier einmalig unter Anführungszeichen gesetzt, um nochmals auf die in dieser Arbeit vertretene Perspektive der Konstruktion von Realität und Fakten hinzuweisen – ausführlich diskutiert im Kapitel der wissenschaftstheoretischen Fundierung.

unausweichlich scheint.⁶² Realistischer (bzw. näher an der Konstruktion von Realität) wirken da integrierende Ansätze, die zwar die Existenz gewisser isolierter Nischenöffentlichkeiten anerkennen, jedoch durch gewisse koordinierte Effekte auch die Emergenz von größeren „networked public spheres“ (Benkler, 2006) identifizieren (empirisch aufgearbeitet am Beispiel österreichischer Netzöffentlichkeit in Ausserhofer & Maireder, 2013; Maireder & Ausserhofer, 2014). Jedoch können diese koordinierenden Effekte wie etwa Filterung, Akkreditierung, Synthese (Benkler, 2006, S. 242) auf diesen großen Ebenen nur auf Algorithmen (Geiger, 2009) bzw. menschlich-algorithmischer Co-Konstruktion basieren. Das Problem dabei ist, dass die meisten dieser algorithmischen Intermediäre, die versprechen, vernetzte Öffentlichkeiten zu produzieren, von privaten Unternehmen hergestellt und reguliert werden; Außenstehenden sind sie unzugänglich (Ausserhofer, 2013). Nicht zuletzt aus diesem Grund gibt es verstärkt Rufe nach algorithmischer Transparenz und Bemühungen, die Produktion und Arbeit von Algorithmen bei der Konstruktion vernetzter Öffentlichkeiten offenzulegen (Diakopoulos, 2014, 2015; Mager, 2012).

Ein weiterer Aspekt in diesem Kontext ist, dass Medienunternehmen ihre Nachrichtenproduktion und -distribution immer stärker an Social-Media-Plattformen ausrichten (Bell & Owen, 2017; Gillespie, 2010). Im Unterschied zu journalistischen Medienorganisationen können sie auf weit mehr Ressourcen zurückgreifen. Sie kontrollieren in zunehmendem Maße, was Publika sehen, welche Organisationen für diese Aufmerksamkeit wie bezahlt werden, und sogar welches Format und welcher Typ von Journalismus sich entwickeln soll (Bell & Owen, 2017, S. 9).

Eine besondere Rolle in dieser soziotechnischen Konstruktion von vernetzten Öffentlichkeiten kommt indes Nischenöffentlichkeiten zu. Manche treten mit dem institutionalisierten Journalismus in Konkurrenz und nutzen die Produktionsmöglichkeiten und Distributionseffekte digitaler Netzmedien. Es lassen sich neuartige Praktiken beobachten; etwas, das Meraz und Papacharissi (2013) aufbauend auf Barzilai-Nahon (2008) bzw. Entman (1993) „networked gatekeeping“ und „networked framing“ genannt haben: kollektive Quellenbeobachtung und -verarbeitung, in der Gruppen gewisse Problemdefinitionen, kausale

⁶² Filter-Bubble- und Echokammer-Effekte sind nur bei bestimmten Themen und generell äußerst moderat vorhanden; darüber hinaus tragen individuelle Handlungen stärker zu ideologischer Segregation bei als algorithmisch herbeigeführte Auswahlentscheidungen (Barberá, Jost, Nagler, Tucker & Bonneau, 2015; Bozdag, 2013; Flaxman, Goel & Rao, 2016; Gentzkow & Shapiro, 2011; Jürgens, Stark & Magin, 2015).

Interpretationen, moralische Einschätzungen, und/oder Handlungsempfehlungen entwickeln. Derartiges passiert auf vielen gängigen Social-Media-Plattformen – auf Imageboards wie *4chan* (Potts & Harrison, 2013), Social-News-Seiten wie *Reddit* (Leavitt, 2016; Leavitt & Robinson, 2017), auf sozialen Netzwerken wie *Facebook* (A. Silver & Matthews, 2016) und *Twitter* (Hermida, Lewis & Zamith, 2014), und der *Wikipedia* (Roessing, 2014). Gehäuft treten diese Praktiken bei großen Nachrichteneignissen und Notfallsituationen auf. Menschen, die meist nicht direkt von dem Vorfall betroffen sind, aggregieren selbständig Informationen, insbesondere aus Onlinequellen (Hornmoen & Backholm, 2017).

Ähnliche Vorgänge sind in der täglichen politischen Kommunikation und insbesondere im Vorfeld von Wahlen und im rechten politischen Spektrum zu beobachten. Hier versucht ein Ökosystem, bestehend aus etablierten Medien, neuen Nachrichtenkanälen, Social-Media-„SerienaktivistInnen“ (Bastos & Mercea, 2016) und Bots, in verschiedensten Öffentlichkeiten einen Zustand der permanenten Empörung und Erregung herzustellen. Das geschieht etwa, indem große Gruppen von Personen sich absprechen und gemeinsam gewisse Artikel kommentieren oder KritikerInnen angreifen (Broderick, 2017). Um diese Aktivitäten möglichst effektiv ausführen zu können, werden kollaborativ Anleitungen und Vorlagen angefertigt (Broderick, 2017). Auch in der Medienproduktion üben sich diese „smart mobs“ (Rheingold, 2002): Sie erstellen etwa ‚Memes‘, Zuspitzungen oder Falschmeldungen und verbreiten sie über soziale Medien. Dabei werden sie von Bots unterstützt, kleine Skripte, die automatisiert vorgegebene Aufgaben erledigen – etwa ‚kapern‘ sie Hashtags oder adressieren andere UserInnen wiederholt mit Botschaften. Albright (2016) nennt derartige Systeme ‚Mikro-Propaganda-Maschinen‘, „an influence network that can tailor people’s opinions, emotional reactions, and create ‚viral‘ sharing episodes around what should be serious or contemplative issues“.

Zusammengefasst steht der „postindustrielle Journalismus“ (C. W. Anderson, Bell & Shirky, 2012) steht vor vielen Herausforderungen: Aspekte wie Vertrauensverlust, abwandernde Publika, zahlreiche parajournalistische, um die Informationshoheit konkurrierende Akteure sowie neue, die Inhaltsproduktion und Distribution dominierende Intermediäre sind nur einige der vielen Phänomene, die die klassische journalistische Profession vor große Herausforderungen stellen. Die Frage ist, ob die Praxis des datenintensiven Journalismus, wie sie in den vorangegangenen Kapiteln beschrieben wurde, etwas dazu beitragen kann, gewisse Effekte dieser Krise zu lindern. Darauf wird im Folgenden eingegangen.

7.4.3 Vertrauensbildung mit datenintensivem Journalismus

Dem Vertrauensverlust gegenüber Medien und dem Vorwurf von gelenkter Berichterstattung oder ‚fake news‘ ist nicht leicht zu begegnen. Eine Maßnahme auf journalistischer Produktionsseite kann die im Datenjournalismus stark vertretene Offenlegung von Quellen und Methoden sein – gleichsam der Anspruch ein ‚Open Science light‘ zu etablieren (zu Open Science siehe den entsprechenden Abschnitt in Kapitel 3). In Anbetracht der Entwicklung, dass der Versuch, sich Objektivität anzunähern, mehr denn je zum Scheitern verurteilt ist, gilt es, Selbsttransparenz der journalistischen Arbeit als Ideal zu stärken (Soffer, 2009; Weinberger, 2011, S. 112–113). Schließlich bleibt im Netz keine Aussage widersprochen und das Gegenargument zu einer Aussage im Netz stets nur ein Kommentar, eine Suchanfrage oder einen Link entfernt (Weinberger, 2011, S. 12). Freilich ist die Idee der Selbsttransparenz nicht völlig neu (z.B. schon diskutiert von Ruß-Mohl, 2001). Meier und Reimer (2011) haben verschiedene Ansätze systematisiert, die bei der Herstellung von redaktioneller Transparenz helfen können: Dabei unterscheiden sie zwischen Transparenz in den redaktionellen Routinen und Entscheidungen sowie der Transparenz journalistischer Beiträge. Die Maßnahmen sind vielfältig, reichen von der Veröffentlichung von Ethik-Kodizes, Videos aus Redaktionsdiskussionen über Lesereinbindung (in Form von Beiräten, Hotlines, Onlinekommentaren) bis hin zur direkten Nutzerpartizipation bei der Produktion eines Beitrags, der Darstellung der Themenkarriere und Links zum Quellenmaterial. Auch die Offenlegung von redaktionellen Entscheidungen und den Entstehungsprozessen ist eine mögliche Maßnahme. Durch Bemühungen wie diese sollen RezipientInnen in die Lage versetzt werden, die Konstruktion, die bei der Produktion eines Beitrags passiert, zumindest potenziell nachzuvollziehen (Meier & Reimer, 2011, S. 138). Der datenintensive Journalismus greift selbst auf einige dieser Maßnahmen zurück, wie in den beiden Ergebniskapiteln beschrieben wurde. Etwa werden NutzerInnen bei der Datensammlung eingebunden, Datensätze verlinkt, Software-Applikationen veröffentlicht oder Werkstattberichte produziert.

Auch wenn derartige Maßnahmen als Ideal hochgehalten werden, so sind sie nicht ‚Standard‘ im datenintensiven Journalismus (und schon gar nicht im klassischen Textjournalismus). Das liegt vor allem an (zu) knappen Ressourcen; auch gibt es Bedenken bezüglich des Verhältnisses zwischen Ressourceneinsatz und Nutzen – schließlich nimmt nur ein Bruchteil der RezipientInnen derartige Angebote wahr. Darüber hinaus fehlt in vielen Redaktionen ein entsprechendes Qualitätsmanagement (Funck, 2016). Weiters befürchten

manche die Gefährdung der Autonomie einer Redaktion; andere haben die Sorge, dass zu viele Informationen die Aufmerksamkeit abziehen, „von dem, was wirklich wichtig ist“ (Meier & Reimer, 2011, S. 136). Schließlich stellt sich die Frage, ob derartige Transparenzbemühungen überhaupt einen Effekt auf die Vertrauensbildung haben? Mit anderen Worten: Rennt man mit derartigen Maßnahmen nicht nur offene Türen bei jenen ein, die ohnehin schon einem Beitrag bzw. einem Medium vertrauen? In einer Untersuchung mit klassischem Textjournalismus zeigte sich, dass Verlinkungen selbst nur moderat das Vertrauen erhöhen, jedoch redaktionelle Offenheit, also etwa die Einladung zum Dialog mit den RedakteurInnen, vertrauensbildend wirkt (Meier & Reimer, 2011, S. 151). Umgelegt auf den datenintensiven Journalismus bedeutet das, dass die Offenlegung von Quelldaten zwar im Hinblick auf die Reproduzierbarkeit und Weiterverwendung sinnvoll ist, aber relativ wenig zu einem Vertrauensaufbau beiträgt. Die Öffnung von Produktionsprozessen und die Einbindung von RezipientInnen, zum Beispiel bei der Datenerhebung („crowdsourcing“), hat wiederum das Potenzial, vertrauensbildend zu wirken.⁶³

Hinsichtlich der Öffnung von redaktionellen Prozessen und der Einbindung des Publikums schlägt Usher (2016, S. 189) einen Perspektivenwechsel vor. Aufbauend auf dem Umstand, dass mit dem datenintensiven Journalismus Routinen der Open-Source-Softwareentwicklung in die journalistische Produktion Einzug gehalten haben, plädiert sie, Nachrichten als Programmcode („news as code“) zu sehen⁶⁴:

For example, news can be thought of as specific chunks that can be dealt with in discrete ways. These small chunks can be fact checked on their own. If news is code, that code can be open to the public, and the public can be invited to contribute to the elements that make up stories. Thus, the story becomes a participatory platform built from small elements into a larger, collectively shaped story.

Obwohl die Umsetzung einer solchen Idee weit davon entfernt ist, realisiert zu werden, ist es durchaus interessant, über deren Konsequenzen zu reflektieren. Ein journalistischer Beitrag, egal ob allein Text oder ein Gemenge aus Text, Daten und Visualisierung, besteht somit aus einzelnen Teilen, die geprüft werden können. Die Öffentlichkeit wird eingeladen, diese

⁶³ Nicht zum Vertrauensaufbau, jedoch zur Absicherung der Legitimation in einer inhaltlichen Diskussion zu einem Artikel sind die besprochenen Erweiterungen jedoch in jedem Fall sinnvoll.

⁶⁴ Holovaty (2006) hatte diese Idee bereits zehn Jahre zuvor in den Diskurs eingebracht und damit stark zum Reframing datenintensiver journalistischer Praktiken als Datenjournalismus beigetragen.

Elemente zu hinterfragen. Für die Überprüfung werden die jahrzehntlang bewährten Methoden des ‚code reviews‘ aus der Softwareentwicklung adaptiert.

Eine weitere Möglichkeit, Publikum über datenintensiven Journalismus einzubinden und damit das Vertrauen zu erhöhen, sind schließlich Beiträge mit algorithmischer Personalisierung (Usher, 2016, S. 190), die die RezipientInnen in ihrer speziellen Situation, ihrem eigenen Wissenstand abholen. Beiträge, die einerseits illustrieren, wie der/die RezipientIn ganz persönlich betroffen ist, die aber auch das ‚große Bild‘ vermitteln (vgl. Latour, Jensen, Venturini, Grauwin & Boullier, 2012). Damit ist nicht gemeint, die Konsequenzen von Daten mit einem Menschen zu illustrieren, quasi ‚den Daten ein Gesicht zu geben‘. Vielmehr meint dies, aufbauend auf Daten, die RezipientInnen freiwillig zur Verfügung stellen, individuelle algorithmische (Text-)Beiträge zu produzieren. Dies würde wohl zu einer größeren Bindung der Leserschaft und zu Vertrauensaufbau führen. Eine Kennzeichnung solcher individuell erzeugten Beiträge wäre im Sinne einer algorithmischen Transparenz unabdingbar. Zugleich könnten derartige Lösungen, massenhaft eingesetzt, die fragilen Sphären vernetzter Öffentlichkeiten unter Umständen weiter dekonstruieren.

7.4.4 Vergesellschaftete data literacy

Ein Problem im Kontext der Selbsttransparenz im datenintensiven Journalismus ist, dass Daten in Rohform eine viel größere Barriere für RezipientInnen darstellen als gewöhnliche Texte.⁶⁵ Ein Rechercheinterview, das zusätzlich in voller Länge veröffentlicht wird, ist weit zugänglicher als ein Datensatz oder ein Scraper, für deren Aneignung weit andere Kompetenzen erforderlich sind. Dies führt unweigerlich zur Frage der Daten- und Medienkompetenz. Diese betrifft nicht allein die RezipientInnen. Wie die Analysen zeigen konnten, handelt es sich bei datenintensivem Journalismus um eine Praxis, die sich zwar in den vergangenen Jahren stark verbreitet hat, die jedoch in Redaktionen nur von einem kleinen Teil der Belegschaft beherrscht wird. Es ist anzunehmen, dass diese Popularisierung auch in den kommenden Jahren nicht abreißen wird und selbst von nicht direkt mit Datenjournalismus befassten RedakteurInnen einschlägige Kompetenzen verlangt werden. Sowohl

⁶⁵ Dies gilt im Übrigen nicht nur für unstrukturierte Daten, sondern auch für deren Visualisierungen. Auch diese können weniger zugänglich sein als gewöhnlicher Text.

redaktionsintern als auch gesamtgesellschaftlich wird die Anforderung, digitale Medienkompetenz und ‚data literacy‘ zu beherrschen, steigen.

Was gilt es also zu beherrschen, um mit datenintensiven journalistischen Inhalten zu interagieren bzw. diese zu produzieren? Die Ergebniskapitel haben gezeigt, dass im Kern drei Professionen an der Produktion datenintensiver Beiträge beteiligt sind: Menschen mit Hintergrund in Programmierung, Design und/oder Onlinejournalismus. Im Kontext der kritischen Interaktion und Produktion im Datenjournalismus wird häufig der Begriff der ‚data literacy‘ genannt. Sie umfasst im Kern jene Kompetenzen, die auch in den Ergebniskapiteln diskutiert wurden: die Rolle von Daten in einer Gesellschaft zu verstehen, Daten zu finden und zu beziehen; Daten zu lesen, zu interpretieren und zu evaluieren; Daten handzuhaben; und schließlich Daten zu verwenden, inklusive der Produktion von Elementen der Datensynthese und des ethischen Umgangs mit Daten (Calzada Prado & Marzal, 2013). Würde man in dieser Definition ‚Daten‘ durch ‚Web‘, ‚Social Media‘ oder ‚neue Medien‘ ersetzen, so würde diese Beschreibung jenen Aufzählungen ähneln, die unter Begriffen wie „web literacies“ (Chung, O’Byrne & Bond Gill, 2016; Wittenbrink & Ausserhofer, 2013), „new media literacies“ (Jenkins, Purushotma, Weigel, Clinton & Robison, 2009) oder „social media literacies“ (Rheingold, 2010) beschrieben werden. Und selbst das „nach wie vor prominenteste“ (Dander, 2014, S. 6) Medienkompetenzmodell, jenes von Dieter Baacke (1973, 1996; Treumann, Baacke, Haacke, Hugger & Vollbrecht, 2002), nennt ähnliche Aspekte: Medienkritik, Medienkunde, Mediennutzung und Mediengestaltung.

An dieser Stelle könnte man jetzt die einzelnen Kompetenzmodelle ‚filetieren‘ und deren Eignung für die datenjournalistische Praxis oder die (Journalistenaus-)Bildung prüfen. Ein solches ‚Abhaken‘ ist jedoch nur bedingt sinnvoll, schließlich wäre, wie Bauer (2014a, S. 14) argumentiert, eine solche Frage falsch gestellt, weil sie allein die „Nachrüstung des Individuums den zunehmend komplexer werdenden gesellschaftlichen Rahmenbedingungen gegenüber, oder nur als Nachrüstung der Strukturen der Gesellschaft dem Individuum gegenüber“ thematisieren würde. Stattdessen soll eine holistischere Perspektive aufgemacht werden (Bauer, 2014a, S. 17): Ausgehend von der Vorstellung, dass Medienorganisationen auch zu den gesellschaftlichen „Basisinstituten“ (so wie Politik, Wirtschaft, Bildung) zu zählen sind (Bauer, 2014a, S. 14) und dass Datenkompetenz als Teil eines Kommunikationssinns (Bauer, 1979, S. 169–179) von verschiedensten gesellschaftlichen Akteuren co-konstruiert wird, haben Medienorganisationen, die datenintensiven Journalismus betreiben, nicht nur eine publizistische sondern auch eine ‚daten-pädagogische‘ Verantwortung. Mit anderen Worten:

Es reicht für Medienorganisationen nicht, zunehmend datenintensive Beiträge zu produzieren. Sie müssen durch die Einbindung von RezipientInnen auch versuchen, kritische ‚data literacy‘ einer Gesellschaft zu steigern. Die anderen gesellschaftlichen Basisinstitutionen haben sich ebenso daran zu beteiligen.

Selbstverständlich kann ein offener datenintensiver Journalismus inklusive einer von mehreren Akteuren getragenen Datenkompetenz nur einen kleinen Beitrag zum Kitten der Risse leisten, die durch die Gesellschaft gehen. Ohne eine Politik, die sich des Problems der wachsenden sozialen Ungleichheit annimmt, die Medien-Oligopole und Distributions-Quasi-Monopole (Prokop, 2002) stärker reguliert und öffentlich-rechtliche und am Gemeinwohl orientierte Medien (Fuchs, 2014; Lorenz-Meyer, 2013) als echte Alternativen etabliert, werden diese Bemühungen vergeblich sein.

8 Literaturverzeichnis

- Abels, H. (2007). *Interaktion, Identität, Präsentation: Kleine Einführung in interpretative Theorien der Soziologie* (4. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Adler, E. S. & Clark, R. (2011). *An invitation to social research: How it's done* (4. Aufl.). Belmont, CA: Wadsworth.
- Aitamurto, T. (2013). Balancing between open and closed: Co-creation in magazine journalism. *Digital Journalism*, 1(2), 229–251. doi:10.1080/21670811.2012.750150
- Aitamurto, T. (2016). Crowdsourcing as a knowledge-search method in digital journalism. *Digital Journalism*, 4(2), 280–297. doi:10.1080/21670811.2015.1034807
- Aitamurto, T. & Lewis, S. C. (2013). Open innovation in digital journalism: Examining the impact of open APIs at four news organizations. *New Media & Society*, 15(2), 314–331. doi:10.1177/1461444812450682
- Aitamurto, T., Sirkkunen, E. & Lehtonen, P. (2011). *Trends in data journalism* (Forschungsbericht). Abgerufen von http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf
- Albright, J. (2016, November 18). The #election2016 micro-propaganda machine. Abgerufen von <https://medium.com/@d1gi/the-election2016-micro-propaganda-machine-383449cc1fba>
- Alvesson, M. & Deetz, S. (2000). *Doing critical management research*. London: Sage.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6. Aufl.). Washington, DC: American Psychological Association.
- Andersen, E. (2013). *Datastøttet journalistikk*. Oslo: Cappelen Damm Akademisk.
- Anderson, C. (2008, Juni 23). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. Abgerufen von <https://www.wired.com/2008/06/pb-theory/>
- Anderson, C. W. (2011a). Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms. *Journalism*, 12(5), 550–566. doi:10.1177/1464884911402451

- Anderson, C. W. (2011b). Notes towards an analysis of computational journalism. *HIIG Discussion Paper Series*, 2012(1). doi:10.2139/ssrn.2009292
- Anderson, C. W. (2013a). *Rebuilding the news: Metropolitan journalism in the digital age*. Philadelphia, PA: Temple University Press.
- Anderson, C. W. (2013b). Towards a sociology of computational and algorithmic journalism. *New Media & Society*, 15(7), 1005–1021. doi:10.1177/1461444812465137
- Anderson, C. W. (2015). Between the unique and the pattern: Historical tensions in our understanding of quantitative journalism. *Digital Journalism*, 3(3), 349–363. doi:10.1080/21670811.2014.976407
- Anderson, C. W., Bell, E. & Shirky, C. (2012). *Post-industrial journalism: Adapting to the present*. New York, NY: Tow Center for Digital Journalism. Abgerufen von <http://towcenter.org/research/post-industrial-journalism-adapting-to-the-present-2/>
- Andrews, K., Traunmüller, T., Wolking, T., Goldgruber, E., Gutounig, R. & Ausserhofer, J. (2016). Styrian diversity visualisation: Visualising statistical open data with a leanweb app and data server. In T. Isenberg & F. Sadlo (Hrsg.), *EuroVis 2016 - Posters*. Groningen: The Eurographics Association. doi:10.2312/eurp.20161150
- Andrews, K., Traunmüller, T., Wolking, T., Gutounig, R. & Ausserhofer, J. (2015). Building an open data visualisation web app using a data server: The Styrian diversity visualisation project. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business* (S. 48:1–48:4). New York, NY: ACM. doi:10.1145/2809563.2809596
- Appelgren, E. (2017). An illusion of interactivity: The paternalistic side of data journalism. *Journalism Practice*. doi:10.1080/17512786.2017.1299032
- Appelgren, E. & Nygren, G. (2014a). Data journalism in Sweden – opportunities and challenges: A case study of Brottspejl at Sveriges Television (SVT). In C. Mills, M. Pidd & E. Ward (Hrsg.), *Proceedings of the Digital Humanities Congress 2012*. Sheffield: HRI Online Publications. Abgerufen von <https://www.hrionline.ac.uk/openbook/chapter/dhc2012-appelgren>
- Appelgren, E. & Nygren, G. (2014b). Data journalism in Sweden: Introducing new methods and genres of journalism into „old“ organizations. *Digital Journalism*, 2(3), 394–405. doi:10.1080/21670811.2014.884344

- Ausserhofer, J. (2008). *Die Vielfalt des Hybriden: Die Akteur-Netzwerk-Theorie Bruno Latours und ihr Beschreibungspotenzial für webbasierte soziale Netzwerke* (Diplomarbeit). FH Joanneum, Graz. doi:10.5281/zenodo.569907
- Ausserhofer, J. (2012). Daten, Daten, Daten! In H. M. Fischer, C. Aichhorn & T. Wolkingner (Hrsg.), *Das Beste aus zehn Jahren Journalismus und Public Relations (PR)* (S. 118–122). Graz: FH Joanneum. doi:10.5281/zenodo.569906
- Ausserhofer, J. (2013). Gute Filter, böse Filter? Algorithmen, der persönliche Informationshaushalt und digitale Öffentlichkeiten. In C. Landler, P. Parycek & M. C. Kettemann (Hrsg.), *Netpolitik in Österreich: Internet, Macht, Menschenrechte* (S. 83–90). Krems: Edition Donau-Universität Krems. doi:10.5281/zenodo.569913
- Ausserhofer, J. (2015). „Die Methode liegt im Code“: Routinen und digitale Methoden im Datenjournalismus. In A. Maireder, J. Ausserhofer, C. Schumann & M. Taddicken (Hrsg.), *Digitale Methoden in der Kommunikationswissenschaft* (S. 87–111). Berlin: Digital Communication Research. doi:10.17174/dcr.v2.5
- Ausserhofer, J. (in Druck). Die Datenbank verdient die Hauptrolle: Bausteine einer Methodologie für Open Digital Humanities. In S. Eichhorn, B. Oberreither, M. Rauchenbacher, I. Schwentner & K. Serles (Hrsg.), *Aufgehoben? Speicherorte, -diskurse und -medien von Literatur*. Würzburg: Königshausen & Neumann.
- Ausserhofer, J. & Maireder, A. (2013). National politics on Twitter: Structures and topics of a networked public sphere. *Information, Communication & Society*, 16(3), 291–314. doi:10.1080/1369118X.2012.756050
- Ausserhofer, J., Goldgruber, E., Gutounig, R. & Matiasek, S. (2017). *Status quo and challenges of data journalism: A literature review on specific domain problems* (Research Project Deliverable Nr. D1.1). Visual Analytics in Data-driven Journalism.
- Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S. & Goldgruber, E. (2017). The datafication of data journalism scholarship: Focal points, methods and research propositions for the investigation of data-intensive newswork. *Journalism*. doi:10.1177/1464884917700667
- Baack, S. (2013). A new style of news reporting: Wikileaks and data-driven journalism. In B. Rambatan & J. Johanssen (Hrsg.), *Cyborg Subjects: Discourses on Digital Culture* (S. 113–122). CreateSpace Independent Publishing.

- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2), 2053951715594634. doi:10.1177/2053951715594634
- Baacke, D. (1973). *Kommunikation und Kompetenz: Grundlegung einer Didaktik der Kommunikation und ihrer Medien*. München: Juventa.
- Baacke, D. (1996). Medienkompetenz: Begrifflichkeit und sozialer Wandel. In A. von Rein (Hrsg.), *Medienkompetenz als Schlüsselbegriff* (S. 112–124). Bad Heilbrunn: Klinkhardt.
- Backholm, K., Ausserhofer, J., Frey, E., Larsen, A. G., Hornmoen, H., Högvåg, J. & Reimerth, G. (in Druck). Crises, rumors and reposts: Journalists' social media content gathering and verification practices in breaking news situations. *Media and Communication*.
- Bakker, P. (2014). Mr. Gates returns: Curation, community management and other new roles for journalists. *Journalism Studies*, 15(5), 596–606. doi:10.1080/1461670X.2014.901783
- Barber, M. (2014). Alfred Schutz. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Spring 2014). Abgerufen von <http://plato.stanford.edu/archives/spr2014/entries/schutz/>
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A. & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531–1542. doi:10.1177/0956797615594620
- Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology*, 59(9), 1493–1512. doi:10.1002/asi.20857
- Bastos, M. T. & Mercea, D. (2016). Serial activists: Political Twitter beyond influentials and the twitterariat. *New Media & Society*, 18(10), 2359–2378. doi:10.1177/1461444815584764
- Batsell, J. (2014, Juli 16). How data journalism can become a revenue supplement for news organizations. Abgerufen von <http://mediashift.org/2014/07/how-data-journalism-can-become-a-revenue-supplement-for-news-organizations/>

- Bauer, T. A. (1979). *Medienpädagogik: Einführung und Grundlegung* (Bd. 1. Theorie – Diskussion: Der Kommunikationssinn). Wien: Böhlau.
- Bauer, T. A. (2005). Making news – faking views? Or: Faking news – making views? News agencies' self-understanding under pressure of change due to radical transitions of systems in media, politics and communities. *deScripto - A Journal of Media in South East Europe*, 4, 10–13. Abgerufen von http://www.univie.ac.at/seemo/deScripto_No4.pdf
- Bauer, T. A. (2008a). Bildung als soziale Praxis: Zum kulturellen Wandel vom Wissen und Bildung im Kontext zunehmend medialisierter Gesellschaften. In T. A. Bauer & G. E. Ortner (Hrsg.), *Bildung für Europa: Politische Ansprüche und Anregungen für die Praxis* (S. 14–35). Paderborn: B+B Medien.
- Bauer, T. A. (2008b). Kultur der Medialität: Medienbildung als das pädagogische Programm von Medialitätskultur. *Medienimpulse*, 65(September), 9–12. Abgerufen von https://www.mediamanual.at/mediamanual/mm2/themen/pdf/MI65_bauer.pdf
- Bauer, T. A. (2014a). Bildungsmedien und Medienbildung im Kontext des Medienwandels. In T. A. Bauer, M. Ivanišin & B. Mikuszeit (Hrsg.), *Medien für die europäische Bildungsgesellschaft: Medienbildung – Medienbewertung – Mediennutzung* (S. 13–46). Frankfurt am Main: Peter Lang.
- Bauer, T. A. (2014b). *Kommunikation wissenschaftlich denken: Perspektiven einer kontextuellen Theorie gesellschaftlicher Verständigung*. Wien: Böhlau.
- BBC. (2017). Global experience language. Abgerufen von <http://www.bbc.co.uk/gel>
- Becker, H. S. (1996). The epistemology of qualitative research. In R. Jessor, A. Colby & R. Shweder (Hrsg.), *Ethnography and human development: Context and meaning in social inquiry* (S. 53–71). Chicago, IL: University of Chicago Press.
- Bell, E. & Owen, T. (2017). *The Platform Press: How Silicon Valley reengineered journalism*. New York, NY: Tow Center for Digital Journalism. Abgerufen von http://towcenter.org/wp-content/uploads/2017/03/The_Platform_Press_Tow_Report_2017.pdf
- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

- Berger, P. L. (1967). *The sacred canopy: Elements of a sociological theory of religion*. Garden City, NY: Doubleday.
- Berger, P. L. & Luckmann, T. (2009). *Die gesellschaftliche Konstruktion der Wirklichkeit: Eine Theorie der Wissenssoziologie*. (M. Plessner, Übers.) (22. Aufl.). Frankfurt am Main: Fischer. (Original erschienen 1966)
- Berlin declaration on open access to knowledge in the sciences and humanities. (2003). Max-Planck-Gesellschaft. Abgerufen von <https://openaccess.mpg.de/Berlin-Declaration>
- Berry, D. M. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, 12, 1–22. Abgerufen von <http://www.culturemachine.net/index.php/cm/article/view/440/470>
- Biermann, K., Blickle, P. & Venohr, S. (2015, Februar 5). Pegida-Anhänger auf Facebook: Wer sind sie? *Die Zeit*. Abgerufen von <http://www.zeit.de/gesellschaft/zeitgeschehen/2015-02/wer-ist-pegida-facebook-daten/>
- Bloor, D. (1991). *Knowledge and social imagery* (2. Aufl.). Chicago, IL: University of Chicago Press. (Original erschienen 1976)
- Bodle, R. (2011). Regimes of sharing: Open APIs, interoperability, and Facebook. *Information, Communication & Society*, 14(3), 320–337.
doi:10.1080/1369118X.2010.542825
- Bogner, A., Littig, B. & Menz, W. (2014). *Interviews mit Experten: Eine praxisorientierte Einführung*. Wiesbaden: Springer. doi:10.1007/978-3-531-19416-5
- Bonß, W. (1982). *Die Einübung des Tatsachenblicks: Zur Struktur und Veränderung empirischer Sozialforschung*. Frankfurt am Main: Suhrkamp.
- Borger, M., van Hoof, A., Costera Meijer, I. & Sanders, J. (2013). Constructing participatory journalism as a scholarly object. *Digital Journalism*, 1(1), 117–134.
doi:10.1080/21670811.2012.740267
- Borges, J. L. (1966). Die analytische Sprache John Wilkins'. In K. A. Horst (Übers.), *Das Eine und die Vielen: Essays zur Literatur* (S. 209–214). München: Carl Hanser.

- Borges-Rey, E. (2016). Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice*, 10(7), 833–843.
doi:10.1080/17512786.2016.1159921
- Borges-Rey, E. (2017). Towards an epistemology of data journalism in the devolved nations of the UK: Changes and continuities in materiality, performativity and reflexivity. *Journalism*. doi:10.1177/1464884917693864
- Bounegru, L. (2012). Data journalism in perspective. In J. Gray, L. Bounegru & L. Chambers (Hrsg.), *The data journalism handbook: How journalists can use data to improve the news* (S. 17–22). Sebastopol, CA: O'Reilly Media. Abgerufen von http://datajournalismhandbook.org/1.0/en/introduction_4.html
- Bowker, G. C. & Star, S. L. (Hrsg.). (1999a). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.
- Bowker, G. C. & Star, S. L. (1999b). Introduction: To classify is human. In G. C. Bowker & S. L. Star (Hrsg.), *Sorting things out: Classification and its consequences* (S. 1–32). Cambridge, MA: MIT Press.
- boyd, danah & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.678878
- Boyer, B., Keefe, J., Lindenberg, F., Park, J. & Wu, C. (2012). A five minute field guide. In J. Gray, L. Bounegru & L. Chambers (Hrsg.), *The data journalism handbook: How journalists can use data to improve the news*. Sebastopol, CA: O'Reilly Media. Abgerufen von http://datajournalismhandbook.org/1.0/en/getting_data_0.html
- Boyles, J. L. & Meyer, E. (2016). Letting the data speak: Role perceptions of data journalists in fostering democratic conversation. *Digital Journalism*, 4(7), 944–954.
doi:10.1080/21670811.2016.1166063
- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3), 209–227. doi:10.1007/s10676-013-9321-6
- Bozeman, B. & Boardman, C. (2014). *Research collaboration and team science: A state-of-the-art review and agenda*. Cham: Springer.
- Bradley, J.-C., Owens, K. & Williams, A. (2008). Chemistry crowdsourcing and open notebook science. *Nature Precedings*. doi:10.1038/npre.2008.1505.1

- Bradshaw, P. (2011, Juli 7). The inverted pyramid of data journalism. Abgerufen von <http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>
- Bradshaw, P. (2012). What is data journalism? In J. Gray, L. Bounegru & L. Chambers (Hrsg.), *The data journalism handbook: How journalists can use data to improve the news*. Sebastopol, CA: O'Reilly Media. Abgerufen von http://datajournalismhandbook.org/1.0/en/introduction_0.html
- Bradshaw, P. (2013). *Scraping for Journalists: How to grab data from hundreds of sources, put it in a form you can interrogate – and still hit deadlines*. Leanpub. Abgerufen von <https://leanpub.com/scrapingforjournalists>
- Brehmer, M., Ingram, S., Stray, J. & Munzner, T. (2014). Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 2271–2280. doi:10.1109/TVCG.2014.2346431
- Brives, C. & Latour, B. (2007). Wissenschaft durch den Gefrierschrank betrachtet. In A. te Heesen & A. Michels (Hrsg.), H. Wittenbrink (Übers.), *auf\ zu. Der Schrank in den Wissenschaften*. (S. 74–79). Berlin.
- Broderick, R. (2017, Jänner 24). Trump supporters online are pretending to be French to manipulate France's election. *BuzzFeed*. Abgerufen von <https://www.buzzfeed.com/ryanhatesthis/inside-the-private-chat-rooms-trump-supporters-are-using-to>
- Bruns, A. (2005). *Gatewatching: Collaborative online news production*. New York, NY: Peter Lang.
- Bryman, A. (2004). *Quantity and quality in social research*. London: Routledge. (Original erschienen 1988)
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97–113. doi:10.1177/1468794106058877
- Bucher, H.-J. (2003). Journalistische Qualität und Theorien des Journalismus. In K.-D. Altmeppen & H.-J. Bucher (Hrsg.), *Qualität im Journalismus: Grundlagen, Dimensionen, Praxismodelle* (S. 11–34). Wiesbaden: Westdeutscher Verlag.

- Bucher, T. (2013). Objects of intense feeling: The case of the Twitter API. *Computational Culture*, (3). Abgerufen von <http://computationalculture.net/article/objects-of-intense-feeling-the-case-of-the-twitter-api>
- Bucy, E. P. (2004). Interactivity in society: Locating an elusive concept. *The Information Society*, 20(5), 373–383. doi:10.1080/01972240490508063
- Budapest Open Access Initiative. (2012, September 12). Abgerufen von <http://budapestopenaccessinitiative.org/>
- Bunz, M. (2011). Das offene Geheimnis: Zur Politik der Wahrheit im Datenjournalismus. In H. Geiselberger (Hrsg.), *Wikileaks und die Folgen: Netz – Medien – Politik* (S. 134–151). Berlin: Suhrkamp.
- Burrows, R. & Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, 1(1). doi:10.1177/2053951714540280
- Callon, M. (1984). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Briec Bay. *The Sociological Review*, 32, 196–233. doi:10.1111/j.1467-954X.1984.tb00113.x
- Calzada Prado, J. & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, 63(2), 123–134. doi:10.1515/libri-2013-0010
- Carey, J. W. (2009). *Communication as culture: Essays on media and society*. New York, NY: Routledge.
- Carney, R. N. & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, 14(1), 5–26. doi:1040-726X/02/0300-0005/0
- Carter, S., Cox, A. & Quealy, K. (2009, November 6). The jobless rate for people like you. *New York Times*. Abgerufen von <http://www.nytimes.com/interactive/2009/11/06/business/economy/unemployment-lines.html>
- Carter, S., Cox, A., Quealy, K. & Schoenfeld, A. (2009, Juli 31). How different groups spend their day. *New York Times*. Abgerufen von

<http://www.nytimes.com/interactive/2009/07/31/business/20080801-metrics-graphic.html>

- Castells, M. (2001). *The Internet galaxy: Reflections on the Internet, business, and society*. Oxford: Oxford University Press.
- Castells, M. (2010). *The information age: Economy, society, and culture* (2. Aufl., Bd. 1: The rise of the network society). Chichester: Wiley-Blackwell.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage.
- Cherubini, F. & Nielsen, R. K. (2016). *Editorial analytics: How news media are developing and using audience data and metrics* (Digital news project report). Oxford: Reuters Institute for the Study of Journalism. doi:/10.2139/ssrn.2739328
- Chung, A.-M., O’Byrne, I. & Bond Gill, I. (2016). Web Literacy 2.0. Abgerufen von <http://mozilla.github.io/content/web-lit-whitepaper/>
- Coddington, M. (2015). Clarifying journalism’s quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), 331–348. doi:10.1080/21670811.2014.976400
- Cohen, N. S. (2015). From pink slips to pink slime: Transforming media labor in a digital age. *The Communication Review*, 18(2), 98–122. doi:10.1080/10714421.2015.1031996
- Cohen, R. S. & Schnelle, T. (1986). Introduction. In R. S. Cohen & T. Schnelle (Hrsg.), *Cognition and fact: Materials on Ludwik Fleck* (S. ix–xxxiii). Dordrecht: Reidel.
- Cohen, S., Hamilton, J. T. & Turner, F. (2011a). Computational journalism. *Communications of the ACM*, 54(10), 66–71. doi:10.1145/2001269.2001288
- Cohen, S., Li, C., Yang, J. & Yu, C. (2011). Computational journalism: A call to arms to database researchers (S. 148–151). Gehalten auf der 5th Biennial Conference on Innovative Data Systems Research (CIDR 2011), Asilomar, CA. Abgerufen von http://cidrdb.org/cidr2011/Papers/CIDR11_Paper17.pdf
- Coleman, E. G. (2013). *Coding freedom: The ethics and aesthetics of hacking*. Princeton, NJ: Princeton University Press.

- Collin, F. (2011). *Science studies as naturalized philosophy*. Dordrecht: Springer.
doi:10.1007/978-90-481-9741-5
- Collins, H. M. (1981). Stages in the empirical programme of relativism. *Social Studies of Science*, 11(1), 3–10. doi:10.1177/030631278101100101
- Collins, H. M. & Pinch, T. J. (1998). *The golem: What you should know about science* (2. Aufl.). Cambridge: Cambridge University Press.
- Cottle, S. (2007). Ethnography and news production: New(s) developments in the field. *Sociology Compass*, 1(1), 1–16. doi:10.1111/j.1751-9020.2007.00002.x
- Cox, M. (2000). The development of computer-assisted reporting. Gehalten auf der Conference of the Newspaper Division, Association for Education in Journalism and Mass Communication (AEJMC), Chapel Hill, NC. Abgerufen von <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.631.6220&rep=rep1&type=pdf>
- Creech, B. & Mendelson, A. L. (2015). Imagining the journalist of the future: Technological visions of journalism education and newswork. *The Communication Review*, 18(2), 142–165. doi:10.1080/10714421.2015.1031998
- Dallas Morning News (2005, Oktober). Graphics Stylebook. Abgerufen von https://knightcenter.utexas.edu/mooc/file/tdmn_graphics.pdf
- Dander, V. (2014). Von der „Macht der Daten“ zur „Gemachtheit von Daten“. Praktische Datenkritik als Gegenstand der Medienpädagogik. *Mediale Kontrolle unter Beobachtung*, 3(1). Abgerufen von <http://www.medialekontrolle.de/wp-content/uploads/2014/09/Dander-Valentin-2014-03-01.pdf>
- Dant, A. & Richards, J. (2011, Dezember 8). Behind the rumours: How we built our Twitter riots interactive. *The Guardian*. Abgerufen von <http://www.guardian.co.uk/news/datablog/2011/dec/08/twitter-riots-interactive>
- Data Journalism Awards. (2017). Categories. Abgerufen von <https://www.datajournalismawards.org/categories/>
- Davenport, L. D., Fico, F. & DeFleur, M. H. (2002). Computer-assisted reporting in classrooms: A decade of diffusion and a comparison to newsrooms. *Journalism & Mass Communication Educator*, 57(1), 6. doi:10.1177/107769580205700103

- Davenport, L. D., Fico, F. & Detwiler, M. (2000). Computer-assisted reporting in Michigan daily newspapers: More than a decade of adoption. gehalten auf der Association for Education in Journalism and Mass Communication (AEJMC) National Convention, Phoenix, AZ. Abgerufen von [http://online.sfsu.edu/jjohnson/Courses&Syllabi/BU-JO807/Bibli&Articles/Davenport\(2000\).pdf](http://online.sfsu.edu/jjohnson/Courses&Syllabi/BU-JO807/Bibli&Articles/Davenport(2000).pdf)
- Davenport, L. D., Fico, F. & Weinstock, D. (1996). Computers in newsrooms of Michigan's newspapers. *Newspaper Research Journal*, 17(3–4), 14–28.
- Davenport, T. H. (1997). *Information ecology: Mastering the information and knowledge environment*. New York, NY: Oxford University Press.
- De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F. & Le Cam, F. (2015). Waiting for data journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism*, 3(3), 432–446.
doi:10.1080/21670811.2014.976415
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2. Aufl.). New York: McGraw-Hill.
- Deuze, M. (2005). What is journalism? Professional identity and ideology of journalists reconsidered. *Journalism*, 6(4), 442–464. doi:10.1177/1464884905056815
- Diakopoulos, N. (2011, April). A functional roadmap for innovation in computational journalism. Abgerufen von <http://www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/>
- Diakopoulos, N. (2012). *Cultivating the landscape of innovation in computational journalism*. CUNY Graduate School of Journalism & Tow-Knight Center for Entrepreneurial Journalism. Abgerufen von http://www.nickdiakopoulos.com/wp-content/uploads/2012/05/diakopoulos_whitepaper_systematicinnovation.pdf
- Diakopoulos, N. (2014). *Algorithmic accountability reporting: On the investigation of black boxes* (A Tow/Knight Brief). New York, NY: Tow Center for Digital Journalism. Abgerufen von http://towcenter.org/wp-content/uploads/2014/02/78524_Tow-Center-Report-WEB-1.pdf
- Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398–415.
doi:10.1080/21670811.2014.976411

- Diakopoulos, N., Naaman, M. & Kivran-Swaine, F. (2010). Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *2010 IEEE Symposium on Visual Analytics Science and Technology* (S. 115–122). doi:10.1109/VAST.2010.5652922
- Dick, M. (2014). Interactive infographics and news values. *Digital Journalism*, 2(4), 490–506. doi:10.1080/21670811.2013.841368
- Domingo, D. & Paterson, C. (Hrsg.). (2011). *Making online news volume 2: Newsroom ethnographies in the second decade of internet journalism*. New York, NY: Peter Lang.
- Domingo, D., Quandt, T., Heinonen, A., Paulussen, S., Singer, J. B. & Vujnovic, M. (2008). Participatory journalism practices in the media and beyond. *Journalism Practice*, 2(3), 326–342. doi:10.1080/17512780802281065
- Dörr, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, 4(6), 700–722. doi:10.1080/21670811.2015.1096748
- Dresing, T. & Pehl, T. (2015). *Praxisbuch Interview, Transkription & Analyse: Anleitungen und Regelsysteme für qualitativ Forschende* (6. Aufl.). Marburg. Abgerufen von <http://www.audiotranskription.de/praxisbuch>
- Drucker, P. F. (1988). The coming of the new organization. *Harvard Business Review*, 66, 45–53. Abgerufen von <https://hbr.org/1988/01/the-coming-of-the-new-organization>
- Eberle, T. S. (1992). A new paradigm for the sociology of knowledge: The „social construction of reality“ after 25 years. *Schweizerische Zeitschrift für Soziologie*, 18(2), 493–502. Abgerufen von <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-24182>
- Eisenstein, E. L. (2005). *The printing press as an agent of change: Communications and cultural transformations in early-modern Europe*. New York, NY: Cambridge University Press. (Original erschienen 1979)
- Eliasoph, N. (1988). Routines and the making of oppositional news. *Critical Studies in Mass Communication*, 5(4), 313–334. doi:10.1080/15295038809366719
- Emerson, T. I. (1976). Legal foundations of the right to know. *Washington University Law Review*, 1976(1), 1–24. Abgerufen von http://openscholarship.wustl.edu/law_lawreview/vol1976/iss1/6

- Engesser, S. (2013). *Die Qualität des Partizipativen Journalismus im Web: Bausteine für ein integratives theoretisches Konzept und eine explanative empirische Analyse*. Wiesbaden: Springer. doi:10.1007/978-3-658-00584-9
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. doi:10.1111/j.1460-2466.1993.tb01304.x
- Esser, H. (1991). Die Rationalität des Alltagshandelns: Eine Rekonstruktion der Handlungstheorie von Alfred Schütz. *Zeitschrift für Soziologie*, 20(6), 430–445. doi:10.1515/zfsoz-1991-0602
- Europäische Kommission. (2016). *Media Pluralism and Democracy* (Report Nr. 452). doi:10.2838/248670
- Fairfield, J. & Shtein, H. (2014). Big data, big problems: Emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics*, 29(1), 38–51. doi:10.1080/08900523.2014.863126
- Fecher, B., Friesike, S. & Hebing, M. (2015). What drives academic data sharing? *PLoS ONE*, 10(2), e0118053. doi:10.1371/journal.pone.0118053
- Feigenbaum, A., Thorsen, E., Weissmann, D. & Demirkol, O. (2016). Visualising data stories together: Reflections on data journalism education from the Bournemouth University Datalabs Project. *Journalism Education*, 5(3), 59–74. Abgerufen von <http://journalism-education.org/2016/10/visualising-data-stories-together-reflections-on-data-journalism-education/>
- Felle, T. (2016). Digital watchdogs? Data reporting and the news media's traditional „fourth estate“ function. *Journalism*, 17(1), 85–96. doi:10.1177/1464884915593246
- Fengler, S. (2016). Journalismus als rationales Handeln. In M. Löffelholz & L. Rothenberger (Hrsg.), *Handbuch Journalismustheorien* (S. 235–248). Wiesbaden: Springer. doi:10.1007/978-3-531-18966-6_13
- Fengler, S. & Russ-Mohl, S. (2005). *Der Journalist als „Homo oeconomicus“*. Konstanz: UVK.
- Fielding, N. G. & Fielding, J. L. (1986). *Linking Data*. Newbury Park, CA: Sage.
- Fink, K. & Anderson, C. (2015). Data journalism in the United States: Beyond the „usual suspects“. *Journalism Studies*, 16(4), 467–481. doi:10.1080/1461670X.2014.939852

- Flaxman, S., Goel, S. & Rao, J. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298–320. doi:10.1093/poq/nfw006
- Fleck, L. (1980). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*. Frankfurt am Main: Suhrkamp. (Original erschienen 1935)
- Fleck, L. (1986). Scientific observation and perception in general. In R. S. Cohen & T. Schnelle (Hrsg.), *Cognition and fact: Materials on Ludwik Fleck* (S. 59–78). Dordrecht: Reidel. (Original erschienen 1935)
- Flew, T., Daniel, A. & Spurgeon, C. L. (2010). The Promise of Computational Journalism. In K. McCallum (Hrsg.), *Media, Democracy and Change: Refereed Proceedings of the Australian and New Zealand Communication Association Annual Conference*. Canberra.
- Flew, T., Spurgeon, C., Daniel, A. & Swift, A. (2012). The promise of computational journalism. *Journalism Practice*, 6(2), 157–171. doi:10.1080/17512786.2011.616655
- Flick, U. (2007). *Designing qualitative research*. London: Sage.
- Flick, U. (2009). *An introduction to qualitative research* (4. Aufl.). London: Sage.
- Flick, U. (2011). *Triangulation: Eine Einführung* (3., aktualisierte Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Flick, U., Kardorff, E. von & Steinke, I. (Hrsg.). (2004). *A companion to qualitative research*. (B. Jenner, Übers.). London: Sage.
- Fodor, J. A. (1984). *The modularity of mind: An essay on faculty psychology* (3. Aufl.). Cambridge, MA: MIT Press.
- Foucault, M. (2003). *Die Ordnung der Dinge: Eine Archäologie der Humanwissenschaften*. (U. Köppen, Übers.). Frankfurt am Main: Suhrkamp. (Original erschienen 1966)
- Franck, G. (1998). *Ökonomie der Aufmerksamkeit: Ein Entwurf*. München: Hanser.
- Friesinger, G. & Herwig, J. (Hrsg.). (2014). *The art of reverse engineering: Open— dissect — rebuild*. Bielefeld: Transcript.
- Fry, J., Lockyer, S., Oppenheim, C., Houghton, J. & Rasmussen, B. (2008). *Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes*. Loughborough University: Centre for

- Strategic Economic Studies. Abgerufen von
<https://dspace.lboro.ac.uk/dspace/handle/2134/4600>
- Fuchs, C. (2014). Social media and the public sphere. *tripleC: Communication, Capitalism & Critique*, 12(1), 57–101. Abgerufen von <http://www.triple-c.at/index.php/tripleC/article/view/552>
- Funck, A. (2016). *Die transparente Redaktion: Ein Ansatz für die strategische Öffentlichkeitsarbeit von Zeitungsverlagen*. Wiesbaden: Springer. doi:10.1007/978-3-658-11156-4
- Galison, P. L. (1997). *Image and logic: A material culture of microphysics*. Chicago, IL: University of Chicago Press.
- Gans, H. J. (2004). *Deciding what's news: A study of CBS Evening News, NBC Nightly News, Newsweek, and Time*. Evanston, IL: Northwestern University Press. (Original erschienen 1979)
- Garrison, B. (1996, September). *Computer-assisted reporting: A national use update*. Gehalten bei der National Society of Professional Journalists, Washington, D.C. Abgerufen von <https://web.archive.org/web/20130914034958/http://com.miami.edu/car/spj-96/index.htm>
- Garrison, B. (1999). Newspaper size as a factor in use of computer-assisted reporting. *Newspaper Research Journal*, 20(3), 72–85. doi:10.1177/073953299902000306
- Gartner, G. (2014). *Finanzierung von Datenjournalismus: Erlösmodelle und Potenziale für datenbasierte Inhalte* (Bachelorarbeit). Fachhochschule Wien der WKW, Wien. Abgerufen von http://geraldgartner.at/wp-content/uploads/2014/06/BA2-2014_Gartner-Kopie.pdf
- Gartner, G., Maan, N., Schmid, F. & Hametner, M. (2016, November 28). Präsidentschaftskandidaten: Was von den Facebook-Seiten verschwindet. *derStandard.at*. Abgerufen von <http://derstandard.at/2000048203012/Praesidentschaftskandidaten-Was-von-den-Facebook-Seiten-verschwindet>

- Geiger, R. S. (2009). Does Habermas understand the Internet? The algorithmic construction of the blogo/public sphere. *Gnovis. A Journal of Communication, Culture, and Technology*, 10(1), 1–29. Abgerufen von <https://ssrn.com/abstract=2734947>
- Gentzkow, M. & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4), 1799–1839. doi:10.1093/qje/qjr044
- Gerhards, J. (1994). Politische Öffentlichkeit. Ein system- und akteurstheoretischer Bestimmungsversuch. In F. Neidhardt (Hrsg.), *Öffentlichkeit, öffentliche Meinung, soziale Bewegungen* (S. 77–105). Opladen: Westdeutscher Verlag.
- Gerlitz, C. & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society*, 15(8), 1348–1365. doi:10.1177/1461444812472322
- Gesellschaft für deutsche Sprache. (2016, Dezember 9). GfdS wählt „postfaktisch“ zum Wort des Jahres 2016. Abgerufen von <http://gfds.de/wort-des-jahres-2016/>
- Gillespie, T. (2010). The politics of ‘platforms’. *New Media & Society*, 12(3), 347–364. doi:10.1177/1461444809342738
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, 12(4), 436–445. doi:10.2307/798843
- Glaser, B. G. & Strauss, A. L. (1967). *The discovery of Grounded Theory: Strategies for qualitative research*. New Brunswick: Aldine Transaction.
- Graefe, A. (2016). *Guide to automated journalism*. New York, NY: Tow Center for Digital Journalism. Abgerufen von <https://www.gitbook.com/book/towcenter/guide-to-automated-journalism/details>
- Gray, J., Bounegru, L. & Chambers, L. (Hrsg.). (2012). *The data journalism handbook: How journalists can use data to improve the news*. Sebastopol, CA: O’Reilly. Abgerufen von <http://datajournalismhandbook.org/>
- Greimas, A. J. (1971). *Strukturelle Semantik: Methodologische Untersuchungen*. (J. Ihwe, Übers.). Braunschweig: Friedr. Vieweg + Sohn. doi:10.1007/978-3-663-06836-5
- Guardian Interactive Team, Procter, R., Vis, F. & Voss, A. (2011, Dezember 7). How riot rumours spread on Twitter. *The Guardian: UK News*. Abgerufen von <http://www.theguardian.com/uk/interactive/2011/dec/07/london-riots-twitter>

- Guthrie, G. & Murthy, V. (2009). Past, present and possible future developments in human capital accounting: A tribute to Jan-Erik Gröjer. *Journal of Human Resource Costing & Accounting*, 13(2), 125–142. doi:10.1108/14013380910968647
- Gutounig, R. (2015). *Wissen in digitalen Netzwerken: Potenziale Neuer Medien für Wissensprozesse*. Wiesbaden: Springer. doi:10.1007/978-3-658-02110-8
- Gutounig, R., Goldgruber, E., Ausserhofer, J., Andrews, K., Traunmüller, T. & Wolkingner, T. (2016). The Styrian diversity visualisation project: Communicating data stories with an open data visualisation web app. In *Tagungsband des 10. Forschungsforum der Österreichischen Fachhochschulen*. Wien. Abgerufen von <http://ffhoarep.fh-ooe.at/handle/123456789/540>
- Gynnild, A. (2014). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism*, 15(6), 713–730. doi:10.1177/1464884913486393
- Haas, H. (1999). *Empirischer Journalismus: Verfahren zur Erkundung gesellschaftlicher Wirklichkeit*. Wien: Böhlau.
- Habermas, J. (1990). *Strukturwandel der Öffentlichkeit: Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft*. Frankfurt am Main: Suhrkamp. (Original erschienen 1962)
- Habermas, J. (2006). Political communication in media society: Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. *Communication Theory*, 16(4), 411–426. doi:10.1111/j.1468-2885.2006.00280.x
- Hack. Def. 4b. (2016). In *Oxford English Dictionary* (3. Aufl.). Abgerufen von <http://www.oed.com/view/Entry/83027#eid1212014640>
- Hamilton, J. T. & Turner, F. (2009, Juli). *Accountability through algorithm: Developing the field of computational journalism*. Report vom Center For Advanced Study in the Behavioral Sciences Summer Workshop. Abgerufen von <http://dewitt.sanford.duke.edu/wp-content/uploads/2011/12/About-3-Research-B-cj-1-finalreport.pdf>
- Hannaford, L. (2015). Computational journalism in the UK newsroom: Hybrids or specialists? *Journalism Education*, 4(1), 6–21. Abgerufen von <http://hdl.handle.net/2173/575953>

- Hanson, N. R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge: Cambridge University Press.
- Heise, C. (2016). *Von Open Access zu Open Science: Zum Wandel digitaler Kulturen der wissenschaftlichen Kommunikation* (Dissertation). Leuphana Universität, Lüneburg. doi:/10.5281/zenodo.31247
- Heise, N. (2015). Big Data – small problems? Ethische Dimensionen der Forschung mit Online-Kommunikationsspuren. In A. Maireder, J. Ausserhofer, C. Schumann & M. Taddicken (Hrsg.), *Digitale Methoden in der Kommunikationswissenschaft* (S. 39–58). Berlin: Digital Communication Research. doi:10.17174/dcr.v2.3
- Hermida, A., Lewis, S. C. & Zamith, R. (2014). Sourcing the Arab Spring: A case study of Andy Carvin’s sources on Twitter during the Tunisian and Egyptian revolutions. *Journal of Computer-Mediated Communication*, 19(3), 479–499. doi:10.1111/jcc4.12074
- Heuguet, G. & Langlais, P.-C. (2014). Un manuel de journalisme au service des „invisibles“? Le cas du Data Journalism Handbook. *Sur le journalisme, About journalism, Sobre jornalismo*, 3(1), 100–113. Abgerufen von <http://surlejournalisme.com/rev/index.php/slj/article/view/133>
- Hewett, J. (2013). Learning in progress: From computer-assisted reporting to data journalism, via freedom of information, open data and more. In J. Mair & R. L. Keeble (Hrsg.), *Data Journalism: Mapping the Future* (S. 3–14). Bury St Edmunds: Abramis Academic Publishing.
- Hewett, J. (2016). Learning to teach data journalism: Innovation, influence and constraints. *Journalism*, 17(1), 119–137. doi:10.1177/1464884915612681
- Hirst, T. (2009, März 10). Using Many Eyes Wikified to visualise Guardian Data Store data on Google Docs. Abgerufen von <https://blog.ouseful.info/2009/03/10/using-many-eyes-wikified-to-visualise-guardian-data-store-data-on-google-docs/>
- Hirst, T. (2014, April 29). Personal recollections of the „data journalism“ phrase. Abgerufen von <https://blog.ouseful.info/2014/04/29/personal-recollections-of-the-data-journalism-phrase/>

- Hofmann, P. & Hirschauer, S. (2012). Die konstruktivistische Wende. In S. Maasen, M. Kaiser, M. Reinhart & B. Sutter (Hrsg.), *Handbuch Wissenschaftssoziologie* (S. 85–99). Wiesbaden: Springer. doi:10.1007/978-3-531-18918-5_7
- Holovaty, A. (2006, September 6). A fundamental way newspaper sites need to change. Abgerufen von <http://www.holovaty.com/writing/fundamental-change/>
- Hornmoen, H. & Backholm, K. (Hrsg.). (2017). *Social media use in crisis and risk communication*. Bingley: Emerald. (Manuskript in Vorbereitung)
- Howard, A. B. (2014). *The art and science of data-driven journalism*. New York, NY: Tow Center for Digital Journalism. Abgerufen von <http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>
- Howard, P. N. (2002). Network ethnography and the hypermedia organization: New media, new organizations, new methods. *New Media & Society*, 4(4), 550–574. doi:10.1177/146144402321466813
- Hoyningen-Huene, P. (1989). *Die Wissenschaftsphilosophie Thomas S. Kuhns: Rekonstruktion und Grundlagenprobleme*. Braunschweig: Vieweg.
- Hoyningen-Huene, P. & Lohse, S. (2012). Die Kuhn'sche Wende. In S. Maasen, M. Kaiser, M. Reinhart & B. Sutter (Hrsg.), *Handbuch Wissenschaftssoziologie* (S. 73–84). Wiesbaden: Springer. doi:10.1007/978-3-531-18918-5
- Hullman, J., Diakopoulos, N., Momeni, E. & Adar, E. (2015). Content, context, and critique: Commenting on a data visualization blog. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (S. 1170–1175). New York, NY: ACM. doi:10.1145/2675133.2675207
- Husserl, E. (1976). *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie: Eine Einleitung in die phänomenologische Philosophie*. (W. Biemel, Hrsg.). Den Haag: Martinus Nijhoff. (Original erschienen 1936)
- Huter, M. (2015, Juli 17). Die neuen Regeln zum Weiterverwenden öffentlicher Dokumente. Abgerufen von <http://okfn.at/2015/07/17/die-neuen-regeln-zum-weiterverwenden-offentlicher-dokumente/>
- Irvine, A. (2011). Duration, dominance and depth in telephone and face-to-face interviews: A comparative exploration. *International Journal of Qualitative Methods*, 10(3), 202–220. doi:10.1177/160940691101000302

- Jackson, J. & Sweney, M. (2015, Mai 13). Trinity Mirror's UsVsTh3m and Amp3d thought to be facing axe as jobs set to go. *The Guardian*. Abgerufen von <https://www.theguardian.com/media/2015/may/13/trinity-mirrors-usvsth3m-and-amp3d-thought-to-be-facing-axe-as-jobs-set-to-go>
- Jahoda, M., Lazarsfeld, P. F. & Zeisel, H. (2014). *Die Arbeitslosen von Marienthal: Ein soziographischer Versuch über die Wirkungen langandauernder Arbeitslosigkeit* (24. Aufl.). Frankfurt am Main: Suhrkamp. (Original erschienen 1933)
- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258–268. doi:10.1080/10580530.2012.716740
- Jenkins, H., Purushotma, R., Weigel, M., Clinton, K. & Robison, A. J. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. Cambridge, MA: MIT Press.
- Joas, H. & Knöbl, W. (2004). *Sozialtheorie: Zwanzig einführende Vorlesungen*. Frankfurt am Main: Suhrkamp.
- Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1), 11–21. doi:10.1108/eb026526
- Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13(1), 72–91. doi:10.1080/19331681.2015.1132401
- Jürgens, P., Stark, B. & Magin, M. (2015). Messung von Personalisierung in computervermittelter Kommunikation. In A. Maireder, J. Ausserhofer, C. Schumann & M. Taddicken (Hrsg.), *Digitale Methoden in der Kommunikationswissenschaft* (S. 251–270). Berlin: Digital Communication Research. doi:10.17174/dcr.v2.11
- Kaefer, F., Roper, J. & Sinha, P. (2015). A software-assisted qualitative content analysis of news articles: Example and reflections. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 16(2). Abgerufen von <http://www.qualitative-research.net/index.php/fqs/article/view/2123>
- Karlsen, J. & Stavelin, E. (2014). Computational journalism in Norwegian newsrooms. *Journalism Practice*, 8(1), 34–48. doi:10.1080/17512786.2013.813190

- Katzenbach, C. (2016). *Governance – Technik – Kommunikation: Perspektiven einer kommunikationswissenschaftlichen Governance-Forschung* (Dissertation). Freie Universität Berlin. Abgerufen von http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000103153
- Kayser-Bril, N. (2013, Mai 20). How to make money with data journalism in the US and Europe. Abgerufen von http://datadrivenjournalism.net/news_and_analysis/How_to_Make_Money_with_Data_Journalism_in_the_US_and_Europe
- Kelle, U. (2008). *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung: Theoretische Grundlagen und methodologische Konzepte* (2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Keller, R., Reichertz, J. & Knoblauch, H. (Hrsg.). (2013). *Kommunikativer Konstruktivismus*. Wiesbaden: Springer. doi:10.1007/978-3-531-19797-5
- Kleining, G. (1982). Umriss zu einer Methodologie qualitativer Sozialforschung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 34(2), 224–253. Abgerufen von <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-8619>
- Knight, M. (2015). Data journalism in the UK: A preliminary analysis of form and content. *Journal of Media Practice*, 16(1), 55–72. doi:10.1080/14682753.2015.1015801
- Knoblauch, H. (1995). *Kommunikationskultur: Die kommunikative Konstruktion kultureller Kontexte*. Berlin: Walter de Gruyter.
- Knoblauch, H. (2013, August 16). Latours Popanz: Über Mißverständnisse des Sozialkonstruktivismus. Abgerufen von <http://soziologie.de/blog/2013/08/latours-popanz-ueber-misverstaendnisse-des-sozialkonstruktivismus/>
- Knoblauch, H. (2014). *Wissenssoziologie* (3., überarbeitete Aufl.). Konstanz: UVK.
- Knoblauch, H. & Schnettler, B. (2007). Konstruktivismus. In R. Buber & H. H. Holzmüller (Hrsg.), *Qualitative Marktforschung: Konzepte – Methoden – Analysen* (S. 127–137). Wiesbaden: Gabler.

- Knorr-Cetina, K. (1977). Producing and reproducing knowledge: Descriptive or constructive? Toward a model of research production. *Social Science Information*, 16(6), 669–696. Abgerufen von <http://ssi.sagepub.com/content/16/6/669.short>
- Knorr-Cetina, K. (1981). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford: Pergamon Press.
- Knorr-Cetina, K. (1989). Spielarten des Konstruktivismus: Einige Notizen und Anmerkungen. *Soziale Welt*, 40(1/2), 86–96.
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
- Kraker, P., Dörler, D., Ferus, A., Gutounig, R., Heigl, F., Kaier, C., ... Vignoli, M. (2016, Juni 15). The Vienna principles: A vision for scholarly communication in the 21st century. Open Access Network Austria. doi:10.5281/zenodo.55597
- Kranzberg, M. (1986). Technology and history: „Kranzberg’s laws“. *Technology and Culture*, 27(3), 544–560. doi:10.2307/3105385
- Kreibich, C. (2016). *Scholar.Py*. Abgerufen von <https://github.com/ckreibich/scholar.py>
- Kroes, N. (2011, Dezember 12). Data is the new gold. Abgerufen von <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/11/872&type=HTML>
- Krotz, F. (2005). *Neue Theorien entwickeln: Eine Einführung in die Grounded Theory, die Heuristische Sozialforschung und die Ethnographie anhand von Beispielen aus der Kommunikationsforschung*. Köln: Herbert von Halem.
- Krotz, F. (2008). Kultureller und gesellschaftlicher Wandel im Kontext des Wandels von Medien und Kommunikation. In T. Thomas (Hrsg.), *Medienkultur und soziales Handeln* (S. 43–62). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-90898-4_2
- Krotz, F. & Hepp, A. (Hrsg.). (2012). *Mediatisierte Welten: Forschungsfelder und Beschreibungsansätze*. Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-94332-9
- Kuhn, T. S. (1970). Reflections on my critics. In I. Lakatos & A. Musgrave (Hrsg.), *Criticism and the growth of knowledge* (S. 231–278). Cambridge: Cambridge University Press.

- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2. Aufl.). Chicago, IL: University of Chicago Press. (Original erschienen 1962)
- Kuhn, T. S. (1976). *Die Struktur wissenschaftlicher Revolutionen* (Zweite revidierte und um das Postskriptum von 1969 ergänzte Aufl.). Frankfurt am Main: Suhrkamp. (Original erschienen 1969)
- Kundra, V. (2012). Digital fuel of the 21st century: Innovation through open data and the network effect. *Joan Shorenstein Center on the Press, Politics and Public Policy Discussion Paper Series*, (70). Abgerufen von http://www.hks.harvard.edu/presspol/publications/papers/discussion_papers/d70_kundra.pdf
- Kuzel, A. J. (1992). Sampling in qualitative inquiry. In B. F. Crabtree & W. L. Miller (Hrsg.), *Doing qualitative research* (S. 31–44). Thousand Oaks, CA: Sage.
- Kvale, S. (2007). *Doing Interviews*. London: Sage. doi:10.4135/9781849208963
- Lamble, S. (2001). Computer-assisted reporting, Philip Meyer and The Emperor's new clothes. *eJournalist*, 1(2), 1–9. Abgerufen von <http://ejournalist.com.au/v1n2/lamble.pdf>
- Latour, B. (1988). *The pasteurization of France*. (A. Sheridan & J. Law, Übers.). Cambridge, MA: Harvard University Press.
- Latour, B. (1994). On technical mediation: philosophy, sociology, genealogy. *Common Knowledge*, 3(2), 29–64.
- Latour, B. (2001). *Das Parlament der Dinge: Für eine politische Ökologie*. (G. Roßler, Übers.). Frankfurt am Main: Suhrkamp.
- Latour, B. (2005). *Reassembling the social: An introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Latour, B. (2007). *Eine neue Soziologie für eine neue Gesellschaft: Einführung in die Akteur-Netzwerk-Theorie*. (G. Roßler, Übers.). Frankfurt am Main: Suhrkamp.
- Latour, B. (2010). *On the modern cult of the factish gods*. Durham: Duke University Press.
- Latour, B. & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts* (2. Aufl.). Princeton: Princeton University Press. (Original erschienen 1979)

- Latour, B., Jensen, P., Venturini, T., Grauwin, S. & Boullier, D. (2012). 'The whole is always smaller than its parts': A digital test of Gabriel Tarde's monads. *The British Journal of Sociology*, 63(4), 590–615. doi:10.1111/j.1468-4446.2012.01428.x
- Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems Practice*, 5(4), 379–393. doi:10.1007/BF01059830
- Law, J. (2004). *After method: Mess in social science research*. London: Routledge.
- Leavitt, A. (2016, August). *Upvoting the news: Breaking news aggregation, crowd collaboration, and algorithm-driven attention on reddit.com* (Dissertation). University of Southern California, Los Angeles, CA. Abgerufen von <http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll40/id/297862>
- Leavitt, A. & Robinson, J. J. (2017). The role of information visibility in network gatekeeping: Information aggregation on Reddit during crisis events. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (S. 1246–1261). New York, NY: ACM. doi:10.1145/2998181.2998299
- Lecheler, S. & Kruike-meier, S. (2016). Re-evaluating journalistic routines in a digital age: A review of research on the use of online sources. *New Media & Society*, 18(1), 156–171. doi:10.1177/1461444815600412
- Lee, K. C. & Fleming, C. A. (1995). Problems of introducing courses in computer-assisted reporting. *Journalism & Mass Communication Educator*, 50(3), 23–34. doi:10.1177/107769589505000304
- Lewis, S. C. (2010). *Journalism innovation and the ethic of participation: A case study of the Knight Foundation and its news challenge* (Dissertation). University of Texas, Austin, TX. Abgerufen von <http://sethleoewis.org/wp-content/uploads/2010/08/Seth-C.-Lewis-dissertation-2010.pdf>
- Lewis, S. C. (2011). Journalism innovation and participation: An analysis of the Knight News Challenge. *International Journal of Communication*, 5, 1623–1648. Abgerufen von <http://ijoc.org/index.php/ijoc/article/view/1140>
- Lewis, S. C. (2015). Journalism in an era of big data. *Digital Journalism*, 3(3), 321–330. doi:10.1080/21670811.2014.976399

- Lewis, S. C. & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture and Society*, 35(5), 602–619. doi:10.1177/0163443713485494
- Lewis, S. C. & Usher, N. (2014). Code, collaboration, and the future of journalism: A case study of the Hacks/Hackers global network. *Digital Journalism*, 2(3), 383–393. doi:10.1080/21670811.2014.895504
- Lewis, S. C. & Westlund, O. (2015). Big data and journalism. *Digital Journalism*, 3(3), 447–466. doi:10.1080/21670811.2014.976418
- Léchenet, A. (2014). *Global database investigations: The role of the computer-assisted reporter*. Oxford: Reuters Institute for the Study of Journalism. Abgerufen von <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Global%20Database%20Investigations%20-%20The%20role%20of%20the%20computer-assisted%20reporter.pdf>
- Lievrouw, L. A. (2002). Determination and contingency in new media development: Diffusion of innovations and social shaping of technology perspectives. In *Handbook of new media: Social shaping and consequences of ICTs* (S. 183–199). London: Sage.
- Lindemann, G. (2011). On Latour's social theory and theory of society, and his contribution to saving the world. *Human Studies*, 34(1), 93–110. doi:10.1007/s10746-011-9178-9
- Loenhoff, J. (2011). Die Objektivität des Sozialen. In B. Pörksen (Hrsg.), *Schlüsselwerke des Konstruktivismus* (S. 143–159). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Loosen, W., Reimer, J. & De Silva-Schmidt, F. (2016). *Wenn aus Daten Journalismus wird: Eine Inhaltsanalyse der für die Data Journalism Awards 2013 bis 2015 nominierten Projekte* (Arbeitspapiere des Hans-Bredow-Instituts Nr. 39). Hamburg: Verlag Hans-Bredow-Institut. Abgerufen von <http://hbi.to/957>
- Loosen, W., Reimer, J. & Schmidt, F. (2015). When data become news: A content analysis of data journalism pieces. Gehalten auf der The Future of Journalism 2015 Conference, Cardiff.
- Lopez, P. (2009). GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications. In M. Agosti, J. Borbinha, S. Kapidakis, C. Papatheodorou & G. Tsakonas (Hrsg.), *Research and Advanced Technology for Digital Libraries* (S. 473–474). Berlin: Springer. doi:10.1007/978-3-642-04346-8_62

- Lorenz, M. (2010). Status and Outlook for Data-Driven Journalism. In European Journalism Centre (Hrsg.), *Data-Driven Journalism: What Is There to Learn?* (S. 8–17). Amsterdam. Abgerufen von http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf
- Lorenz, M. (2012). Business models for data journalism. In J. Gray, L. Bounegru & L. Chambers (Hrsg.), *The data journalism handbook: How journalists can use data to improve the news*. Sebastopol, CA: O'Reilly Media. Abgerufen von http://datajournalismhandbook.org/1.0/en/in_the_newsroom_10.html
- Lorenz-Meyer, L. (2011, September 29). Was ist Datenjournalismus? Abgerufen von <http://www.scarlatti.de/?p=861>
- Lorenz-Meyer, L. (2013, Juni 10). Jagen, Sammeln, Pflegen: Alte und neue Aufgaben für die Öffentlich-Rechtlichen im Internet. Abgerufen von <http://www.scarlatti.de/?p=1524>
- Los Angeles Times Data Desk. (2017). Crime L.A. – Mapping L.A. *Los Angeles Times*. Abgerufen von <http://maps.latimes.com/crime/>
- Lotan, G. (2013, April 25). When your Twitter friend turns out to be the Boston bomber. *Digg*. Abgerufen von <http://digg.com/originals/dzhokhar-tsarnaev-twitter-map>
- Löffelholz, M. & Rothenberger, L. (2016). Einführung und Überblick. In M. Löffelholz & L. Rothenberger (Hrsg.), *Handbuch Journalismustheorien* (S. 11–25). Wiesbaden: Springer. doi:10.1007/978-3-531-18966-6_1
- Luckmann, T. (1970). On the boundaries of the social world. In M. Natanson (Hrsg.), *Phenomenology and social reality: Essays in memory of Alfred Schutz* (S. 73–100). Den Haag: Martinus Nijhoff. doi:10.1007/978-94-011-7523-4_5
- Luckmann, T. (2002). Moral communication in modern societies. *Human Studies*, 25(1), 19–32. doi:10.1023/A:1014838423896
- Luckmann, T. (2013). Predigten, Moralpredigten und Moral predigen. In J. Bergmann & T. Luckmann (Hrsg.), *Kommunikative Konstruktion von Moral* (Bd. 2, S. 80–111). Mannheim: Verlag für Gesprächsforschung.
- Lugo-Ocando, J. & Brandão, R. F. (2016). Stabbing news: Articulating crime statistics in the newsroom. *Journalism Practice*, 10(6), 715–729. doi:10.1080/17512786.2015.1058179

- Luhmann, N. (1996). *Die neuzeitlichen Wissenschaften und die Phänomenologie*. Wien: Picus.
- Luhmann, N. (1997). *Die Gesellschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- Lupton, D. (2013, Oktober 29). Swimming or drowning in the data ocean? Thoughts on the metaphors of big data. Abgerufen von <http://simplysociology.wordpress.com/2013/10/29/swimming-or-drowning-in-the-data-ocean-thoughts-on-the-metaphors-of-big-data/>
- Lyman, P. & Varian, H. R. (2003). How much information? Abgerufen von <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/index.htm>
- Lynch, M. (1985). *Art and artifact in laboratory science*. London: Routledge & Kegan Paul.
- Lyotard, J.-F. (2015). *Das postmoderne Wissen: Ein Bericht*. (P. Engelmann, Hrsg., O. Pfersmann, Übers.) (4., unveränderte Neuaufl.). Wien: Passagen Verlag. (Original erschienen 1979)
- Maan, N. & Schmid, F. (2016, Juli 9). Was Pegida für die Wahrheit hält. *Der Standard*, S. 4–5. Wien.
- Maan, N., Schmid, F. & Hametner, M. (2016, Oktober 1). „Zur Info“: Straches Facebook-Welt. *Der Standard*: 4-5. Wien.
- Mager, A. (2012). Algorithmic Ideology: How capitalist society shapes search engines. *Information, Communication & Society*, 15(5), 769–787. doi:10.1080/1369118X.2012.676056
- Mahle, I. (2016). *Digitale Bezahlhalte im World Wide Web als Geschäftsmodell: Eine quantitative Befragung der ePaper-Abonnenten von Stuttgarter Zeitung und Stuttgarter Nachrichten*. Baden-Baden: Nomos. doi:10.5771/9783845266220
- Mahrt, M. (2015). Mit Big Data gegen das „Ende der Theorie“? In A. Maireder, J. Ausserhofer, C. Schumann & M. Taddicken (Hrsg.), *Digitale Methoden in der Kommunikationswissenschaft* (S. 23–37). Berlin: Digital Communication Research. doi:10.17174/dcr.v2.2
- Maireder, A. & Ausserhofer, J. (2014). Political discourses on Twitter: Networking topics, objects and people. In K. Weller, A. Bruns, J. Burgess, M. Mahrt & C. Puschmann

- (Hrsg.), *Twitter and Society* (S. 305–318). New York, NY: Peter Lang.
doi:10.5281/zenodo.569914
- Manovich, L. (2013). *Software takes command*. New York, NY: Bloomsbury.
- Markham, A. (2013). Remix cultures, remix methods: Reframing qualitative inquiry for social media contexts. In N. K. Denzin & M. D. Giardina (Hrsg.), *Global dimensions of qualitative inquiry* (S. 63–81). Walnut Creek, CA: Left Coast Press.
- Marres, N. (2012a). *Material participation: Technology, the environment and everyday publics*. Basingstoke: Palgrave Macmillan. doi:10.1057/9781137029669
- Marres, N. (2012b). The redistribution of methods: On intervention in digital social research, broadly conceived. *The Sociological Review*, 60(S1), 139–165. doi:10.1111/j.1467-954X.2012.02121.x
- Marsh, D. & Hodsdon, A. (2017). The Guardian and Observer Style Guide. Abgerufen von <http://www.theguardian.com/info/series/guardian-and-observer-style-guide>
- Marvin, C. (1988). *When old technologies were new: Thinking about electric communication in the late nineteenth century*. New York: Oxford University Press.
- Massaro, M., Dumay, J. & Garlatti, A. (2015). Public sector knowledge management: A structured literature review. *Journal of Knowledge Management*, 19(3), 530–558. doi:10.1108/JKM-11-2014-0466
- Massaro, M., Dumay, J. & Guthrie, J. (2016). On the shoulders of giants: Undertaking a structured literature review in accounting. *Accounting, Auditing & Accountability Journal*, 29(5), 767–801. doi:10.1108/AAAJ-01-2015-1939
- Matiasek, S. (2015). *Mit Daten Geschichten erzählen: Redaktionelle Prozesse im Datenjournalismus am Beispiel von derstandard.at* (Bachelorarbeit). FH Joanneum, Graz.
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston, MA: Houghton Mifflin Harcourt.
- Meier, K. & Reimer, J. (2011). Transparenz im Journalismus: Instrumente, Konfliktpotentiale, Wirkung. *Publizistik*, 56(2), 133–155. doi:10.1007/s11616-011-0116-7

- Meraz, S. & Papacharissi, Z. (2013). Networked gatekeeping and networked framing on #egypt. *The International Journal of Press/Politics*, 18(2), 138–166.
doi:10.1177/1940161212474472
- Merton, R. K. (1968). *Social theory and social structure*. New York, NY: The Free Press.
(Original erschienen 1949)
- Meuser, M. & Nagel, U. (2002). ExpertInneninterviews: Vielfach erprobt, wenig bedacht. In A. Bogner, B. Littig & W. Menz (Hrsg.), *Das Experteninterview: Theorie, Methode, Anwendung* (S. 71–94). Wiesbaden: Springer. doi:10.1007/978-3-322-93270-9
- Meyer, P. (2002). *Precision journalism: A reporter's introduction to social science methods* (4. Aufl.). Oxford: Rowman & Littlefield. (Original erschienen 1973)
- Meyer, P. (2012, Oktober 18). Precision journalism and narrative journalism: Toward a unified field theory. Abgerufen von <http://www.nieman.harvard.edu/reports/article-online-exclusive/100044/Precision-Journalism-and-Narrative-Journalism-Toward-a-Unified-Field-Theory.aspx>
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2. Aufl.). Thousand Oaks, CA: Sage.
- Moretti, F. (2013). *Distant Reading*. London: Verso.
- Muthukumaraswamy, K. (2010). When the media meet crowds of wisdom. *Journalism Practice*, 4(1), 48–65. doi:10.1080/17512780903068874
- Münch, R. & Schmidt, J. (2005). Medien und sozialer Wandel. In M. Jäckel (Hrsg.), *Mediensoziologie: Grundfragen und Forschungsfelder* (S. 201–218). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Narodoslawsky, B. (2016, Juli 6). Die Datenrevolution im Journalismus. *Falter*, 27a(16), 18–20. Abgerufen von <https://www.falter.at/falter/e-paper/lesen/434/okonomie#&gid=1&pid=18>
- Neuberger, C. (2004). Journalismus als systembezogene Akteurkonstellation. Grundlagen einer integrativen Journalismustheorie. In M. Löffelholz (Hrsg.), *Theorien des Journalismus: Ein diskursives Handbuch* (2., vollständig überarbeitete und erweiterte Aufl., S. 287–303). Wiesbaden: Springer. doi:10.1007/978-3-663-01620-5_13
- Neuberger, C., Nuernbergk, C. & Rischke, M. (2007). Weblogs und Journalismus: Konkurrenz, Ergänzung oder Integration? *media Perspektiven*, 2007(2), 96–112.

Abgerufen von http://www.centrestage.de/podpress_trac/web/356/0/02-2007_neuberger.pdf

New York Times. (2010). The war logs. Abgerufen von

<http://www.nytimes.com/interactive/world/war-logs.html>

Niederer, C., Aigner, W. & Rind, A. (2015). Survey on visualizing dynamic, weighted, and directed graphs in the context of data-driven journalism. In H.-J. Schulz, B. Urban & U. Freiherr von Lukas (Hrsg.), *Proceedings of the International Summer School on Visual Computing 2015* (S. 49–58). Rostock: Fraunhofer Verlag.

Niederer, C., Rind, A., Aigner, W., Ausserhofer, J., Gutounig, R. & Sedlmair, M. (2016). Visual exploration of media transparency for data journalists: Problem characterization and abstraction. In *Tagungsband des 10. Forschungsforum der Österreichischen Fachhochschulen*. Wien. Abgerufen von <http://ffhoarep.fh-ooe.at/handle/123456789/542>

North, K. (2016). *Wissensorientierte Unternehmensführung: Wissensmanagement gestalten* (6., aktualisierte und erweiterte Aufl.). Wiesbaden: Springer Gabler. doi:10.1007/978-3-658-11643-9

Obermayer, B. & Obermaier, F. (2016). *Panama Papers: Die Geschichte einer weltweiten Enthüllung*. Köln: Kiepenheuer & Witsch.

Oxford Dictionaries. (2016). Word of the Year 2016 is... Abgerufen von

<https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2016>

Parasie, S. (2015). Data-driven revelation? Epistemological tensions in investigative journalism in the age of „big data“. *Digital Journalism*, 3(3), 364–380. doi:10.1080/21670811.2014.976408

Parasie, S. & Dagiral, E. (2013). Data-driven journalism and the public good: „Computer-assisted-reporters“ and „programmer-journalists“ in Chicago. *New Media & Society*, 15(6), 853–871. doi:10.1177/1461444812463345

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York: Penguin.

Parks, W. (1957). Open government principle: Applying the right to know under the Constitution. *The George Washington Law Review*, 26(1).

- Passig, K. (2013). *Standardsituationen der Technologiekritik*. Berlin: Suhrkamp.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Beverly Hills, CA: Sage.
- Peters, J. D. (2001). „The only proper scale of representation“: The politics of statistics and stories. *Political Communication*, 18(4), 433–449. doi:10.1080/10584600152647137
- Plaue, C. & Cook, L. R. (2015). Data journalism: Lessons learned while designing an interdisciplinary service course. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education* (S. 126–131). New York, NY: ACM. doi:10.1145/2676723.2677263
- Popper, K. (1935). *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft*. Wien: Springer. doi:10.1007/978-3-7091-4177-9
- Potts, L. & Harrison, A. (2013). Interfaces as rhetorical constructions: Reddit and 4chan during the Boston Marathon bombings. In *Proceedings of the 31st ACM International Conference on Design of Communication* (S. 143–150). New York, NY: ACM. doi:10.1145/2507065.2507079
- Powers, M. (2012). „In forms that are familiar and yet-to-be invented“: American journalism and the discourse of technologically specific work. *Journal of Communication Inquiry*, 36(1), 24–43. doi:10.1177/0196859911426009
- Poynter Institute for Media Studies. (1999). *When nerds and words collide: Reflections on the development of computer assisted reporting*. St. Petersburg, FL: Poynter.
- Primo, A. & Zago, G. (2015). Who and what do journalism? An actor-network perspective. *Digital Journalism*, 3(1), 38–52. doi:10.1080/21670811.2014.927987
- Procter, R., Vis, F. & Voss, A. (2013). Reading the riots on Twitter: Methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, 16(3), 197–214. doi:10.1080/13645579.2013.774172
- Prokop, D. (2002). *Der Medien-Kapitalismus: Das Lexikon der neuen kritischen Medienforschung* (aktualisierte Neuaufl.). Hamburg: VSA.
- Pulimood, S. M., Shaw, D. & Lounsbury, E. (2011). Gumshoe: A model for undergraduate computational journalism education. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education* (S. 529–534). New York, NY: ACM. doi:10.1145/1953163.1953314

- Puschmann, C. (2015, November). *Sozialwissenschaften durch Big Data? Ein „systematic review“ zu Methoden, Daten und Fragestellungen der Computational Social Science*. gehalten auf der gemeinsamen Jahrestagung der Fachgruppen Computervermittelte Kommunikation und Soziologie der Medienkommunikation der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft, Berlin.
- Puschmann, C. & Ausserhofer, J. (2017). Social data APIs: Origin, types, issues. In M. T. Schäfer & K. van Es (Hrsg.), *The datafied society: Studying culture through data* (S. 147–154). Amsterdam: Amsterdam University Press. doi:10.5281/zenodo.344887
- Puschmann, C. & Burgess, J. (2014). Metaphors of big data. *International Journal of Communication*, 8, 1690–1709. Abgerufen von <http://ijoc.org/index.php/ijoc/article/view/2169>
- Puschmann, C. & Scheffler, T. (2016). Topic modeling for media and communication research: A short primer. *HIIG Discussion Paper Series*, 2016(5). doi:10.2139/ssrn.2836478
- Puschmann, C., Ausserhofer, J., Maan, N. & Hametner, M. (2016). Information laundering and counter-publics: The news sources of Islamophobic groups on Twitter. In *Proceedings of Social Media in the Newsroom (SMnews 2016) (Workshop at the 10th International AAAI Conference on Web and Social Media (ICWSM16))* (S. 143–150). Menlo Park: AAAI Press. Abgerufen von <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13224>
- QSR International. (2016). *Nvivo*. Abgerufen von <http://www.qsrinternational.com/>
- Raabe, J. (2005). *Die Beobachtung journalistischer Akteure: Optionen einer empirisch-kritischen Journalismusforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-322-80463-1
- Radchenko, I. & Sakoyan, A. (2014). The view on open data and data journalism: Cases, educational resources and current trends. In D. I. Ignatov, M. Y. Khachay, A. Panchenko, N. Konstantinova & R. E. Yavorsky (Hrsg.), *Analysis of Images, Social Networks and Texts* (S. 47–54). Cham: Springer.
- Ragin, C. C. (1994). *Constructing social research: The unity and diversity of method*. Thousand Oaks, CA: Pine Forge Press.

- Rheinberger, H.-J. (1997). *Toward a history of epistemic things: Synthesizing proteins in the test tube*. Stanford, CA: Stanford University Press.
- Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Reading, MA: Addison-Wesley.
- Rheingold, H. (2002). *Smart mobs: The next social revolution*. Cambridge, MA: Perseus.
- Rheingold, H. (2010). Attention, and other 21st-century social media literacies. *EDUCAUSE Review*, 45(5), 14. Abgerufen von <http://er.educause.edu/articles/2010/10/attention-and-other-21stcentury-social-media-literacies>
- Rieder, B. & Röhle, T. (2012). Digital Methods: Five Challenges. In D. M. Berry (Hrsg.), *Understanding Digital Humanities* (S. 67–84). Basingstoke: Palgrave Macmillan.
- Rind, A., Pfahler, D., Niederer, C. & Aigner, W. (2016). Exploring media transparency with multiple views. In W. Aigner, G. Schmiedl, K. Blumenstein & M. Zeppelzauer (Hrsg.), *Proceedings of the 9th Forum Media Technology 2016* (S. 65–73). St. Pölten: CEUR-WS. Abgerufen von <http://ceur-ws.org/Vol-1734/>
- Roessing, T. (2014). Enzyklopädie-Amateure als Amateur-Journalisten: Wikipedia als Gateway für aktuelle Ereignisse. *Studies in Communication | Media*, 3(2), 205–227. doi:10.5771/2192-4007-2014-2-205
- Rogers, R. (2010). Internet research: The question of method – A keynote address from the YouTube and the 2008 election cycle in the United States conference. *Journal of Information Technology & Politics*, 7(2–3), 241–260. doi:10.1080/19331681003753438
- Rogers, R. (2013). *Digital methods*. Cambridge, MA: MIT Press.
- Rogers, R. (2014). Political research in the digital age. *International Public Policy Review*, 8(1), 73–87.
- Rogers, S. (2008, Dezember 18). Turning official figures into understandable graphics, at the press of a button. Abgerufen von <http://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata>
- Rogers, S. (2011, April 7). Data journalism broken down: What we do to the data before you see it. Abgerufen von <http://www.theguardian.com/news/datablog/2011/apr/07/data-journalism-workflow>

- Rogers, S. (2013). *Facts are sacred: The power of data*. London: Faber and Faber.
- Rosenberg, D. (2013). Data before the fact. In L. Gitelman (Hrsg.), „*Raw data*“ is an *oxymoron* (S. 15–40). Cambridge: MIT Press.
- Rost, L. C. (2016, Mai 17). One chart, twelve charting libraries. Abgerufen von <https://lisacharlotterost.github.io/2016/05/17/one-chart-code/>
- Roy Rosenzweig Center for History and New Media. (2015). *Zotero*. Abgerufen von <https://www.zotero.org/>
- Royal, C. (2012). The journalist as programmer: A case study of the New York Times Interactive News Technology department. #ISOJ the Official Research Journal of the *International Symposium on Online Journalism*, 2(1). Abgerufen von <https://online.journalism.utexas.edu/2010/papers/Royal10.pdf>
- Royal, C. & Blasingame, D. (2015). Data journalism: An explication. *International Symposium on Online Journalism*, 5(1), 24–46. Abgerufen von https://online.journalism.utexas.edu/journals/ISOJ_Journal_V5_N1_2015_Spring.pdf
- Rubin, H. J. & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* (3. Aufl.). Thousand Oaks, CA: Sage.
- Rusch, T., Hofmarcher, P., Hatzinger, R. & Hornik, K. (2013). Model trees with topic model preprocessing: An approach for data journalism illustrated with the WikiLeaks Afghanistan war logs. *The Annals of Applied Statistics*, 7(2), 613–639. doi:10.1214/12-AOAS618
- Russ-Mohl, S. (2006). The economics of journalism and the challenge to improve journalism quality: A research manifesto. *Studies in Communication Sciences*, 6(2), 189–208.
- Ruß-Mohl, S. (2001). Benchmarking Transparenz und Interaktivität bei führenden amerikanischen Zeitungen. In H. J. Kleinsteuber (Hrsg.), *Aktuelle Medientrends in den USA: Journalismus, politische Kommunikation und Medien im Zeitalter der Digitalisierung* (S. 109–125). Springer. doi:10.1007/978-3-663-07786-2_6
- Sady, W. (2012). Ludwik Fleck. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Summer 2012). Abgerufen von <http://plato.stanford.edu/archives/sum2012/entries/fleck/>

- Schäfer, L. & Schnelle, T. (1980). Ludwik Flecks Begründung der soziologischen Betrachtungsweise in der Wissenschaftstheorie. In L. Fleck, *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv* (S. VII–XLIX). Frankfurt am Main: Suhrkamp.
- Schmid, F., Maan, N., Hametner, M. & Šlerka, J. (2016, Dezember 1). Was Hofers und Van der Bellens Anhänger auf Facebook sehen. *derStandard.at*. Abgerufen von <http://derstandard.at/2000048426887/Was-Hofers-und-Van-der-Bellens-Anhaenger-auf-Facebook-sehen>
- Schmidt, R. (2012). *Soziologie der Praktiken: Konzeptionelle Studien und empirische Analysen*. Berlin: Suhrkamp.
- Scholl, A. (2011). Der unauflösbare Zusammenhang von Fragestellung, Theorie und Methode: Die reflexive Bedeutung der Methodologie (nicht nur) in der Journalismusforschung. In O. Jandura, T. Quandt & J. Vogelgesang (Hrsg.), *Methoden der Journalismusforschung* (S. 15–32). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks: Sage.
- Schudson, M. (2001). The objectivity norm in American journalism. *Journalism*, 2(2), 149–170. doi:10.1177/146488490100200201
- Schudson, M. (2005). Four approaches to the sociology of news. In J. Curran & M. Gurevitch (Hrsg.), *Mass Media and Society* (4. Aufl., S. 172–197). London: Hodder Arnold.
- Schudson, M. (2010). Political observatories, databases & news in the emerging ecology of public information. *Daedalus*, 139(2), 100–109. doi:10.1162/daed.2010.139.2.100
- Schumann, C., Ausserhofer, J., Maireder, A. & Taddicken, M. (2015). Digitale Methoden in der Kommunikationswissenschaft: Ansätze zur Analyse öffentlicher Kommunikation im Internet. In A. Maireder, J. Ausserhofer, C. Schumann & M. Taddicken (Hrsg.), *Digitale Methoden in der Kommunikationswissenschaft* (S. 9–19). Berlin: Digital Communication Research. doi:10.17174/dcr.v2.1
- Schütz, A. (1932). *Der sinnhafte Aufbau der sozialen Welt: Eine Einleitung in die verstehende Soziologie*. Wien: Springer. doi:10.1007/978-3-7091-3108-4

- Schütz, A. (1971). Über die mannigfaltigen Wirklichkeiten. In *Gesammelte Aufsätze I: Das Problem der sozialen Wirklichkeit* (S. 237–298). Den Haag: Martinus Nijhoff. doi:10.1007/978-94-010-2858-5_9 (Original erschienen 1945)
- Schütz, A. (1971). Wissenschaftliche Interpretation und Alltagsverständnis menschlichen Handelns. In *Gesammelte Aufsätze I: Das Problem der sozialen Wirklichkeit* (S. 3–54). Den Haag: Martinus Nijhoff. doi:10.1007/978-94-010-2858-5_1 (Original erschienen 1953)
- Schütz, A. (1972). Die soziale Welt und die Theorie der sozialen Handlung. In A. Brodersen (Hrsg.), A. von Baeyer (Übers.), *Gesammelte Aufsätze II: Studien zur soziologischen Theorie* (S. 3–21). Den Haag: Martinus Nijhoff. doi:10.1007/978-94-010-2849-3_1 (Original erschienen 1960)
- Schütz, A. & Luckmann, T. (2003). *Strukturen der Lebenswelt*. Stuttgart: UVK. (Original erschienen 1973)
- Schützeichel, R. (2012). Wissenssoziologie. In S. Maasen, M. Kaiser, M. Reinhart & B. Sutter (Hrsg.), *Handbuch Wissenschaftssoziologie* (S. 17–26). Wiesbaden: Springer. doi:10.1007/978-3-531-18918-5_1
- Segel, E. & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1139–1148. doi:10.1109/TVCG.2010.179
- Siebenhofer, A. & Kanya, E. (2015, August 26). Polizei-PR versus Wirklichkeit. *Falter*, (25), 22–24. Abgerufen von <https://www.falter.at/archiv/wp/polizei-pr-versus-wirklichkeit>
- Siles, I. & Boczkowski, P. (2012). At the intersection of content and materiality: A text-material perspective on the use of media technologies. *Communication Theory*, 22(3), 227–249. doi:10.1111/j.1468-2885.2012.01408.x
- Silver, A. & Matthews, L. (2016). The use of Facebook for information seeking, decision support, and self-organization following a significant disaster. *Information, Communication & Society*. doi:10.1080/1369118X.2016.1253762
- Silver, N. (2014, März 17). What the fox knows. *FiveThirtyEight*. Abgerufen von <http://fivethirtyeight.com/features/what-the-fox-knows/>
- Simmel, G. (1908). *Soziologie: Untersuchungen über die Formen der Vergesellschaftung*. Leipzig: Duncker & Humblot.

- Singer, J. B. (2007). Contested autonomy: Professional and popular claims on journalistic norms. *Journalism Studies*, 8(1), 79–95. doi:10.1080/14616700601056866
- Singer, J. B., Domingo, D., Heinonen, A., Hermida, A., Paulussen, S., Quandt, T., ... Vujnovic, M. (2011). *Participatory journalism: Guarding open gates at online newspapers*. Chichester: Wiley-Blackwell.
- Smit, G., Haan, Y. de & Buijs, L. (2014). Visualizing News: Make It Work. *Digital Journalism*, 2(3), 344–354. doi:10.1080/21670811.2014.897847
- Soffer, O. (2009). The competing ideals of objectivity and dialogue in American journalism. *Journalism*, 10(4), 473–491. doi:10.1177/1464884909104950
- Spinner, J. (2014, September 24). The big conundrum: Should journalists learn code? *American Journalism Review*. Abgerufen von <http://ajr.org/2014/09/24/should-journalists-learn-code/>
- Splendore, S., Di Salvo, P., Eberwein, T., Groenhart, H., Kus, M. & Porlezza, C. (2016). Educational strategies in data journalism: A comparative study of six European countries. *Journalism*, 17(1), 138–152. doi:10.1177/1464884915612683
- Stavelin, E. (2013). *Computational journalism: When journalism meets programming* (Dissertation). University of Bergen.
- Steensen, S. (2010). *Back to the feature: Online journalism as innovation, transformation and practice* (Dissertation). University of Oslo. Abgerufen von https://www.researchgate.net/profile/Steen_Steensen/publication/270793186_Back_to_the_feature_Online_journalism_as_innovation_transformation_and_practice/links/5656d53e08aeafc2aac09d66.pdf
- Steensen, S. (2016). What is the matter with newsroom culture? A sociomaterial analysis of professional knowledge creation in the newsroom. *Journalism*. doi:10.1177/1464884916657517
- Steensen, S. & Ahva, L. (2015). Theories of Journalism in a Digital Age. *Journalism Practice*, 9(1), 1–18. doi:10.1080/17512786.2014.928454
- Stingl de Vasconcelos Guedes, T. (2011). *Begehrtes Wissen: Eine systemtheoretische Reflexion zu Wissens- und Entscheidungskulturen in Organisationen* (Dissertation). Universität Wien. Abgerufen von <http://othes.univie.ac.at/15758/>

- Strauss, A. & Corbin, J. (1996). *Grounded Theory: Grundlagen qualitativer Sozialforschung*. (S. Niewiarra & H. Legewie, Übers.). Weinheim: Beltz. (Original erschienen 1990)
- Strübing, J. (2008). *Grounded Theory: Zur sozialtheoretischen und epistemologischen Fundierung des Verfahrens der empirisch begründeten Theoriebildung* (2., überarbeitete und erweiterte Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Sturges, J. E. & Hanrahan, K. J. (2004). Comparing telephone and face-to-face qualitative interviewing: A research note. *Qualitative Research*, 4(1), 107–118.
doi:10.1177/1468794104041110
- Suber, P. (2012). *Open access*. Cambridge, MA: MIT Press.
- Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton, NJ: Princeton University Press.
- Tabary, C., Provost, A.-M. & Trottier, A. (2016). Data journalism's actors, practices and skills: A case study from Quebec. *Journalism*, 17(1), 66–84.
doi:10.1177/1464884915593245
- Tandoc, E. C. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media & Society*, 16(4), 559–575. doi:10.1177/1461444814530541
- Tandoc, E. C. & Oh, S.-K. (2015). Small departures, big continuities? Norms, values, and routines in The Guardian's big data journalism. *Journalism Studies*.
doi:10.1080/1461670X.2015.1104260
- Tankard, J. W. & Lasorsa, D. L. (2000). Teaching strategies for analysis in computer-assisted reporting. *Journalism & Mass Communication Educator*, 5(3), 14–26.
doi:10.1177/107769580005500303
- Tauberer, J. (2014, August). 14 principles of open government data. Abgerufen von <https://opengovdata.io/2014/principles/>
- Thelitz, N., Kohli, A., Fichter, A., Kobler, S., Preusse, T., Schudel, B., ... Interactive Things. (2014, September 8). Ungleiches Gezwitscher aus dem Parlament. *Neue Zürcher Zeitung*. Abgerufen von <http://www.nzz.ch/schweiz/ungleiches-gezwitscher-aus-dem-parlament-1.18378048>
- Tranfield, D., Denyer, D. & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207–222. doi:10.1111/1467-8551.00375

- Treadwell, G., Ross, T., Lee, A. & Lowenstein, J. K. (2016). A numbers game: Two case studies in teaching data journalism. *Journalism & Mass Communication Educator*, 71(3), 297–308. doi:10.1177/1077695816665215
- Treumann, K. P., Baacke, D., Haacke, K., Hugger, K. U. & Vollbrecht, R. (2002). *Medienkompetenz im digitalen Zeitalter*. Wiesbaden: Springer. doi:10.1007/978-3-663-10774-3
- Tröger, J. (2012, Oktober 11). Wie die Berliner Abgeordneten Twitter nutzen. *Berliner Morgenpost*. Abgerufen von <http://www.morgenpost.de/berlin-aktuell/article109759023/Wie-die-Berliner-Abgeordneten-Twitter-nutzen.html>
- Tuchman, G. (1978). *Making news: A study in the construction of reality*. New York, NY: The Free Press.
- Tuomi, I. (1999). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Information Systems*, 16(3), 103–117. doi:10.1080/07421222.1999.11518258
- Ubaldi, B. (2013). *Open government data: Towards empirical analysis of open government data initiatives* (OECD Working Papers on Public Governance Nr. 22). Paris: OECD Publishing. doi:10.1787/5k46bj4f03s7-en
- Usher, N. (2014). *Making news at The New York Times*. Ann Arbor, MI: University of Michigan Press. doi:10.3998/nmw.12848274.0001.001
- Usher, N. (2016). *Interactive journalism: Hackers, data, and code*. Urbana, IL: University of Illinois Press.
- Uskali, T. & Appelgren, E. (2015). The Third Nordic Data Journalism Conference: Call for papers. Abgerufen von <http://www.noda2016.fi/callforpapers.html>
- Uskali, T. & Kuutti, H. (2015). Models and streams of data journalism. *The Journal of Media Innovations*, 2(1), 77–88. doi:10.5617/jmi.v2i1.882
- van den Hoonaard, W. C. (2004). The epistemological framework of qualitative research: Is it still a Cinderella to quantitative research? In J. Fikfak, F. Adam & D. Garz (Hrsg.), *Qualitative research: Different perspectives, emerging trends* (S. 35–49). Ljubljana: Založba ZRC.

- Vehkoo, J. (2013). *Crowdsourcing in investigative journalism*. Oxford: Reuters Institute for the Study of Journalism. Abgerufen von http://reutersinstitute.politics.ox.ac.uk/sites/default/files/Crowdsourcing%20in%20Investigative%20Journalism_0.pdf
- Wahl-Jorgensen, K. & Hanitzsch, T. (2009). Introduction: On why and how we should do journalism studies. In K. Wahl-Jorgensen & T. Hanitzsch (Hrsg.), *The handbook of journalism studies* (S. 3–16). New York: Routledge.
- Wallace, W. L. (1971). *The logic of science in sociology*. Chicago, IL: Aldine and Atherton.
- Warnes, S. (2017, Februar 23). Is there room for „tabloid data journalism“? Abgerufen von <https://medium.com/@sophiewarnes/is-there-room-for-tabloid-data-journalism-496121268c46>
- Weber, W. (2013). Typen, Muster und hybride Formen: Ein Typologisierungsmodell für interaktive Infografiken. In W. Weber, M. Burmester & R. Tille (Hrsg.), *Interaktive Infografiken* (S. 25–37). Berlin: Springer. doi:10.1007/978-3-642-15453-9_2
- Weber, W. & Rall, H. (2012). Data visualization in online journalism and its implications for the production process. In *Proceedings of the 16th International Conference on Information Visualisation (IV)* (S. 349–356). Montpellier: IEEE. doi:10.1109/IV.2012.65
- Weber, W. & Rall, H. (2013). „We are journalists“: Production practices, attitudes and a case study of the New York Times newsroom. In W. Weber, M. Burmester & R. Tille (Hrsg.), *Interaktive Infografiken* (S. 161–172). Berlin: Springer. doi:10.1007/978-3-642-15453-9_9
- Wehling, P. (2007). Wissensregime. In R. Schützeichel (Hrsg.), *Handbuch Wissenssoziologie und Wissensforschung* (S. 704–712). Konstanz: UVK.
- Weinacht, S. & Spiller, R. (2014). Datenjournalismus in Deutschland: Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. *Publizistik*, 59(4), 411–433. doi:10.1007/s11616-014-0213-5
- Weinberg, S. (1998). The revolution that didn't happen. *The New York Review of Books*, 45(15), 48–52. Abgerufen von <http://www.nybooks.com/articles/1998/10/08/the-revolution-that-didnt-happen/>

- Weinberger, D. (2011). *Too big to know: Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room*. New York, NY: Basic Books.
- Weinstein, M. (2006). TAMS Analyzer: Anthropology as cultural critique in a digital age. *Social Science Computer Review*, 24(1), 68–77. doi:10.1177/0894439305281496
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.
- Wenger, E., McDermott, R. A. & Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Boston, MA: Harvard Business School Press.
- Westie, F. R. (1957). Toward closer relations between theory and research: A procedure and an example. *American Sociological Review*, 22(2), 149–154. doi:10.2307/2088851
- Whyte, A. & Pryor, G. (2011). Open science in practice: Researcher perspectives and participation. *International Journal of Digital Curation*, 6(1), 199–213. doi:10.2218/ijdc.v6i1.182
- Wilke, J. (2003). Zur Geschichte der journalistischen Qualität. In K.-D. Altmeyden & H.-J. Bucher (Hrsg.), *Qualität im Journalismus: Grundlagen, Dimensionen, Praxismodelle* (S. 40–54). Wiesbaden: Westdeutscher Verlag.
- Willke, H. (1998). *Systemisches Wissensmanagement*. Stuttgart: Lucius & Lucius.
- Wittenbrink, H. & Ausserhofer, J. (2013). Web Literacies und offene Bildung. In C. Landler, P. Parycek & M. C. Kettmann (Hrsg.), *Netpolitik in Österreich: Internet, Macht, Menschenrechte* (S. 225–336). Krems: Edition Donau-Universität Krems. doi:10.5281/zenodo.569917
- Wuchty, S., Jones, B. F. & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039. doi:10.1126/science.1136099
- Yarnall, L., Johnson, J. T., Rinne, L. & Ranney, M. A. (2008). How post-secondary journalism educators teach advanced car data analysis skills in the digital age. *Journalism & Mass Communication Educator*, 63(2), 146–164. doi:10.1177/107769580806300204

Young, M. L. & Hermida, A. (2015). From Mr. and Mrs. Outlier to central tendencies: Computational journalism and crime reporting at the Los Angeles Times. *Digital Journalism*, 3(3), 381–397. doi:10.1080/21670811.2014.976409

Zanchelli, M. & Crucianelli, S. (2012). *Integrating data journalism into newsrooms*. International Center for Journalists. Abgerufen von http://www.icfj.org/sites/default/files/integrating%20data%20journalism-english_0.pdf

Zwick, D. & Denegri Knott, J. (2009). Manufacturing customers: The database as new means of production. *Journal of Consumer Culture*, 9(2), 221–247. doi:10.1177/1469540509104375

9 Anhang

9.1 Codebuch

9.1.1 Codierprinzipien

- Codiert wird nach bestem Wissen. Das bedeutet, dass Vorwissen und Kontext aus dem Text für die Vergabe von Codes berücksichtigt werden.
- Die Standard-Codiereinheit ist ein Satz, außer im Codebuch ist etwas anderes vorgegeben.
- Eine Codiereinheit kann mehrfach codiert werden.
- Gruppen (z.B. ,10', ,20') werden nicht codiert, nur Codes (z.B. ,11', ,25').

9.1.2 Codetabelle

Tabelle 6. Codebuch

G Dimension	C Thema	Definition	Codierhinweis	Beispiel(e)
10 Individuum	11 Ausbildung	Professioneller Hintergrund, Karriere des Individuums. Etwa Studien, Weiterbildungen, Praktika, frühere Beschäftigungen oder andere professionelle Engagements.		„Many data journalists began as politics or business reporters and gradually picked up data skills as they became useful to particular stories.“ „The graphics people have degrees in cartography, statistics, graphic design and journalism, but others have different backgrounds, for example the graphics editor for Science has a degree in East Asian Studies, and the senior editor got his degree in Urban Studies, and the interactive designer majored in Economics.“
	13 Skills	Individuelle fachliche Fertigkeiten, die für DIJ benötigt werden.		„Most of these journalists have a background in CAR and have acquired sound skills in programming languages (Python, Ruby, PHP, etc.).“ „These steps require new competences and capabilities of those creating data stories: mathematical, visual and verbal competences, in addition technical and multimedia skills. Depending on the intended final outcome this

			might also include knowledge of modeling virtual characters, animating them, film editing and the addition of interactive elements, not to forget the obvious necessity to write text(s), incorporate sounds, and provide journalistic integrity and authenticity.“
15 Sozialisation und Selbstverständnis	Ereignisse, Erlebnisse, Umstände, die den Werdegang des Individuums geprägt haben. Keine Ausbildung. Was brachte das Individuum abseits von Ausbildung und Vorberufen zum DIJ? Auch: Persönliche Motive, die mit diesen Ereignissen verknüpft sind.		„Joe Germuska perfectly embodies this: when this programmer involved in open software left his previous job for a commercial company, he began to attend open government meetings more regularly. That was how he met Brian Boyer and Adrian Holovaty, and finally got hired as a programmer-journalist at the Tribune in 2009.“
20 Allgemeine Aspekte			
21 Definition	Kompakte Beschreibungen von DIJ und ähnlichen Praktiken.		„Lorenz describes it as a process of refinement, raw data is transformed into something meaningful: As a result the value to the public grows, especially when complex facts are boiled down into a clear story that people can easily understand and remember.“ „For one of the editors, data journalism is mostly about the competence to process raw data and to visualize this data so that it is understandable to the audience. Furthermore, this editor describes data journalism as a possibility for creating news that potentially can be agenda setting. In contrast to the daily work in the newsroom, where news, accidents and press releases dominate the flow of information, data journalism is about finding news in an area where no one has found a story.“
22 Berufsbezeichnungen	Bezeichnungen für DIJ und dessen PraktikerInnen.	Codiereinheit = der konkrete Ausdruck. Jede Bezeichnung ist nur einmal pro Dokument zu codieren.	„Datenjournalismus“, „programmer-journalists“

23 Organisationen	Organisationen hinter DIJ-Projekten. Medienorganisationen und Nicht-Medienorganisationen (wie z.B. NGOs und Unternehmen).	Codiereinheit = der Ausdruck. Jede Bezeichnung ist nur einmal pro Dokument zu codieren.	„The New York Times“, „The Guardian“, „Wikileaks“,
24 Beispiele	Beispiele für konkrete DIJ-Projekte seit 2009. Keine historischen Beispiele von vor 2009 (-> 093 Geschichte).	Codiereinheit = gesamte Äußerung zum Beispiel. Ausnahme: Bezieht sich ein ganzes Dokument auf ein Beispiel, dann ist nur die Bezeichnung des Beispiels zu codieren.	„The War Logs contained 391,832 field reports from soldiers. Since each report describes only a single incident, visualizations are extremely helpful to see patterns and get a bigger picture. (...)“ „The interactive graphic ‘The Guantánamo Docket’ is based on documents and research related to the 779 people who have been sent to the Guantánamo Bay prison since 2002.“
25 Inhaltscharakteristika	Charakteristika, Eigenschaften von DIJ-Beiträgen, z.B. im Vergleich zu klassischem ‚Textjournalismus‘. Besonderheiten der DIJ- ‚Darstellungsform‘.		„Most of the crowdsourcing projects were visualized using maps, and when working with public data from authorities, a common tool for visualization was to use maps. To summarize, in almost all of the interviews the editors described maps as the standard visualizing method for data journalism projects and crowdsourcing projects in particular.“ „Stories covering social issues (poverty, the environment, education and housing) were disproportionately likely to contain data elements (...), followed by world and news stories.“
27 Geschäftsprozesse und Infrastrukturen	Betriebswirtschaftliche Aspekte von DIJ, z.B. Ist DIJ profitabel? Wie sind Organisationen, die DIJ betreiben, ausgestattet und strukturiert, etwa in Bezug auf MitarbeiterInnen? Was sind unterstützende Strukturen sowohl inner- als auch außerhalb der Organisation?		„Larger organizations were more likely to undertake data work that involved a division of labor, with computer-assisted reporters, graphic designers, statisticians, and programmers working on teams. Smaller organizations were more likely to have ‘one-man bands’ who acquired data skills as needed or due to their own initiative.“ „Apart from advertising, news organisations envision some other future possibilities for: business models in data journalism, such as the following: - Premium model: Making people to pay for better visualisations. - Becoming

			the go-to store for data for organisations as well as for individual citizens. At the store, extra value would be created by compiling mega databases of, for example, government data, looking for common identifications in data sets and making these databases able to talk to each other. The data store could sell access to these databases and provide analysis services and visualisations.“
28 Kultur, Politik und Epistemologie	Kulturelle, politische, epistemologische Aspekte von DIJ. Z.B.: Wie sind die Verbindungen zu Hacker-Communities? Sehen sich die Beitragenden als gesellschaftliche AufklärerInnen? Wie ist das Verhältnis zur Open-Data-Bewegung? Welche Rolle spielen Daten in der Argumentation? Objektivität und Präzision von DIJ. Verhältnis zu Wissenschaft.		„Die Leute, die hier arbeiten, sind gesellschaftspolitisch interessiert und schätzen es eben auch, andauernd etwas Neues zu machen.“ „Also du kannst damit, glaube ich, Geschichten machen, die objektiver sind.“ „Für unsere Ergebnisse können wir oft nur Methoden benutzen, die nicht an die Wissenschaft heranreichen. Deswegen können wir nur deskriptive Ergebnisse formulieren, Deutungsansätze anbieten, vielleicht Experten stärken.“
29 Ortsbezug	Lokale, regionale oder nationale Besonderheiten von DIJ.	Codiereinheit = die Bezeichnung des Ortes plus Kontext.	„In Germany, where the open government movement was (and still is) much weaker, Der Spiegel covered the Afghanistan War Logs in a much more ‘traditional’ way, using no interactive illustrations at all and focusing on the print version. The experience in Britain and the United States to work with huge amounts of data was clearly an advantage for the coverage and made newspapers from other countries aware of the potential.“ „In the study of Norway, on the other hand, all organizations ‘have their base in Oslo or Bergen (or both).’ The lack of local news organizations, the authors note, may be an artifact of the snowball sampling methodology or, importantly, ‘the possibility that very few local newsrooms practice computational journalism on a regular basis’.“

30 Arbeitsabläufe

31 Themenfindung	Anstoß, Ideengebung für einen Beitrag. Vorarbeiten bis zur Auftragserteilung.	„At the FT there is a mixed approach to commissioning. Simple commissions that can be done in a day (such as timelines, or profile pictograms) are readily undertaken, but larger projects will only be started upon if they are likely to be carried on a main index page, or if the team stands to learn something from doing the project (which will, it is felt, in turn improve or streamline future coverage). Unlike the BBC Specials team, there is a propensity (and indeed a perceived necessity) to undertake ‘fishing’ exercises in large data sets, and so undertake original investigative work.“ „Putting developers and reporters in the same area of the newsroom also facilitates another component of these data-journalism teams: ideas for stories that come from either the developers/data miners or the journalists.“
32 Datensammlung	Erhebung und Beschaffung von strukturieren Informationen.	„Access to data is perceived as good by the Norwegian journalists interviewed in this study.“ „The data journalism presented relied heavily on institutional sources, especially government agencies. In addition, there is evidence of the rise of data-based press releases: a substantial proportion of the stories showed evidence of a body of data being released wholesale by research institutes and other vested bodies.“
33 Datensäuberung, -verifikation, -analyse	Reinigung der Daten, Quellenkritik, Datenauswertung.	„As Daniel O’Neil, one of the co-founders of EveryBlock explains, the combination of heterogeneous data sets may generate useful knowledge: ‘Okay, let’s see. You got a liquor license and then crime went up. I think that it’s fair to mention that. I think it’s a reasonable piece of data – not even data – insight to use in planning, right?’“ „The messiness and inaccuracy of the data also

		<p>raised an ethical concern. As Agustin told us, it was very important for the CIR to avoid making wrong claims about the safety of a school. Falsely claiming that a school is unsafe for children might provoke unfounded reactions from people and ruin the reputation of the news organization: No information is ever clean. No data is ever perfect. I'm willing to accept that. But given that limitation, it's really important and really necessary and really hard to figure out what we can responsibly say with this information."</p>
34 Produktion und Visualisierung	<p>Die Produktion des Beitrags. Auch: Das Schreiben und Gegenchecken sowie die Entwicklung von Visualisierungen.</p>	<p>„When it comes to info graphics and storytelling, our results show that most of the journalists interviewed for this study have ended up preferring traditional linear storytelling, emphasizing computation as useful for research rather than presentation.“</p> <p>„Another challenge for producing visualizations is the lack of common graphic strategies at newsrooms. There are no guidelines for making visualizations for journalistic stories.“</p>
35 Publikation und Evaluierung	<p>Die Veröffentlichung des Beitrags und damit verbundene Prozesse wie die Zurverfügungstellung von Quellcode und Daten.</p>	<p>„At the Tribune, Brian Boyer and his team also put the emphasis on giving out the data – through friendly applications – so that the readers can do their own research on how an issue affects their personal situation.“</p> <p>„Data-driven journalists could share their code and data on open source repositories like GitHub for others to inspect, replicate, and extend.“</p>
37 Kollaboration und Routinen	<p>Zusammenarbeit im Team. Auch: Position des Arbeitsplatzes. Teamgröße und -zusammensetzung. Dauer von Projekten. Interne oder öffentliche Dokumentation der Arbeit (Werkstattberichte).</p>	<p>„These departments are often separated, and cross-departmental cooperation is rare. One respondent told us that: ‘the relationship to IT is increasingly institutionalized and formalized and alienated. It is a sad tendency, but that’s that.’ An effort to bypass organizational ICT</p>

			infrastructure was found in all the newsrooms. Extra software that needed installing, databases that needed to be created, and the set-up of servers to host Web applications, are all examples of tasks that were done without the help of the ICT department. This said, the respondents also reported that in extraordinary cases, e.g. large WikiLeaks dumps, this non-cooperation state could be overcome.“
38	Fördernde und hemmende Faktoren	Aspekte, die die Produktion von DIJ fördern oder behindern. Herausforderungen im Arbeitsalltag.	„The technology needed to undertake computational journalism is relatively cheap and available and for the most part already exists in newsrooms. The limiting factors are not the technical infrastructure but according to one of the interviewees, ‘time and goodwill’ granted from the editors.“
39	Werkzeuge, Programme, technische Infrastruktur	Infrastruktur(en), Technologien und nichtmenschliche Akteure. Inkludiert Datenlieferanten, Werkzeuge und Quellen.	Codiereinheit = der „ManyEyes“ konkrete „Google Fusion Tables“ Ausdruck. Jede Bezeichnung ist nur einmal pro Dokument zu codieren.
80 Forschungspublikation			Betrifft nur Literaturanalyse.
81	Disziplin, Forschungsgebiete, Theoretische Rahmen, Methodologien	Welcher Disziplin bzw. welchem Forschungsgebiet sind der/die AutorInnen bzw. das den Beitrag veröffentlichende Medium zugehörig? Nicht: Forschungsfokus oder -methode (->082).	„To frame and analyse these findings we are going to develop an understanding of computational journalism as a rhetorical craft, using the Aristotelian concept of techné, building on theory from design, writing and science studies.“ „In this article, following a tradition that considers journalism as a form of knowledge (Park, 1940), we analyze how the contemporary integration of programmers in newsrooms challenges established epistemologies of how data can support investigative journalism.“
82	Forschungsfokus und -methode	Aspekte rund um den durchgeführten Forschungsprozess, wie z.B. Forschungsfragen, Forschungsdesign, Forschungsfokus, Interviewfragen, etc.. Nicht: Theorie, Methodologie (-> 81).	„Drawing on the analysis of a particular investigation performed by a San Francisco-based news organization, this article shows the existence of epistemological tensions in the making of journalistic revelations that involve the

		processing of vast amounts of data.“ „Zu diesem Zweck werden hier zunächst aus den Science and Technology Studies und der Digital-Methods-Forschung heraus die Grundzüge einer datenjournalistischen Methodologie entwickelt.“
84 Forschungslücken	Identifizierte Forschungslücken. Vorschläge für künftige Forschung.	„Our assumption that any number of these observations may be specific to small media markets should be further explored, if only by conducting similar research in neighbouring countries with different market sizes (France, the Netherlands, Germany, or the United Kingdom) that would include larger newsrooms with a priori more resources to devote to data journalism.“ „The study of software design and the use of this software should shed some light on the complexity in which news is created.“
90 Andere Aspekte		
92 Gender	Genderaspekte des DIJ, zum Beispiel Gender Gaps, Seilschaften, fehlende Ausgewogenheit.	„Diversity concerns are particularly relevant in the data journalism space, given the broader issues with women in technology that have become evident in recent years. Online and off, misogyny and discrimination endure in the industry, along with subtler sexism and racism. The challenge that editors face in hiring a diverse team of data journalists is structural, reflecting broader societal issues.“ „And, it should be noted, that in 2009 during the PI’s visit, each of these positions was filled by a male. Later in the year, the department hired its first female employee, a highly experienced coder.“
93 Geschichte	Meilensteine des DIJ. Wichtige Entwicklungsschritte. Projekte und Praktiken aus der Zeit vor 2006.	„In 1972, Burnham from the New York Times analyzed crime reports and arrest statistics from the New York City Police department, revealing discrepancies among the rates of crimes reported in the city and the arrests made

		<p>in the precincts. He pointed out in 1973 that a black person in New York City was 40 times more likely to be murdered than a white person. Here, computer data and statistics were used to reveal issues concerning the public good.“</p> <p>„Another case among many others is Bill Dedman, from The Atlanta Journal-Constitution, who won the Pulitzer Prize in 1989. Combining data from the US Census Bureau and the Federal Financial Institution Examination Council, he revealed racist policies in lending in Atlanta-area financial institutions.“</p>
95	Algorithmen, Big Data	<p>Die Kontextualisierung der Begriffe ‚Algorithmus‘/, ‚Algorithmen‘ und ‚Big Data‘.</p> <p>„First, amid the rise of big data across many fields and industries, there has been a rise in data-driven approaches to journalism that can only be handled by complicated computer algorithms, database processing, and programming languages and interpretation.“</p> <p>„With the growth and spread of data-processing artifacts, several institutions have come to question the grounds of their knowledge production.“</p>
96	Privatsphäre	<p>Wie die Praxis und die Beiträge des DIJ Aspekte von Privatsphäre berühren.</p> <p>„Others said that they believed there are times when data should be withheld in the interest of protecting people’s privacy.“</p>

G = Gruppe, C = Code, DIJ = Datenintensiver Journalismus.

9.2 Interviewleitfaden

9.2.1 Ziele der Datenerhebung

Kenntnis über...

- datenintensive Arbeitsabläufe im Journalismus,
- die redaktionellen Produktionsbedingungen,
- die Kompetenzanforderungen für involvierte Personen,
- die Rolle von nicht-menschlichen Akteuren in diesen Prozessen sowie
- weitere mögliche InterviewpartnerInnen

zu erlangen.

9.2.2 Leitprinzipien

- Dies ist ein generischer Leitfaden, der für jedeN InterviewteN angepasst wurde. Auch die Formulierung der Fragen wurde im Laufe der Zeit verändert. Er dient der Orientierung, ob alle Aspekte abgefragt wurden.
- Der Leitfaden wird variabel gehandhabt. Nachfragen sind möglich und sollten möglichst offen gestellt werden.

9.2.3 Eröffnung des Interviews

- Im Rahmen meines Dissertationsvorhabens untersuche ich Arbeitsabläufe im Datenjournalismus bei verschiedenen Redaktionen. Vielen Dank, dass Sie sich bereit erklärt haben, mir für ein Interview zur Verfügung zu stehen.
- Zur Erleichterung der Analyse möchte ich das Interview gerne aufzeichnen. Die Tonaufnahme und das Transkript, das ich im Anschluss anfertigen werde, sind nur für mich bestimmt. Sind Sie damit einverstanden? Sie können das Interview jederzeit unterbrechen oder beenden. Auch können Sie ‚off the record gehen‘, wenn Sie mir Hintergrundinformationen mitteilen möchten, die nicht veröffentlicht werden sollen.
- Ich würde Sie gerne namentlich zitieren. Vor einer Veröffentlichung bekommen Sie sämtliche Stellen, die Sie betreffen, noch einmal vorgelegt. Sie können sie kommentieren,

oder eine Anonymisierung verlangen. Gerne können wir jederzeit eine komplette Anonymisierung vereinbaren. Ist diese Vorgehensweise in Ordnung für Sie?

9.2.4 Hintergründe zur Person

- Wie sind Sie zum Datenjournalismus gekommen?
- Was ist Ihr fachlicher Hintergrund?
- Was war Ihr erster journalistischer Beitrag, den Sie als Datenjournalismus bezeichnen würden?
- Was war Ihre erste Berührung mit Datenjournalismus?
- Wie sieht Ihr typischer Arbeitstag aus?

9.2.5 Arbeitsabläufe bei der Produktion von datenintensiven Beiträgen

Mich interessieren besonders die Herausforderungen und Schwierigkeiten, wenn man einen datenjournalistischen Beitrag in einer Redaktion oder einer redaktionsähnlichen Konstellation umsetzt. Dazu habe ich mir Fragen überlegt, die verschiedene Ebenen betreffen: die Teamebene, die Organisationsebene, die Methoden- und Technikebene. Ich würde Sie bitten, einerseits diese Fragen aufbauend auf Ihren bisherigen Erfahrungen zu beantworten („so ist es gelaufen“), andererseits gerne auch in eine externe Expertenrolle zu schlüpfen („so sollte es sein“). Bitte bringen Sie gerne Beispiele aus laufenden oder abgeschlossenen Projekten ein – sowohl von Ihnen oder von anderen.

9.2.5.1 Teamebene

- Mit wem arbeiten Sie bei der Produktion von datenintensiven Beiträgen zusammen?
- Wie ist die Gruppe, die datenintensive Projekte umsetzt, zusammengesetzt?
- Welche Kompetenzen haben die einzelnen Personen darin?

—

- Wie sollte ein Team im Idealfall aufgebaut sein?
- Welche Kompetenzen benötigen die einzelnen Mitglieder?
- Können Sie mir die Kompetenzen im Detail beschreiben?

—

- Wie sieht bei einem Projekt die Arbeitsteilung aus?
- Wer macht was?
- Braucht es jemanden für die Koordination?
- Wie finden Sie Ihre Themen?

—

- Wo sitzen Sie in der Redaktion?
- Wo sitzen die TeamkollegInnen?
- Welche Vor- und Nachteile ergeben sich aus der/den von Ihnen geschilderten Position(en)?
- Wo und in welcher Konstellation würde Sie und Ihre KollegInnen im Idealfall sitzen?

—

- Wann und wie oft gibt es Treffen?
- Wie verlaufen diese?
- Wann und wie oft sollte es im Idealfall Treffen geben?
- Gibt es informelle Treffen, bei denen sich Leute austauschen? Was passiert dort?

—

- Wie sieht die technische Infrastruktur aus?
- Gibt es gemeinsame Ressourcen, in denen Inhalte festgehalten werden? Ich denke etwa an eine Tafel in der Redaktion, ein Wiki, in dem Methoden und Tools festgeschrieben sind, oder einen Server, auf dem wichtige Datensets liegen?

9.2.5.2 Organisationsebene

- Wie arbeiten Print und Online bei datenintensiven Projekten zusammen?
- Welche Freiräume erlauben und Vorgaben machen die Leitungspositionen (Chefredaktion, RessortleiterInnen, etc.)?
- Wer sind die größten SkeptikerInnen oder VerhinderInnen von datenjournalistischen Projekten? Wer die größten FörderInnen?

9.2.5.3 Methoden- und Technikebene

- Können Sie mir einen typischen Arbeitsablauf für ein datenjournalistisches Projekt anhand eines Beispiels beschreiben?
- Welche Prozesse nehmen am meisten Zeit im Workflow in Anspruch? Welche Abläufe gehen rasch?
- Was sind auf der technischen Ebene die größten Schwierigkeiten und Herausforderungen?
- Inwieweit werden Erklärungen zur Recherche mitveröffentlicht, und wenn ja bzw. nein: Warum?
- Inwieweit werden Daten und selbst geschriebene Programme gleich mitveröffentlicht, und wenn ja bzw. nein: Warum?
- Welchen Einfluss haben die Werkzeuge und die Software auf die Arbeit und Beiträge?
- Inwieweit unterscheidet sich Datenjournalismus von wissenschaftlichem Arbeiten?

9.2.5.4 Abschluss

- Ich bin mit meinem Fragenkatalog am Ende angelangt. Haben Sie noch eine Frage oder gibt es etwas, das Sie hinzufügen wollen?
- Können Sie mir noch weitere Personen nennen, mit denen es sich zu sprechen lohnt?
- Kann ich mich bei etwaigen Nachfragen noch einmal bei Ihnen melden?
- Vielen Dank für das Gespräch.

9.3 Zusammenfassung

Die zunehmende gesamtgesellschaftliche Orientierung an Daten hat in den vergangenen Jahren auch den Journalismus erfasst. Datenintensiver Journalismus oder Datenjournalismus, hier verstanden als (die manuelle Produktion von) Nachrichten, die überwiegend auf der konzentrierten Sammlung, Analyse und Visualisierung von strukturierten Informationen beruhen, hat seit 2008 eine starke Popularisierung erfahren. Zugleich haben Medienorganisationen Schwierigkeiten, diese journalistische Spezialdisziplin zu etablieren, denn sie erfordert hohe Ressourcen sowie neue Kompetenzen und Abläufe in den Redaktionen. Die Forschung zu Datenjournalismus stellt sich bislang als granular und fragmentiert dar. Obwohl einige der europäischen Datenjournalismus-Leuchttürme in Deutschland und Österreich zu finden sind, ist der deutschsprachige Raum ein weitestgehend blinder Fleck in der Forschungslandschaft.

Vor diesem Hintergrund untersucht die vorliegende Arbeit in einem mehrstufigen Mixed-Methods-Verfahren die redaktionellen Routinen und Produktionsbedingungen der Entstehung von datenjournalistischen Beiträgen. Anschließend an eine wissenschaftstheoretische und methodologische Diskussion geschieht das einerseits mittels einer Forschungssynthese, andererseits durch eine Interviewstudie, ergänzt durch Daten aus (teilnehmenden) Beobachtungen. Die Forschungssynthese, ein systematischer Überblick über empirische Arbeiten von 1996 bis 2015 charakterisiert das Forschungsfeld der ‚data journalism studies‘. Dieses ist geprägt von einem starken Anstieg seit 2010, einer gewissen Armut an sowohl Theorie als auch quantitativen Forschungsdesigns. In der Literatur stellt sich die Kultur, Identität und Praxis des datenintensiven Journalismus quer über Medien und Länder hinweg als relativ homogen dar. Dieser Eindruck wird durch die Interviewstudie bestätigt, die unter anderem auf Redaktionsbesuchen bei mehreren Medien in Deutschland und Österreich beruht. Dort bewegt sich das technologie-spezifizierte Feld zwischen dem Ideal, ein ‚digitaler Wachhund‘ der Demokratie zu sein, und knappen redaktionellen Ressourcen. Die Praxis ist dabei geprägt von ständigem Experimentieren – meist in Teams – und vom Wunsch, das technisch Mögliche immer weiter auszureizen. Die Arbeit schließt mit einer theoretischen Perspektive, die die Praxis des datenintensiven Journalismus im Kontext menschlich-algorithmisch co-konstruierter Öffentlichkeiten und ‚data literacy‘ verortet.

9.4 Abstract

In recent years, society's orientation towards data has affected journalism. Data-intensive journalism or data journalism – here understood as (the manual production of) news based predominantly on the concentrated collection, analysis and visualization of structured information – has become increasingly popular since 2008. Despite this interest, establishing this special journalistic discipline is difficult for media organizations because it requires significant resources, as well as new competencies and practices in the newsroom. To date, research on data journalism has, in general, been granular and fragmented, and within the German-speaking regions there has been almost no research on the subject even though Germany and Austria produce some of the most-acknowledged data journalism in Europe.

Against this background, this work examines the editorial routines and conditions that produce works of data journalism via a multi-stage, mixed-method approach. Following an epistemological and methodological discussion, this is realized with a research synthesis and an interview study, supplemented by data from (participating) observations. The research synthesis describes the field of 'data journalism studies', providing an overview of empirical work from 1996 to 2015. While the number of contributors has increased significantly since 2010, works in the field tend to exhibit a certain lack of both theory and quantitative research designs. The literature has found the culture, identity and practice of data-intensive journalism to be relatively homogeneous across media organizations and countries, an impression confirmed by the interview study. Based on visits to several newsrooms in Germany and Austria, participants in the study described data-intensive journalism as a technologically specific field that oscillates between the ideal of being a 'digital watchdog' for democracy and the reality of competing for scarce editorial resources. The practice is characterized by constant experimentation – mostly in teams – and the desire to continue to extend the technical possibilities. The thesis concludes with a theoretical perspective that locates the practice of data-intensive journalism in the context of human-algorithmically co-constructed publics and 'data literacy'.

