



RAHEL ESTERMANN

46

MIT DATEN SPRECHEN

50

PRAKTIKEN, EXPERTISEN
UND VISUALISIERUNGSMODI
IM DATENJOURNALISMUS

54

55

56

57

58

59

60

61

62

63

64

65

[transcript] DIGITALE SOZIOLOGIE

Aus:

Rahel Estermann

Mit Daten sprechen

Praktiken, Expertisen und Visualisierungsmodi
im Datenjournalismus

Januar 2023, 300 S., kart., 13 SW-Abb., 18 Farbbabb.

44,00 € (DE), 978-3-8376-6590-1

E-Book: kostenlos erhältlich als Open-Access-Publikation

PDF: ISBN 978-3-8394-6590-5

Datenjournalist*innen verwenden für ihre Berichterstattung digitale Daten, welche sie verarbeiten, visualisieren und journalistisch publizieren. Rahel Estermann erarbeitet auf Praxis- und Feldebene, welche Expertisen, Materialitäten, Kulturen und sozialen Prozesse sich im Datenjournalismus verschränken. Datenvisualisierungen stellen sich dabei als gemeinsames Vokabular heraus, welches ermöglicht, in zweifacher Weise *mit den Daten zu sprechen*: einerseits zur Produktion von Erkenntnissen, andererseits für das Darstellen von Erkenntnissen mittels Daten. So werden die Prozesse der Neuaushandlung von epistemologischen und professionellen Kulturen in einer datafizierten Gesellschaft sichtbar.

Rahel Estermann, geb. 1987, ist im Luzerner Kantonsrat, für die Grünen Schweiz und in der Zivilgesellschaft politisch aktiv. Die Beziehung von Digitalisierung und Gesellschaft ist für sie dabei genauso relevant wie in ihrer soziologischen Forschung zu Datenjournalismus und Datenvisualisierungen.

Weitere Informationen und Bestellung unter:
www.transcript-verlag.de/978-3-8376-6590-1

© 2023 transcript Verlag, Bielefeld

Inhalt

Dank	11
Abbildungsverzeichnis	13
1 Einleitung	17
1.1 Beitrag der Arbeit	20
1.2 Beautiful Data, Big Data	23
1.3 Abgrenzungen	25
1.4 Aufbau der Arbeit	26
2 Forschungsdesign	31
2.1 Erkundung des Feldes	31
2.2 Die Vorgehensweise anleitende Konzepte	34
2.2.1 Der abduktive Ansatz	34
2.2.2 Praktiken und Relationen: <i>zooming in and out</i>	36
2.2.3 Relationale Ethnografie	37
2.3 Teamarbeit untersuchen	39
2.3.1 Datenerhebung	40
2.3.2 Datenauswertung	47
2.4 Das Feld analysieren	49
2.4.1 Datenerhebung	49
2.4.2 Datenauswertung	53
2.5 Überblick	54
3 Datenjournalismus: theoretische und empirische Einordnung	57
3.1 Die datenjournalistische Community	58
3.1.1 Hacks und Hackers: »bridge the two worlds«	58
3.1.2 Die »Ankommenshalbstunde«: informell verbinden und kooperieren	60

3.1.3	Bedeutungsvolle Visualisierung: die Rolle eines Design-Studios	62
3.1.4	Flaggschiffe ennet des Atlantiks: über die Grenzen hinweg	63
3.1.5	Datenjournalismus als interpretative Community	65
3.2	Datenjournalismus als versammelte Expertisen	68
3.2.1	Kulturelle Rahmung: Ethos, Mindset und <i>Open</i> -Kultur	68
3.2.2	Werken mit Datensätzen, geometrischen Formen und <i>Tools</i>	77
3.2.3	Quantitativ, qualitativ, iterativ, visualisiert – epistemologische (Ent-)Spannungen	85
3.2.4	Die Expertisen des datenjournalistischen Feldes	91
3.3	Zirkulierende Narrative	95
3.3.1	Narrative über Datenjournalismus	96
3.3.2	Das lokale Editieren von Narrativen	111
3.4	Fazit und aufgeworfene Fragen	113
4	Heuristiken der Aushandlung und Visualisierung	115
4.1	Die »trading zone«	115
4.2	Durch Visualisierungen Daten erkennen und Evidenz befestigen	117
4.2.1	Das Labor als Fabrik der Erkenntnis	118
4.2.2	Visualisierung als Erkenntnismaschinerie des Sehens	120
4.2.3	Visualisierung als Befestigung von Evidenz	127
4.3	Visualisierungen in der <i>trading zone</i> : Kombination der Heuristiken	133
5	Daten erkennen: Visualisierungen als Arbeitsoberfläche	137
5.1	Arbeitspraktiken des Erkennens von Daten	137
5.1.1	Datafizierung: Geschichten in Daten übersetzen	138
5.1.2	Saubere Daten: Granularität, Ordnung, Interoperabilität	140
5.1.3	Geschichten drehen	145
5.1.4	Auswertungen plausibilisieren	158
5.2	Erkenntnisproduktion als Aushandlung	171
5.2.1	Koordinationsmuster in den Arbeitspraktiken	173
5.2.2	<i>Traders I</i>	181
5.2.3	Fazit: koordinierte Expertisen für das Erkennen von Daten	183
5.3	Visuelles Vokabular	185
5.4	Tools	190
5.4.1	Formalisierte Visualisierungssysteme	191
5.4.2	Computational notebooks	195
5.4.3	Papier	196
5.4.4	Fazit	198

5.5	Ideen imitieren	199
5.6	Fazit	202
6	Evidenz befestigen: Visualisierungen als Montage	205
6.1	Arbeitspraktiken der Evidenzbefestigung	205
6.1.1	Hauptaussage hervorheben	206
6.1.2	›Humanizing data‹: Personalisierung und Konkretisierung	216
6.1.3	Datenquellen und Methoden transparent machen	220
6.1.4	Ausdifferenzierte Expertisen ausweisen	223
6.2	Evidenzbefestigung als Aushandlung	227
6.2.1	Koordinationsmuster in den Arbeitspraktiken	227
6.2.2	<i>Traders II</i>	237
6.2.3	Fazit	238
6.3	Leitlinien und <i>Tools</i>	239
6.3.1	<i>Styleguide</i>	240
6.3.2	Q und Datawrapper	242
6.4	Ausweichen auf neuartige Formate	251
6.4.1	Der <i>Explainer</i> : Ereignisse verknüpfen	251
6.4.2	Der <i>Longread</i> : ein Format für Datenvisualisierungen	256
6.5	Fazit	258
7	Schlussbetrachtung: Mit den Daten sprechen	261
8	Literaturverzeichnis	273
9	Anhang	289

1 Einleitung

Etwas zurückhaltend, fast brav hält Charlie Smart den Award in den Händen. Erst als er sieht, dass die weiteren Ausgezeichneten neben ihm auf der Bühne ihren Preis enthusiastisch und mit in die Höhe gereckter Trophäe feiern, drückt sich seine Freude mit einem Lachen aus.¹

Smart befindet sich an jenem Abend des 14. Juni 2019 für seinen damaligen Arbeitgeber, The Pudding, auf der Bühne.² Im Rahmen des Jahrestreffens des Global Editors Network (GEN Summit) vereinigen sich die Chefredaktor*innen wichtiger Medienorganisationen und verleihen (unter anderem) die jährlichen Data Journalism Awards, die wichtigste Auszeichnung des Feldes. Sie werden in einem Dutzend Kategorien vergeben, unter anderem gehen die Preise 2019 an Reuters, die New York Times, La Nación aus Argentinien, den deutschen Tagesspiegel, Vox, AP, Quartz und die bekannte Datenjournalistin und Aktivistin Eva Belmonte. The Pudding gewinnt die Kategorie »Data Journalism website of the year«³ und setzt sich dabei unter anderem gegen Reuters und das Wall Street Journal durch.⁴ Die Jury begründet ihre Entscheidung für The Pudding folgendermassen:

-
- 1 Video-Kurz-Zusammenfassung der GEN Data Journalism Awards 2019: <https://www.youtube.com/watch?v=yjykUEOvoE8> (30.01.2021), Sekunden 0:50 bis 0:56. Smart ist als zweiter von rechts zu sehen.
 - 2 Seit Ende 2019 arbeitet Smart für die New York Times: <https://www.nytimes.com/by/charlie-smart> (30.01.2021)
 - 3 Die Kategorie wird folgendermassen beschrieben: »Best journalism website or section of a website that is focused on data journalism – the collection, presentation, analysis of data and the creation of content around that data and its presentation. [...]« <https://datajournalismawards.org/category/> (14.11.2019, inzwischen nur noch über das Web Archive verfügbar)
 - 4 <https://datajournalismawards.org/2019-shortlist/> (14.11.2019, inzwischen nur noch über das Web Archive verfügbar)

»The site is a remarkable assortment of data visualisations, typified by extraordinary data collection (as in ›Women's pocket sizes‹), sophisticated data processing (as in ›Vogue covers‹) and a design that is both sumptuous (like ›Population mountains‹) and straightforward (like ›American anxieties‹). Taken together, The Pudding sets the high-water mark of the medium.«⁵

Es ist ein in mehrfacher Hinsicht denkwürdiger Abend in Athen. Einerseits ist es, wie sich ein halbes Jahr später herausstellt, der letzte GEN Summit und die achte bleibt somit die letzte Verleihung der Data Journalism Awards: Das GEN hat Finanzierungsprobleme und geht ein.⁶ Andererseits: Charlie Smart mit Award auf der Bühne. Auf den ersten Blick erscheint es nicht erstaunlich, dass The Pudding einen Award holt – 2019 ist The Pudding in der gesamten datenjournalistischen Community bekannt für seine ausgeklügelten Arbeiten. Überraschend ist jedoch, dass es Charlie Smart ist, welcher aus den USA nach Europa an den GEN Summit fliegt. Erst einige Wochen vorher – und befristet auf ein halbes Jahr – hat Smart bei The Pudding zu arbeiten begonnen.⁷ An den in der Jury-Begründung erwähnten Projekten war er nicht beteiligt. Die prägenden Köpfe des Pudding-Teams bleiben zuhause. Stattdessen Smart, der an den Verdiensten wenig beteiligte Neuling und Kurzzeit-Mitarbeiter, inmitten des ›Who is Who‹ des Datenjournalismus.

Auf den zweiten Blick jedoch ist auch schon die Tatsache bemerkenswert, dass The Pudding eine datenjournalistische Auszeichnung erhält. Erst zwei Jahre zuvor, Anfang 2017, hatten einige Mitarbeitende der US-amerikanischen Design-Agentur Polygraph The Pudding ins Leben gerufen. Während sie mit den bezahlten Aufträgen von Polygraph ihr Auskommen finanzieren, sollte The Pudding eine Spielwiese sein, wo sie ihren eigenen Interessen nachgehen, ihre »passion project[s]«⁸ verwirklichen und frei veröffentlichen können. Mit ihren »visual essays«, mit welchen »The Pudding explains ideas debated in culture«⁹, hat sich das Team innerhalb kurzer Zeit zu einer der meistbeachteten Publikationsplattformen der Community gemauert.

5 <https://datajournalismawards.org/2019-winners/> (14.11.2019, inzwischen nur noch über das Web Archive verfügbar)

6 <https://www.globoeditorsnetwork.org/about-us/farewell/> (14.11.2019, inzwischen nur noch über das Web Archive verfügbar)

7 <https://www.linkedin.com/in/charliesmart/> (30.01.2021) Es scheint, dass Smart sein Profil seit Ende 2019 nicht mehr aktualisiert hat.

8 Interview D17

9 <https://pudding.cool/about/#team> (30.01.2021)

Und nun, an jenem 14. Juni 2019 – vor den Augen der globalen Medienelite, welche nach neuem Vertrauen und Geschäftsmodellen in der digitalen Welt sucht –, erhält das Team die Auszeichnung mit der Begründung, die Exzellenz in Sachen Datensammlung, -verarbeitung und -visualisierung zu verkörpern. Ein Team, das im Kern aus nur acht Köpfen besteht. Und vor allem gängige journalistische Prozesse und Konventionen auf den Kopf stellt¹⁰: Es gibt keine Redaktionsräume, alle Teammitglieder arbeiten verstreut im englischen Sprachraum; es gibt keine*n Chefredaktor*in, keine formellen Hierarchien, sondern ein »collective«; die Löhne sind öffentlich; es gibt keinen Termindruck durch Publikationsfristen; auf der Webseite gibt es weder eine Bezahlschranke noch Werbung – stattdessen finanziert sich das Team durch bezahlte Aufträge von Unternehmen an eine eng verflochtene Schwester-Firma¹¹; sämtliche Datensätze und Auswertungsmethoden der Projekte sind auf Github transparent und reproduzierbar, teilweise sogar in langen Youtube-Videos aufgezeichnet; das Team ist offen für Themenvorschläge und Zusammenarbeit mit Interessierten, welche eigene Ideen und Datensätze mitbringen; Geschichten im Ideenstadium finden sich in einem öffentlichen »Backlog« aufgeführt – in der Hoffnung, interessierte Mitstreiter*innen zu finden; Preisverleihungen verfügen über keinen grossen Stellenwert.

In der Episode rund um Charlie Smart und den Award für The Pudding kondensieren sich viele interessante Charakteristika eines Phänomens, das sich seit einem guten Jahrzehnt etabliert hat: Datenjournalismus. The Pudding steht nicht stellvertretend für das gesamte Feld oder für alle Datenjournalismus-Teams, die aufgezählten Eigenschaften treffen auf die meisten anderen Organisationen nur sporadisch oder überhaupt nicht zu. Trotzdem verdeutlicht sich: Datenjournalismus meint zwar auch neue Technologien; Datenjournalismus ist aber vor allem ein soziales und kulturelles Phänomen. Diese Aspekte zu erforschen, insbesondere ihre Interaktion mit neuen Werkzeugen, hat mich in rund dreieinhalb Jahren Forschung angetrieben.

10 Die Quellen für die folgenden Beschreibungen sind die Webseite von The Pudding (<https://pudding.cool/about/>, 30.01.2021), ihr Github-Repositorium (<https://github.com/the-pudding>, 30.01.2021), ihr Youtube-Kanal (<https://www.youtube.com/channel/UCFuV9vMFVluW9CAOdS4Oluw/videos>, 30.01.2021) sowie das Interview D17 mit einem Kern-Mitglied des Teams.

11 Wie erwähnt arbeiten alle Teammitglieder auch für die Schwester-Agentur Polygraph, welche bezahlte Aufträge ausführt. Mehr zur Verflechtung von Polygraph und The Pudding: <https://pudding.cool/faq/#polygraph> (30.01.2021)

1.1 Beitrag der Arbeit

Das initiale Anliegen des Forschungsprojektes bestand darin, das im Laufe der 2010er-Jahre entstandene organisatorische und professionelle Feld des Datenjournalismus zu untersuchen. Der Soziologe Schudson (2000) postuliert drei verschiedene Ansätze, um journalistische Felder und Praktiken zu untersuchen: erstens aus einer politisch-ökonomischen Perspektive; zweitens aus einer (organisations-)soziologischen Perspektive (Routinen, Redaktionsorganisation, Weltbilder der Journalist*innen); drittens aus einer (soziologisch-)kulturellen Perspektive. Die vorliegende Arbeit verbindet soziologische Perspektiven aus der Journalismusforschung und den Science and Technology Studies (STS) und untersucht damit Datenjournalismus in einer Kombination von Schudsons zweitem und drittem Ansatz.

Damit leistet die Dissertation einen Beitrag zu aktuellen Diskussionen in verschiedenen wissenschaftlichen Disziplinen. Sie geht auf soziologische Fragen nach der Datafizierung und Quantifizierung der Gesellschaft ein, indem sie das Zusammenspiel von Technologie und sozialen Strukturen untersucht. Für die STS ist die Diskussion der Wissensbildung mit digitalen Daten als neuartiger Materialität ein Anschluss an die Tradition der Laborstudien und die Überlegungen zur Rolle von Visualisierungen in der Produktion wissenschaftlichen Wissens. Diese Arbeit erweitert die Journalismusforschung um einen Beitrag zu sozialen Prozessen in den Redaktionen, welche die bestehende journalistische Kultur im Zusammenhang mit der Materialität digitaler Daten neu verhandeln: Datengetriebene journalistische Berichterstattung über die Welt entwickelt sich hin zu einem Projekt mit zahlreichen beteiligten Expertisen, einer anspruchsvollen Koordination ebenjener, organisiert durch das Adaptieren wissenschaftlich geprägter Konventionen wie Peer-Review und Reproduzierbarkeit. Die Dissertation bietet durch die Untersuchung der epistemologischen Hintergründe und der Überzeugungskraft von Bildlichkeit aber auch Anschlusspunkte an die generelleren Diskurse der Medien- und Kommunikationswissenschaften, welche die zunehmende Nutzung von Datenvisualisierungen diskutieren.

Diese Beiträge kondensieren im Schema eines datenjournalistischen Dreiecks der Expertisen, deren Aushandlung strukturiert ist durch bestimmte Modi von Datenvisualisierungen. Die vorliegende Arbeit hat mit dem Datenjournalismus ein im digitalen 21. Jahrhundert entstandenes Feld als Fall erforscht, deren Erkenntnisse aber für die genannten wissenschaftlichen Diskurse einen über den Fall hinausweisenden Beitrag leisten.

Die Arbeit untersucht insbesondere zwei Fragestellungen:

a) Warum gibt es den Datenjournalismus?

Beim Vorhaben, das Feld zu definieren, zeigte sich schnell, wie vieldeutig der Begriff *Datenjournalismus* Verwendung findet – sowohl in der Theorie wie in der Empirie. Es stellt sich folgerichtig die Frage: Warum ist von *dem* Datenjournalismus die Rede, warum erscheint dies als relevante soziale Kategorie, obwohl damit eine Vielzahl an Phänomenen und Praktiken verbunden wird? Oder andersherum: Es gilt zu ergründen, was dieses Feld über den Begriff hinaus verbindet – sodass sich eine Community formiert, welche sozial eng verflochten ist und über geteilte epistemologische Vorgehensweisen und kulturelle Vorstellungen verfügt. Und sodass die Diskussion darüber und die Auszeichnung derer, welche *the high-water mark of the medium* setzen, erst möglich ist.

Bisher sind die Deutungsangebote aus der Forschungsliteratur dazu lückenhaft geblieben. Dies hat damit zu tun, dass die Kategorie *des* Datenjournalismus dabei meist vorausgesetzt und definiert, aber selten hinterfragt und im Zusammenspiel seiner Teilelemente untersucht wurde – auch wenn schon mehrfach Autor*innen darauf verwiesen, dass die verschiedenartigen epistemologischen Zugänge und professionellen Kulturen im Datenjournalismus einer ethnografischen Untersuchung bedürfen (siehe beispielsweise Parasie und Dagiral 2012). Die Dissertation nimmt diesen Faden auf und folgt einer ethnografischen Vorgehensweise, die sich über mehrere Ebenen des Sozialen aufspannt – und ihre Interpretationen darauf abstützt, die verschiedenartigen Expertisen der Akteur*innen – Programmierung, Visualisierungen, Journalismus – zu differenzieren und Datenjournalismus als Manifestation ihrer sozio-technischen Aushandlung zu betrachten.

b) Welche Rolle spielen Datenvisualisierungen?

Die datenjournalistische Produktion von Erkenntnissen ist ein oft untersuchter Prozess in der Journalismusforschung. Ein Element blieb dabei aber bisher unterbeleuchtet: Datenvisualisierungen. Ihre Rolle als Darstellungsform ist zwar anerkannt und die vorliegende Arbeit wird diesen Aspekt vertiefen. Gleichzeitig wird sie aufzeigen, dass Datenvisualisierungen auch bereits den Produktionsprozess von Berichterstattung prägen. Denn Datenvisualisierungen spielen an der aus kultureller Sicht entscheidenden Schnittstelle eine Rolle: »[...], the cultural view finds symbolic determinants of news in the relations

between ›facts‹ and symbols« (Schudson 2000:189). Die symbolische Form bestimmt, welche Art von Erkenntnissen der Datenjournalismus überhaupt über die Welt produzieren kann – und wann diese als evident anerkannt sind. Wenn The Pudding von *visual essays* als ihren Produkten spricht, so impliziert dies nicht bloss ein Visualisierungen inkludierendes Präsentationsformat, sondern eine andersartige Produktion der Inhalte.

Für die Untersuchung der Rolle von Datenvisualisierungen kann dabei auf einen grossen Wissensbestand der STS zurückgegriffen werden. Obwohl in den Forschungskontexten der in dieser Arbeit referenzierten STS-Autor*innen Daten in der Regel nicht in digitaler Form vorlagen, bieten deren grundlegende theoretische Überlegungen eine gute Basis für die Einordnung auch von computergestützt produzierten Datenvisualisierungen. Denn die Frage nach dem Verhältnis von Daten, Visualisierungen und Evidenz ist die gleiche, und die STS-Forschung stellte sie in den Achtzigerjahren in den Naturwissenschaften: »But are sense data problematic? Consider that in the natural sciences *evidence appears to be embodied in visibility; in a literal sense, it is embodied in what we can see on a data display*. Thus understood, the notion of evidence is built upon the difference between what one can see and what one may think, or have heard, or believe« [Hervorhebung RE] (Amann und Knorr-Cetina 1988:134).

Es kann kaum überschätzt werden, wie weitreichend die hervorgehobene Überlegung ist: Sie postuliert, dass die Evidenz von Daten gebunden ist an eine Abbildung. Datenvisualisierungen nehmen also nicht nur *eine* Rolle ein in der Produktion von Erkenntnis und Evidenz aus Daten, sondern *die* zentrale. Diese Dissertation wird zeigen, dass Datenvisualisierungen die strukturierende Materialität der Sprache des Datenjournalismus sind. Die Sprachmetapher ist dabei mit einer doppelten Bedeutung versehen: Datenvisualisierungen dienen einerseits dazu, *mit* den Daten zu sprechen – in einem Sinne der Produktion von Erkenntnissen durch die Mathematisierung und Geometrisierung von Datensätzen, welche in der Überblendung mit einem visuellen Vokabular resultiert.¹² Andererseits ist es dank Visualisierungen möglich, *mittels*

12 Die Sprachmetapher wird auch im Diskurs des datenjournalistischen Feldes oft verwendet, wenn von *conversations with data* die Rede ist. So nennt sich der Newsletter von DataJournalism.com, der gemeinsamen Datenjournalismus-Plattform des European Journalism Centre und der Google News Initiative, »conversations with data« (siehe <https://datajournalism.com/read/newsletters>, 19.07.2020). Ebenso wurde die Metapher genutzt, um Konferenzpanels über Datenjournalismus zu betiteln (siehe <https://www.journalismfestival.com/programme/2018/conversations-with-data>, 19.07.2020). Al-

(beziehungsweise anhand von) Daten zu sprechen, um eine Aussage zu treffen – dabei kommt diesen Aussagen aufgrund ihrer visuellen Montage eine besondere Überzeugungskraft als Evidenz zu.

1.2 Beautiful Data, Big Data

Die vorliegende Dissertation betrachtet die zwei Fragestellungen nicht getrennt – im Gegenteil zeigt sich, wie eng die Antworten auf die beiden zusammenhängen. Die Entstehung immer neuer Schnittstellen und Überlappungen durch gesellschaftliche Ausdifferenzierung und Querbezüge (Passoth und Rammert 2016) spiegelt sich im Datenjournalismus in der Verflechtung von Expertisen, unterschiedlichen Narrativen und ausdifferenzierten Übersetzungsskripten ebenjener. Dass diese Diversität trotzdem durch *einen* Begriff gefasst wird, erscheint auf den ersten Blick erstaunlich. Die Arbeit erschließt die Erklärung dafür, gerade indem sie die zweite Fragestellung zu Datenvisualisierungen miteinbezieht. Denn die Verflechtung der Expertisen basiert auf Koordinationsprozessen zwischen ihnen, wofür das Konzept der *trading zone* (Galison 1997) eine analytische Fassung bietet. Es verweist in letzter Konsequenz darauf: Erfolgt die Aushandlung nicht nur zwischen zwei, sondern drei Expertisen, so kann eine neue Kultur mit einer eigenen Sprache entstehen und sich festigen. Diese Arbeit postuliert, dass es an diesem Punkt Datenvisualisierungen sind, welche als Materialität institutionalisierend und strukturierend wirken.

Die Rolle von Datenvisualisierungen als Materialitäten in gesellschaftlichen Prozessen mitzubetrachten erscheint nicht erst, aber gerade im Zeitalter der »Data Revolution« (Kitchin 2014b) angezeigt. Die Referenz auf die visuelle Dimension von Daten ist in den letzten Jahren zweideutig erfolgt. Einerseits in einem direkten, praktischen Sinne die Abbildung von Daten meinent, die als zugehörige Dimension der zunehmenden Quantifizierung aufscheinen (Espeland und Stevens 2008). Andererseits erfolgt die Evaluation des Werts von Daten mittels Verweis auf deren Ästhetik (Mützel et al. 2018). Halpern erklärt:

lerdings verwendet das Feld die Metapher, ohne spezifisch auf die Rolle von Datenvisualisierungen hinzuweisen.

»These narratives come with labels such as ›Beautiful Data‹ and ›Beautiful Evidence‹. Opening with the premise that the web today is above all about the collection of personal data, many data visualization sites and textbooks urge the designers, engineers, and programmers of our future to address the important aesthetic component of making this data useful, which is to say, ›beautiful‹. But data is not always beautiful. It must be crafted and mined to make it valuable and beautiful.« (Halpern 2014:5)

Auch wenn in dieser Arbeit eher die praktische und weniger die metaphori-sche Ebene im Zentrum steht, verweist die zweite darauf: In der ästhetischen Komponente liegt eine grosse Macht, welche auf das Publikum wirkt, indem sie Evidenz erzeugt und die Arbeit zu deren Erzeugung gleichzeitig verbirgt. Diese Überlegung wird sich auch auf einer praktischen Ebene reflektieren, wie diese Arbeit später zeigt.

Über die beiden ausgelegten Forschungsfragen hinaus leistet die Arbeit einen Beitrag zum Verständnis, wie die digitalen Technologien des 21. Jahrhunderts mit gesellschaftlichen Prozessen, insbesondere denjenigen der Erkenntnisproduktion, interagieren. Sie hat dabei den Anspruch, aufzuzeigen, dass die Werkzeuge einhergehen mit neuen sozialen und kulturellen Verflechtungen. Dies zeigt sich im Phänomen des Datenjournalismus, welcher als Teil des Journalismus entstand und diesen gleichzeitig erweiterte.¹³

»I think it's generally good?«, lautet die Antwort, als ich ein Teammitglied von The Pudding frage, weshalb sie alles Mögliche teilen und transparent machen.¹⁴ Das Verpacken der Antwort in eine rhetorische Frage deutet an, dass diese kulturellen Vorstellungen über ›guten Datenjournalismus‹ keine schon seit langem institutionalisierten Handlungsanleitungen, sondern umstrittene Aushandlungsprozesse jüngerer Datums sind. Sie gehen einher mit der Frage, wie die gesellschaftliche Selbstbeobachtung sich in einer zunehmend digitalisierten Welt weiterentwickelt – indem sie auch im Journalismus neue Akteur*innen mit ihren Expertisen versammelt und sich Sinngebungs- und Kommunikationsmuster verändern.

13 Auch wenn das Verhältnis von Datenjournalismus und Journalismus, Ersteres verstanden als Teil des Zweiteren, sowohl in diesen einleitenden Absätzen wie auch im Folgenden immer wieder mitschwingt, so ist dieses Verhältnis doch nicht das Analyseobjekt der Forschung (es wurde schon andernorts eingehend beleuchtet, siehe beispielsweise Hermida und Young 2019).

14 Interview D17

»Journalism is the socially valued and structured work of crafting and distributing socially significant news and discussion«, definiert Vos (2018:4) Journalismus, rudimentär und gekürzt. Als Berichterstatter – und somit Produzent von Erkenntnissen – über die Geschehnisse in der Welt ist der Journalismus von den Digitalisierungsprozessen der letzten Jahrzehnte geprägt: Digitale Daten und computergestützte Auswertungsmethoden, gemeinhin mit dem Schlagwort Big Data versehen, bergen das Potenzial, die journalistischen Erkenntnisprozesse zu verändern. Lewis und Westlund (2015:453) verweisen auf die beiden von Vos erwähnten Aspekte der journalistischen Tätigkeit – »crafting news« und »distributing news« – und folgern mit Bezug auf Big Data: »Big data, as a set of technological processes as well as a key source of power, thus opens new paths for imagining how journalistic investigations develop epistemologically relevant revelations. More broadly, big data has implications for rethinking the epistemologies of news production and news distribution.« Die Dissertation greift durch ihre Forschungsfokuse folgerichtig auch die Fragestellung nach dem epistemologischen Potenzial von Big Data auf: Datenvisualisierungen nehmen genau in den beiden genannten Bereichen – »news production and news distribution« – eine signifikante Rolle ein, welche diese Dissertation untersucht und einordnet. Dabei ermöglicht es der Fokus auf das Phänomen des Datenjournalismus, aufzuzeigen, wie ein Feld von sozialen und kulturellen Strukturen – wie der Community und dem visuellen Vokabular – erst stabilisiert und neue Standards – wie beispielsweise das Teilen oder Peer-Review-Prozesse – institutionalisiert werden. Und diese dann erklären, wieso ein Team wie The Pudding die Erfolgsgeschichten eines Feldes erzählt, das nach »gutem Datenjournalismus« strebt.

1.3 Abgrenzungen

Visualisierungen im Journalismus können sehr viele verschiedene Formate annehmen. Die Überlegungen dieser Arbeit beziehen sich auf »wissenschaftlich-technische Bilder« (Burri 2008:350), das heisst die (interaktive oder nicht-interaktive) Visualisierung von Daten, gespeichert in Datenbanken, durch kalkulierende Instrumente – seien dies Programmiersprachen oder spezifische Visualisierungssoftwares.

Die Dissertation hat nicht weiter die Rolle von anderen Formen von Visualisierungen im Journalismus untersucht – zum Beispiel Infografiken, also gezeichnete visuelle Elemente, welche bereits seit Jahrzehnten im Journalis-

mus verbreitet sind, aber keine Datenbanken als Grundlage und Kalkulationen in der Herstellung benötigen. Weitere visuelle Formate, die ich ausgeklammert habe, sind Fotos, Illustrationen oder Karikaturen. Auch der genuin visuelle Journalismus des Fernsehens und der Einsatz von Videos gehören nicht zum Untersuchungsfeld. Auch wenn diese Formate ab und zu auch im Rahmen von datenjournalistischen Praktiken zum Einsatz kommen – immerhin gehören auch einige der bekanntesten datenjournalistischen Teams zu öffentlich-rechtlichen (Fernseh-)Anstalten –, so sind doch ihre technischen und epistemologischen Dimensionen sehr verschieden (Wintsch 2006). Dies spiegelt sich auch darin, dass Videos im Rahmen datenjournalistischer Projekte die digitalen Daten meist doch wieder in Formen von Diagrammen oder Karten integrieren.¹⁵

1.4 Aufbau der Arbeit

Im Aufbau der Arbeit widerspiegelt sich das induktive, felderkundende und theoriegenerierende Vorgehen dieser Dissertation. Das Phänomen des Datenjournalismus erforschte ich theoretisch imprägniert, aber stark von der Empirie geprägt. Deshalb folgt gleich nach dieser Einleitung das Forschungsdesign (Kapitel 2), um den iterativen Prozess auszulegen, der die Arbeit rahmt. Das Design zeigt auf, dass ich die Forschung auf verschiedenen Ebenen und mit verschiedenen Fokussen gleichzeitig betrieb. Dabei leiteten mich drei theoretisch-methodische Konzepte an: der abduktive Ansatz, das Changieren der Fokusse zwischen Praktiken und Relationen (*zooming in and out*) sowie die relationale Ethnografie als methodisches Konzept, um diese beiden Ebenen zu integrieren. Kapitel 2 erläutert detailliert mein Vorgehen sowie die Quellen, aus welchen ich meine empirischen Daten gewann. Dabei rückt es die ethnografische Untersuchung von zwei datenjournalistischen Teams – NZZ Storytelling und BR Data – ins Zentrum.

Kapitel 3 nimmt eine Einordnung des datenjournalistischen Feldes anhand dreier soziologischer Konzepte vor: Community, Expertisen und Narrative. Das Skizzieren der lokalen Community (3.1) verweist auf eine Gruppierung, welche sich unter dem Dach von Hacks/Hackers regelmässig trifft und dabei

15 Beispielhaft dieses Video des BR, welches auf der Datenauswertung des Datenjournalismus-Teams aufbaut: <https://www.br.de/mediathek/video/das-verbaute-land-alt-ernativen-zum-neubaugebiet-av:5abc23cb7659070018e784bo> (25.11.2020)

enge informelle Beziehungen in lockerer Atmosphäre aufbaut. Dabei spielen Visualisierungsaspekte, insbesondere ein Design-Studio, eine grosse Rolle – genauso wie die internationale Vernetzung und Orientierung, häufig über Twitter, und die omnipräsenten Flaggschiff-Medienorganisationen aus dem englischen Sprachraum. Der Abschnitt beschreibt, wie diese *interpretative community* Datenjournalismus trägt, indem sie die Sinnggebung teilt und über gemeinsam ausgehandelte Bewertungsstandards verfügt. Und dass die Community dafür gemeinsame Kanäle und Plattformen des Austauschs rege pflegt.

Dann zeige ich beschreibend die Diversität und Vielschichtigkeit des Feldes auf (3.2): Im Datenjournalismus versammeln sich drei Expertisen – Visualisierung, Journalismus, Programmierung – rund um drei verschiedene Dimensionen. Erstens die professionelle Kultur: ein Ethos des gemeinsamen Lernens, den die Datenjournalist*innen teilen; eine Ausbildung, die zu einem *Mindset* befähigen soll, um in der Community mitsprechen zu können; und der Spirit der *Open*-Kultur mit Werten wie Partizipation, Bastelei, Teilen und Transparenz. Zweitens das Werken mit verschiedenen Materialitäten: mit strukturierten und interoperablen Datensätzen; mit einem visuellen Vokabular aus einer Palette von geometrischen Formen, von einfachen Diagrammen bis hin zu *Data Art*; mit computergestützten *Tools*, welche die Auswertungen möglichst reproduzierbar und transparent machen. Und drittens Spannungen, welche sich zwischen den epistemologischen Grundsätzen des konventionellen Journalismus und den quantitativen Verfahren des Datenjournalismus ergeben – und die mittels iterativer Vorgehensweisen (*data-driven vs. story-driven*) sowie durch Datenvisualisierungen (teilweise) entspannt werden. Diese drei Dimensionen, eine Triade, nuancieren das Expertisen-Dreieck von Journalismus, Visualisierung und Programmierung. Das Dreieck der Triaden ordnet die empirischen Beobachtungen meiner Forschung, insbesondere der Teams, ein.

Danach zeichne ich die Narrative des datenjournalistischen Feldes nach (3.3). Ihre Unterschiedlichkeit ist sehr auffällig, sie schliessen sich aber weder auf der Feld- noch auf der Teamebene aus: Es kann beiderorts sehr verschiedenartige und gleichzeitig zirkulierende Narrative und Erfolgsgeschichten geben. Das Konzept des Narrativs baut dabei auf den theoretischen Überlegungen aus der Organisationssoziologie auf (Sahlin-Andersson 1996), welche darauf hinweisen, dass auf lokaler Ebene die Narrative editiert (und nicht kopiert) werden. Das heisst, die strukturellen und institutionellen Gegebenheiten von Datenjournalismus-Teams prägen, welche und wie sie die Narrative in eige-

ne datenjournalistische Praktiken umsetzen. Zum Schluss des dritten Kapitels (3.4) rolle ich aus der Forschung auf Feld-Ebene relevante Fragen für die beiden zentralen und ausführlichen Kapitel mit den Team-Ethnografien aus – es sind die beiden bereits hier in der Einleitung erwähnten Fragestellungen nach der Persistenz der vielfältigen sozialen Kategorie Datenjournalismus sowie der Rolle von Datenvisualisierungen.

Das Kapitel 4 greift die aufgeworfenen Fragen auf und entwickelt sie theoretisch weiter – mit dem Ziel, Gerüste zu entwickeln, mit welchen die Team-Ethnografien die Fragen zu beantworten vermögen. Dabei sind theoretische Überlegungen aus den STS, der Wissenschaftsforschung und ihrer Laborstudien, der Anthropologie und den Medienwissenschaften zentral: Die Heuristiken der *trading zone*, von *seeing data* und *becoming evident* kombiniere ich schliesslich zu zwei Modi der Visualisierung, welche Aushandlungen von Expertisen strukturieren.

Die beiden Kapitel 5 und 6 bilden das Herzstück dieser Arbeit. Sie untersuchen in den zwei datenjournalistischen Teams die zwei Modi: die Produktion von Erkenntnis auf Datenvisualisierungen als Arbeitsoberfläche (5) respektive die Befestigung von Evidenz auf Datenvisualisierungen als Montage (6). Dabei lassen sich die Aushandlungen von Expertisen in Diskussionen, aber auch Werkzeugen und Formaten beobachten. Je nach Visualisierungsmodus stehen dabei andere Konstellationen und Verschränkungen im Vordergrund. Der ethnografische Ansatz ermöglicht mir, mit einer reichen Beschreibung des datenjournalistischen Alltags die Strukturen der Praktiken und die subtilen Verbindungen zur Community, ihrer Kultur, ihren Werkzeugen und den Narrativen der Feld-Ebene zu erschliessen.

In Abschnitt 5.1 beschreibe ich zuerst vier zentrale Praktiken zur Produktion von Erkenntnis im Datenjournalismus: die Datafizierung, welche Geschichten in Datensätze übersetzt; die Herstellung sauberer Daten; das Drehen von Geschichten rund um die durch Visualisierungen produzierten Muster; das Plausibilisieren der Auswertungen durch Samples, Einzelfälle und Vergleiche, welche die qualitativen und quantitativen Welten passend machen. Die folgenden Abschnitte analysieren diese Praktiken dann als Aushandlungen zwischen den Expertisen – zuerst in Diskussionen (5.2), in welchen die *traders* eine koordinierende Rolle spielen; dann in Relationierungen durch das visuelle Vokabular (5.3). Abschnitt 5.4 beleuchtet dann, wie sich diese Koordinationsprozesse in Werkzeugen reflektieren: in formalisierten Visualisierungssystemen wie beispielsweise der Programmiersprache R, in *computational notebooks*, aber auch auf Papier. Schliesslich weist 5.5 darauf

hin, wie das Imitieren von Ideen aus der Community – Erfolgsgeschichten, Projekte – direkt datenjournalistische Praktiken in den Teams anleitet.

Kapitel 6 folgt dann demselben Aufbau, bezieht ihn aber auf den Modus der Evidenzbefestigung. Abschnitt 6.1 beschreibt die Praktiken der Hervorhebung der Hauptaussage durch Leseführung sowie den Vergleich; die Praktik des ›humanizing data‹ durch Personalisierung und Konkretisierung; Transparenzpraktiken rund um Datenquellen sowie Methoden und Visualisierungen; das differenzierte Ausweisen von Expertisen in den Autor*innen-Angaben. Wie im ersten Modus koordinieren die Teams diese Praktiken in Aushandlungen durch Diskussionen, wieder vermögen die *traders* eine wichtige Stellung einzunehmen (siehe Abschnitt 6.2). Die Aushandlungen spiegeln sich in Leitlinien und Werkzeugen (Abschnitt 6.3), welche die Teams nutzen – und institutionalisieren sich insbesondere in datenjournalistischer Visualisierungssoftware wie Q und Datawrapper, welche sich an ein Publikum ausserhalb des eigenen Feldes richtet. Das Ausweichen auf neuartige Formate (Abschnitt 6.4) durch die beiden Teams weist dann nochmals auf die materielle, epistemologische und kulturelle Abgrenzung datenjournalistischer Praktiken in den Redaktionen hin.

Die Schlussbetrachtung in Kapitel 7 webt zuerst den roten Faden des Forschungsvorhabens nochmals durch die Kapitel hindurch, um dann darauf aufbauend Datenjournalismus als Kreol-Sprache zu umreissen. Sie rahmt, wie Datenjournalist*innen anhand der zwei Visualisierungsmodi *mit* den und *mittels* der Daten sprechen. Zum Schluss ordne ich die Erkenntnisse dieser Arbeit in einen weiteren Forschungskontext ein und weise auf offene Fragen hin, welche als Desiderat meiner Dissertation verbleiben.