

KAPITEL 2

Datenexzensionalismus

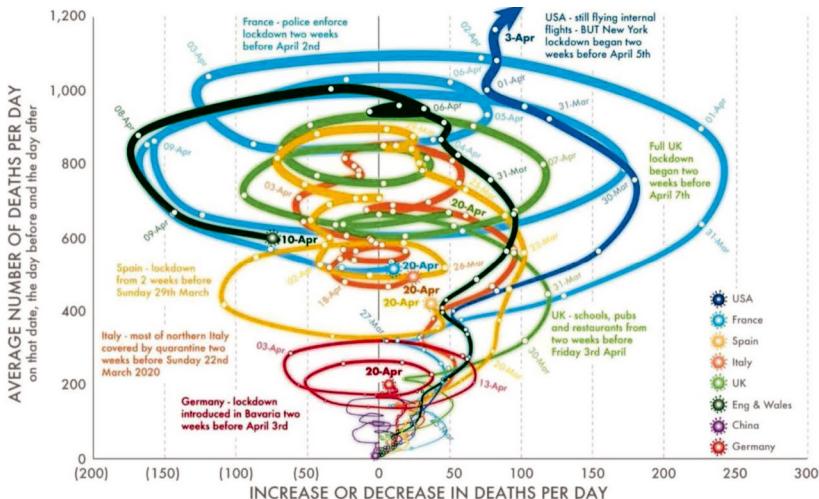
1. Daten als Ausnahmeerscheinung

Nachdem im vorangegangen Kapitel durch die Denkfigur des Scheiterns vier Ansatzpunkte – Datenzentrismus, Unbeweglichkeit, Anthropozentrismus und Skalierungsverweigerung – für das Beschreiben der konzeptionellen Schwächen der Datenvisualisierungspraxis gefunden wurden, setzen sich die zwei folgenden Kapitel genauer mit den grundsätzlichen Konzeptionen um Datenvisualisierung auseinander. Bereits das Wortkomposit Datenvisualisierungen setzt zwei Konzeptionen voraus: Daten und Visualisierung. Ganz grundsätzlich folgen Datenvisualisierungen ihrem Namen nach einer bestimmten Idee von Daten als Gegenstand ihrer Darstellung einerseits und Visualisierung als eine Beschreibung des vollzogenen gestalterischen Prozesses andererseits. In diesem Kapitel liegt der Schwerpunkt zunächst auf der Einordnung der vorliegenden Abhängigkeit von Daten.

Mit Abhängigkeit meine ich die vielen Gestaltungsprozesse, die an der Datengeneration beteiligt sind. Ich denke dabei beispielsweise an die Entstehung der nationalen Fallzahlen von Coronavirus-Infektionen. Zunächst braucht es eine konkrete Vorstellung von einem Gegenstand, um ihn methodisch erfassen zu können. Das Coronavirus „SARS-CoV-2“ wurde erstmals im Dezember 2019 sequenziert. Zur quantitativen Erfassung der Fallzahlen braucht es dann skalierbare Nachweismethoden, die vor allem in PCR- und Antigen-Schnelltests (eine erste Datenvisualisierung an sich) ihre verbreitete Anwendung gefunden haben. Diese Nachweise wiederum werden durch eine Meldung statistisch formalisiert. Entweder das Testlabor oder die infizierte Person selbst kontaktiert nach einem positiven Befund das örtliche Gesundheitsamt. Die Landesbehörden kommunizieren dann, im Fall von Deutschland, an das Robert Koch-Institut, wo alle Falldaten zentral zusammenkommen. Jedoch werden nicht alle wirklich infizierten Personen erkannt, sodass die Dunkelziffer dementsprechend einem Vielfachen der gemessenen Fallzahlen entspricht.¹³⁶ Ein weiteres Problem ist, dass sich durch die Meldeketten teilweise mehrtägige Verzögerun-

¹³⁶ Basierend auf den Studienergebnissen wird geschätzt, dass es etwa 1,8-mal so viele SARS-CoV-2-Infizierte in Deutschland gab, wie für diesen Zeitraum nach Infektionsschutzgesetz an die Gesundheitsämter gemeldet wurden. Vgl. Robert Koch-Institut, 2021.

gen einstellen, weswegen Datenjournalist:innen entschieden haben auch alternative Fallzahlenermittlungen, wie etwa die der Johns-Hopkins-Universität, zu berücksichtigen. Die Covid-Falldaten sind demnach ein Konstrukt, dass einem komplexen Phänomen einen numerischen Wert gibt, dabei aber auch Kompromisse eingeht.



Wie kurz illustriert, ist es nicht selbstverständlich Daten vorzufinden, zu prozessieren und zu visualisieren. Schon die primäre Idee von Daten als Analysegegenstand begründet sich nicht aus sich selbst heraus. Es braucht eine konkrete Absicht oder eine vorliegenden Motivation, die nötig ist, um ein Interesse an Daten zu entwickeln. Dazu kommt, dass Daten nicht natürlich vorkommen, sondern erst in verschiedenen Arten und Weisen hergestellt bzw. zur Weiterbearbeitung aufbereitet werden müssen. Es bedarf oft vielzähliger, mühsamer und teils brachialer Transformationsprozesse, um einen Datensatz herzustellen bzw. zu bearbeiten. Sowohl das Interesse an, das Vorkommen von und die Aufbereitung der Daten ist durchgängig künstlich. Das ist an sich keine besondere Neugkeit.¹³⁷ Es zeigt aber, dass mit dem Konzept der Daten auch eine bestimmte Erwartungshaltung verbunden ist, die die Arbeit mit und an Daten legitimiert. Anders formuliert: Es werden Daten generiert und bestehende Daten analysiert, weil man sich etwas von ihnen erhofft.

Abb. 20 „Phase-portrait diagrams showing mortality rates of Covid-19 virus“; McClure, Kirsten, 2021.

¹³⁷ Vgl. Forschungsansätze innerhalb der „Science and Technology Studies“ (STS), beispielsweise Latour, 2002 oder Daston und Galison, 2007.

Daten sind aufgrund ihrer strukturellen Form sehr vielseitig und werden so in einer Vielzahl von Auswertungs- und Anwendungskontexten eingesetzt.¹³⁸ Datensätzen können sehr klein sein. Beispielsweise ist ein Geburtstag ein Datenwert in einer Jahresordnung. Sie können aber aber sehr groß sein. So benötigte es zur Berechnung des ersten Datenbildes eines schwarzen Lochs 5 Petabyte (5000 TB oder 5.000.000 GB) an Daten. Daten können einen relativ überschaubaren Gegenstand abbilden, etwa Bücher in einem Regal. Sie werden aber auch dazu benutzt, um vollkommen ästhetisch unzugängliche Gegenstände, wie etwa das Klima, zu abstrahieren. Es gibt Datensätze zu einzelnen Individuen, bspw. durch medizinische Untersuchungen. In meiner Arbeit im britischen Statistikamt prozessierte ich hingegen nationale bis über nationale Datensätze, bspw. zu Sterbe- oder Geburtenraten. Letztlich wird beinahe gesellschaftliche Bereich durch Daten abstrahiert. In epistemischen, medizinischen, ökonomischen, ökologischen, soziopolitischen, technischen oder rein individuellen Interessensfelder finden Daten statt.



Mich interessiert jedoch weniger die artentypische Vielfalt und Anwendung, sondern die zuvor angedeutete Ebene der Erwartungen, Vorstellungen und Überzeugungen von Visualisierer:innen gegenüber Daten. Ich möchte im Kern überprüfen, inwiefern sich kulturelle Modelle von Daten in die Praxis mit Daten einschreiben. Wer spricht wie über Daten? Meine Vermutung ist, dass die Art und Weise, wie über Daten gedacht bzw. gesprochen wird, beeinflusst, inwiefern Daten als kulturelle Artefakte wahrgenommen,

Abb. 21 Datenanalystin Katie Bouman mit Festplatten, die die fünf Petabyte des datengenerierte Bild eines schwarzen Lochs beinhalten, 2019.

¹³⁸ Später werde ich diese Eigenschaft von Daten als „Egalität“ definieren, siehe Kapitel 2: 3.1.

bearbeitet und hinterfragt werden.¹³⁹ Als einen ersten Hinweis darauf habe ich im vorangegangen Kapitel der Datenzentrismus als eine Dimension des Scheiterns vorgeschlagen. Durch die bisherige Betrachtung ließ sich mit dem Datenzentrismus eine Praxis ableiten, die allein Datenartefakte in das Zentrum einer Betrachtung stellt und die kontextuellen Bedingungen ihrer künstlichen Konstruktion ignoriert. Die zugrunde liegenden Beweggründe oder Modelle zu hinterfragen, ist dem Datenzentrismus unzugänglich. Für die Denkfigur der Zentrierung möchte ich im folgenden Kapitel eine Alternative entwerfen. Zwar kann ich mir durch die Zentrierung erklären, dass Daten eine gewisse Sonderstellung, nämlich den Hauptfokus, erhalten, jedoch weniger, wodurch sich diese Zentrierung begründet. Mit dem Datenzentrismus können die strukturellen Folgen einer zentrierten Idee von Daten zwar skizziert, aber weniger die Motivation für diesen Datenzentrismus abgeleitet werden.

Im Folgenden führe ich dazu das Konzept des Datenexzeptionalismus ein. Der Exzeptionalismus verweist auf das Narrativ der Ausnahme und der außerordentlichen Erscheinung.¹⁴⁰ Im Bezug auf Daten forme ich die Figur des Datenexzeptionalismus, die es mir ermöglicht Vorstellungen der Besonderheit von Daten bis hin zu einer Überlegenheit der Daten (Datenhybris) offenzulegen. Im Datenexzeptionalismus werden Daten eine zentrale Erscheinung einer zunehmend technologisierten Gesellschaft verstanden. Datenexzeptionalist:innen verleihen Daten dabei eine solche Relevanz, dass das Hauptaugenmerk verschiedener Wissens- und Gesellschaftsbereiche der Anreicherung, Auswertung und Synthese von Daten gilt. Daten gelten darin als eine solche Ausnahmeerscheinung, dass jegliche, auch bildgebende, Kapazitäten aufgewandt werden, um scheinbar veränderte Datensituation zu nutzen. Eine Überhöhung die Folgen hat.

Mein Konzept des Datenexzeptionalismus schärfe ich in den folgenden Unterkapiteln in drei Aspekten: Datennarrative, Datenaffinität und Datenkritik. Zuerst ist der Datenexzeptionalismus ein Phänomen, welcher vor allem durch Narrative geprägt

¹³⁹ Damit folgt dieses Kapitel zunächst einem sozialkonstruktivistischen Schwerpunkt, wie er beispielsweise innerhalb der „Science and Technology Studies“ herausgebildet wurde. Vgl. Potthast, 2017.

¹⁴⁰ Beispiele für die bisherige Wortverwendung beim nationalen (Fluck, 2016) und sexuellen Exzeptionalismus (Dietze, 2019), aber auch schon im technischen Bereich bei Sascha Lobos Internet-Exzeptionalismus (Lobo, 2013).

wird. Es wird viel über Daten gesprochen, aber der Datenbegriff an sich bleibt bemerkenswert leer. In Fallbeispielen zeige ich, inwiefern die Sprache mit und um Daten auch die Praxis mit ihnen prägt. Zweitens ist es die Datenaffinität, die das Anspruchsdenken gegenüber Daten begründet. Doch woher kommt diese Neigung? Eine Antwort suche ich in der Erweiterung des Datenbegriffs weniger als computertechnische, sondern gesellschaftliche Kategorie. In diesem Unterkapitel führe ich auch meine eigene Datendefinition ein.¹⁴¹ Der letzte Aspekte der Datenkritik zeigt, dass sich selbst in Diskursen, die sich kritisch mit Datenphänomenen auseinander setzen, sich das Konzept der Datenexzceptionalismus als Beanspruchung eines Ausnahmezustandes wiederfindet. Der Dreiklang aus Datennarrativen, Datenaffinität und Datenkritik führt mich letztlich zur Beschreibung des Datenexzceptionalismus als Modell. Die zentrale Position dieses Kapitels ist, dass Daten in ihren Erwartungen und Prämissen vor allem durch eine Intention modelliert sind. Daten an sich, aber auch ihre Beschreibung als Ausnahmeherrscheinung sind gewollt. Daten passieren nicht einfach, sondern sie sind das Produkt einer spezifischen Absicht. Diese Absicht zu verschleiern ist der Kern des Datenexzceptionalismus als eine Haltung über und mit Daten zu denken.

¹⁴¹ Siehe Kapitel 3: 3.1.

2. Datennarrative

Wie wird über Daten gesprochen?

Zunächst gilt es sich dafür die zeitgenössischen Datennarrative im Kontext der Datenvisualisierung zu vergegenwärtigen. Dabei drücken Narrative über Daten vielmehr eine Hoffnung gegenüber der Datenverwendung aus, als dass sie Datenstrukturen an sich adressieren. Im vorangegangen Kapitel wurde dazu bereits in das Motiv Big Data als ein prägendes Datennarrativ eingeführt. Als Sammelbegriff für verschiedene technologische Interventionen mit Daten zu arbeiten einerseits, verweist die auf menschliche Dimensionen reduzierte Beschreibung *Big* seit 2012 andererseits auf eine kulturelle Idee im Umgang mit Daten.¹⁴² Datenmengen werden als groß bzw. zu groß bezeichnet, weil sie durch ihr vermehrtes und übermäßiges Auftreten vorherige Konventionen, Maßnahmen und Annahmen übertreffen. Sie werden beschrieben als zu viel, zu komplex oder zu variabel, als dass sie vom Menschen oder von bisherigen Methoden konzipiert und verarbeitet werden könnten. Daten erscheinen dadurch im Übermaß.

Weitere Narrative über die Besonderheit der Daten, zeigen sich beispielhaft auch in konkreten konzeptionellen Veränderungen innerhalb etablierter Fachdisziplinen. Durch die weitreichende Verbreitung von Datentechnik und den daraus resultierenden Daten können verschiedene Gegenstände mit statistischen Mitteln analysiert werden. In einigen Wissenschaften führte das zu fundamentalen Neuausrichtungen, die sich an Konzepten und Methoden der angewandten Mathematik oder Informatik orientieren.¹⁴³ Dadurch entstanden neue Zweige von etablierten Wissenschaftsdisziplinen, wie die Bioinformatik, die Computerlinguistik oder die Computational Social Sciences, die schon in ihrer Namensgebung klar ihr Abhängigkeitsverhältnis angeben.¹⁴⁴

Die Idee des Datenexzceptionalismus – Daten als Ausnahmeerscheinung – skizziere ich im Folgenden durch prominente Narrative innerhalb spezifischer Fachdisziplinen. Diese Disziplinen eint eine Ausrichtung auf die Vermehrung und Verfügbarkeit von Datenartefakten. Ich möchte anhand von drei Beispielen zeigen,

¹⁴² Lohr, 2012.

¹⁴³ Geiselberger und Moorstedt, 2013.

¹⁴⁴ Lazer et al., 2009.

wie sich aus disziplinären Entwicklungen Datennarrative der Ausnahme ableiten lassen. Näher betrachte ich dafür den Datenjournalismus, die Digital Humanities und die Data Science. Es gibt zahlreiche andere Disziplinen, deren konzeptionelle Entwicklung durch Datenphänomene beeinflusst wurde. Wahrscheinlich liegt die Schwierigkeit vielmehr darin eine Disziplin zu finden, die nicht durch Daten in irgendeiner Weise beeinflusst worden ist. Ich wähle jedoch diese drei Beispiele, da ich in meiner eigenen Laufbahn mit Datenjournalist:innen, *digitalen* Geisteswissenschaftler:innen und Data Scientists gearbeitet habe und dementsprechende Erfahrungen berufen kann. In den drei Ausrichtungen weise ich drei weitere Datennarrative – *Datengetriebenheit*, *Digitalisierung* und das *Lernen von Daten* – nach.

2.1 Im Datenjournalismus

Symptomatisch wird die Neuorientierung in der Datenkondition oft durch die Hinzufügung des Substantivs Daten zum Namen der eigentlichen Disziplin. So auch im Falle des Datenjournalismus, welcher als ein erstes Beispiel dienen soll. Innerhalb einer journalistischen Traditionen wird sich dabei mit dem Zugänglichmachen und der verständlichen (teilweise visuellen) Aufbereitung von Datenartefakten beschäftigt.¹⁴⁵ Im Datenjournalismus zeigt sich insbesondere das Narrativ der *Datengetriebenheit*.

Zunächst ist der Begriff des Datenjournalismus ein Phänomen des 21. Jahrhunderts, obwohl sich je nach Datendefinition eine längere Historie des journalistischen Arbeitens mit Daten aufzeigen lässt. Wenn die Erscheinung von Daten mit Daten-technik gleichgesetzt wird, lassen sich die ersten Tendenzen eines sogenannten „computer-assisted reporting“ mit der Etablierung von Computertechnologie auf die 1950er Jahre datieren.¹⁴⁶ In den darauffolgenden konzeptionellen Neuausrichtungen des Journalismus in den 1960er und 1970er Jahren wurden dem „new journalism“ mit seiner literarischen Erweiterung der journalistischen Methoden vermeintlich wissenschaftlichere Ansätze in der Suche

¹⁴⁵ Errea, 2017.

¹⁴⁶ Cox, 2000.

nach Wahrheit und Objektivität entgegengesetzt.¹⁴⁷ Philip Meyers Konzept des „precision journalism“ steht Pate für diesen Ansatz.¹⁴⁸ Nach Dekaden der journalistischen Arbeit mit Computern wurde ab 2009 der Begriff „data-driven journalism“ vor allem durch den „datablog“ des Guardians populär gemacht, indem sich dort auf die effizientere Verarbeitung durch verbesserte Computertechnologie und die Verfügbarkeit von Webtechnologie berufen wird.¹⁴⁹ Das Narrativ der *Datengetriebenheit* des Journalismus findet sich demnach erst in den späten 2000er Jahren ein. In den 2010er Jahren wurde die nun als Datenjournalismus benannte Datenpraktik, zum Beispiel durch die Etablierung von Datenjournalismus-Teams innerhalb von Newsrooms, wie beim genannten Guardian, der New York Times oder in Deutschland bei ZEIT Online, weiter gesellschaftlich etabliert und professionalisiert.¹⁵⁰ Mit den Entwicklungen um die Covid-19-Pandemie entwickelte sich der Datenjournalismus durch die *datengetriebene* Nachrichtenmeldung über Inzidenzwerten, Übersterblichkeit und Impfquoten zu einem zentralen Element der täglichen Berichterstattungen.¹⁵¹

Anhand dieses sehr verkürzten historischen Abrisses wurde deutlich, dass die Wortschöpfung des „data-driven journalism“ bzw. „Datenjournalismus“ nicht im Zuge der ursprünglichen Entwicklung der Datentechnik entstanden ist und somit keine Veränderung der journalistischen Arbeit durch Computertechnologie an sich markieren will. Vielmehr scheinen die Begriffe eine Reaktion auf die veränderte Verfügbarkeit, Menge und Geschwindigkeit von Daten durch die Entwicklungen innerhalb der Informations- und Kommunikationstechnologie zu sein. Es soll nicht abgestritten werden, dass der Journalismus durch datensteigernde Phänomene, wie etwa die schnelle Adaption von Smartphones,¹⁵² die Verbreitung von Social Media¹⁵³ oder die Entwicklung von IoT-Geräten,¹⁵⁴ auch qualitative Änderungen erfährt. Es bedarf allerdings einer

¹⁴⁷ Gray, Chambers und Bounegru, 2012, S. 35.

¹⁴⁸ Meyer, 2002.

¹⁴⁹ Rogers, 2011.

¹⁵⁰ Gray und Bounegru, 2019, S. 4.

¹⁵¹ Mau, 2020.

¹⁵² 14 Jahre nach Einführung des iPhones gibt es nun über 1 Milliarde aktive Geräte. Vgl. Kastrenakes, 2021.

¹⁵³ Mehr als die Hälfte der Weltbevölkerung benutzt Social-Media-Awendungen, siehe Kemp, 2020.

¹⁵⁴ Dignan, 2019.

spezifische Vorstellung, diese Datenmengen dann auch für journalistische Zwecke analysieren zu wollen. Dass große Datenmengen das journalistische Arbeiten verändern, ist keine rein technische Konsequenz. Der Begriff des Datenjournalismus verweist in dieser Lesart weniger auf eine technische, sondern eine kulturelle Änderung der journalistischen Arbeit durch das Versprechen der Daten:

„Information asymmetry – not the lack of information, but the inability to take in and process it with the speed and volume it comes to us – is one of the most significant problem that citizens face in making choices about how to live their lives. [...] Good data journalism helps to combat information asymmetry.“

– Gray, Chambers und Bounegru, 2012, S. 23.

„There is a promise in data, and this is what excites newsrooms, making them look for a new type reporter.“

– Gray, Chambers und Bounegru, 2012, S. 20.

„Data-driven journalism is the future.“

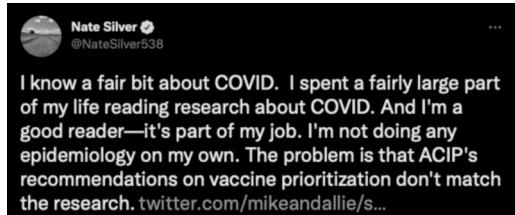
– Tim Berners-Lee zitiert in Gray, Chambers und Bounegru, 2012, S. 22.

Es scheint also vielversprechend sich auf die Datennarrative im Datenjournalismus zu konzentrieren, um nachvollziehen zu können, wie sich konzeptionelle Schwerpunkte verschieben. In ihrem Beitrag zu einem Standardwerk des Datenjournalismus, dem „Data Journalism Handbook 2“, differenziert Nikki Usher dazu zwei Typen des Datenjournalismus: den investigativen und den alltäglichen Datenjournalismus.¹⁵⁵ Der investigative Datenjournalismus steht dabei für die bisherigen Erfolgsgeschichten des Datenjournalismus, wie insbesondere die Arbeiten über die Panama Papers und die Paradise Papers von 2016 bzw. 2017. Die Qualität dieses investigativen Journalismus besteht in der akkurate Aufbereitung großer Datenleaks,¹⁵⁶ durch eine kollaborative Zusammenarbeit verschiedener Expert:innen und Zuhilfenahme von Software mit dem Zweck die Öffentlichkeit über verborgene Wirtschaftsaktivitäten zu informieren. Das Attribut Daten wird dabei gerechtfertigt durch den Aufwand, die Computertechnologie und die Expertise, die nötig sind, mehrere Terabyte große Datensätze zu bearbeiten.

¹⁵⁵ Usher, 2019, S. 154.

¹⁵⁶ Verstanden als von Whistleblowern bereitgestellte Datensätze durch persönliche Einsendungen oder Plattformen, wie beispielsweise wikileaks.

Auf der anderen Seite steht für Usher der „daily data journalism“, „which might mean anything from an effort at ASAP journalistic cartography to turning public opinion polling or a research study into an easily shareable meme with the veneer of journalism attached“.¹⁵⁷ Sie beschreibt damit eine Seite des Datenjournalismus, die an der zunehmenden Professionalisierung und Routinisierung ihre Zweckmäßigkeit verliert oder zum Selbstzweck wird. Eine Facette dieses Selbstzwecks lässt sich anhand des Statistikers Nate Silvers und seiner Datenjournalismus-Plattform „FiveThirtyEight“ beschreiben.¹⁵⁸ Sein übermäßig Vertrauen in quantitativen Daten und deren Auswertung wird meistens am Beispiel der statistischen Modellierung zur US Wahl 2016 porträtiert. Wie andere Datenjournalist:innen konnte auch Nate Silver mit einer reinen Dateninterpretation nicht die Wahl von Donald Trump zum 45. Präsidenten der Vereinigten Staaten voraussagen, sondern prognostizierte, trotz Angabe einer Ungewissheit, einen Sieg von Hillary Clinton.¹⁵⁹ Ein weiteres Beispiel sind Silvers Äußerungen zu seinen Arbeiten im Rahmen der Covid-19-Pandemie, in denen er seinen statistischen Ansatz über die Kenntnisse von Fachexpert:innen zu pandemischen Ereignissen hebt:



Mit Dylan Byers gesprochen, vertritt Silver eine datenjournalistische Haltung, die quantitative Zugänge übermäßig propagiert und subjektive Zugänge zum Journalismus verschmäht.¹⁶⁰ Für Leon Wieseltier stehen Nate Silvers Projekte gar im Kontrast zu den Erfolgen des investigativen Datenjournalismus beispielhaft für die

Abb. 22 Tweet von Nate Silver, 2020.

¹⁵⁷ Usher, 2019, S. 154.

¹⁵⁸ Gray und Bounegru, 2019, S. 4.

¹⁵⁹ Mutalik, 2016.

¹⁶⁰ Byers, 2014.

Fehlschläge des Datenjournalismus.¹⁶¹ Und auch für Mona Chalabi vertritt Silvers „FiveThirtyEight“ den Prototyp des elitären Statistikers, der jegliche quantitative Evidenz zum Selbstzweck erhebt.¹⁶² Der Fokus auf Daten alleine begründet aber noch keine journalistische Motivation. Trotz aller Datenmengen muss auch ein datengetriebener Journalismus inhaltliche Fragen formulieren, zu denen dann die Datenanalyse herangezogen werden kann. So resümiert Usher:

„In the future, I imagine data journalism as unshackled from the term ‚data‘ and instead focused on the word journalism.“

– Usher, 2019, S. 155.

Nun könnte man wie Jonathan Gray und Liliana Bounegru in der Einleitung zum „Datajournalism Handbook“ argumentieren, dass Daten und Datenjournalismus in diesem fragilen „post-truth“-Moment nicht mehr als selbstverständlich betrachtet werden können und sich ihre subjektive Natur und Abhängigkeit von politischen, wirtschaftlichen und kulturellen Interessen zeigt.¹⁶³ Im Gegenteil würde ich argumentieren, dass Daten nicht so sehr die Ausnahme sind, auf die es nun zu reagieren gilt, sondern schon immer Bestandteil der journalistischen Arbeit waren. So verwies der damalige Leiter des Datenjournalismus beim Guardian, Simon Rogers, darauf, dass Datenjournalismus zwar im Trend liege, aber bei weitem nicht neu sei. Schon in der allerersten Ausgabe des Guardians von 1821 findet sich eine Datentabelle.¹⁶⁴ Rogers Relativierung bezieht sich sogar immer noch auf rein quantitative Daten. Mit einer Erweiterung des Datenbegriffs auf qualitative Daten, auch *weiche Daten* genannt, wie sie beispielsweise durch Beobachtungen, Interviews oder Dokumente entstehen, lässt sich ein Großteil des Journalismus, wenn nicht gleich jeglicher Journalismus, als daten- bzw. quellenbasiert beschreiben.

Ich möchte an dieser Stelle nicht versuchen, Journalismus an sich anders oder neu zu bestimmen, was an sich ein strittiges und komplexes Unterfangen wäre. Der Punkt, auf den ich hinaus möchte, ist, dass sich mit dem Datenjournalismus nicht viel am Grundkon-

¹⁶¹ Wieseltier, 2014.

¹⁶² Chalabi, 2019.

¹⁶³ Gray und Bounegru, 2019, S. 4.

¹⁶⁴ The Guardian Datablog, 2011.

zept der journalistischen Arbeit geändert hat. Zwar können mit Computern mehr Daten schneller prozessiert werden, aber genauso wie vor der Erscheinung von Datentechnik müssen Ereignisse kontextualisiert und verifiziert werden. Auch Simon Rogers versucht die Dringlichkeit aus dem Datenjournalismus zu nehmen:

„It's just journalism.“

– Rogers, 2011.

The image shows a full-page spread of the first issue of the *Guardian* newspaper from May 5, 1821. The page is filled with dense, two-column printed text. In the center, there is a prominent rectangular data table with multiple columns and rows, containing numerical and descriptive information. The table appears to be a ledger or a record of some kind, possibly related to the financial operations of the newspaper or a specific event. The surrounding text discusses various topics of the time, including political figures like George Canning and Sir John Scott, and other news items typical of early 19th-century journalism.

Mit diesem Beispiel sollte gezeigt werden, wie Daten und Datentechnik in der Idee des Datenjournalismus eine weder technische noch historische, sondern kulturell begründete Sonderstellung einnehmen. Im Falle eines investigativen Journalismus kann kritisch mit den *neuen* Datenmengen umgegangen werden und, wie im Falle der Panama/Paradise Papers, gesellschaftliche Abhängigkeiten offengelegt werden. Das Versprechen der Besonderheit von Daten führt allerdings auch zu den beschriebenen Situationen, in denen quantitative Daten allein und deren Bearbeitung durch computergestützte Methoden zum journalistischen Selbstzweck erhoben werden. „Data-driven journalism“ übersetzt sich so in einen von Daten getriebenen Journalismus, der die Datentraditionen seines Fachs verkennt. In der *Datengetriebenheit* vergisst der Journalismus, dass er schon immer mit Daten arbeitet. Der Datenjournalismus als disziplinäres Phänomen lässt sich demnach nicht datentechnisch verstehen, sondern viel mehr als eine kulturelle Veränderung, die an die Besonderheit von Daten glaubt. Die *Datengetriebenheit* des Datenjournalismus versteht sich als eine beispielhafte Erscheinung des Datenexzensionalismus.

Abb. 23 Datentabelle in der ersten *Guardian* Ausgabe vom 5. Mai 1821.

2.2 In den Digital Humanities

Ein weiteres Beispiel für die Neukonzeption einer Disziplin durch eine vermehrte Aufmerksamkeit gegenüber Daten findet sich in den Entwicklungen um die *Digital Humanities* (DH). Seit den 1990er Jahren wird darüber geforscht und diskutiert, inwiefern sich die geisteswissenschaftliche Praxis und Theorie durch sogenannte *digitale* Technologien verändert. Genau so divers, wie das geisteswissenschaftliche Themenspektrum, sind auch die Forschungsansätze der DH, jedoch hat sich die Projektlandschaft über die Jahrzehnte spezialisiert und professionalisiert. Populäre Schwerpunkte sind beispielsweise die Computerlinguistik, die Digitalisierung von kulturellen Sammlungen und Archiven oder die digitale Kunstgeschichte. Doch inwiefern beeinflusst das Narrativ der Digitalisierung die Arbeit der Geisteswissenschaften?

Durch den Zusatz digital soll gekennzeichnet werden, dass sich durch Computertechnologien die Bedingungen und Arbeitsweisen der Geisteswissenschaften geändert haben und sich weiterhin ändern. Daten stehen hierbei im Mittelpunkt: Als Voraussetzungen, als Spuren und Ergebnisse von Computerprogrammen treten sie an die Stelle des bisherigen (analogen) Forschungsgegenstandes. Die Forschungsfragen ändern sich teilweise fundamental, wenn beispielsweise Kunsthistoriker:innen keine Einzelbilder mehr betrachten, sondern Datenbanken von tausenden prozessierten Bildern einsehen können.¹⁶⁵ Die bisherigen Modellierungen anzupassen und zu überdenken scheint dabei die Kernaufgabe bei der vermehrten Benutzung von computergestützten Methoden zu sein.

Jedoch finden sich schon in der Namensgebung der Digital Humanities Probleme in der geforderten Modellreflexion. Ursprünglich noch klar als Computational Humanities bezeichnet, zeigt sich mit *digital* eine programmatiche Uneindeutigkeit. Digital verweist im Wortursprung (lateinisch: „*digitus*“ – Finger) auf eindeutig bestimmbar/zeigbare und diskrete Einheiten.¹⁶⁶ In dieser ursprünglichen Definition sind bspw. der genetische Code der DNA (mit eindeutiger Basenabfolge), die Klaviatur (mit eindeutiger Tastenzuordnung zu einem Ton) oder das Alphabet (mit

¹⁶⁵ Ein Beispiel für einen solchen methodischen Ansatz bildet Franco Morettis Konzept des Distant Reading, was unter anderem bei Johanna Drucker kritisch besprochen wird. Vgl. Drucker, 2017a.

¹⁶⁶ Vgl. Wenzel, 2003.

eindeutig festgelegten Buchstaben) Beispiele für digitale Systeme. Durch die Popularisierung von Digitaltechnik wurde der Begriff digital als Referenz für das binäre elektronische Digitalsystem (Strom aus und an, symbolisiert als 0 und 1) etabliert. Im allgemeinen Sprachgebrauch wurde digital primär zum Synonym von Computertechnologie, was eine drastische Verkürzung der längeren Geschichte von diskreten Auszeichnungstechnologien darstellt. Es gibt einen etablierten Diskurs innerhalb der Geisteswissenschaften, der auf die tieferreichende Dimension des Digitalbegriffs verweist. Im nächsten Abschnitt gehe ich näher auf diesen Diskurs und auf die Relation von Digitalisierung und Datafizierung ein.

An dieser Stelle interessiert mich zunächst noch die Frage, warum trotz dieser Kenntnisse um die konzeptionellen Verkürzungen das Schlagwort *Digital Humanities* Anklang gefunden hat. Gerade da Geisteswissenschaften seit jeher mit Texten und Schrift und damit digitalen Kulturtechniken per se arbeiten, erscheint die Abgrenzung gegenüber den traditionellen Geisteswissenschaften mithilfe der digitalen Strukturierung als unpassende Einordnung. Die populärwissenschaftliche Verwendung des Wortes digital als Referenz zu Computertechnologien verschleiert, dass Daten als Phänomen nicht allein den neuen Technologien zugeschrieben werden können. Schon vor dem Computer wurde in den Geisteswissenschaften mit numerischen Daten (sei es über Tabellen und Listen) und mit qualitativen Daten (Interviews, Aufzeichnungen und Bildern) gearbeitet.¹⁶⁷

In ihrer Namensgebung zeigen die DH somit einen Umgang mit Digitalisierung und Computertechnologien, die ihre Neuigkeit und Sonderstellung bestärken soll und dafür eine grundätzliche und kritische Betrachtung der Implikationen von Computertechnologie aufgibt. Es gibt einige Wissenschaftler:innen, die innerhalb eines DH-Kontexts arbeiten und – sich dieser konzeptionellen Schwächen bewusst – einen kritischen Umgang und eine Weiterentwicklung der Ideen fordern. Nach den ersten zwei Generationen ist es vor allem die dritte Generation an DH-Forscher:innen, die zu Reflexionen dieser Art und erweiterten Praktiken forschen.¹⁶⁸ Als prägende Forscherin dieser Generation sei an dieser Stelle die Visualisierungstheoretikerin Johanna Drucker genannt, deren Ideen später im Kapitel noch diskutiert werden.

¹⁶⁷ Vgl. Rosenberg, 2013.

¹⁶⁸ Berry, 2012, S. 3-4.

Ähnlich zu den Entwicklungen im Datenjournalismus zeigt sich bei den Digital Humanities eine bestimmte kulturelle Idee im Umgang mit computergestützten Methoden und der Menge an quantitativen Daten, die ihre Neuigkeit und Sonderstellung betont und weniger auf die Kontextualisierung und Hintergründe dieses Phänomens eingeht – ein klassisches Fallbeispiel des Datenexzptionalismus. Wie im vorherigen Kapitel bereits angedeutet, sind es naturalisierende bzw. anthropomorphisierende Narrative, wie *Big Data*, *Data Mining*, *Data Cleaning* oder *Data Harvesting*, die diesen Umgang mit Daten sprachlich legitimieren. Durch solche Schlagworte soll neben der pragmatischen Hoffnung auf Fortschritt, Effizienz und Optimierung ganz allgemein der soziale und kulturelle Wandel in einer zunehmend digitalen Gesellschaft markiert werden. Weiterhin thematisieren Begriffe wie Datafication (Datafizierung)¹⁶⁹ oder Dataveillance (Datenüberwachung)¹⁷⁰ die technische Aspekte und Konsequenzen dieser zunehmenden Quantifizierung der Gesellschaft. Knapp formuliert ist das Narrativ der Digitalisierung als absolute Gleichsetzung von Datenstrukturen mit Datentechnologie ein Phänomen des Datenexzptionalismus.

2.3 In der Data Science

Nach Datenjournalismus und den Digital Humanities betrachte ich die relativ junge Disziplin der Data Science als ein drittes Beispiel auf ihre datennarrative Besonderheiten. Im Gegensatz zu den anderen Beispielen wird dabei nicht eine spezielle Ausrichtung bestehender Fachdisziplinen oder Wissenschaftsbereiche unter einer Datenperspektive formiert, sondern es wird aus dem Datenphänomen heraus eine ganz eigene Wissenschaft postuliert. Während andere Disziplinen die Datafizierung ihres bisherigen Untersuchungsgegenstandes diskutieren, werden bei der Data Science die Daten selbst zum Kern der Betrachtungen erhoben. Doch was wird sich konkret in der Beschäftigung mit den Daten erhofft? Bei populärwissenschaftlichen Definitionsversuchen dieser Wissenschaft, werden erste Hinweise sichtbar:

¹⁶⁹ Mayer-Schönberger und Cukier, 2013a.

¹⁷⁰ Clarke, 1988.

„Data science combines the scientific method, math and statistics, specialized programming, advanced analytics, AI, and even storytelling to uncover and explain the business insights buried in data.“
– IBM, 2020.

„Data science combines multiple fields including statistics, scientific methods, and data analysis to extract value from data.“
– Oracle, 2021.

Beispielhaft zitiere

ich hier zwei große Computertechnologie-Unternehmen, IBM und Oracle, für die die *Data Science* gleichzeitig ein Geschäftsfeld und gleichzeitig ein Forschungssinteresse ist. In den Definitionen wird ein interdisziplinäres Forschungsfeld abgesteckt, dass sich zwischen einer Kombination aus wissenschaftlichen Methoden, angewandter Mathematik bzw. Statistik und der computergestützten Analyse von quantitativen Daten bewegt. Als Ziel formulieren beide Unternehmen Einsichten und Nutzen aus den Daten zu gewinnen. Welche das im Detail sein sollen oder könnten, bleibt unklar.

In einem ersten Deutungsansatz scheint es nicht verwunderlich, dass große Technologieunternehmen, die Zugang zu den vermehrt produzierten Datenmengen haben, ihr Produkt- und Service-Portfolio durch die vorherrschenden Datennarrative aufwerten. Allerdings ist die Begriffsgeschichte und der Anwendungskontext zu komplex, um nur ein Marketing-Schlagwort zu sein. Beispielsweise etabliert sich die Datenwissenschaft als Disziplin in akademischen Institutionen. So bietet die Universität Potsdam, an der auch diese Dissertation verfasst wurde, einen Master-Studiengang mit dem Namen *Data Science* an und wird wie folgt beschrieben:

„Das interdisziplinäre Fach [...] befasst sich mit Methoden für die automatisierte Gewinnung von Wissen, Einsichten, Prognose-, Risiko- und Handlungsmodellen aus Daten.“
– Universität Potsdam, 2021.

Bemerkenswerterweise ändert sich an der grundsätzlichen Beschreibung trotz des akademischen Kontexts wenig. Auch hier findet sich der Wunsch Einsichten und sogar Wissen aus der Analyse von Daten zu gewinnen.¹⁷¹ An der mathematischen-naturwissenschaftlichen Fakultät angesiedelt, wird das Programm durch eine „einzigartige, interdisziplinäre Kombination von

¹⁷¹ Vgl. Albrecht, 2021.

maschinellem Lernen [und] statistischer Datenanalyse“ neben fachspezifischen Modulen beworben.¹⁷² Vereinfacht kategorisiert, könnte man dahinter einen mathematisch-statistischen Studiengang vermuten, der verstärkt auf die Kombination mit computergestützten Methoden aus der angewandten Informatik setzt.

Tatsächlich äußern einige Statistiker, wie auch der schon erwähnte Datenjournalist Nate Silver, dass die Datenwissenschaft kein wirklich neues Feld sei, sondern im Grunde genommen nur ein anderes Wort für Statistik:

,Data scientist is just a sexed up word for statistician.“

– Nach: Leek, 2013.

Die Diskussion um die Abhängigkeitsbeziehung von Statistik und Data Science ist jedoch etwas umfassender und darum mehr als eine bloße Umbenennung. Dies umso mehr, weil die konzeptionellen Grabenkämpfe, durch welche Data Science als Konzept erst etabliert werden konnte, von etablierten Statistikern selbst geführt wurden.

Da sich dieses Kapitel für Narrative der Ausnahme bezüglich Daten interessiert, ist es daher zweckdienlich, vor allem nach den historischen Gründen und Motiven zu fragen, wegen denen sich Fürsprecher der Datenwissenschaft gegen die traditionelle Statistik stellen. Stellvertretend für diese Perspektive kann der Text „50 Years of Data Science“ von David Donoho gelten.¹⁷³ Donoho, seines Zeichens Professor für Statistik an der Standford University, beschreibt darin in einer historischen Perspektive die ursprüngliche Entwicklung der Datenwissenschaft und gibt einen Ausblick, wie sich die Disziplin als Wissenschaft seiner Meinung nach verändern müsse.

Anlass des Textes war der 100. Geburtstag des Mathematikers John W. Tukey. Dieser ist – neben seiner Entwicklung von Methoden der automatisierten Fourier-Transformation und Datenvisualisierung – von besonderem Interesse für Donoho, weil er sich schon recht früh für eine inhaltliche Reform zugunsten der intensiveren Beschäftigung mit Daten innerhalb der akademischen Statistik einsetzte. So schrieb Tukey 1962 in seiner Schrift „The Future of Data Analysis“:

¹⁷² Universität Potsdam, 2021.

¹⁷³ Donoho, 2017.

„For a long time I have thought I was a statistician, interested in inferences from the particular to the general. [...] All in all, I have come to feel that my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data.“

Für Tukey bedarf es einer Datenanalyse als eigenständige Disziplin, um den

Herausforderungen bei der Arbeit mit den nicht genau definierten Datenstrukturen gerecht zu werden. Die theoretische Statistik spielt für ihn dabei nur eine ergänzende Rolle im Methodenportfolio. Donoho referiert auf weitere Statistiker, welche für ihn stellvertretend für die Bewegung stehen, die auf Tukeys Impuls hin seit den 2000er Jahren sich für die thematische Erweiterung der akademischen Statistik um eine dezidiertere Beschäftigung mit Daten aussprechen. So etwa William S. Cleveland, dem Donoho auch die Prägung des Begriffs Data Science zuschreibt. In der Schrift „Data Science: an Action Plan for Expanding the Technical Areas of the Field of Statistics“ der erstmaligen Nennung schreibt Cleveland:

„This document describes a plan to enlarge the major areas of technical work of the field of statistics. Because the plan is ambitious and implies substantial change, the altered field will be called “data science”. The focus of the plan is the practicing data analyst. A basic premise is that technical areas of data science should be judged by the extent to which they enable the analyst to learn from data.“

Cleveland beschreibt ein *learning from data* als das

Ziel dieser Datenwissenschaft, das vor allem durch eine technische Reform erreicht werden soll. Zumindest die Vorstellung von den *Daten zu lernen* findet sich auch bei Donoho als Abgrenzung zur bisherigen akademischen Tradition, aber auch zu den aktuellen Entwicklungen der Datenwissenschaft als Disziplin. Er benutzt im fortlaufenden Text die viel diskutierte Relation von Statistik und Data Science als Schablone, um seine Vision von Datenwissenschaft gegenüber wiederkehrenden Narrativen abzugrenzen.

Donoho differenziert drei Kriterien, die bei der Diskussion wiederkehrend als Argument eingebracht werden: Big Data, methodische Fähigkeiten und der Arbeitsmarkt. Big Data als Phänomen ist für ihn kein belastbares Kriterium für die Unterscheidung von Statistik und Datenwissenschaft, da „statisticians have been comfortable with large datasets for a long time, and have been holding conferences gathering together experts in ‚large datasets‘ for several decades, even as the definition of large was ever expanding“.¹⁷⁴ Auch die veränderten Anforderungen an die Fähigkeiten von Datenwissenschaftler:innen sind für ihn nur ein Vorwand für einen Unterschied, da sich die spezifischen Fertigkeiten nicht an die realen Probleme bei der Analyse von Daten richten, sondern vielmehr rein computertechnische Kompetenzen sind. Im dritten Schritt zweifelt er auch an, dass Studierende mit nur primär statistischer und ohne informative Fachkenntnisse nicht vorbereitet seien, um ein Jobangebot im Bereich der Data Science zu bekommen, wie er anhand einer Definition eines prototypischen Datenwissenschaftlers verdeutlicht:

„*Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.*“

– Donoho, 2017, S. 754.

Interessant an dieser Stelle ist für meine Betrachtung, dass sich Donoho dem Narrativ Big Data oder deren kulturell konstruierten Erwartungen nicht nur bewusst ist, sondern sie argumentativ projiziert, um seine *would be*-Vorstellung von der zeitgenössischen primär technisch getriebenen Datenwissenschaft abzugrenzen:

„*However, a science doesn't just spring into existence simply because a deluge of data will soon be filling telecom servers, and because some administrators think they can sense the resulting trends in hiring and government funding.*“

– Donoho, 2017, S. 748.

„*A broad collection of technical activities is not a science; it could simply be a trade such as cooking or a technical field such as geotechnical engineering.*“

– Donoho, 2017, S. 758.

¹⁷⁴

Donoho, 2017, S. 747.

Nach Donohos Ansicht legitimiert sich eine Wissenschaft nicht aus der Betrachtung von Folgeerscheinungen einer technologischen Entwicklung und kann auch nicht nur mit rein technischen Methoden bewältigt werden. Um seine *would be*-Datenwissenschaft zu formen, folgt er einer Unterscheidung von „Lesser“ und „Greater Statistics“ nach

John Chambers:

„Greater statistics can be defined simply, if loosely, to learning from data, from the first planning or collection to the last presentation or report. Lesser statistics is the body of specifically statistical methodology that has evolved within the profession.“

– Chambers, 1993, S. 182.

Bei Donoho wird parallel daraus die „Lesser (LGS) and Greater Data Science (GDS)“. Während LGS die bisherigen methodischen Aktivitäten zur Datenanalyse beschreibt, postuliert die GDS eben jene Datenwissenschaft die das *learning from data* fundiert ermöglicht. Neben methodischen Fragestellungen steht hierbei im Vordergrund, dass die Methoden der Datenanalyse und -verarbeitung selbst reflektiert werden und Technologien zur Verbesserung der Methoden vorgeschlagen werden sollen. In dieser Variante der Datenwissenschaft sei die Kernmotivation intellektueller Natur.¹⁷⁵

Das zentrale Narrativ des *Lernens von den Daten* meint also eine Validierung der Methoden zur Untersuchung der Daten – eine Wissenschaft der Datenwissenschaft. Die Datenartefakte an sich verbleiben jedoch unhinterfragt. Sowohl die rein praktisch handelnde LDS, als auch die reflektierende GDS interessieren sich in dieser Vorstellung von Datenwissenschaft dafür, was man mit und aus Daten machen kann, aber weniger für den Gegenstand ihrer Untersuchungen an sich: Was sind Daten und warum gibt es sie? Trotz seines Wissens und seiner Argumentation um die aktuellen Big Data-Narrative verbleibt Donohos Betrachtung zu 50 Jahren Datenwissenschaft ohne Aussagen, warum Daten überhaupt im Interesse einer Datenwissenschaft stehen sollten. Hier zeigt sich beispielhaft eine weitere Facette des Datenexzptionalismus. Gerade durch die Abgrenzung von zeitgenössischen Narrativen wird die Datenwissenschaft durch die Allgegenwart der Daten gerechtfertigt, aber die Erscheinung der Daten als Phänomen nicht hinterfragt. Zusammengefasst interessiert sich die Datenwissenschaft mehr für das Lernen von den Daten, als für die Daten selbst.

¹⁷⁵ Donoho, 2017.

2.4 Zweifel an den Daten

Anhand der drei besprochenen Beispiele des Datenjournalismus, der Digital Humanities und der Datenwissenschaft und den zugeteilten Narrativen, beschrieben als *Datengetriebenheit*, *Digitalisierung* und des *Lernens von Daten*, sollten Aspekte des praktizierten Datenexzptionalismus skizziert werden. Deutlich wurde, dass etablierte Disziplinen, traditionelle Wissenschaften und interdisziplinäre Ansätze mit einer kulturellen Projektion von Daten im konzeptionellen Ungleichgewicht sind. Die aufgezeigten Daten-narrative verweisen dabei nicht so sehr auf Daten selbst, sondern auf eine veränderte Datendynamik. Vor allem durch die Etablierung und Popularisierung von Informations- und Kommunikationstechnologie (IKT) in den letzten Jahren nimmt die Ausbreitung von Daten in vielen Anwendungsbereichen von privaten bis hin zu wirtschaftlichen, wissenschaftlichen und politischen Kontexten zu. Vor allem das Internet und das mittlerweile 30 Jahre alte World Wide Web gelten als ein Katalysator der sogenannten Datenrevolution. Zugespitzt kann man mit Dirk Baecker sagen: „Alle Gesellschaftsfelder sammeln Daten.“¹⁷⁶

Es entsteht der Eindruck, dass sich Daten von jeglicher realen Erscheinung bilden lassen. Tatsächlich sind Daten durch ihre formale Struktur in der Lage zumindest alles Diskretisierbare abzubilden. Sobald ich einen Gegenstand konkret beschreibe falle ich eineindeutige Unterscheidungen, die wiederum in Daten münden. Ich schaue beispielsweise auf eine Obsttheke und konzipiere die verschiedenen Obstsorten nach meinem kulturell geprägten Ordnungssystem. Eine eineindeutige Zuordnung eines Obstes ist ein Datum. Daten sind medientheoretisch gesehen Vermittler, die einen Gegenstand sortierbar und operabel für die menschliche Kognition werden lassen. Als solche „Boten“ treten sie „zwischen“ den Gegenstand und ihren Betrachter.¹⁷⁷ Prägnanter formuliert: Alles, was sich messen, aufzeichnen und anderweitig ordnen lässt, kann man mit Daten darstellen. Alle Daten eint ihre diskrete Ordnung. Ich denke damit alle Daten allgemein als Ordnungsstrukturen. Ich werde im nächsten Unterkapitel *Datenaffinität* weiter auf die grundlegenden Definitionen und Eigenschaften von Daten zu sprechen kommen.

¹⁷⁶ Geiselberger und Moorstedt, 2013, S. 16.

¹⁷⁷ Vgl. Krämer, 2020a.



Durch die enorme Verbreitung und globalen Vernetzung der IKT (Dateninfrastruktur, wie etwa Datenzentren) mit ihren entsprechenden Netzwerkprotokollen (zum Beispiel IP/TCP-Protokolle) und Software-Anwendungen (smarte Technologien) gibt es unzählige technische Quellen für die automatisierte Erstellung von Daten. Die Computerisierung wird auf medialer Ebene auch mit dem streitbaren Begriff *New Media* in Abgrenzung zu bisherigen medialen Erscheinungen markiert.¹⁷⁸ Der Begriff wurde vor allem durch Lev Manovich in „Languages of New Media“ differenziert. Bei ihm steht er in folgender Relation zu Daten:

„New media today can be understood as the mix between older cultural conventions of data representation, access and manipulation and newer conventions of data representation, access and manipulation. The ‚old‘ data are representations of visual reality and human experience [...] the ‚new‘ data is numerical data.“

– Manovich, 2003, S. 15.

Bei Manovich wird der Ursprung von Daten nicht mit dem Computer gleichgesetzt, aber den Daten durch ihre Computerisierung bzw. Umsetzung in ein rein numerisches Format eine neue Qualität zugeschrieben. Die numerische Repräsentation ist bei Manovich eine von fünf Prinzipien der neuen Medien, die ihre vielfältige Transformierbarkeit garantiert.¹⁷⁹ Dazu kommt, dass Daten generell durch ihre

Abb. 24 Brennendes Datenzentrum der Firma OVHcloud in Strasbourg, 2021.

¹⁷⁸ Vgl. Galloway, 2011b.

¹⁷⁹ Manovich, 2001, S. 27-30.

formale Struktur eine große Vielfalt an Themen abbilden können, womit nicht vollkommen unbegründet die Neigung entsteht sie allumfassend zu gebrauchen. Architekturtheoretiker Georg Vrachliotis beschreibt beide Motive als „universellen Entstehungs- und Verwendungspluralismus“.¹⁸⁰ Mit Big Data entsteht abermals eine Vorstellung eines neutrales Abbildes der Welt,¹⁸¹ das für erkenntnisgetriebene Untersuchungen genutzt werden kann und muss. Damit einher geht die Hoffnung auf ein universelles Wissen innerhalb der erworbenen Datenbestände.¹⁸²

Daten als goldenes Kalb

„In Daten schlummert Gold“,¹⁸³ so formulierte Gabriele Hornsteiner die Motivation für Sozialwissenschaftler:innen, aber auch bei Helbing¹⁸⁴ und Bollier¹⁸⁵ finden sich ähnliche Vorstellungen. In letzter Konsequenz gelten Datennarrative, wie *Big Data*, *Datengetriebenheit* oder *Lernen von den Daten*, als neue Versprechen der Allwissenheit.¹⁸⁶ Ein weiterer Trend nach dem Aufkommen von Big Data ist die intensive Entwicklung des sogenannten maschinellen Lernens (oft konzipiert als Sub-Kategorie von *künstlicher Intelligenz*) zur automatisierten Verwertung, der durch Big Data entstandenen Datenmengen.¹⁸⁷ Sowohl Big Data wie auch maschinelles Lernen – als zeitgenössischen Narrative über und mit Daten – beschäftigen sich demnach mit ihrer scheinbar erhöhtem Aufkommen. Im sogenannten Datenzeitalter steht weniger die Frage im Vordergrund was Daten eigentlich sind und warum sie gebraucht werden,¹⁸⁸ sondern vielmehr die Hoffnung, Nützliches und Innovatives mit den anfallenden Datenmengen anfangen zu können. Die Erzählung ist, dass ihr vermehrtes Aufkommen eine grundlegend neue Situation einleite, weswegen auf

¹⁸⁰ Vrachliotis und Weibel, 2019, S. 17.

¹⁸¹ Wie bereits beim Physikalismus, Vgl. Neurath, 1931.

¹⁸² Bächle, 2016, S. 13.

¹⁸³ Hornsteiner, 2012, S. 1.

¹⁸⁴ Helbing, 2015.

¹⁸⁵ Bollier, 2010.

¹⁸⁶ Geiselberger und Moorstedt, 2013.

¹⁸⁷ Etwa an der HfG Karlsruhe, 2021.

¹⁸⁸ Anderson, 2008.

ihr außergewöhnliches Auftreten eine möglichst produktive Verwertung folgen müsse.

Solche Erzählungen münden in eine Datengläubigkeit, einen Datenfetisch, einen Datenchauvinismus oder einem Datenkultismus, um nur ein paar Gegenbegriffe zu nennen. Daten narrative nähren sich vor allem durch einen Glauben an die Besonderheit der Datensituation – den Datenexzptionalismus. Meine Formulierung von Gläubigkeit bezieht sich auf eine dogmatische Vorstellung vom Zweck und Nützlichkeit der Daten als ein soziokulturelles Phänomen mit einer langen Traditionsgeschichte. Das gesellschaftliche Vertrauen in vor allem quantitative Daten wurde mittlerweile über Jahrhunderte aufgebaut und ist somit bei weitem keine Neuigkeit.¹⁸⁹ Im Folgenden gehe ich noch auf die Entwicklung der Quantifizierung ein,¹⁹⁰ jedoch finden sich frühe Formen der Datenverarbeitung und Statistik bereits mit dem Aufkommen der modernen Staaten im 19. Jahrhundert oder auch durchaus früher,¹⁹¹ wie Rob Kitchin vermerkt:

„Societies have collected, stored and analyzed data for a couple of millennia as a means to record and manage their activities. For example, the ancient Egyptians collected administrative records of land deeds, field sizes and livestock for taxation purposes, the 1086 Do-mesday Book captured demographic data, double entry bookkeeping was used by bankers and insurers in the fourteenth century, and the first national registry was undertaken in Sweden in the seventeenth century.“

– Kitchin, 2014a, S. 3.

Daten als Ordnungsstruktur und ihr Einfluss auf die Gesellschaft sind also keine Neuigkeit. Trotzdem rekurrieren die zeitgenössischen Datennarrative, wie Datafizierung oder eben Big Data auf die radikal andere Datensituation durch die Entwicklungen der IKT,¹⁹² also letztlich die elektronische Datenverarbeitung. Dass der Datenbegriff so weitreichende Anwendung gefunden hat liegt also auch an der Gleichsetzung von Daten und Computertechnologie. „[Data] production and nature is being transformed“, wie es Rob

¹⁸⁹ Vgl. Porter, 1995.

¹⁹⁰ Siehe Kapitel 2: 3.3.

¹⁹¹ Nassehi, 2019, S. 45 u. 59.

¹⁹² Mayer-Schönberger und Cukier, 2013b.

Kitchin formuliert, um die „data revolution“ zu begründen.¹⁹³ Die Idee der Besonderheit von Daten resultiert aus diesem konzeptuellen Zusammenschluss von Daten als kulturelles Phänomen und den komplexer werdenden Computertechnologien. In dieser Techno-Narration nehmen Daten dann die Rolle von etwas Einzigartigem oder Außergewöhnlichem an. In der Kombination wird die Schwelle zu einem radikalen Umbruch durch Begriffssphänomene wie digitale Revolution/Transformation/Wandel oder die Industrie 4.0 markiert.

„What we have witnessed over the past year is the dawn of a second wave of digital transformation sweeping every company and every industry.“

– Nadella, 2021.

Es lassen sich weitere Beispiele für Begriffsschöpfungen in Bezug auf die gestiegene Verfügbarkeit von Daten finden, die ihre besondere Betrachtung rechtfertigen sollen: *data deluge*, *data flooding*, *data abundance*, *data smog* oder *data exhaust* sind nur ein paar der vielen Naturalisierungen der Datensituation. Wenn Daten dann noch eine direkte Relation zu Information unterstellt wird zeigen sich folgende rhetorische Erscheinungen: *information explosion*, *information overload*, *infobesity*, *infoxication* oder *information anxiety*. Dies sind weitere Beispiele für die Narrative des Datenexzptionalismus.

Ich argumentiere hier nicht, dass die enorm gestiegene Quantität der Daten keinen Einfluss auf die Qualität im Umgang mit ihnen hat. Ganz im Gegenteil sind kritische Befragungen zu Bedingungen und Abhängigkeiten, die vor allem durch die Verbreitung der Datentechnologien und die dadurch resultierenden Datenmengen entstanden sind, wertvolle Beiträge für den Daten-diskurs.¹⁹⁴ Solche Ansätze setzen allerdings immer einen Datenbegriff voraus, der gerade durch die Dominanz technonarrativer Perspektiven ungenügend hinterfragt zu sein scheint. Ungenügend in dem Sinne, dass sich solche Diskurse nur selten für den Ursprung und Grund von Daten abseits rein technischer Normierungen interessieren. Prinzipiell sind es zwei sehr unterschiedliche Fragen, ob man sich für die politischen, erkenntnistheoretischen, ethischen

¹⁹³ Kitchin, 2014a, S. 24.

¹⁹⁴ Derart gerichtete Diskurse finden sich in Bewegungen wie Open Data (<https://data.gov.uk>), Public Data (<https://www.publicdatalab.org>), Slow Data (<http://beautifuldata.net/2015/02/slow-data>) oder den Critical Data Studies (Iliadis und Russo, 2016).

und anders gearteten Folgen von Datentechnik interessiert oder ob man sich fragt aus welchen Gründen und für welche Zwecke Daten generell Anschluss in der Gesellschaft finden konnten und somit die Datentechnik an sich erst ermöglichten. So positioniert, könnte eine Fragestellung der zweiten Art auch hilfreiche Ergänzung für kritische Fragestellungen der ersten Art bieten.

Daten als Computerphänomen

Dahingehend möchte ich an dieser Stelle eine medienarchäologische Position dieses Kapitels formulieren: Daten sind kein exklusives Phänomen der elektronischen Datenverarbeitung und können daher auch nicht in einem verkürzten Zusammenschluss als Folge der Computerisierung gerahmt werden.¹⁹⁵ Daten sind ein Teilphänomen einer technologisierten Gesellschaft, die ohne Zweifel drastischen Veränderungen unterliegt. Jedoch vergibt man die Chance, den Einfluss der Daten auf diese Gesellschaften zu verstehen, wenn sie als alleiniges Computerresultat abgetan werden. Die Kritik, die ich mit der Figur des Datenexzptionalismus formulieren möchte, richtet sich an die Vorstellung einer Sonderstellung von Daten, die ihre Aufmerksamkeit, Verwertbarkeit und ihr epistemisches Potenzial aus sich selbst heraus begründen und die sich einer Befragung ihres Wertemodells entsagen. Damit Daten zu der besonderen Erscheinung werden konnten, zu der sie kulturell erhoben wurden, brauchte es Modelle, also Vorstellungen und Motivationen, die über die Technologie hinaus bestehen. Diese Modelle wiederum ermöglichen erst die Entwicklung von Computertechnologie, die dann eine weitreichende Anwendung dieser Datenidee darstellt. Die weitreichendere Frage ist für mich weniger, wie sich Datenstrukturen in die „nächste Gesellschaft“ im Detail einschreiben,¹⁹⁶ da Daten und andere Quantifizierungen kein Alleinstellungsmerkmal für diese Gesellschaftsentwicklung darstellen. Vielmehr stellt sich die Fragen, wie sich die Affinität um Daten und Quantifizierung als politisches, soziales und wissenschaftliches Modell generell etablieren konnte.

Im kritischen Diskurs zur Zustandsbeschreibung, den Konsequenzen und Grenzen von primär datengetriebenen Projektionen, haben sich in den letzten Jahren verschiedene

¹⁹⁵ Vgl. Gießmann und Burckhardt, 2014.

¹⁹⁶ Baecker, 2018, S. 14.

Ismen herausgebildet. Zu nennen sind zum Beispiel „Dataismus“, „Datenpositivismus“, „Datenzentrismus“ und „Datenuniversalismus“.¹⁹⁷ Sie bieten – auf Grundlage verschiedener Ansätze – Analysen zum zeitgenössischen Umgang mit Computertechnologien und der spezifischen Verschränkung mit Datenartefakten an. Sicherlich kann infrage gestellt werden, ob bei der Vielzahl dieser Perspektiven noch eine weitere Differenzierung nötig ist, um die Kritik gegenüber der Datenhybris zu konkretisieren. Mein Interesse an dem Datenphänomen begründet sich weniger in der Besonderheit der technischen Entwicklungen, sondern den soziokulturellen Bedingungen, die solche Technologien erst ermöglichen. Im Gegensatz zu den genannten Schwerpunkten möchte in diesem Kapitel das Aufkommen der Daten nicht aus einer technologischen Innovation heraus, sondern in einer soziologischen Tradition verstehen. Die Wechselwirkung von gesellschaftlichen Interessen und technologischen Entwicklungen ist ein etabliertes Thema der Medienarchäologie, den Science and Technology Studies, aber auch in der Soziologie. Insbesondere makrosoziologische Theorien haben versucht die Ursprünge von Digitalisierungs- und Datenphänomene in einem gesellschaftlichen Kontext zu verorten. So liegt ein Schwerpunkt meiner nachfolgenden Betrachtung zur Datenaffinität in den disziplinären Strukturen der Soziologie. Eine Differenzierung dieser Perspektiven wird dabei auch der Schwerpunkt meiner folgenden Betrachtung sein.

¹⁹⁷ Vgl. Brooks, 2013.

3. Datenaffinität

*Was ist und woher kommt
der Fokus auf Daten?*

Als einen zweiten Aspekt des Datenexzptionalismus nach den Datennarrativen beschreibe ich folgend die Datenaffinität. Wenn der Datenexzptionalismus eine künstliche Erscheinung sein soll, dann muss es Gründe und Motivationen dafür geben, warum auf Daten in der Form gehofft wird. Woher röhrt die Affinität zu den Daten? Um diese Frage anzugehen, bedarf es zunächst einer konkreteren Bestimmung des Datenbegriffs. Ein affiner Bezug kann nur beschrieben werden, wenn das Spektrum der Datenerscheinungen zumindest greifbar wird. Vor allem da Datennarrative, wie vorab gezeigt, kaum darauf eingehen, was Daten sind und wie sie auftreten können, werde ich zuerst meinen Datenbegriff schärfen.

Im Abgleich mit dieser Datendefinition wird im Folgenden deutlich, dass im Datenexzptionalismus eine bewusste Reduktion des Datenbegriffs vorgenommen wird. Die Datenaffinität richtet sich auf einen ganz spezifischen Datentyp, den der numerischen und auch durch Datentechnik prozessierbaren Daten. Durch diese Simplifizierung tritt eine Ursache-Folge-Umkehrung ans Licht. In der computertechnischen Konzeption werden Daten zur Folge von Computertechnologie – *es gibt Daten nur, weil es Computer gibt*. Vergessen wird dabei, dass es ohne die Datenlogik keine Rechnerarchitektur geben könnte. Ich zeige auf, wie die Datenaffinität Daten zur Folge anstatt zur Bedingung eines Digitalisierungsdiskurses stilisiert.

Als Gegenentwurf führe ich in den soziologischen Diskurs zur digitalen Gesellschaft ein. Insbesondere makrosoziologische Perspektiven spüren nach, inwiefern Daten- und Digitalmodelle die moderne Gesellschaft prägten. Anstatt Datenaffinität in ihrer konzeptionelle Reduktion zu belassen, soll im philosophischen Sinne von Affinität eine grundsätzliche Verwandtschaft zwischen digitalen Strukturen und moderner Gesellschaft nachvollzogen werden. Daten sind in einer solchen Auslegung keine Ausnahmeherscheinung mehr, sondern eine Grundbedingungen der modernen Gesellschaft. An die Stelle des Datenexzptionalismus tritt eine Datennormalisierung.

3.1 Datendefinition

Die im Februar 2021 veröffentlichte Datenstrategie der Bundesregierung verdeutlicht die Notwendigkeit der Diskussion des Datenbegriffs. Es ist bemerkenswert, dass das Papier sowohl im Titel als auch in den 122 Seiten über 1.600 Mal das Wort Daten nennt, nur um im abschließenden Glossar zu verkünden, dass das grundsätzliche Konzept von Daten gar nicht eindeutig sei. Es ist zu vermuten, dass die konzeptionelle Unschärfe eventuell die Verwendung als politischen Begriff begünstigt. Jedenfalls verweist das Bundeskanzler:innenamt letztlich auf ein Verständnis von digitalen Daten. In ihrer Diskretheit sind jedoch alle Daten digital, sprich eindeutig unterscheidbar. Das Presse- und Informationsamt meint mit digital vermutlich computertech-nische Erzeugnisse, aber wie schon vorab betont, erschöpft sich die Begriffsbestimmung von Daten nicht im Kontext von Computertechnologie bzw. negiert dadurch somit sogar den längeren Verwendungskontext. Daten bleiben in der Datenstrategie der Bundesregierung also gänzlich unbestimmt.

„Eine einheitliche Definition von Daten gibt es nicht. Mit Daten werden u. a. die einzelnen Werte innerhalb eines Datensatzes bezeichnet. In der Literatur wird unter Daten jedwede Art von Elementen verstanden, die durch einen Computer interpretierbar ist. Grundsätzlich umfasst der Datenbegriff elektronisch gespeicherte und nicht-elektronisch gespeicherte Zustände oder Wiedergaben von Sachverhalten (Erhebungen, Berechnungen, Messungen, Texte). In dieser Strategie wird der Ausdruck Daten primär im Sinne von digitale Daten verwendet.“

– Presse- und Informationsamt der Bundesregierung, 2021, S. 107.

Daten als Fakten?

Die obligatorische etymologische Betrachtung erscheint als Startpunkt für eine nähere Bestimmung des Datenbegriffs hilfreich. Im Deutschen Wörterbuch der Brüder Grimm wird der Begriff Datum vor allem durch die Verwendung als „gegebene Größe“ definiert.¹⁹⁸ Daten sind in diesem Ursprung das Gegebene. Dort findet sich auch, neben der alltäglichen Verwendung als Kalenderangabe, die Gleichsetzung mit dem Faktum. Durch diese etymologischen Spur begründet sich die synonyme Verwendung von Daten

¹⁹⁸ Grimm und Grimm, 1965.

und Fakten im allgemeinen Sprachgebrauch. Faktum ist jedoch im Wortursprung das Partizip von „*facere*“, bedeutet soviel wie *das Gemachte*, und kontrastiert damit vielmehr das Datum eher als damit synonym zu sein. Die Verbindung von Daten und Fakten scheint daher ihrem etymologischen Ursprung nach problematisch.

Eine wissenschaftshistorische Aufarbeitungen der Be- griffsrelation von Daten und Fakten findet sich etwa bei Hans-Jörg Rheinberger.¹⁹⁹ In seinem Text „Wie werden aus Spuren Daten, und wie verhalten sich Daten zu Fakten?“ bespricht er neben der epistemologischen Kategorie der Spur die Begriffsentwicklung von Daten und Fakten in den neuzeitlichen Wissenschaften.²⁰⁰ In diesem Kontext kehrt sich die Bedeutung von Daten und Fakten um:

„Daten werden als etwas von Instrumenten Gemachtes und von Prozeduren Hervorgebrachtes angesehen, während Fakten als etwas Gegebenes angenommen und unterstellt werden.“

– Rheinberger, 2007, S. 117.

Rheinberger zeigt am Beispiel der methodischen Entwicklungen in der Molekularbiologie, dass in einem traditionellen Wissenschaftsverständnis die Suche nach Daten nur in Relation zu einem gegebenen Faktum Bedeutung erhielt. Ein vermuteter Sachverhalt, an dem sich die Daten messen müssen, steht im Zentrum der wissenschaftlichen Betrachtung. Im Gegensatz dazu beschreibt er die neueren Entwicklungen der computergestützten Bioinformatik. Dort werden die vielfach vorhandenen Daten aufgefunden und Fakten daraus konstruiert. In Essenz seiner wissenschaftshistorischen Betrachtungen der Biowissenschaften formuliert er ein „Primat der Daten“.²⁰¹ Man sei von einer „*hypothesengeleiteten*“ zu einer „*datengeleiteten*“ Forschung übergegangen.²⁰² Rheinberger beschreibt damit exemplarisch eine weitere Instanz des Datenex- ceptionalismus. Das Paradigma lautet: Daten sind gegeben und aus ihnen müssen nun Fakten gemacht werden.

¹⁹⁹ Rheinberger, 2007; Vgl. auch Kuhn, 1976 und Canguilhem, 2006.

²⁰⁰ Der Begriff der Spur wird noch im Kapitel 6 von Bedeutung sein, siehe Kapitel 6: 2.1.

²⁰¹ Rheinberger, 2007, S. 124.

²⁰² Ebd.

Paradigm	Nature	Form	When
First	Experimental science	Empiricism; describing natural phenomena	Pre-Renaissance
Second	Theoretical science	Modelling and generalisation	Pre-computers
Third	Computational science	Simulation of complex phenomena	Pre-big data
Fourth	Exploratory science	Data-intensive; statistical exploration and data mining	Now

Der Datenexzeptionalismus als eine Naturalisierung der Datenerrscheinung führt wieder zum etymologischen Ursprung des Begriffs Daten. Durch die Beschreibung des dynamischen Bedeutungsverhältnisses wurde deutlich, dass beide Begriffe, Daten und Fakten, nicht gleichzusetzen sind. Zwar wechselten die Begriffsbedeutungen regelmäßig, aber überlagerten sie sich nie. Pragmatisch formuliert ist eine weitere Begriffsbestimmung von Daten negativer Natur: Daten sind keine Fakten.

Daten als Informationen?

Insbesondere im Kontext von Datenvisualisierungen spannt sich neben Daten und Fakten ein größeres Begriffsfeld auf, welches auch die thematisch verwandten Begriffe Information und Wissen beinhaltet. Nicht nur, dass aus Datenvisualisierungen Informationen und letztlich Wissen abgeleitet werden soll, sogar in der Benennung sind Informationsvisualisierung und Wissensvisualisierung übliche Kategorien. Diese Begriffsrelation ist insbesondere geprägt durch das Modell der „DIKW-Pyramide“. Der Ursprung dieses Konzeptes ist bis heute nicht eindeutig geklärt, jedoch wird eine Beschreibung in „From Data to Wisdom“ von Russell L. Ackoff aus dem Jahr 1989 oft als die erstmalige ausführliche Beschreibung referenziert.

Dort schreibt er:

„Wisdom is located at the top of a hierarchy of types, types of content of the human mind. Descending from wisdom there are understanding, knowledge, information, and, at the bottom, data.“
– Ackoff, 1989, S. 3.

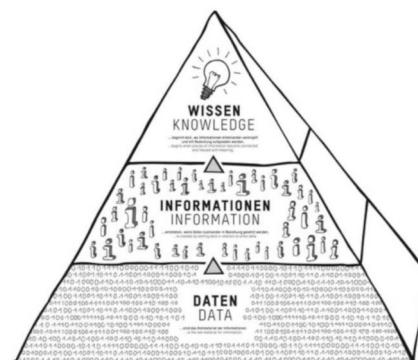


Abb. 25 Vier Paradigmen der Wissenschaft nach Jim Gray.

Abb. 26 Wissenspyramide, 2015.

Dieses Konzept beschreibt eine hierarchische Verbindung zwischen data, information, knowledge und wisdom – „DIKW“. Durch diese Ordnung soll vermittelt werden, dass es einen natürlichen Verlauf von Daten zu Informationen, von Informationen zu Wissen und von Wissen zu Weisheit gibt. Stark verkürzt bieten Daten in dieser Lesart das Fundament für nicht nur Wissen, sondern auch für ein fundamentaleres Verständnis von Zusammenhängen in der Welt. Das ist eine rhetorisch wirksame Projektion des Datenexzptionalismus und damit Motivation für das Ziel der Datenvisualisierung oder generell aller Bestreben, mit Daten einen Mehrwert aus der Beschäftigung mit ihnen zu ziehen. Daten werden in diesem Gefüge von Ackoff wie folgt definiert:

„Data are symbols that represent properties of objects, events and their environments. They are products of observation. [...] Information, as noted, is extracted from data by analysis in many aspects of which computers are adept. [...] In either case, information is inferred from data.“

– Ackoff, 1989, S. 3.

Daten sind in seiner Definition zunächst abstrahierte Zeichen, aber auch das Material, was durch Analysen extrahiert wird, um Informationen aus ihnen zu gewinnen. Informationen werden aus Daten abgeleitet, was Daten zur Bedingung für Informationen werden lässt. Diese Extraktion und Ableitung setzt allerdings eine funktionale Verbindung von Daten und Informationen voraus, die durch Ackoff nicht begründet wird. Es wird an keiner Stelle in seinen Ausführungen benannt, wie man von den Daten zu Informationen gelangt. Die funktionale Verbindung wird einfach vorausgesetzt.

Ich will damit nicht andeuten, dass Menschen aus der Arbeit mit Daten keine Bedeutung ableiten können. Es kann beispielsweise eine Erkenntnis sein, dass sich ein Datensatz nur ungenügend für eine Betrachtung eignet und demnach ein anderer Datensatz generiert werden muss. Es kann auch gerade ein unvollständiger oder fehlleitender Datensatz zu einer Erkenntnis zu einem Gegenstand führen. Es kann aber auch eine Erkenntnis sein, dass der betrachtete Gegenstand gar nicht in Datenform abgebildet werden kann. In allen Fällen wurden eine Erkenntnis erlangt, indem Daten nicht zu Informationen gewandelt wurden. Was ich mit diesen Grenzfällen verdeutlichen will ist, dass die

Wandlung von Daten zu Informationen ist keine lineare und garantierte, sondern eine äußert komplexe und keine unbedingt umsetzbare Prozedur darstellt.

Ähnliche Zweifel an dem Abhängigkeitsverhältnis von Daten, Informationen und Wissen formuliert auch Rafael Capurro:

„Putting the three concepts (*data, information, knowledge*) as done here, gives the impression of a logical hierarchy: information is set together out of data and knowledge comes out from putting together information. This is a fairytale.“

– Capurro, 2003.

Capurro beschäftigte sich aus philosophischer Perspektive eingehend mit dem Informationsbegriff. Information ist ein komplexer Begriff, der durch die Ideengeschichte hinweg vielen Wendungen unterlag und vor allem in zeitgenössischen computertechnologischen Kontexten meist in Zusammenhang mit der mathematischen Informationstheorie, besonders geprägt durch Claude Shannon, verwendet wird.²⁰³ Wichtig ist mir, mit Capurro zu betonen, dass Informationen eine eigene epistemische Kategorie sind, die unabhängig von Daten besteht.²⁰⁴ Daten sind keine Informationen und Daten sind keine zwingende Voraussetzung für Informationen.

Als ein Beispiel ziehe ich wieder die Infektionszahlen in der Covid-19-Pandemie heran. Die Daten zur Infektionslage sind ein spezifisches Ordnungssystem, mit dem die Pandemie gesundheitspolitisch strukturiert wird. Die Daten an sich geben aber keine Informationen zur pandemischen Lage. Die Daten müssen interpretiert, kontextualisiert, verglichen und kritisiert werden, damit mit ihnen eine Information für die pandemische Lage entstehen kann. Diese Prozesse sind so vielfältig und abhängig von bspw. individualpsychologischen Prägungen der betroffenen Menschen oder gesundheitspolitischen Modellen der verantwortlichen Regierung, dass man beinahe sagen könnte, dass Informationen trotz Daten entstehen. In jedem Fall gibt es keine eindeutige Relation zwischen Daten und Informationen.

Ich will hier nicht argumentieren, dass Daten und Informationen in keinem Zusammenhang zueinander stehen, nur ist diese Relation eben nicht so klar, wie es das DIKW-Modell vorzu-

²⁰³ Weitere Lektüre zum Begriff der Information bei Blair et al., 2021.

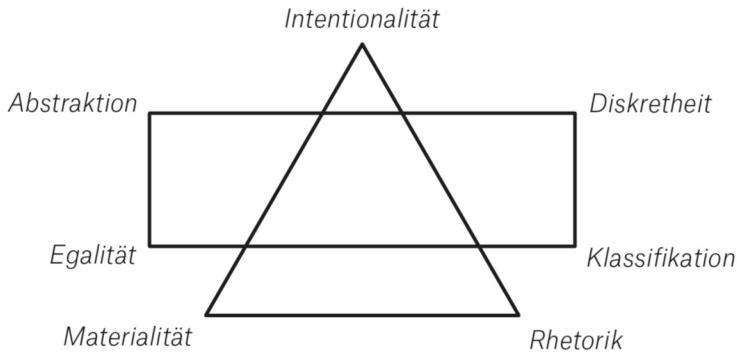
²⁰⁴ Capurro, 2003.

geben scheint. Wenn Informationen aus Daten gewonnen werden, dann geschieht dies nicht zwangsläufig durch die Anhäufung und die automatisierte Analyse von Daten. Bedenken gilt es immer dann anzumelden, wenn beinahe wahllos zwischen Daten- und Informationsbegriffen gewechselt wird, wie es bei der Daten- bzw. Informationsvisualisierung der Fall ist.

Datenbegriff

In der bisherigen Bestimmung habe ich ausgeschlossen was Daten schon einmal nicht sind: Fakten und Informationen. Im Folgenden gilt nun zu klären was Daten sind. Die Datenaffinität als Aspekt des Datenexzceptionalismus lässt sich nur in der Verengung und Vereinfachung des Datenbegriffs verstehen. Dagegen möchte ich versuchen einen grundlegenden Datenbegriff als Ausgangspunkt zu formulieren.

Ganz grundsätzlich verstehe ich Daten in dieser Arbeit als Ordnungsstrukturen. Daten sind vor allem Konzepte, die eine Ordnung veranlassen. Als ein Beispiel dient mir das Bücherregal neben dem Schreibtisch, an dem ich diesen Text schreibe. Ich kann mir ein eindeutiges Entscheidungskriterium überlegen, von dem ich einen Datensatz abstrahiere. Beispielsweise sortiere ich alle Bücher nach ihrem Anfangsbuchstaben, welcher eindeutig unterscheidbar ist, und habe danach einen spezifischen Datensatz aus meinen Büchern abstrahiert. Wichtig dabei ist, dass nicht die realen Bücher meine Daten sind, sondern die ordnende Struktur, die ich ihnen auflege. Als ein weiteres Beispiel für die Ordnungsfunktion betrachte ich wieder die Covid-19-Infektionsdaten. Die Pandemie ist ein komplexes Phänomen. Die Erkrankung ist ein Entscheidungskriterium, welches sich durch einen Nachweis eindeutig bestimmen lässt. Mit dem Ordnungskonzept der Infektion erzeugt man einen Datensatz von infizierten und nicht-infizierten Teilen einer Bevölkerung. Wieder ist wichtig, dass nicht die Menschen die Daten sind, sondern die ordnende Struktur, die über sie gelegt wird.



Neben der ordnenden Grundfunktion beschreibe ich im weiteren sieben spezifischen Eigenschaften von Daten: Diskretheit, Egalität, Abstraktion, Materialität, Rhetorik, Klassifikation und insbesondere ihre Intentionalität. Vorab beschrieb ich bereits Diskretheit und Egalität als zwei Eigenschaften der Ordnungsstrukturen. Mit Medienwissenschaftler Thomas Bächle möchte ich die weiteren Eigenschaften von Daten beschreiben. Zunächst bestätigt Bächle aber meine zwei vorab definierten Kategorien der Diskretheit und der Egalität.

1 – Diskretheit

Diese Bemerkung ist wichtig, denn oft wird diskret mit abzählbar übersetzt, was davon ablenken kann, dass diskret nebenzählbaren Zahlenwerten prinzipiell die eindeutige Unterscheidung jeglicher Art bedeutet. Ein Beispiel für eine diskrete Struktur ist die Binärstruktur. Binäre Anordnungen finden sich beispielsweise in der Sprache bei Dichotomien, wie Natur/Kultur, Mensch/Maschine oder analog/digital, aber auch im mathematischen Binärsystem (0 und 1) und bei den elektrischen Zuständen der Digitaltechnik (Strom aus und ein). Ein weiteres Beispiel findet sich wieder in den Covid-19-Inzidenzzahlen. Dort ist es die binäre Struktur *infiziert/nicht infiziert*, die das Entscheidungskriterium ausmacht. Über eine Nachweismethode ist eine Infektion eineindeutig bestimmbar und damit ein diskreter Datensatz. Letztlich bedeutet das, dass Daten letztlich von jedem Gegenstand abstrahiert werden können, der in eineindeutige Strukturen geordnet werden kann.

„Daten liegen zwar nicht immer in numerischer Form, aber stets als diskrete Einheiten vor.“

– Bächle, 2016, S. 124.

Abb. 27 Aspekte des Datenbegriffs.

2 – Egalität

Die Egalität der Daten beschreibt Bächle als *aggregativ* anhand eines Zitates von Lisa Gitelman:

„They are collected in assortments of individual, homologous data entries and are accumulated into larger or smaller data sets.“

– Gitelman, 2013, S. 8.

Da alle Daten die diskrete Struktur gemeinsam haben, sind sie auch frei miteinander kombinierbar – ein struktureller Universalismus. Diese Egalität ermöglicht die beliebige Kombination von verschiedenen Datensätzen unabhängig von semantischen Bedingungen. Die Kombination von Daten muss keinen Sinn ergeben. In einer computertechnischen Konzeption spricht man dann von Datenbanken, die innerhalb des Digitaldiskurses einen eigenen Forschungsbereich markieren.²⁰⁵ Auch im Beispiel der Covid-19-Infektionszahlen wird diese Egalität genutzt, wenn diese Daten mit anderen Statistiken kombiniert werden. Üblich ist u.a. die Kombination mit Einwohner:innenzahlen zu einer *je 100.000 Einwohner:innen*-Statistik. Weiterhin ist die Verknüpfung mit Daten zu Hospitalisierungen, Todesfällen und Impfungen Teil der Gesundheitspolitik. In der weiteren Analyse werden Inzidenzzahlen beispielsweise mit wirtschaftlichen Daten kombiniert, um gesellschaftliche Auswirkungen zu modellieren.

Bächle leitet seine weiteren Dateneigenschaften an Lisa Gitemans Standardwerk zur Datenkritik „Raw Data‘ is an Oxymoron“ ab.²⁰⁶ Der titelgebende Ausspruch stammt wiederum von Geoffrey Bowker. Die Bezeichnung *rohe Daten* zielt auf die Idee, dass Daten als natürliche Vorkommen nur auf eine bestimmte Art und Weise verarbeitet werden müssen, um einen Wert zu erhalten. Bowker verweist mit der Bezeichnung „Oxymoron darauf hin, dass Daten als artifizielle Ordnungsstrukturen niemals roh sein können. Daten sind niemals einfach da bzw. *rob*. Der Umstand, dass diese Korrektur in der Erzählung über Daten notwendig ist, verweist auf eine von Bächles Dateneigenschaften.

²⁰⁵ Vgl. Manovich, 1998 und Vesna, 2007. Vgl. auch die Forschung von Hunger, 2018 zu relationalen Datenbanken.

²⁰⁶ Gitelman, 2013, S. 6-8.

3 – Rhetorik

Als eine weitere Eigenschaft benennt Bächle demnach dezidiert die Rhetorik von Daten. Die Relevanz von Daten ist demnach vor allem kulturell erzeugt. Ganz ähnlich beschrieb ich diese Position im vorherigen Unterkapitel *Datennarrative*.²⁰⁷ Bächle referiert dazu wieder Lisa Gitelman:

„This aggregative quality of data helps to lend them their potential power, their rhetorical weight (More is better, isn't it?).“
– Gitelman, 2013, S. 8.

Auch in der Corona-Pandemie galt die nationale Aggregationsfähigkeit möglichst vieler Nachweistestungen als Erfolgsmerkmal der Gesundheitspolitik. Wer viele bzw. möglichst flächendeckende Inzidenzzahlen aufbereiten konnte, hatte ein vermeintlich besseres Verständnis vom Verlauf der Pandemie und einen konkreteren Handlungsspielraum. Datenstrukturen waren und sind essentiell in der Aushandlung der pandemischen Lage. Die Gewichtung, die ihnen beigegeben wird, ist jedoch kein Selbstzweck. Daten alleine übernehmen keine politische Verantwortung. Die Relevanz von Fallzahlen wurde über wissenschaftliche und politische Narrative hergestellt. Daten werden rhetorisch legitimiert.

4 – Abstraktion

Die nächste Dateneigenschaft nach Bächle ist die Abstraktion. Daten sind abstrahiert von einem vom Menschen beobachteten Gegenstand und liegen dann in symbolischer Form, also als Zeichen, vor.²⁰⁸ Diese Abstraktion ist immer verlustbehaftet. Sie deckt sich niemals komplett mit dem realen Objekt der Abstraktion, wenn der Betrachtungsgegenstand nicht sogar an sich fiktionalisiert ist. Im Falle des Bücherregals werden meine Bücher auf bestimmte Kriterien hin reduziert. Im Falle von Daten zum Klimawandel wird beispielsweise ein für den Menschen nicht wahrnehmbarer Gegenstand zwar auch vereinfacht, aber dadurch auch erst in einer Form greifbar. Ebenso sind die Infek-

²⁰⁷ Siehe Kapitel 2: 2.

²⁰⁸ Bächle, 2016, S. 124.

tionszahlen zur Corona-Pandemie ein Weg eine nicht sichtbare Viruserkrankung konkret und prozessierbar zu machen. Dementsprechend sind Datenabstraktionen immer nur eine menschliche Perspektive auf einen Gegenstand und müssen in ihrer Eindimensionalität auch so besprochen werden.

5 – Materialität

Neben der semiotischen Ebene deutet Bächle als weitere Eigenschaft auf die dafür notwendige materielle Ebene hin. Daten müssen erhoben, ausgewertet, bereitgestellt, archiviert und sichtbar gemacht werden, was arbeits- und ressourcenintensiv ist. Ein bekanntes Beispiel sind die wartungsaufwändigen und energieintensiven Datenzentren.²⁰⁹ Wieder auf die Infektionszahlen bezogen, zeigt sich die materielle Dimension auf einen ganz individuellen Ebene in der Anhäufung des Mülls zu den Antigen-Schnelltests. Materialität bedeutet aber auch den menschlichen Arbeitsaufwand in Form der überlasteten Fachkräfte in den Testlaboren und den Gesundheitsämtern. Die Erstellung von Inzidenzzahlen ist ein enormer materieller Kraftakt. Der Aufwand und die Folgen der Datenkultur auf einer materiellen Ebene ist so signifikant, dass die Materialität eine kritische Eigenschaft von Daten ist.

6 – Klassifikation

Weiterhin definiert Bächle die Hierarchie als Dateneigenschaft und meint damit, dass alle Daten auf einem Klassifikationssystem basierend geschaffen und interpretiert werden. Diese Abhängigkeit ist bedeutend, denn es gibt keine Daten ohne eine zugrundeliegende Hierarchie. Daten müssen in ein existierendes Ordnungssystem gepresst werden, um realisiert zu werden. Die Abstraktion geschieht nur im Rahmen eines gegebenen Systems; alles was außerhalb dieses Systems liegt, kann nicht abstrahiert werden. Die Datenerstellung ist damit auch immer gewollt repressiv – letztlich ein Akt der Gewalt gegenüber dem Gegenstand der Klassifikation. Die Ordnungssysteme sind allerdings nach der Erstellung der Daten in ihnen nicht eindeutig ersichtlich, wie auch Bächle wieder im Rückgriff auf Gitelman referiert:

²⁰⁹ Dommann, Rickli und Stadler, 2020.

„It follows that the imagination of data is in some measure always an act of classification [...] though the underlying principles at work can be hard to recover. Once in place, classification schemes are notoriously difficult to discern and analyse.“

– Gitelman, 2013, S. 9.

Im Falle der Covid-19-Fallzahlen verdeutlicht sich die uneinsichtige Ordnungsstruktur hinter den Datenwerten

positiv oder *negativ*. Im Falle von sogenannten PCR-Tests beruht das diagnostische Verfahren auf der Vervielfältigung des genetischen Materials des Corona-Virus. Ab einem von Mediziner:innen festgelegten Schwellenwert der Häufigkeit einer Vervielfachung gilt eine Person als positiv. Ein Schulkind, eine Studentin oder eine berufstätige Mutter werden ab diesem Zeitpunkt für den Datensatz nur noch unter einem Unterscheidungskriterium bewertet. Das ist die strukturelle Gewalt des Ordnungssystems, indem nur die Kriterien wert haben, die für diese Ordnung festgelegt wurden. Ungeordnetes bleibt für Daten wertlos.

Klassifikationssystem als Grundlage der Modellierung von Daten sind als diese hierarchischen Systeme oftmals uneinsichtig und essenziell als Eigenschaft von Daten zu markieren. Es braucht demnach eine intensive qualitative Auseinandersetzung mit den Daten, um ihre klassifikatorischen Bedingungen freizulegen. Susan Leigh Star und Geoffrey Bowker (von dem auch der Ausspruch „Raw Data is an Oxymoron“ stammt) beschreiben in „Sorting Things Out“ einen solchen Ansatz, der die signifikanten Auswirkungen von Klassifikationssystemen und Standardisierungen auf soziokultureller Ebene sichtbar macht.²¹⁰

7 – Intentionalität

Neben ihrem klassifikatorischen Fundament müssen Daten aber auch bewusst durch dieses Ordnungssystem für einen bestimmten Zweck gerahmt werden. Daten können nur konstruiert werden, wenn ein Modell mit Vorsatz auf sie angewandt wird. Die Arbeit mit und an Daten begründet sich nicht aus sich selbst heraus, sondern bedingt eine Absicht. Ohne eine Intention gibt es keine Daten. Diese performative Ebene möchte ich als letzte Eigenschaft der Intentionalität von Daten ergänzen. Eine Spur dieser Begriffsdeutung findet sich schon im historischen Sprachgebrauch. Im

²¹⁰ Star und Bowker, 1999.

bereits zitierten Wörterbuch der Brüder Grimm findet sich auch ein Eintrag, der eine Verwendung des Begriffs „datum“ im 16./17. Jahrhundert als „zur Bezeichnung einer intention“ beschreibt.²¹¹ Eine Rückbesinnung auf diesen etymologischen Zweig erachte ich als eine wichtige Gegenerzählung zum lateinischen Ursprungs als das Gegebene.

Wieder am Beispiel der Corona-Pandemie sind die Infektionszahlen keine natürliche Begleiterscheinung, sondern ein gewolltes und wohl intendiertes Artefakt einer bestimmten Gesundheitspolitik. So betont auch Architekturtheoretiker Georg Vrachliotis die intentionalen Aspekte von Daten:

„Daten gelten als epistemische Konstrukte, die unter bestimmten Bedingungen und mit bestimmten Absichten hergestellt werden – politisch, sozial, ökonomisch. [...] Daten tragen daher nicht nur ein prognostisches, sondern vor allen Dingen auch ein utopisches Potential in sich.“

– Vrachliotis und Weibel, 2019, S. 17.

„Daten sind nie etwas Gegebenes, sondern etwas Ge schaffenes. Daten sind sogenannte epistemische Konstrukte, die unter ganz bestimmten Bedingungen und mit ganz bestimmten Absichten hergestellt und auch verwertet werden – politisch, sozial, ökonomisch.“

– Vrachliotis, 2019, S. 37.

Alles in allem sind
Daten in meiner
Auslegung Ordnungs-

strukturen mit spezifischen Eigenschaften. Daten etablieren demnach eine spezifische Ordnung auf ihren Betrachtungsgegenstand. In meiner Begriffsdefinition haben diese ordnenden Strukturen sieben Eigenschaften: Diskretetheit, Egalität, Rhetorik, Abstraktion, Materialität, Klassifikation, Intentionalität. Daten können von allem gebildet werden, was eindeutig unterscheidbar ist. Daten lassen sich bei ähnlicher Struktur beliebig miteinander kombinieren. Daten sind kein Selbstzweck, sondern werden inszeniert. Daten sind immer Vereinfachungen eines Gegenstandes. Daten herzustellen und zu bearbeiten bedeutet einen materiellen Aufwand. Die Ordnung der Daten erfolgt durch ein vorab etabliertes Klassifikationssystem. In erster Linie ereignen sich Daten nicht. Sie sind keine natürliche Erscheinung. Es gibt keine *raw data*, kein *data flooding* und kein *data exhaust*. Daten sind eine bewusste Entscheidung. Daten sind gewollt.

²¹¹ Grimm und Grimm, 1965.

3.2 Typisierung von Daten

Die Absicht in den Daten zeigt sich auch in der etablierten Typisierung von Datenarten. Entgegen meiner generell gehaltenen Datendefinition, die letztlich alle Daten betrifft, versuchen diese Typisierungen den Datenbegriff nach bestimmten Verwendungszwecken zu strukturieren. Im Folgenden zeige ich wie Motive der Datenaffinität auf diese Einordnung wirken. Die Datenaffinität richtet sich auf ganz spezifischen Datentypen, die der Vorstellung einer Verwertbarkeit der Daten entgegen kommt. Daten sind in dieser Perspektive also nicht gleich Daten. Es gibt unterschiedliche Arten mit verschiedenen Ausrichtungen.

In diesen unterschiedenen Datenarten zeigen sich Wünsche und Erwartungen gegenüber Daten, die eine weitere Ausdifferenzierung des Datenexzptionalismus ermöglichen. Im Abgleich mit meiner Datendefinition mache im Folgenden deutlich, dass der Datenexzptionalismus eine bewusste Reduktion des Datenbegriffs vornimmt. Anhand einer Zusammenfassung wiederkehrender Datentypisierungen von Humangeograph Robert Kitchin möchte ich zeigen, wie der reduktive Ansatz bestimmte Datenerscheinungen favorisiert.²¹² Im Detail bespreche ich fünf Unterscheidungskategorien nach Kitchin: Form, Struktur, Quelle, Produzent und Typ.

Form: qualitativ oder quantitativ

In der empirischen Forschung wird zwischen quantitativen und qualitativen Daten unterschieden. Quantitative Daten meint numerische, alsozählbare, Werte. Diese können dann in einer vereinfachten Ordnung von nominaler, ordinaler oder metrischer Skalierung angegeben werden.²¹³ Daten sind aber nicht auf diese quantitative Konzeption, vor allem in einer technischen Lesart von Daten, allein beschränkt. So bilden qualitative Daten in dieser Differenzierung das nicht-numerische Gegenstück. Mit qualitativen Daten sind dann zum Beispiel textuelle (Schrift),

²¹² Kitchin, 2014a, S. 31ff.

²¹³ Nominal meint kategoriale Anordnungen wie Geschlechtereinteilungen oder Postleitzahlen. Ordinal verweist auf Rangordnungen, wie Noten oder andere Einstufungen. Metrisch meint dann die klassischen Messungen als Zahlenwerte, entweder in Intervallskala ohne Nullpunkt (z.B. Zeitmaße) oder in einer Verhältnisskala mit Nullpunkt (z.B. Preise).

visuelle (Bild und Video) oder auditive Datenerscheinungen (Sprache und Töne) gemeint. Durch meine vorangegangene Definition von Daten in Abhängigkeit von Klassifikationssystemen als Eigenschaft der Klassifikation wird allerdings deutlich, dass jede Datenform von einem subjektiv-intendierten Ordnungssystem aus erstellt wird und somit generell als qualitativ gilt. Prägnanter formuliert: Es gibt keine nicht-qualitativen Daten. Die Unterscheidung qualitativ und quantitativ ist im Datenkontext simplifizierend und unzureichend im Sinne der generellen Eigenschaften von Daten.

Struktur: strukturiert oder unstrukturiert

Eine weitere formale Unterscheidung findet sich auch in der strukturellen Beschreibung nach Kitchen wieder. Als *strukturiert* werden solche Daten bezeichnet, die einem einheitlichen Datenmodell folgen. Die Konsistenz dieses Formates erlaubt die einfache Kombination, Prozessierung und Analyse solcher Daten. Strukturierte Daten sind die Bedingung für eine maschinenlesbare Aufbereitung, die wiederum Voraussetzung für die automatisierte Datenanalyse ist. Als *unstrukturiert* werden im Gegensatz dazu solche Daten bezeichnet, die kein einheitliches Modell als Grundlage haben, was verdeutlichen soll, dass solche Daten schwer miteinander zu kombinieren sind. Durch die Entwicklungen der vernetzen Informations- und Kommunikationstechnologie sollen solche unstrukturierte Daten mittlerweile die große Mehrheit bilden, die es gilt, vor allem im Rahmen der Entwicklungen von Anwendungen wie Machine Learning automatisiert zu restrukturieren.²¹⁴ Kein einheitliches Modell heißt aber nicht, dass diese nicht nachzuvollziehen wären. Es gibt keine unstrukturierten Daten in meiner grundsätzlichen Definition von Daten. Die Idee der unstrukturierten Daten verweist vielmehr auf eine kulturelle Ignoranz gegenüber der Komplexität von verschiedenen Ursprüngen von Daten.

Quelle: captured, exhaust und derived data

Die Unterscheidung der Quelle interessiert sich für die verschiedenen Arten der Datengeneration. Kitchen übernimmt mit der Bezeichnung „captured“ und „exhaust“ Narrative des Datenexzeptionalismus.

²¹⁴ Lakshmikanthan, 2019.

lismus. Weder sind Daten eine natürliche Ressource, die gefangen werden kann, noch ist die Datengeneration ein natürlicher Prozess, der von sich aus Daten generiert. Nochmals zur Erinnerung meine vorherige Beschreibung zu den Eigenschaften von Daten: „Daten sind eine bewusste Entscheidung. Daten sind gewollt.“ Unter diesem Vorbehalt ist seine Unterscheidung dennoch hilfreich, um die Narrative und die Arbeitsweise mit Daten zu verstehen. Mit „Captured data“ meint Kitchen eine absichtliche Datengeneration durch verschiedene Arten von Messungen, wie beispielsweise durch manuelle Experimente im Labor oder die automatisierte Messung durch Sensoren.

Im Gegensatz dazu beschreibt er mit „exhaust data“ solche Daten, die als Nebenprodukt und nicht als Hauptziel von Systemen oder Apparaten entstanden sind. Sein Beispiel ist die elektrische Registrierkasse, die neben den primär notwendigen Daten zum Einkauf und zur Bezahlung auch Daten zum Warenbestand, der Arbeitsleistung und dem Konsumentenverhalten produziert. An diesem Beispiel wird deutlich, dass die Idee von Daten als Nebenprodukt nur mit einem naiven Verständnis von technischen Funktionsweisen aufrechtzuerhalten ist. Sicherlich kann eine Hauptfunktion eines Apparates oder Systems definiert werden, aber der tatsächliche Funktionsumfang und Datenhaushalt fällt nicht einfach als Nebenprodukt an, sondern wird bewusst und mit Absicht erdacht, gestaltet und umgesetzt. In dieser Lesart sind Kitchens „exhaust data“ eher bewusst verborgene Datensätze.

Als konzeptionelles Gegenstück führt Kitchen neben den zwei Quellenarten, die Daten an sich erst generieren, mit „derived data“ eine dritte Art ein. Dabei werden Daten durch die Weiterverarbeitung von bereits existierenden oder „captured data“ produziert. Der Begriff ist gut geeignet, um die komplexen Transformationsebenen von Datenverarbeitung und -analyse zu greifen. Wenn man aber auch die „captured data“ nur als eine Ableitung von einem bestimmten Modell von Daten begreift, dann ließen sich alle Daten generell als „derived“ bezeichnen.

Produzent: primary, secondary und tertiary

Die vierte Kategorie bietet Anlass über die Produktionsbedingungen von Daten nachzudenken. Bei Kitchen werden drei Ordnungsstufen – „primary“, „secondary“ und „tertiary“ – definiert. „Primary“ sind Daten, die von einem Subjekt für ein bestimmtes Anliegen

generiert worden sind. Die „secondary“ Daten sind die von einer anderen Partei verwendeten und ausgewerteten „primary“ Daten der Ursprungspartei. „Tertiary“ Daten sind abgeleitete Kategorien und statistische Zusammenfassungen der „primary“ und „secondary“ Daten. Daten des „primary“-Typs gelten als Idealtyp, da sich nur bei ihnen die Validität ihrer Entstehungsmethodik überprüfen und reflektieren lässt. Auf der anderen Seite sind die „secondary“ und „tertiary“ Daten aus der beschriebenen Definition heraus nie für ihren spezifischen Anwendungszweck intendiert, werden also „zweckentfremdet“. Allerdings sind „secondary“ und „tertiary“ Daten wesentlich *möbiler*, d.h., sie können kombiniert, geteilt und angereichert werden oder sie werden gar kommodifiziert.²¹⁵ Die Datennarrative des Datenexzptionalismus, also beispielsweise Big Data, bauen auf der Idee der Aggregation von „secondary“ und „tertiary“ Daten.

Typ: Attributdaten, Indexdaten und Metadaten

Als letzte Kategorie führt Kitchen in generelle Datentypen ein. Er unterscheidet zwischen „attribut“, „indexical“ und „metadata“. „Attribut data“ nach Kitchen sind Daten, die Aspekte eines Phänomens wiedergeben, welche künstlich mit ihnen assoziiert wurden. Darunter fallen kulturelle Vorstellungen wie Alter, Geschlecht oder Gewicht bei personenbezogenen Beschreibungen. Die meisten Daten beschreiben solche rein deskriptiven Zuschreibungen von Attributen. Dagegen stellt er die Idee von „indexical data“, die ihren Namen nach zum relationalen Verweis oder zur Identifikation bzw. Authentifizierung dienen. Solche eineindeutigen Datenkennzeichnungen finden ihre Anwendung zum Beispiel bei Kreditkartennummern oder IP-Adressen. Im Falle der Personenbeschreibungen wären Fingerabdrücke etwa „indexical data“. Nach der Deskription und der Relation als Datentypen folgen die Metadaten. Sie sind vereinfacht gesagt Daten über Daten selbst. Sie vereinen deskriptive Elemente zur Beschreibung der Datenelemente und deren Ursprung, strukturelle Elemente über die Organisation der Daten und administrative Elemente, die technische Details, wie Erstellungsdatum und Datenformate, dokumentieren. Das Wissen der Metadaten über die Daten verleiht manchen zu großen Schlussfolgerungen:

²¹⁵ Srnicek, 2016.

„Wenn man genug Metadaten hat, braucht man eigentlich keinen Inhalt mehr.“

– Stewart Baker (NSA), zitiert in Pasquinelli, 2019a, S. 78.

So sind es weniger die Unterscheidungen der Datentypen selbst, die bestimmten Datennarrativen unterlegen sind, als viel mehr der Gebrauch dieser Datentypen.

Insgesamt konnte ich durch das Aufzeigen der konzeptionellen Fragilität von üblichen Datenunterscheidungen Aspekte ihrer kulturellen Projektion herausstellen. Während Unterscheidungen bezüglich ihres generellen Typs auf besondere Akzente in der Datenverwendung verweisen, sind die gezogenen Unterscheidungen – in ihrer Form, ihrer Struktur, ihrer Quelle und ihrer Produktion – vor allem Ausprägungen eines bestimmten Datennarratives. Teilweise sind die Datentypisierungen an sich widersprüchlich oder verweisen auf naturalisierende Konzepte von Daten.

Obwohl in der von Rob Kitchin strukturierten Typisierung die Arten der Daten näher beschrieben werden sollen, wird Datenbegriff durch simplifizierende Konzeptionen verwässert. Typisierungen, wie „captured data“, helfen demnach nicht die Komplexität der Datenerstellung besser zu strukturieren, sondern reduzieren den Datenbegriff auf Datennarrative, wie eben das natürliche Vorkommen der Daten. Die Datentypisierungen tendieren zur Priorisierung von Dateneigenschaften, die der Datenaffinität gerecht werden sollen. Es geht in solchen Typisierungen nicht darum, sich allgemeinen Eigenschaften von Daten, wie ich sie vorab definierte, zu vergegenwärtigen. Es soll viel mehr ein sehr spezifischer Datentyp abgegrenzt werden. Kurzum: Der Datenexzensionalismus ist eine bewusste Reduktion des Datenbegriffs auf bestimmte Dateneigenschaften und Datentypen.

3.3 Von Messwerten zu elektronisch gespeicherten Zeichen

Im Datenexzptionalismus wird ein bestimmter Datenkontext

favorisiert: numerische und computertechnische Datenstrukturen. Im Gegensatz zu meinem vorgeschlagenen Datenbegriff, der letztlich jede eineindeutige Abstraktion potenziell als Datum zulässt, sind Daten in diesem Verständnis vor allem Zahlenwerte. Ob als empirische Messung, statistische Betrachtung oder Computerdatei: Als Daten versteht man dort Zahlen.²¹⁶ Natürlich können Daten auch in numerischer Form vorliegen. Das ist nicht mein Kritikpunkt. Ein Problem sehe ich, wenn ein Datenbegriff primär oder gar ausschließlich von einer Zahlenvorstellung geprägt wird. Diese Formung des Datenbegriffs ist eine spezifische kulturelle Entwicklung. Die Reduktion des Datenbegriffs auf statistische und computertechnische Datenkategorie folgt einer längeren Traditionsgeschichte, auf die ich im Folgenden aufmerksam machen will.

Im historischen Blick auf die Quantifizierung ist der Datenbegriff durch die Idee des Messwertes und der statistischen Auswertung geprägt. Tiefergehende Betrachtungen zur Entwicklung dieser Idee wurden bereits durch die Disziplinen der Statistikgeschichte und Wissenschaftsgeschichte etabliert. Diese Perspektiven sind wichtig, um zu verstehen, unter welchen Bedingungen und mit welchen Maßnahmen sich die Affinität zu Daten entwickelte. Die Standardwerke für diese Betrachtung sind zum Beispiel Theodore M. Porters „Rise of Statistical Thinking, 1820-1900“,²¹⁷ indem er die Etablierung der Statistik durch die Sozialwissenschaften im 19. Jahrhundert beschreibt. Alain Desrosières’ zeigt in „The Politics of Large Numbers“ die Verbindung von modernen Regierungen und der Entwicklung statistischer Methoden.²¹⁸ In einem ähnlichen Modus betrachtet James C. Scott in „Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed“ die Ursprünge der numerischen Standardisierung und Rationalisierung im Entstehen der frühen modernen Nationalstaaten.²¹⁹ Auch Scott betont dabei eingehend die Intentionalität der zahlenbasierten Staatsmethoden:

²¹⁶ Vgl. Gießmann und Burkhardt, 2014.

²¹⁷ Porter, 1986.

²¹⁸ Desrosières, 1998.

²¹⁹ Scott, 1998.

„These state simplifications, the basic givens of modern statecraft, were, I began to realize, rather like abridged maps. They did not successfully represent the actual activity of society they depicted, nor were they intended to, they represented only that slice of it that interested the official observer.“

– Scott, 1998, S. 3.

Als Kontrast zu den einheitlich männlichen Stimmen aus dem Diskurs der

1980er und 1990er Jahren soll noch Michelle Murphys Publikation „The Economization of Life“ als weiteres Werk der Wissenschaftsgeschichte zur Beschreibung der Entwicklung der statistischen Verhältnisse genannt werden.²²⁰ Darin beschreibt sie das Aufkommen von statistischen Techniken zur Regulierung und Ökonomisierung von Bevölkerungen unter einer biopolitischen Perspektive. Obwohl alle vier genannten Stimmen im Detail unterschiedliche Ansätze verfolgen, eint alle die Idee des Ursprungs der Statistik und damit auch der Datenverarbeitung im Zusammenhang mit dem Aufkommen der souveränen Nationalstaaten im 18. Jahrhundert und die Etablierung dieser Methoden im 19. Jahrhundert. Der etymologische Ursprung des Begriffs Statistik (lat. „statisticus“, was soviel wie staatswissenschaftlich bedeutet) unterstützt den historischen Verweis auf die Staatstheorie.²²¹

„Die Statistik ist die erste politische Wissenschaft! Ich kenne den Kopf eines Menschen, wenn ich weiß, wieviel Haare er produziert.“

– Marx, 1842.

Aus dem statistisch-empirischen Verständnis entwickelte sich eine weitere

dominante Zahlenkonzeption gegenüber Daten. Statistiken bedeuten eine Vielzahl von Metriken über die Gesellschaft. In demographischen, ökonomischen, medizinischen oder politischen Daten werden parametrisierte Faktoren gesellschaftlichen Lebens gemessen und analysiert.²²²

Durch die Etablierung der Computertechnologie werden Daten nicht nur als Zahlenwerte an sich, sondern als Zahlensystem für die Operationen des Computers konzipiert. Daraus folgt die Beschreibung von Daten als elektronische Form und daran anschließende Hoffnungen bezüglich ihrer Kommunizierbarkeit.

²²⁰ Murphy, 2017.

²²¹ Desrosieres, 1993, S. 200.

²²² Vgl. Destatis, 2021.

Daten werden in solchen Konzeptionen dann als exklusives Phänomen der Informations- und Kommunikationstechnologie gedeutet:

„Durch neue Möglichkeiten der informationstechnologischen Speicherung, Verwaltung, Bearbeitung und Übertragung setzte sich nach der Kybernetisierungswelle in der Mitte des 20. Jahrhunderts der Datenbegriff gegenüber dem wissenschaftstheoretischen tradierten Messwertbegriff durch und verselbstständigte sich.“

– Ngo, 2019, S. 2.

Dieser

spezielle

Datenbegriff als Begleiteffekt der Computerisierung ist eine wichtige Komponente zur Schärfung meines Konzepts des Datenexzepationalismus. Natürlich sind Daten der elementare Baustein für die Prozesse der Digitaltechnik. In ihrer binären Logik werden Daten dabei als Dualzahl gespeichert. Bei Digitalrechnern passiert das auf Grundlage der diskreten Zahlenrepräsentation üblicherweise mit 0 und 1, was übersetzt auf die meist in Digitalrechnern verwendete Elektrotechnik auch *Strom an* und *Strom aus* bedeutet. Eine Unterscheidung in 0 oder 1 ist ein sogenannter bit, ein Kofferwort aus *binary digit*, und ist damit die kleinstmögliche Datenmenge. (Ein Byte, oft acht Bit aber nicht zwingend, ist kleinste adressierbare, also funktionale, Datenmenge eines technisches Systems.) Für den Datenbegriff des Computers (im Folgenden vereinfacht verwendet für Digitalrechner mit Elektrotechnik) gilt mit Peter Weibel gesprochen:

„Datentheoretisch gilt also als Realität, was mathematisch repräsentierbar und elektronisch schaltbar ist.“

– Weibel in Vrachliotis und Weibel, 2019, S. 17.

Noch prägnanter

formuliert es Friedrich Kittler:

„Nur was schaltbar ist, ist überhaupt.“

– Kittler, 1993, S. 182.

Demnach ist für den Computer ein

Gegenstand nur erfassbar, wenn er datentechnisch prozessiert werden kann. Ein Teil dieser Reduktion des Datenbegriffs auf quantitative Formen ist bereits aus der Idee von Daten als statistischer Messwert bekannt geworden. Ein zweiter Aspekt nach Weibels Definition ist die elektronische Datenverarbeitung. Wie William Aspray in „Computing Before Computers“ darstellt ist die Ideengeschichte von Computertechnologie allgemein um einiges länger

als die ihrer elektrotechnischen Umsetzung.²²³ Dort werden zum Beispiel frühe Konzepte der Kalkulation auch abseits der westlichen Kultur,²²⁴ erste logische Maschinen, Lochkartentechnik und analoge Computer besprochen.²²⁵

Die elektronische Schaltung ist erst mit den Computern, die ab den 1940er Jahren gebaut wurden, möglich. Solche auch *stored-program computers* genannten Computer speichern Maschinenbefehle elektrisch lesbar.²²⁶ Diese Maschinenbefehle dienen zur Programmierung des Computers. Programmierung meint die Manipulation der in der Hardware des Rechners gespeicherten Daten und wird unter dem Begriff Software zusammengefasst.²²⁷ Die Programmierung wurde zunächst manuell erledigt, aber später durch Programmiersprachen automatisiert. Letztlich werden die Programmieranweisungen, in Summe Programm genannt, auch als Daten in der Hardware des Computers gespeichert.²²⁸ Die Aufbereitung von analogen Daten für Digitaltechnik und ihre Verarbeitung und Speicherung innerhalb solcher Systeme wird auch als „Digitalisierung“ bezeichnet.

3.4 Von der Datafizierung zur Digitalisierung

In einer vereinfachten Gleichsetzung wird im allgemeinen Sprachgebrauch die Digitalisierung und die Anwendung bzw. Verbreitung von Computertechnologie zum Synonym. Die digitale Grundlogik der modernen Rechnerarchitektur wird mit der realpraktischen

²²³ Aspray, 1990.

²²⁴ Siehe dazu auch den Einfluss der traditionellen chinesischen Schrift „I Ching“ auf Leibniz’ Entwicklung der binären Logik, bspw. bei Perkins, 2009.

²²⁵ Siehe dazu auch Schneider, 2007.

²²⁶ Die heute populärste Variante dieser Anordnung ist die Von-Neumann-Architektur, die Daten und Maschinenbefehle, vereinfacht ausgedrückt, am selben Ort speichert.

²²⁷ Wie wichtig die materielle Hardware ist und die Software nur den Blick auf diesen verstellt, merkte vor allem Friedrich Kittler im Kapitel „Es gibt keine Software“ seines Buches „Draculas Vermächtnis“ (Kittler, 1993, S. 225) an.

²²⁸ Charles Petzold beschreibt in „Code - The Hidden Language of Computer Hardware and Software“ anschaulich die prinzipiellen Wirkungsweise von Computern auf Hardware- und Softwareebene; Petzold, 1999.

Dimension technischer Infrastrukturen zusammengeworfen – *alles was den Computer betrifft ist digital*. Eine solche Konzeption des Digitalisierungsbegriff erweist sich als folgenreich für die Vorstellung von Daten. Anstatt die Digitalisierung als konzeptionelle Struktur von Daten zu sehen, werden Daten in diesem Narrativ als Folgeerscheinung von Digitaltechnik gerahmt. Der Datenexzensionalismus macht die allgegenwärtige Computertechnologie zur Hauptursache von Daten. Tatsächlich sind Daten und ihre Digitallogik aber die Bedingung für die Operationen von Computerprozessen. Im Folgenden weise ich auf die Implikationen einer solchen Gleichsetzung hin.

Im computertechnischen Verständnis von Digitalisierung werden analoge Werte, verstanden als aufbereiteter zeit- und wertkontinuierlicher Input, innerhalb einer technischen Umgebung (Analog-Digital-Wandler) digitalisiert, sprich in eine zeit- und wertdiskrete Abfolge quantisiert, und somit für die Digitaltechnik zugänglich gemacht. Durch diese technische Anordnung bestärkt sich eine bereits im Diskurs viel besprochene Opposition von Analogem und Digitalem, die nicht in dieser Technik begründet liegt, aber deutlich prominent wird. In Schröters und Böhnkes „Analog/Digital – Opposition oder Kontinuum?“ (als eine deutschsprachige Referenz dieses Diskurses) wird die „medienhistorische und -theoretische Leitdifferenz der zweiten Hälfte des 20. Jahrhunderts“ eingehend verortet; durch Claus Pias werden weitere begriffliche Folgeentwicklungen nachvollzogen:

„So wirbeln die regelmäßig wiederkehrenden Unterhandlungen um ‚analog‘ und ‚digital‘ jedes Mal neue Begriffspaare an die Oberfläche: Entropie versus Information, kontinuierlich versus diskontinuierlich, linear versus nichtlinear, Ereignis versus Wiederholung, Wahrscheinlichkeit versus Unwahrscheinlichkeit, Reales versus Symbolisches, Natur versus Artefakt, usw.“
– Pias, 2004, S. 299.

In Analog-

Digital-Übertragungen werden üblicherweise natürliche Prozesse als prinzipiell analog (kontinuierlich) gegenüber den Datenprozessen des Computers als digital (diskret) konzipiert. Eine solche Differenzierung hält sich jedoch nur auf theoretischer Ebene, da jede digitale Struktur für eine menschliche Wahrnehmung im Analogen existieren muss:

„Weil auch der menschliche Sinnesapparat analog funktioniert, müssen alle digitalen Botschaften irgendwann wieder ins Analoge umgesetzt werden.“
– Winkler, 2008, S. 129.

Digitale Phänomene sind dieser Definition nach nicht wahrnehmbar. Jedes digitale System braucht einen analogen Träger, um für eine menschliche Rezeption bzw. den Gebrauch durch Menschen nutzbar zu sein. Das digitale Notensystem eines Tascheninstruments ist beispielsweise ohne Klaviatur nicht bedienbar. So schrieb auch Norbert Wiener als Denker der Kybernetik, eine Disziplin in deren ersten Zusammenkünften der historische Ursprung der Analog-Digital-Unterscheidung liegt:

„Every digital device is really an analogical device.“
– Norbert Wiener in Pias, 2003b, S. 158.

Trotz dieser Einsichten werden Daten durch die Analog-Digital-Distinktion durch ihre digitale Prägung, wie in der zitierten Definition in der Datenstrategie der Bundesregierung, als synonym mit Computertechnologie betrachtet. Mit Friedrich Kittler gesprochen wird die „Welt des Symbolischen“ eine „Welt der Maschine“.²²⁹ Eine solche verkürzte Konzeption zeigt sich in Wortanordnungen wie *digitale Daten*, *digitale Medien*, oder die besprochenen *Digital Humanities*. Wie bereits am Anfang des Kapitels am Beispiel der Digital Humanities diskutiert, ist digital im Gegensatz zu seiner ursprünglichen Bedeutung dort eine Kennzeichnung, die veränderte Bedingungen und Arbeitsweisen durch Computertechnologie markieren will. Durch die zunehmende Verbreitung von Digitaltechnik in der Gesellschaft (Computerisierung) wurde der Begriff digital nicht nur als Synonym für die binäre Logik der elektronischen Digitalrechner, sondern weitergehend zum Synonym für jegliche Phänomene mit einer willkürlichen Verbindung zu allen Formen technisch vernetzter, digitaler Kommunikation verkürzt. Diese doppelte Reduktion begleitet durch das Narrativ des Neuen und des Fortschritts führt zu einem simplifizierten Begriff des Digitalen,²³⁰ der zu sehr flexiblen Wortanordnungen zur Beschreibung von gesellschaftlichen Prozessen, wie etwa *digitale Transformation*, *digitaler Wandel* oder *digitale Revolution*, führt:

²²⁹ Kittler, 1993, S. 58.

²³⁰ Distelmeyer, 2014, S. 167.

„Etliche Kombinationen mit ‚digital‘ sind sowohl semantisch als auch grammatisch fragwürdig oder grenzen an Ironie.“
– Baier et al., 2017, S. 35.

Für meine

Argumentation ist es an dieser Stelle wichtig zu prüfen, was für die weitere Betrachtung als digital gelten kann. Digital verweist im lateinischen Wortursprung „*digitus*“ („Finger“) oder „*digitalis*“ („zum Finger gehörend“) generell auf diskrete, also eindeutig bestimmbare, oder etymologisch abgeleitet auf *zeigbare* Einheiten.²³¹ Der Prozess der „Digitalisierung“²³² meint dann ganz allgemein die „symbolische Umschreibung physischer Eigenschaften in eine symbolische, meist numerische Repräsentation“.²³³ Im Bezug zu Daten sprach ich bereits über ihre grundsätzlich symbolische Form, die jedoch nicht zwangsläufig numerisch sein muss. So formuliert es auch Nelson Goodman im Zuge seiner Symboltheorie:

„Plainly, a digital system has nothing special to do with digits [...] The characters of a digital system may have objects or events of any kinds as their inscriptions.“

– Goodman, 1976, S. 160.

In ihrer Betrachtung zur Kulturgeschichte der Digitalisierung betont Sybille Krämer in ähnlicher Weise die generelle Diskretisierung als Kern der Digitalisierung.

Sie nennt dabei das Alphabet als Prototyp eines digitalen Systems. Die gesprochene Sprache wurde in eine digitale Zeichensequenz transformiert, die sich als Schrift etabliert hat.²³⁴ Digitale Systeme können aber auch Zählsysteme, wie das Binärsystem, sein, die ebenso eine lange Traditionsgeschichte haben.²³⁵ Die Bedeutung von digital mit Bezug auf Digitaltechnik wurde dagegen erst seit

²³¹ Vgl. Boden und Edmonds, 2019.

²³² Erste Nennung im Dt. 1985: „Wieviel Informatik braucht der Mensch? Ist der Computer im Klassenzimmer die neue große Lernhilfe, die gequalten Pädagogen zu der beglückenden Erfahrung verhilft, daß es doch noch fleißige, begeisterungsfähige, konzentriert lernende Schüler gibt? Oder macht er, im Gegenteil, jede pädagogische Anstrengung zunicht, förderte er die schleichende Digitalisierung des Bewusstseins, produziert er, als Medium der Welterfahrung, Menschen, die, wie der Computer nur zwischen eins und null, ihrerseits nur zwischen schwarz und weiß zu unterscheiden vermögen?“, Knüpfer, 1985.

²³³ Bächle, 2016, S. 75.

²³⁴ Krämer, 2020b.

²³⁵ Wichtige Konzepte der Ideengeschichte des Computers finden sich bei Boole, Leibniz und Llull, siehe beispielsweise Uckelman, 2010.

Mitte des 20. Jahrhunderts etabliert.²³⁶ In dieser historischen Linie betrachtet ist das gezählte Symbolische eher die Ausnahme, zumindest eine relativ junge Erscheinung, als die Regel von digitalen Anordnungen.

Neben der rein symbolischen Ebene des Digitalen, die vor allem Computerphänomenen zugerechnet wird, werden innerhalb des Digitaldiskurses auch die zeigenden bzw. indexikalischen Eigenschaften des Digitalen besprochen. So verweist beispielsweise Benjamin Peters im Sinne der Peirce'schen Zeichentriade darauf, dass Digitalsysteme, verstanden als Zeichensysteme, neben ihrer symbolischen Abstraktion eines realen Objekts auch auf die realen und konstituierenden Ordnungssysteme außerhalb eines Ähnlichkeitsbezugs verweisen.²³⁷ Bei Computermedien wird zur Beschreibung der indexikalischen Anordnung oft auf die inneren Logiken und Prozesse des Computers (Software-Protokolle, Hardware-Schaltungen oder grundsätzlich Strom etc.) verwiesen, die das digitale Symbol erst ermöglichen. Als ein Beispiel für digitale Medien pre-computatorischer Natur nennt Peters die Abbildungen von politischen Führungspersonen auf Geldwährungen, die in ihrer Bildsprache dann auf politische Institution als ihren Ursprung und Ordnungshintergrund verweisen.²³⁸ Ohne an dieser Stelle wesentlich tiefer in den weiten Digitaldiskurs eintauchen zu wollen, ist für meine Argumentation wichtig, dass sich der Digitalbegriff nicht im rein technischen Bezug zu Computerphänomenen erschöpft. Neben der symbolischen Abstraktion verweist das Digitale auch auf real-weltliche bzw. materielle Anordnungen.

Um dieses Geflecht von symbolischer und materieller Abhängigkeit auf der Ebene von Digitaltechnik zu reflektieren wird im Digitaldiskurs eine bestimmte Auslegung des Begriffs der *Digitalität* benutzt. In „Kultur der Digitalität“ definiert Felix Stalder Digitalität als „jenes Set von Relationen, das heute Basis der Infrastruktur digitaler Netzwerke in Produktion, Nutzung und Transformation materieller und immaterieller Güter sowie in der Konstitution und Koordination persönlichen und kollektiven Handelns realisiert wird.“²³⁹ In seiner „Kritik der Digitalität“ defi-

²³⁶ Distelmeyer, 2014, S. 165.

²³⁷ Peters, 2016, S. 98.

²³⁸ Ebd.

²³⁹ Stalder, 2016, S. 18.

niert Distelmeyer ähnlich: „[Digitalität] bezeichnet die Gesamtheit und Eigenart der Bedingungen und Folgen elektronischer Digitalcomputer in all ihren Formen.“²⁴⁰ Für Distelmeyer ist die abzubildende Komplexität des Begriffs Digitalität eine „Zumutung“, vor allem durch die „mythische“ (nach Roland Barthes) Aufladung des Begriffs, dem er durch die Einführung des Neologismus „Digitalität“ begegnen will. Diese neueren Diskurse um den Digitalbegriff sind gewinnbringend, um die kulturellen Neuprägungen des Begriffs im Kontext seines Bezugs zur Digitaltechnik und die daraus resultierenden Folgen zu reflektieren.

Für mein Anliegen der Verortung des Datenbegriffs will ich jedoch noch vor den Implikationen der Digitaltechnik ansetzen. Daten in ihrer Digitallogik verstehe ich als Bedingung für Digitaltechnik. Mein Verständnis ist damit ein Gegensatz zu einer technikdeterministische Auslegung von Daten, die diese als Folge der Digitaltechnik positioniert. So lässt sich durch die Besprechung der Narrative des Digitalen für meinen Datenbegriff definieren, dass Daten als diskrete symbolische Form grundsätzlich immer digital sind. Daten liegt immer ein diskretes, also auch digitales, Ordnungssystem zugrunde. Die Wortschöpfung *digitale Daten* ist eine Tautologie und gleichzeitig ein Symptom einer technozentristischen Perspektive, die den Datenbegriff auf Computerphänomene reduzieren will. Die Umkehrung dieser Digital-Daten-Relation gilt allerdings nicht bedingungslos. Nicht alle digitalen Anordnungen sind gleichzeitig auch Daten. Die Digitalisierung – verstanden als Transformation in symbolische Formen – ist nur eine Voraussetzung, die dann zum Beispiel durch die Intentionalität der Datengenerierung auch angewandt werden muss.



Abb. 28 FDP-Wahlplakat, 2017.

²⁴⁰ Distelmeyer, 2021, S. 2; neben rein technischen Formen meint dies auch diskursive und ideologische Dimensionen der „Digitalisierung“.

Die besprochenen Narrative des Digitalen, wie die genannte *digitale Revolution* oder der *digitale Wandel*, zeigen Aspekte des Datenexzptionalismus, indem sie diese Abhängigkeit umkehren. Statt Digitalisierung als konzeptionelle Struktur von Daten zu sehen, werden Daten in diesem Narrativ als Folgeerscheinung von Digitaltechnik gerahmt. Wenn digital mit jeglichen Computerphänomenen synonym wird, dann entstehen monokausale Vorstellungen, wie die von *digitalen Daten*, deren Erscheinen exklusiv auf den Rechnern zurückzuführen ist. Eine folgenreiche Simplifikation, die im Zuge der digitalen Transformation die uneingeschränkte Analyse von Daten zu jedem Zweck rechtfertigt. Nach dem Motto: Die Daten sind *jetzt da* und müssen daher auch *genutzt* werden. Durch die Gleichsetzung von Digitaltechnik und Daten werden Ursache und Effekt der Digitalisierung unklar. Ist die Datafizierung der Gesellschaft, in eine „Datengesellschaft“²⁴¹ eine Folge der Digitalisierung im Sinne der Computerisierung? Oder ist die Digitalisierung die Bedingung damit Digitaltechnik sich derart in der Gesellschaft etablieren konnte? Ich möchte an dieser Stelle nicht argumentieren, dass Digitaltechnik und ihre Verwendung keine weitreichenden Folgen auf jeglicher gesellschaftlichen Ebene hätte. Es gibt zweifelsohne einen deutlichen gesellschaftlichen Wandel bzw. Transformationen, die sich allerdings nicht allein durch Technik bedingen. Es bedarf daher eines komplexeren Digitalisierungsbegriffs, denn die simple Gleichsetzung mit der Verbreitung von Digitaltechnik verstellt den Blick auf kontextuelle Bedingungen (sozial, politisch, wie ökonomisch) und hat damit weitreichende Folgen für die Rezeption und Diskussion der Digitalisierung. Solche Erweiterungen finden sich bei Forscher:innen, die sich Nahe der „Science and Technology Studies“ verorten.²⁴²

Mein Interesse gilt vor allem den Voraussetzungen für eine Etablierung von Digitaltechnik und der grundsätzlichen Wechselwirkung zwischen Digitalisierung und Gesellschaft. Es ist vor allem der Blick der Soziologie auf die Datengesellschaft, die für mich die Gründe der Datenaffinität herausarbeiten kann. Gerade jüngere soziologische Betrachtungen fragen nach solchen Formen der Erscheinung und danach, wie Digitalisierung auf gesellschaftlicher Ebene Anklang gefunden hat und zu verstehen ist. Im Folgenden soll nicht der gesamte soziologische Diskurs, sondern vor allem prominente Stimmen des deutschsprachigen Diskurs für eine Erweiterung des Digitalisierungsbegriff eingeführt werden.

²⁴¹ Prietl und Houben, 2018.

²⁴² Vgl. Chun, 2021; Distelmeyer, 2021; Galloway, 2021 und Peters et al., 2021.

3.5 Zugänge zur digitalen Gesellschaft

Digitalisierung wird unter soziologisch verorteten Perspektiven nicht als exklusives Phänomen der Digitaltechnik betrachtet, sondern es wird versucht sie als Eigenschaft von sozialen Praktiken zu greifen. Im Folgenden beschreibe ich Ansätze, die eine Datenaffinität nicht aus einer technischer Anordnung heraus begründen, sondern Gesellschaft grundsätzlich als datenaffin beschreiben – als Datengesellschaft. Anstatt Datenphänomene als Ausnahmeerscheinung von Computertechnologien zu begreifen, werden digitale Ordnungsprinzipien als konstituierend für die moderne Gesellschaft gedacht. Solche Theorien zur digitalen Gesellschaft bilden für mich produktive Gegenkonzepte zum Datenexzentialismus.

In der Veröffentlichung „Digital Constellation“ des Weizenbaum Instituts wird der soziologische Forschungsstand zu einem nuancierteren Bild des Wechselspiels zwischen Technologie und Gesellschaft gerahmt. Die Publikation bestätigt, dass das Konzept der digitalen Gesellschaft in der Soziologie bereits als ein zentrales Forschungsthema etabliert ist. Während es schon länger Diskurse gibt, die sich an den gesellschaftlichen Folgen der etablierten Informations- und Kommunikationstechnologie für die Gesellschaft abarbeiten, konzentrieren sich diese neueren Betrachtungen abseits technologischer Determinismen um soziale Veränderung und andere Aspekte der Digitaltechnik.²⁴³

Das Weizenbaum Instituts teilt den deutschsprachigen Diskursraum in zwei grundsätzliche Ansätze: erstens handlungstheoretisch aus mikrosoziologischer Perspektive und zweitens systemtheoretisch aus makrosoziologischer Perspektive. Stellvertretend wird der Diskurs mit vier Autoren dargestellt: Andreas Reckwitz und Felix Stalder als Vertreter der handlungstheoretischen Ansätze, Dirk Baecker und Armin Nassehi mit einem systemtheoretischen Hintergrund. Alle Perspektiven eint eine soziale Konzeption des Digitalisierungsbegriffs abseits rein technischer Bestimmungen. Ich möchte im Folgenden diese Einteilung aufnehmen, mich aber vor allem auf die makrosoziologische Perspektiven fokussieren, da ich mir von diesen Stimmen erhoffe, die digitale Struktur im Gesellschaftlichen im Besonderen herausarbeiten zu können.

²⁴³ Berg et al., 2020.

Medienkatastrophen bei Dirk Baecker

In der Publikation des Weizenbaum Instituts werden mit Dirk Baecker und Armin Nassehi zwei Repräsentanten von systemtheoretisch orientieren Ansätzen genannt, die sich auf die gesellschaftliche Strukturform in Relation zur Digitalisierung fokussieren. So unterscheidet zunächst Baecker, wie auch hier früher im Kapitel vorgeschlagen, zwischen einer engeren (Analog-Digital-Aufbereitung für Digitaltechnik) und einer erweiterten Definition des Digitalisierungsbegriffs:

„Digitalisierung im weiteren Sinne des Wortes – Digitalisierung der Gesellschaft durch Gesellschaft – ist die Erarbeitung und Erprobung abzählbarer und berechenbarer Daten im Medium analoger Widersprüchlichkeit für die Zwecke der Kommunikation von und mit Maschinen.“

– Baecker, 2018, S. 59.

Baecker beschreibt

in „4.0 oder die Lücke die der Rechner lässt“ allerdings weniger die Bedingungen, als die Zukunft einer digitalen Gesellschaft. Mit 4.0 deutet er auf vierte Medienepoche durch die Digitalisierung nach Oralisation (1.0), Alphabetisierung (2.0) und Buchdruck (3.0). Die Digitalisierung versteht er als „Medienkatastrophe“, die sich dadurch definiert, dass die elektronischen Medien, wie auch die anderen alten neuen Medien zuvor, etablierte Strukturen herausfordern und umwandeln.²⁴⁴ Ganz konkret sei die Strukturform der „nächsten“ Gesellschaft nicht mehr die funktionale Differenzierung der Epoche des Buchdrucks, sondern das Netzwerk.²⁴⁵ Ohne weiter auf die Konzipierung des Netzwerkbegriffs eingehen zu wollen,²⁴⁶ zeigt sich in Baeckers Argumentation zur Medienkatastrophe eine Nähe zum Datenexzptionalismus. Auch wenn sein Datenbegriff in Relation zur Netzwerkidee noch genauer zu bestimmen wäre, wird bei Baecker eine ähnliche Idee von radikalem Umbruch durch die Digitaltechnik deutlich.

Baecker übernimmt dieses Modell einer Mediengenealogie von Marshall McLuhans Medienepochen aus „The Guten-

²⁴⁴ Baecker, 2018, S. 29.

²⁴⁵ Ebd., S. 26.

²⁴⁶ Eine weitere Diskussion findet sich bei Gießmann, 2016 und Stäheli, 2021 der die Grenzen der Vernetzung in seiner „Soziologie der Entnetzung“ zeigt.

berg Galaxy“,²⁴⁷ welches wiederum auf Harold Innis’ Forschung basiert.²⁴⁸ Ausgegangen wird dabei von einem medialen „a priori“, welches durch die Dominanz bestimmter (u.a. technischer) Medien die Handlungsmöglichkeiten einer Gesellschaft soweit beeinflusst sieht, dass diese zeitgeschichtliche Epochen prägen. Obwohl diese Idee des medialen „a priori“ in den Medienwissenschaften bereits als „deterministisch, technozentristisch [und] monokausal“ kritisiert wurde,²⁴⁹ konnte sich die „Behauptung der mentalitätsprägenden Kraft von Medientechnologien“ dennoch vor allem im Diskurs um „digitale Medien“ folgenreich etablieren, was Claus Pias bereits weitergehend diskutiert hat.²⁵⁰

Digitalisierung als Problemlösung bei Armin Nassehi

Im Gegensatz zu Dirk Baeckers Zukunftsprojektionen geht Armin Nassehi in seiner Publikation „Muster – Theorie der digitalen Gesellschaft“ der Frage nach Gründen für die Digitalisierung grundsätzlich nach: Wenn Digitalisierung die Lösung sein soll, für welches gesellschaftliche Problem ist es dann eine Lösung?²⁵¹ Für Nassehi geht diese Frage keiner genauen Definition oder den Problemen von Digitalisierung nach, sondern zielt auf die gesellschaftlichen Funktion von Digitalisierung.²⁵² Im Diskurs um das Digitale, Digitalisierung und ihre Folgen grenzt er seine Fragestellung gegenüber solchen Reflexionen innerhalb des sozial- und kulturwissenschaftlichen Umfeldes ab, die sich vornehmlich mit den technischen Möglichkeitsräumen und gesellschaftlichen Folgen von Digitalisierungsphänomenen beschäftigten.²⁵³

Während solche Perspektiven wichtig und essenziell für ein komplettes Bild der Digitalisierung sind, verschließen sich solche Beiträge tendenziell gegenüber den Gründen und Bedingun-

²⁴⁷ McLuhan, 1962.

²⁴⁸ Ruchatz, 1997.

²⁴⁹ Pias, 2020, S. 61.

²⁵⁰ Ebd.

²⁵¹ Nassehi, 2019, S. 12.

²⁵² Ebd.

²⁵³ Nassehi nennt dabei folgende Akteur:innen: Sherry Turkle („Leben im Netz“), Deborah Lupton („Digital Sociology“), Rob Kitchin („Data Revolution“), Dirk Helbing („Towards Digital Enlightenment“) und Shoshanna Zuboff („Age of Surveillance Capitalism“).

gen der Digitalisierung, wenn sie nur die Anwendungsfolgen der Datentechnologie betrachten. Nassehi formuliert es drastischer:

„Es sind aber Perspektiven, die sich letztlich für die Frage der Digitalisierung selbst überhaupt nicht interessieren, sondern diese als technische, gesellschaftliche und kulturelle Infrastruktur bereits voraussetzen.“

– Nassehi, 2019, S. 14.

Es ist unbestritten, dass

Digitaltechnik weitreichende Auswirkungen auf die gesellschaftliche Lebensrealität hat. Nassehi bemerkt jedoch, dass Diskussionen, wie solche über die Zukunft der Arbeit, wissenschaftliche Erkenntnisse, Fremd- und Selbstkontrolle, Macht- und Kapitalkonzentration oder kulminiert in gesellschaftlichen Narrativen und Trends wie bei der künstlichen Intelligenz, oft Digitalisierung an sich als gegeben annehmen:

„Denn all diese Diskussionen über Störungen gesellschaftlicher Routinen durch die ausgreifende Digitaltechnik kommen letztlich ohne eine fundierte Theorie der Digitalisierung aus – sie setzen die Digitalisierung als Phänomen letztlich voraus.“

– Nassehi, 2019, S. 26.

Ich teile Nassehis Einschätzung der „Digitalisierungsvergessenheit“²⁵⁴ dass konzeptionelle Verkürzungen, wie die Gleichsetzung mit Digitaltechnik, die technischen und theoretischen Bedingungen für Digitalisierung ignorieren oder unreflektiert mittragen. Nassehi verortet die Ausrichtung seiner Forschungsperspektive in der expliziten Reflexion dieser Grundbedingungen von Technologien und ihrer Anwendungskontexte, wie es bereits in den „Science and Technology Studies“ erarbeitet wurde.²⁵⁵ Da ich in Nassehis Gesellschaftstheorie ein potenzielles Gegengewicht zum Datenexzeptionalismus vermute, möchte im Folgenden in die Kerpunkte seiner „Theorie der digitalen Gesellschaft“ einführen, um dadurch auch meinen erweiterten Datenbegriff besser positionieren zu können.

²⁵⁴ Ebd., S. 15.

²⁵⁵ Ebd.

Das Bezugsproblem der Digitalisierung

Grundsätzlich beruft sich Nassehi auf eine „techniksoziologische Intuition“, die im Kontrast zum technologischen Determinismus vermutet, dass Technologien sich nur dann etablieren können, wenn sie kompatibel mit einer Struktur einer Gesellschaft sind.²⁵⁶ Sein Beispiel ist die Rundfunktechnik, die sich nur durch bereits gegebene menschliche (Publikum), politische (Infrastruktur) und ideelle Strukturen (Idee von Erreichbarkeit) einrichten konnte. Die Übertragung dieser Intuition auf Digitalisierung fragt nach den gesellschaftlichen Strukturen, die diese erst ermöglicht haben. Dahingehend arbeitet sich Nassehi, wie alle vorher genannten soziologischen Positionen, an der gesellschaftlichen Moderne ab. Eine seiner wichtigsten Thesen ist, dass die moderne Gesellschaft schon immer und damit lange vor der Etablierung von Digitaltechnik eine digitale Struktur hatte.²⁵⁷ Die Digitaltechnik ist so gedacht nur eine Konsequenz aus einer digitalen Struktur der Gesellschaft an sich.²⁵⁸ Nassehi versucht die Digitalisierung damit als gesellschaftliches Kulturphänomen zu verstehen. In seiner Theorie will er das Problem zur Lösung („Bezugsproblem“) der Digitalisierung in der Struktur der Gesellschaft selbst auffinden, wodurch sich auch der Untertitel der Arbeit „Theorie der digitalen Gesellschaft“ erklärt.²⁵⁹

Dieses „Bezugsproblem“ der Digitalisierung liegt für Nassehi in der Komplexität und in der Regelhaftigkeit der Gesellschaft selbst.²⁶⁰ Aus einer systemtheoretischen Perspektive setzt er voraus, dass alle sozialen Systeme, wie auch die Gesellschaft eines ist, durch Ordnungen („Begrenzungsformen“), wie Regelmäßigkeiten, Regeln, Traditionen, Gewohnheiten und Entscheidungen, geprägt sind.²⁶¹ Mit der steigenden Komplexität von sozialen Systemen werden diese Begrenzungsformen immer uneinsichtiger.²⁶² In der gesellschaftlichen Moderne, in der sich im Gegensatz zu den eher eindeutigeren und sichtbareren hierarchischen Ordnun-

²⁵⁶ Ebd., S. 12.

²⁵⁷ Ebd., S. 19.

²⁵⁸ Ebd., S. 11.

²⁵⁹ Ebd., S. 26-27.

²⁶⁰ Ebd., S. 36.

²⁶¹ Ebd., S. 36-37.

²⁶² Ebd., S. 37.

gen der Vormoderne (vor allem innerhalb von Institutionen) die Ordnungen vielmehr an spezifischen Funktionen (Politik, Ökonomie, Wissenschaft usw.) ausdifferenzieren, ereignete sich eine solche Zunahme der Uneinsichtigkeit in die Ordnungen.²⁶³ Dass Ordnung somit zu einem zentralen Thema der Modernisierung wurde, ist für Nassehi die strukturelle Bedingung für die Entstehung des Digitalen.²⁶⁴

Er rekonstruiert diese Entwicklung anhand von drei historischen Bruchpunkten, an denen die Gesellschaft diese Ordnungsstrukturen an sich selbst „entdeckt“ hat:

Eine erste Entdeckung der Gesellschaft als Gesellschaft vollzog sich nach Nassehi im 18. und 19. Jahrhundert mit der Entstehung der Nationalstaaten. Im Zuge der Nationalisierung etablierten sich die ersten digitalen Selbstbeobachtungen der Gesellschaft in Form von Statistiken, quantitativen Erhebungen, aber auch durch das digitale Medium des Geldes.²⁶⁵ Mit den modernen Staaten ergab sich somit auch die früheste Form der digitalen Datenverarbeitung als Sozialstatistik im 19. Jahrhundert.²⁶⁶

Die zweite Entdeckung zentriert sich für Nassehi um die Liberalisierungs- und Pluralisierungsbewegungen des 20. Jahrhunderts als radikale Komplexitätssteigerung innerhalb der Industriegesellschaft, vor allem nach dem 2. Weltkrieg. Zum Zweck der Selbstsichtbarkeit in dieser Komplexität brauchte es digitale Abstraktionen von analogen Sichtbarkeiten, die in Datenform relativiert und analysiert werden konnten.²⁶⁷

Die dritte und neueste Entdeckung der Gesellschaft durch die Datentechnologie ist für Nassehi ihre digitale Entdeckung. Durch die gestiegenen Möglichkeiten der digitalen Selbstbeobachtung durch Digitaltechnik (mehr denn je wird digitalisiert und damit auch gezählt) ergeben sich Hinweise auf soziale Ordnungen innerhalb der Gesellschaft („gesellschaftliche Muster“).²⁶⁸ Die Digitaltechnik nutzt in Nassehis Leseart die verschiedenen Regelmäßigkeiten der Gesellschaft („Gesellschaftlichkeit“) bzw. die Abweichungen davon, ohne sie erfunden zu

²⁶³ Ebd., S. 41.

²⁶⁴ Ebd., S. 39.

²⁶⁵ Ebd., S. 45–46.

²⁶⁶ Ebd., S. 31.

²⁶⁷ Ebd., S. 45–46.

²⁶⁸ Ebd., S. 50–53.

haben.²⁶⁹ Jedoch ereignet sich die digitale Entdeckung nicht von selbst. So betont Nassehi, dass die Wahrnehmung dieser Regelmäßigkeiten erst durch die Modellierung von Klassifikationssystemen möglich wird.²⁷⁰ In ähnlicher Ausrichtung beschrieb ich bereits die Abhängigkeit der Datenbildung von Klassifikation, Typisierung und Taxonomie. Zur Entdeckung ihrer eigenen und dynamischen Ordnungssysteme entwickelt die Gesellschaft einen „digitalen Blick“.²⁷¹ Zu beachten ist, dass die zugrundeliegende Digitaltechnik auch eigene Abhängigkeiten hat. Ihre operativen Programme und materielle Beschaffenheit prägen in eigener Weise die Gesellschaft. Jedoch immer nur auf Basis gesellschaftlicher Ordnungen, die sich in die Technik einschreibt und ihr sichtbar wird.

Zusammenfassend lässt sich Nassehis Ausgangsfrage nach dem Problem, welches durch die Digitalisierung gelöst werden soll, beantworten:

„Wenn die Form der Lösung, die die Digitaltechnik anbietet, darin besteht, dass man Daten miteinander verknüpft und sich Ordnungsstrukturen erst darin sichtbar machen lassen, dann ist die Antwort auf die Frage nach dem Bezugsproblem nicht die Digitaltechnik, sondern die Digitalität der Gesellschaft selbst.“

– Nassehi, 2019, S. 62.

Die moderne Gesellschaft ist also mit Nassehi selbst nur als digitales Phänomen zu verstehen, weshalb sich Digitaltechniken in ihr überhaupt erst etablieren können.²⁷² Den Beginn der Digitalisierung lokalisiert Nassehi durch seine Beschreibung der historischen Bruchpunkte mit dem Auftauchen der Nationalstaaten in die Frühzeit der Moderne. Eine monokausale Gleichsetzung von Digitalisierung und Digitaltechnik, wie in den erwähnten technodeterministischen Begründungen, ist mit dieser Perspektive nicht möglich:

„Nicht der Computer hat die Datenverarbeitung hervorgebracht, sondern die Zentralisierung von Herrschaft in Nationalstaaten, die Stadtplanung und der Betrieb von Städten, der Bedarf für die schnelle Bereitstellung von Waren für eine abstrakte Anzahl von Betrieben, Verbrauchern und Städten/Regionen.“

– Nassehi, 2019, S. 62.

²⁶⁹ Ebd., S. 44; Differenztheorie als Erklärung der Musterhaftigkeit von Gesellschaft.

²⁷⁰ Ebd., S. 58.

²⁷¹ Ebd., S. 59.

²⁷² Ebd., S. 62.

Nassehis Datenbegriff

In Nassehis Theorie der digitalen Gesellschaft, die Digitaltechnik als Konsequenz einer digitalen Grundstruktur der Gesellschaft konzipiert, ist es gewinnbringend zu schauen, wie der für mein Argument des Datenexzptionalismus wichtige Datenbegriff gestaltet wird. Zunächst betont er, dass neben ihrer Diskretheit, die Egalität („unspezifischer Charakter“) von Daten letztlich das Kernattribut ist,²⁷³ was alles Digitale („Digitaler Universalismus“) verbindet.²⁷⁴

„Die Einfachheit der Daten ist der Schlüssel für ihre Wirksamkeit. Je einfacher die Grundcodierung, desto größer sind die Rekombinationsmöglichkeiten.“

– Nassehi, 2019, S. 145.

Eine wichtige Bemerkung Nassehis ist, dass unter Daten- und Digitalisierungsprozessen nicht

nur Quantifizierungen und Zahlenformen zu verstehen sind. Eine solche Gleichsetzung findet sich u.a. in der soziologischen Analyse über das „metrische Wir“ von Steffen Mau, für den Digitalisierung einen „Zahlentreiber“ darstellt.²⁷⁵ Während Maus Analyse äußerst umfassend über die weitreichenden gesellschaftlichen Folgen eines Glaubens in die Zahlen informiert, werden statt dem Phänomen an sich vielmehr die Sekundärfolgen der Digitalisierung besprochen.²⁷⁶ Zudem findet sich bei Mau keine klare Abgrenzung des Datenbegriffs, wodurch Daten zu einer reinen Begleiterscheinung von Quantifizierungs- und Computerprozessen reduziert werden. Jedoch zeigt auch Nassehi Momente der konzeptionellen Verkürzung des Datenbegriffs in seiner Argumentation, wenn Daten als Rohmaterial des Digitalen eine bestimmte Passivität verliehen wird und Daten in ein nicht weiter begründetes Abhängigkeitsverhältnis zu Informationen gerückt werden.²⁷⁷ Zunächst scheint es so, als ob der Datenbegriff zugunsten der erweiterten Digitalisierungstheorie bei Nassehi mit bestimmten Narrativen des Datenexzptionalismus in der Diskussion passiv mitgeführt wird oder zumindest keine wesentlichen neuen Akzente für den Datenbegriff gesetzt werden.

²⁷³ Ebd., S. 35.

²⁷⁴ Ebd., S. 31.

²⁷⁵ Mau, 2017, S. 40ff.

²⁷⁶ Vgl. Nassehi, 2019, S. 146.

²⁷⁷ Ebd., S. 31.

Daten als Verdopplung

Jedoch führt Nassehi später im Buch die Figur der „Verdopplung“ als exklusive Beschreibung der Daten ein:

„Wenn man das Digitale irgendwie auf den Begriff bringen will, dann ist es letztlich nichts anderes als die Verdopplung der Welt in Datenform.“

– Nassehi, 2019, S. 32-33.

Diese Verdopplung meint, dass Datenpraktiken ihren Gegenstand in Form von eigens erzeugten Daten betrachten, der nicht der Gegenstand selbst ist.²⁷⁸ Ich sprach bereits im vorangegangenen Kapitel über das Scheitern von Datenvisualisierungen über die verwandte Idee der „Map-Territory-Relation“ anhand von Borges’ „On Exactitude in Science“. Auch Nassehi teilt die Annahme, dass dieses Verhältnis von Modell und Realität ein grundlegendes Thema der Epistemologie des 20. Jahrhunderts geworden ist.²⁷⁹ Unter einer zeichentheoretischen Einordnung beschreibt er die Dynamik der Geschlossenheit von Daten:

„Daten kennen die Welt nur datenförmig und sind unhintergebar auf sich selbst verwiesen, weil nur Datenförmiges registriert werden kann.“

– Nassehi, 2019, S. 104.

Diese Selbstreferenz und Geschlossenheit kennzeichnen für Nassehi die Daten als Verdopplungsmedium. Im Gegensatz zu Frieder Nake, der 1984 die Verdopplung als „Computer-Prinzip“ beschrieb,²⁸⁰ argumentiert er, dass die menschliche Kommunikation per se mit diesen Verdopplungen konfrontiert ist.²⁸¹ So setzt Nassehi die Schrift als das erste digitale Medium,²⁸² aber auch die Sprache stellt eine (selbstreferentielle und damit in sich geschlossene) Verdopplung der Welt dar:

„Die Welt liegt zu Beginn der menschlichen Existenz schon in Datenform vor [...] die durch das Sprechen prozessiert werden.“

– Nassehi, 2019, S. 115.

²⁷⁸ Ebd., S. 68.

²⁷⁹ Ebd., S. 78.

²⁸⁰ Vgl. Distelmeyer, 2021, S. 5.

²⁸¹ Vgl. auch Coy, 2012.

²⁸² Nassehi, 2019, S. 114.

Grundsätzlich sind es diese Verdopplungen, die die Routinen einer modernen Gesellschaft ausmachen und parallel in verschiedenen Perspektiven je nach gesellschaftlicher Funktion (Ökonomie, Politik, Bildung, Wissenschaft etc.) auftauchen.²⁸³ Wenn Digitalisierung, verstanden als soziales Phänomen in Verbindung mit Digitaltechnik, als Besonderheit auffällt, dann nicht, weil es eine vollkommen neuartige Erscheinung, sondern eine wiederholte Verdoppelung ist, die mit den Routinen der bisherigen Verdopplungsmedien

bricht:

„Die moderne Gesellschaft verdoppelt sich selbst nochmal in Form von Daten, die permanent anfallen und die quer zu den geübten Praktiken der Gesellschaft liegen.“
– Nassehi, 2019, S. 140.

Mit Nassehis

Figur der Verdopplung kann meine Definition von Daten weiter strukturiert werden. Die Verdopplung stellt jedoch keine neue Eigenschaft dar, sondern erweitert vielmehr die Eigenschaft der Abstraktion. Daten sind in ihrer Abstraktion definiert als symbolische Zeichen auf Grundlage einer menschlichen Wahrnehmung bzw. Intention. Die Figur der Verdopplung ergänzt, dass diese Abstraktionen selbstreferenziell und damit auch in sich geschlossen sind. Diese Abstraktionen sind nicht auf eine Zahlenform oder auf technische Abstraktionen beschränkt, sondern können, wie bei Nassehi angedeutet, auch innerhalb jeglicher menschlicher Abstraktion, wie Sprache oder Schrift, gedacht werden. In einer erweiterten Deutung kann somit jede menschliche Abstraktion zumindest als datenförmig beschrieben werden. In der Figur der Verdopplung zeigt sich somit ein Gegenentwurf zur Projektion, dass Digitalisierung und die vermehrte Erscheinung von Daten eine radikale und besondere Entwicklung seien.

Eine Alternative zum Datenexzptionalismus

Digitalisierung und Digitaltechnik kann in dieser Lesart nicht mehr als Zustand der Ausnahme oder der radikalen Veränderung beschrieben werden, sondern zeigt sich als logische Fortsetzung von gesellschaftlichen Strukturen. So gesehen bricht keine neue Gesellschaft an, vielmehr setzt sich die funktional ausdifferenzierte Gesellschaft konsequent fort. Nassehis Verortung der Digitalisie-

²⁸³ Ebd., S. 115.

rungssphänomene zeigt sich in diesem Moment als sehr produktiv, da sie entgegen technodeterministischer Narrative (aufgrund der Digitaltechnik folgt ein soziokultureller Wandel) die Diskussion zu einem sozialdeterministischen Blick hin (aufgrund der digitalen Struktur der modernen Gesellschaft konnte eine Digitaltechnik folgen) verschiebt. Die digitale Gesellschaft ist nicht das Resultat der Computerisierung, sondern Digitaltechnik bedingt sich aus der digitalen Grundstruktur der modernen Gesellschaft. Daraus folgt auch, dass ein kritischer und reflektierter Umgang mit den Prozessen der Digitalisierung (digital literacy) sich weniger über eine Expertise in den zahlenbasierten Grundlagen der Digitaltechnik (numeracy) definiert, als in der Auseinandersetzung mit bereits existierenden digitalen Phänomenen in der Gesellschaft liegt (Mein Vorschlag: *societal literacy*). Einen weiteren Hinweis zur Deutung dieser Situation

gibt Benjamin
Bratton:

„Although the twentieth century invented computers, it did not invent computation so much as it discovered it as a general force, and offered some initial basic tools to work with it more directly. We are, like everything else, also its product.“

– Bratton, 2016, S. 77.

Mit Bratton versteht sich abseits beider Determinismen, sozial („discovered“) wie technisch („invented“), die Digitaltechnik als epistemische Technologie, denn gerade durch die Digitalisierung wurde die digitale Grundstruktur der modernen Gesellschaft kenntlich gemacht. Denn die Leistung der Datentechnologie war es weniger Muster in den Daten über die Gesellschaft zu finden, als die Muster in der Gesellschaft selbst aufzudecken. Erst die Digitaltechnik ermöglicht die weitere Ausbildung eines *digitalen Blicks*, der es ermöglicht die digitalen Strukturen der Gesellschaft sichtbar zu machen. Der *digitale Blick* schaut auf das Modell der modernen Gesellschaft. Als Konsequenz lese ich Digitalisierung als einen Visualisierungsprozess, der die Ordnungssysteme der modernen Gesellschaft sichtbar macht.²⁸⁴ Erst durch die zunehmende Loslösung von institutionellen Rahmungen und der Verstärkung von Klassifikation und Standardisierungen durch Digitaltechnik konnten diese Ordnungssysteme besser wahrnehmbar werden. Das gesellschaftliche Potenzial von Digitalisierung liegt also nicht so sehr in der Datafizierung der Gesellschaft, sondern in der Erkenntnis, dass die moderne Gesellschaft schon immer datafiziert operiert.

²⁸⁴ Vgl. Tresch, 2012.

Nicht minder vielversprechend, aber im Rahmen dieser Arbeit nur andeutbar, sind Diskurse zur individuellen Motivation von Menschen bei der Arbeit mit Daten. Auch Nassehi betont abschließend seiner Betrachtung die spiegelartige Wirkung der Digitalisierung auf die Gesellschaft. Er beschreibt die Idee einer „Sozioanalyse“, eine gesellschaftliche Form der Psychoanalyse, in der sich die Gesellschaft ihrer bisher verdrängten digitalen Strukturen bewusst wird.²⁸⁵ In der Tat gibt es neben der rein soziologischen Betrachtung bereits Forschungsansätze, die mithilfe der Psychoanalyse eine weitere Deutungsebene der Datenphänomene bearbeiten.²⁸⁶ Der psychoanalytische Blick erweitert den Daten-diskurs, was sich bei Johanssen im Begriff der „Datenperversion“ definiert.²⁸⁷ Analog dazu sehe ich eine datenspezifische Erweiterung des sadistischen Schemas von Jacques Lacan zum „Datensadismus“ als vielversprechend.²⁸⁸ So kann ich nur als Vorschlag für zukünftige kritische Auseinandersetzungen formulieren, dass Theorien der Psychoanalyse eine gewinnbringende Perspektive auf das Phänomen des Datenexzptionalismus darstellen.

Am Anfang des Kapitels stand die Frage, woher die Affinität gegenüber den Daten stammt. Mit Nassehis Theorie der digitalen Gesellschaft konnte abgeleitet werden, dass diese Affinität kein Begleitphänomen der Etablierung von Digitaltechnik in der Gesellschaft ist, sondern spätestens die moderne Gesellschaft selbst aufgrund ihrer Komplexitätssstruktur mit Daten operiert. Obwohl die genaue Relation von Daten und Modernisierung in einem anders gerichteten Forschungsansatz noch weiter zu interpretieren wäre, wurde für meine Argumentation deutlich, dass Ansätze, die sich fundamental mit dem Grund, den Bedingungen und der Struktur von Daten auseinandersetzen, solchen Narrativen der Besonderheit von Daten und deren Ausnahmeerscheinung diametral gegenüberstehen. Anders formuliert: Die Datafizierung der Gesellschaft kann nur diagnostiziert und als Effekt der Computerisierung besprochen werden, wenn auch die gesellschaftliche Ursachen für das Aufkommen von Daten ignoriert werden. Wenn Daten mit Digitaltechnik begründet und gleichgesetzt werden, stellt dies eine bewusste historische und soziokulturelle Verkürzung dar. Es kann so über Daten gesprochen und deren Folgen besprochen werden, ohne Frage darüber zu stellen, was Daten eigentlich sind. Analog

²⁸⁵ Nassehi, 2019, S. 185.

²⁸⁶ Vgl. Johanssen, 2018.

²⁸⁷ Johanssen, 2021.

²⁸⁸ Vgl. Baas, 1995.

zu Nassehi „Digitalisierungsvergessenheit“ im Digitaldiskurs,²⁸⁹ möchte ich den Datenexzptionalismus als „datenvergessen“ weiter ausdifferenzieren.

Der Datenexzptionalismus eint so paradox anmutend „Datenversessenheit“ und „Datenvergessenheit“.²⁹⁰ Gerade die Steigerung der Verfügbarkeit der Daten durch Digitaltechnik potenziert die Vorstellungen darüber, was mit den Daten, die letztlich von allem Abstrahierbarem gebildet werden können, analysiert werden soll. Diese kulturell projizierte Exploitation der Datenversessenheit begründet sich auf dem Grundverständnis einer entsubjektivierten Idee, dass alles zu Daten werden kann und unter einem technisch-objektivierten Weltbild erschlossen werden soll.²⁹¹ *Datenvergessenheit* bedeutet im Kontrast sich in den Möglichkeitsräumen der Daten zu bewegen, ohne zu fragen, was Daten grundsätzlich ausmacht. Sie verweist auf eine ignorierte Realität von Daten in der Gesellschaft abseits computertechnisch-numerischer Konzeptionen. Dabei bedingt die Datenversessenheit die Datenvergessenheit, denn eine die Idealisierung von einer ausgewählten Idee von Daten beruht auch auf der Negation konkurrierender Vorstellungen. Diese Bedingung lässt auch Umkehrungen zu. So zeigen sich beispielsweise in Momenten der Kritik an der Versessenheit gegenüber den Daten, auch Eigenheiten in der Praxis mit Daten, die von einem Positivismus gegenüber Daten nicht beachtet werden können.²⁹²

²⁸⁹ Nassehi, 2019, S. 15.

²⁹⁰ Vgl. Klemm, 2016, S. 31.

²⁹¹ Bächle, 2016, S. 23.

²⁹² Analog zum Positivismus in der Wissenschaftstheorie, geht ein Datenpositivismus davon aus, dass wissenschaftliche Theorie durch Datenartefakte falsifizierbar sind; vgl. Habermas, 1968.

4. Datenkritik

Wie kann die Datenhybris negiert werden?

Wie formuliert sich demnach eine solche Kritik an Datenphänomenen? Im Folgenden betrachte ich kritischen Perspektiven gegenüber Datenprozessen, um abseits einer Datenaffirmation zu verstehen, wie der Datenbegriff in der Datenkritik konzipiert werden. Eine solche Datenkritik verweist unter anderem auf die konzeptionellen Limitationen in der Beschäftigung mit Daten. Es gilt, sich solcher Aspekte von Daten zu vergegenwärtigen, die durch die Datenvergessenheit des dominanten Datenexzessionalismus abhanden gekommen sind. Einen wichtigen Beitrag stellt Hito Steyerl in ihrem Essay „A Sea of Data“, indem zwei wesentliche Aspekte einer Datenkritik formuliert werden: „Dirty Data“ und „Datenapophanie“.²⁹³ Steyerl rahmt zunächst die veränderten Konditionen von Sichtbarkeit in Relation zu den ästhetischen Dimensionen von großen

Datenmengen:

„Seeing is superseded by calculating probabilities. Vision loses importance and is replaced by filtering, decrypting, and pattern recognition.“

– Steyerl, 2016.

Eine Eigenschaft dieser Datenmengen, die Steyerl einführt, ist ihre Ungenauigkeit, Unvollständigkeit und Inkonsistenz, was in der technischen Fachsprache entweder als „dirty data“ oder „rogue data“ bezeichnet wird.²⁹⁴ In meiner Definition von Daten als verlustbehaftete Abstraktionen sind grundsätzlich alle Daten als ungenau, unvollständig und inkonsistent zu verstehen. *Dirty Data* sind also keine Sonderform von Daten, die es zu vermeiden oder zu *reinigen* gilt, sondern eine generelle Grunderscheinung von Daten. Es gibt keine Daten, die realitätsgetreu sind, ansonsten wären es per Definition keine Daten mehr. Die Zuschreibung von *Dirty Data* und der Wunsch nach *Reinigung* sind demnach vielmehr eine kulturelle Projektion, die Daten zum Zweck einer optimierten Datenanalyse möglichst einheitlich, lückenlos und damit übersichtlich vorliegen haben möchte.

²⁹³ Steyerl, 2016.

²⁹⁴ Ebd.

Wenn dementsprechend Daten in diese kulturelle Projektion geraten, die sie rein konzeptionell nie erfüllen können, entstehen problematische Erwartungsgefälle. So bescheinigt Steyerl, dass sich mit dem drastisch vermehrten Aufkommen von Daten in dieser

Erwartungshaltung

auch Probleme bezüglich ihrer Sichtung und Deutung einstellen. Sie beschreibt zum kulturellen Umgang mit dieser Unübersichtlichkeit die Methode der Apophänie. Wie bereits im Kapitel zum Scheitern der Datenbildern angedeutet, ist Apophänie zunächst ein Ende der 1950er Jahre von Klaus Conrad geprägter Begriff zur Beschreibung einer schizophrene Erfahrung, wenn in Dingen augenscheinliche Muster gesehen werden, die keine direkte Verbindung zur Umwelt haben.²⁹⁵ 2013 erweitert dann Benjamin Bratton diese Grundlage zu einer Definition, auf die sich auch Steyerl bezieht:

,This apophenia, a seeing of patterns where there is actually only noise, is neither a failure of imagination nor a virtue, but rather an unavoidable qualification of our predicament and its (only partially decipherable) aftermath.“

– Bratton, 2013.

So kann Datenanalyse mit Steyerl auch als ein Mustersehen beschrieben werden, wo eigentlich keine Muster vorzufinden sind. In meiner Datendefinition ist diese Aussage nicht weiter verwunderlich. Zum einen sind Daten, wie erwähnt, immer Abstraktionen, die keine klare Verbindung mehr zu ihrem Abstraktionsgegenstand vorweisen. Zum anderen ließ sich mit Nashehis erweitertem Digitalisierungsbegriff ableiten, dass Mustererkennung nicht Strukturen innerhalb der Daten, sondern vielmehr den Strukturen der Gesellschaft, in denen Daten auftauchen, selbst gilt. Das Problem, auf das Steyerl zielt, ist also nicht so sehr die Apophänie, die kulturelle Projek-

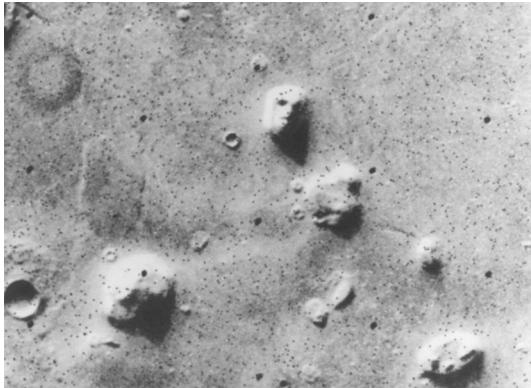


Abb. 29 Mondoberfläche, aufgenommen von Viking 1 am 25.07.1976.

²⁹⁵ Conrad, 1958; Vgl. auch Stäheli, 2021, S. 71-78.

tion von Mustern in Daten an sich, sondern vielmehr die daraus folgende politische Dimension:

„While today statisticians and other experts routinely acknowledge that their findings are mostly probabilistic projections, policymakers of all sorts conveniently ignore this message.“

– Steyerl, 2016.

In dieser Perspektive dreht sich die Narration von einer generellen Abhängigkeit die richtigen Daten zu finden und gewinnbringend auszuwerten, zu einer Abhängigkeit von der Entscheidung, Muster in Daten finden zu wollen.

Verkürzt formuliert könnte man sagen, dass die Datengesellschaft nicht durch die steigende Quantität der Daten geprägt wird, sondern vielmehr durch die kulturelle Projektion einer Datenhybris und der Entscheidung, die Methode der Apophanie zu instrumentalisieren – beides Aspekte eines Datenexzptionalismus. Nochmals: Das Problem ist nicht, dass Muster in Daten gesehen werden, die an sich nicht gegeben sind. Dies ist eine Konsequenz aus der Struktur von Daten an sich. Das Problem, das auch Steyerl adressiert, ist, dass die Apophanie durch die Digitaltechnik und die computergestützte Datenanalyse intentional automatisiert wird.

Bemerkenswert ist, dass Steyerls Perspektive dann weniger eine Kritik an den Daten an sich formuliert, sondern eine Beschäftigung mit den gesellschaftlichen Folgen der Digitaltechnik darstellt. Während sie ihre Forschung nicht dezidiert als Datenkritik benennt, sondern bewusst auf die Sekundärfolgen der Digitaltechnik zielt, gibt es ganze Forschungsstränge, wie die „Critical Data Studies“, die sich hauptsächlich mit der Formulierung einer Datenkritik auseinandersetzen. Mein Verdacht ist, dass sich auch solche Ansätze auf die sekundären Folgen der Digitaltechnik stützen und damit den Datenbegriff selbst kaum thematisieren. Ich vermute, dass sich der Datenexzptionalismus auch in datenkritische Diskurse einschreibt, wenn Daten darin grundsätzlich als grundsätzlich gegeben gesehen werden.

4.1 Critical Data Studies

Die relativ junge Disziplin der „Critical Data Studies“ (im Folgenden als CDS bezeichnet) fokussiert sich auf Fragen „about the nature of data, how they are being produced, organized, analyzed and employed, and how best to make sense of them and the work they do“.²⁹⁶ Im Jahr 2014 verkündeten mehrere Vertreter:innen einen dezidierten Bedarf nach einer solchen Disziplin, wie beispielsweise Dalton und Thatcher in „The Promise of Big Data and the Need for a Critical Data Studies“²⁹⁷ oder Kitchin in „Short Presentation on the Need for Critical Data Studies“.²⁹⁸ Welcher Begriff von Kritik diese neue Disziplin leiten soll, wird nicht weiter erläutert. Jedoch finden sich in der Beschreibung zur praktischen Dimension dieser Datenkritik von Liliana Bounegru und Jonathan Gray²⁹⁹ Hinweise auf die konzeptionelle Nähe zu Phil Agres Konzept der „critical technical practice“.³⁰⁰ In ihrem Beitrag „What Does A Critical Data Studies Look Like, And Why Do We Care?“ zur Rahmung dieser zu entwerfenden Disziplin, die die Kritik an Daten systematisch fassen sollte, nennen Dalton und Thatcher fünf Leitfragen aus bereits existierender Forschung:

- „1. *What historical conditions lead to the realization of ‚big data‘ such as it is?* (Barnes and Wilson, forthcoming; Dalton, 2013)
2. *Who controls ‚big data‘, its production and its analysis? What motives and imperatives drive their work?* (Thatcher, 2014)
3. *Who are the subjects of ‚big data‘ and what knowledges are they producing?* (Haklay, 2012)
4. *How is ‚big data‘ actually applied in the production of spaces, places and landscapes?* (Kitchin and Dodge, 2011)
5. *What is to be done with ‚big data‘ and what other kinds of knowledges could it help produce?* (Shah, 2014).“

– Dalton und Thatcher, 2014b.

²⁹⁶ Kitchin, 2014a.

²⁹⁷ Dalton und Thatcher, 2014a.

²⁹⁸ Kitchin, 2014b.

²⁹⁹ Vgl. Gray und Bounegru, 2019.

³⁰⁰ Agre, 1997.

Wie aus diesen Leitfragen deutlich wird, sollen sich diese CDS vor allem als Kritik am Phänomen Big Data und seinen Folgen abarbeiten, wie auch Iliadis und Russo in ihrer Einführung zu CDS bestätigten.³⁰¹ So wird auch die Zunahme datenpositivistischer Forschungen in den Natur- und Sozialwissenschaften kritisiert.³⁰² Diese Haltung etablierte sich bereits Anfang der 2010er Jahre. Als frühe Referenz werden oft die Provokationen der Microsoft-Forscherinnen Kate Crawfords und Dana Boyds genannt:³⁰³

- „1. Big Data verändert die Definition von Wissen*
- 2. Der Anspruch auf Objektivität und Genauigkeit führt in die Irre*
- 3. Mehr Daten bedeuten nicht automatisch bessere Daten*
- 4. Außerhalb des Ursprungskontext verlieren große Datenmengen ihre Aussagekraft*
- 5. Nur weil Daten zugänglich sind, heißt das noch lange nicht, dass es ethisch vertretbar ist, sie auszuwerten*
- 6. Eingeschränkter Zugang zu Daten lässt eine neue digitale Kluft entstehen.“*

– Übersetzung in Geiselberger und Moorstedt, 2013, S. 192-209.

Sie definieren dabei Big Data als ein kulturelles, technisches, wie wissenschaftliches Phänomen, was auf dem Zusammenspiel von Technologie, Analyse und Mythologie basiert.³⁰⁴ Big Data verstehen sie damit als Ansatz, bei dem mit der Zusammenführung von Digitaltechnik Muster in Daten gefunden werden sollen, deren Relevanz bezüglich ihrer Neutralität vor allem kulturell erzeugt ist. Auch bei Crawford und Boyd, wie auch bei den anderen genannten frühen Andeutungen der CDS, wird deutlich, dass sich ihre Kritik gegen eine bestimmte Auslegung und Praxis von Daten, verstanden als Big Data, richtet. Dabei werden weniger die von Kitchen formulierten Fragen nach den grundsätzlichen Bedingungen und Strukturen von Daten selbst, sondern vordergründig die Folgen der Digitaltechnik im Bezug auf Datenpolitik diskutiert. Solche Analysen sind, nur um es nochmals zu betonen, essenzielle Beiträge zu einem umfassenden Verständnis von Anwendungsfolgen von Digitaltechnik und parallel dazu laufenden daten-

³⁰¹ Iliadis und Russo, 2016.

³⁰² Kitchin, 2015.

³⁰³ Crawford und Boyd, 2011.

³⁰⁴ Ebd., S. 188-189.

positivistischen Narrativen. Jedoch werden Daten dann an sich nicht besprochen, sondern in Gleichsetzung zu digitaltechnischen Phänomenen vorausgesetzt. Ein solcher Diskurs folgt mehr dem Muster einer Technologiekritik, die sich auf die Folgen der Anwendung einer Datenlogik oftmals innerhalb von Computertechnologien fokussiert, als eine holistische Datenkritik anzubieten.

Diese Beobachtung lässt sich nicht auf alle Ansätze innerhalb der CDS übertragen. So diente beispielsweise Lisa Gitelmann und Virginia Jacksons Einführung zu „Raw Data‘ Is an Oxymoron“ als Grundlage für den Datenbegriff,³⁰⁵ der früher in diesem Kapitel bestimmt wurde. Eben jene von Lisa Gitelman zusammengetragene Textsammlung gilt als wichtiger Beitrag der CDS der frühen 2010er Jahre. Im Gegensatz zu Publikationen, die ihren Datenbegriff nur am Phänomen Big Data ausrichten, werden dort Essays versammelt, die die konzeptionelle Geschichte des Datenbegriffs versuchen abzubilden. Datentechnik ist dabei ein, aber nicht der einzige Aspekt der Betrachtung von Datenphänomenen. Schon am Titel „Raw Data‘ Is an Oxymoron“ wird diese wissenschaftshistorische Perspektive deutlich. Das Zitat stammt ursprünglich, wie bereits erwähnt, von Bowker:

„Raw data is both an oxymoron
and a bad idea; to the contrary,
data should be cooked with care.“

– Bowker, 2005, S. 184.

Seine weiteren Ausführungen in „Memory Practices in the Sciences“ beschreiben, wie sich Klassifikationssysteme und deren Technisierung bzw. Automation seit dem 19. Jahrhundert auf die Wissensproduktion auswirken.³⁰⁶ Die Gegenüberstellung von „raw“ und „cooked“ gründet sich auf Lévi-Strauss’ „The Raw and the Cooked“.³⁰⁷ Diese Gegenüberstellung von Natürlichem („raw“) und dem Sozialen („cooked“) relativiert Bowker in einem eigenen Beitrag des Readers „Raw Data‘ Is an Oxymoron“, da alles Natürliche von sich aus schon sozial konstruiert sei.³⁰⁸ Ebenso sind Daten als aktive Projektionen eines Gegenstandes niemals gegeben, natürlich und damit „raw“.

³⁰⁵ Gitelman, 2013, S. 1-14.

³⁰⁶ Bowker, Memory Practices in the Sciences, 2005.

³⁰⁷ Lévi-Strauss, 1969.

³⁰⁸ Gitelman, 2013, S. 168.

Daten sind Abstraktionen, die den Konventionen folgen, die ihre Gestalter:innen ihnen auferlegen. Daten sind demnach immer hoch gradig künstlich und sozial.

Mit Bowkers Forschungsschwerpunkt auf die Wissenschaftsgeschichte und seiner Popularität im datenkritischen Diskurs spiegelt sich hier eine disziplinäre Verwurzelung der CDS. Fraglich bleibt, wie sich CDS im Bezug auf die kritische Betrachtung von Datenkonzeptionen von etablierten wissenschaftshistorischen Perspektive wirklich absetzen, dass sich dadurch eine eigene neue Disziplin begründet. Neben diesen wissenschaftshistorischen Zugängen, die sich für das Wesen der Daten im Wandel interessieren, liegt die Neuigkeit der CDS in anderen Schwerpunkten. So betont Kitchin, dass neben dem historischen, epistemischen und „philosophischen“ Überdenken des Datenbegriffs,³⁰⁹ auch andere Perspektiven den Diskurs der CDS prägen:

„Beyond this philosophical rethinking of data, scholars have begun to make sense of data ethically, politically/economically, spatially/temporally, and technically.“

– Kitchin 2014b.

So gibt es eine Vielzahl von Forschern, die sich mit den verschiedenartigen Folgen der weitreichenden Vervielfältigung von Datenphänomenen und der Etablierung von Digitaltechnik beschäftigen. Die ethisch-moralischen Perspektiven veranschaulichen die Voreingenommenheit (bias) der Daten und die daraus folgenden soziale Ungerechtigkeiten durch ihre datentechnologische Anwendung und Durchsetzung.³¹⁰ So werden Daten auch zum Politikum und ihre Zugänglichkeit und Verwertbarkeit weitreichend diskutiert.³¹¹ Oft steht dabei die Kommodifizierung der Daten und ihre Konzentration innerhalb der großen IT-Unternehmen (GAFAM, FAANG oder MAMAA) zur Diskussion.³¹² Letztlich ruft das auch rein technische Fragen hervor, wie mit den Datenmengen im Rahmen sich entwickelnder Praktiken um die Metapher Machine Learning umgegangen werden soll.³¹³

³⁰⁹ Vgl. auch Halpern, 2015.

³¹⁰ Vgl. Couldry und Mejias, 2019 sowie Amoore, 2020.

³¹¹ Vgl. Viljoen, 2020.

³¹² Vgl. Srineck, 2016.

³¹³ Vgl. Crawford, 2021 sowie Vickers und Allado-McDowell, 2020.

Wie bereits angedeutet, sind solche Ansätze essenziell für ein umfassendes Verständnis der Folgen der gesellschaftlichen Etablierung von Daten und die Multiplizierung dieser Effekte durch Datentechnologie. Jedoch wird in solchen Diskursen oftmals ein bestimmter Datenbegriff vorausgesetzt anstatt grundlegend diskutiert. Daten sind in einer solchen Perspektive eher eine Begleiterscheinung von digitaltechnischen Prozessen, die oftmals das eigentliche Ziel der Analyse sind. Es sind dann Diskurse, die sich weniger kritisch mit dem Status von Daten an sich (*Critical Data Studies*) als mit den sekundären Folgen ihrer Anwendung (*Critical Data Impact Studies*) beschäftigen. Mit Armin Nassehi ließ sich aufzeigen, dass es die digitale Grundstruktur der Gesellschaft ist, die die Erscheinung von Daten begünstigt, und es dadurch weniger die Digitaltechnik ist, die uns die Daten aufzwingt. Ähnlich sozialdeterministisch formuliert es Katharina Hausladen:

„Die Asymmetrien, die Daten hervorrufen, sind so verstandene Asymmetrien, die in der Gesellschaft ja bereits vorhanden sind; ja, die in dieser vorherrschenden und durch Technologien, die in spezifischen sozialen Kontexten [...] entwickelt wurden, weiter reproduziert werden.“

– Hausladen, 2020, S. 25.

Ein solcher Blick wird mit dem Fokus auf die Sekundärfolgen von Daten und Datentechnologien tendenziell verschränkt. Daten werden dann beispielsweise nicht als Bedingung, sondern technodeterministisch als Teilproblem der Digitaltechnik besprochen. Nicht die gesellschaftliche Struktur, die sich bereits als datenaffin kennzeichnet und durch die sich Digitaltechnik etablieren konnte, stehen im Zentrum der Diskussionen. Vielmehr werden die Möglichkeits- und Einschränkungsräume der Digitaltechnik als Verhandlungsort der Datenprobleme markiert. Meine These wäre nun, dass sich durch diese Verwechslung der Problemzuschreibung ein Aspekt des Datenexzptionalismus in Teilen des kritischen Diskurses der CDS selbst zeigt. Wenn Daten als primäre Erscheinung und Folge von Digitaltechnik konzipiert und kritisiert werden, wiederholt sich dabei das Narrativ der Ausnahme, die Daten nicht als kulturelle Struktur der Gesellschaft wahrnimmt und sich damit auch potenziell gegenüber wirklichen Lösungen zum durchaus problematischen Umgang mit Daten jenseits rein technischer Projektionen verschließt.

Für mich sind die CDS daher ein interessantes Beschäftigungsfeld, weil sie einerseits durch die kritischen Perspektiven wertvolle Einsichten im Umgang mit Daten herausbilden, andererseits weil auch sie nicht frei davon sind in ihrer Kritik genau die Narrative weiterzutragen, die es mir in meiner Argumentation zu isolieren gilt. Als Beispiel für diese Zwiespältigkeit möchte ich im Folgenden die Datenkonzeption von Yanni Alexander Loukissas in „All Data Are Local“, als eine Stimme der angewandten CDS, betrachten.

„All knowledge is local, all truth is partial.“

– Le Guin, 2016, S. 354.

Schon 1995 skizzierte Ursula

Le Guin innerhalb eines fiktiven Dialogs in ihrer Kurzgeschichte „A Man of the People“ die räumliche Abhängigkeit von Wissensstrukturen. 15 Jahre später beschrieb dann Bowker in „All Knowledge is Local“ weiter, dass im Gegensatz zur Idee eines universellen Wissens jegliches Wissen auf lokalen und damit soziokulturellen Gegebenheiten, wie beispielsweise lokal spezifischen Klassifikationssystemen, beruht.³¹⁴ 2019 überträgt Loukissas diesen argumentativen Strang auf den Datendiskurs, ohne jedoch direkt auf die beiden genannten Vorläufer:innen zu verweisen. Daten sind für Loukissas kulturelle Artefakte, die als Indizes dementsprechend auf lokales Wissen und ihren Entstehungskontext verweisen.³¹⁵ Diese Sichtweise ermöglicht seinen methodischen Zugang, den er als „local readings“ bezeichnet und der es ihm ermöglicht Daten als Texte zu lesen, die Hinweise auf die lokalen Bedingungen („data settings“) ihrer Herstellungen bereithalten.³¹⁶ Dieser Perspektivwechsel, der nicht auf die Daten an sich, sondern durch sie hindurch auf ihre Entstehungskonditionen schaut, ist ein wesentlicher Beitrag zu einem kritischen Umgang mit Daten. Der Mehrwert von Daten liegt in dieser Betrachtung – im Gegensatz zu Narrativen der Objektivierung und einer neutralen Abbildungsidee – gerade in ihrer Unvollständigkeit, die Rückschlüsse auf die Wissenssysteme zulässt, die diese Daten hervorgebracht haben.³¹⁷

³¹⁴ Bowker, 2010.

³¹⁵ Loukissas, 2019, S. 1,4.

³¹⁶ Ebd., S. 7-8.

³¹⁷ Ebd., S. 2, 10.

Während das Prinzip und die Praxis des Lokalen bei Loukissas im Vordergrund stehen, wird die Konzeption des Datenbegriffs nur begleitend und anhand von Abgrenzungen thematisiert.³¹⁸ Er beschreibt letztlich Daten anhand von fünf Eigenschaften: Daten sind „plural, embedded, small, operational, and material“.³¹⁹ Während die Eigenschaften „plural“, „embedded“, „small“ und „material“ sich mit meiner bisherige Arbeitsdefinition von Daten decken oder sie in bestimmten Aspekten akzentuieren, stellt die Beschreibung „operational“ einen bisher ungenannten Aspekt meiner Beschreibung dar. „Operational“ bedeutet für ihn „part of socio-technical systems from which [data] can't be separated“.³²⁰ Der Aspekt der Operationalität bei Loukissas, verstanden als Zugehörigkeit der Daten zu einem gestalteten Prozess, der ihre Existenz zumindest potenziell verschleiert, bedingt aber auch eine Intention Systeme zu entwerfen, die Daten involvieren. Ich argumentiere daher genau anders gerichtet als Loukissas, der die bewusste interpretative Handlung als Opposition zur Operationalität verortet.³²¹ Erst durch die Intentionalität wird die Etablierung von Daten in soziotechnischen Systemen und damit ihre Operationalität ermöglicht. Nochmals: Daten, aber auch ihre praktische Anwendung in gesellschaftlichen Systemen, ist gewollt und somit keine natürlichen Erscheinung.

Es erscheint ein Datenbegriff, der relativ unberührt von der konzeptionellen Erweiterung der „Lokalisierung“ verbleibt. Zudem bleibt unklar, wie sich das „local data“ von „local knowledge“ bei Bowker (der auch das Vorwort zu „All Data Is Local“ verfasste) abgrenzt. Ich sprach bereits über die Schwierigkeit der Relation von Daten und Wissen in einem epistemischen Gesamtgefüge, aber der existierende Diskurs um räumliche Dispositionen in der Wissenschaftstheorie lässt eine tiefergehende Einordnung des Datenbegriffs bei Loukissas vermissen. Wenn man das Konzept „local data“ als datenkritischen Ansatz in zwei Schritten verstehen möchte, wird die erste Säule der „Lokalisierung“ bei Loukissas umfassend als Figur eingeführt, jedoch die zweite Säule, verstanden als grundsätzliche Datenkritik, nur zweckmäßig bedient. Anhand des Untertitels „Thinking Critically in a Data-Dri-

³¹⁸ Ebd., S. 13-18.

³¹⁹ Ebd., S. 18.

³²⁰ Ebd., S. 17.

³²¹ Ebd.

ven Society“ wird mein Punkt deutlicher: Das „thinking critically“ wird überzeugend geschildert und in sechs Prinzipien im praktischen Umgang mit Daten überführt, nur bleibt die Beschreibung der „data-driven society“ und damit der gesamtgesellschaftliche Kontext in dem diese Datenpraxis stattfinden soll, schattenhaft.

Weil eine grundsätzliche Auseinandersetzung mit dem Datenbegriff nicht im Fokus von Loukissas Auseinandersetzung steht, kommt es zu Aussagen, die latent datenpositivistische Tendenzen offenlegen:

„*Yet data is too widely used to be abandoned just now.*“

– Loukissas, 2019, S. 13.

Es erscheint zumindest zweifelhaft, wenn innerhalb einer Argumentation, die sich innerhalb der CDS verortet,³²² an einem bestimmten Datenbegriff festgehalten wird, nur weil er sich gesellschaftlich etabliert hat. Generell steht die Frage im Raum, warum ein Datenbegriff innerhalb einer Diskussion so ins Zentrum rückt, wenn er für die zentrale Diskussion nur als Platzhalter für die Sekundärfolgen der Datenerscheinungen gebraucht wird.

Eine ähnliche Entwicklung vermute ich im Konzept des Datenfeminismus, jüngst geprägt von Catherine D'Ignazio und Lauren Klein.³²³ Auch hier zeigt sich wieder eine reichhaltige Forschung zu ethischen Fragestellungen bezüglich machtpolitischer Ungleichheiten durch die Anwendung von Daten als soziotechnische Struktur, vorangetrieben durch Theorien des intersektionalen Feminismus. Auch wenn es eine klare historische Parallelentwicklung zwischen einem modernen Datenbegriff und der Idee von Rasse und Geschlecht gibt,³²⁴ haben Daten dieses gesellschaftliche Ungleichgewicht nicht erfunden, sondern, um wieder mit Hausladen zu sprechen, nur weiterhin verstärkt. Es ist fraglich, ob der feministische Diskurs durch die konstruierte Abhängigkeit zu den Sekundärfolgen von Datenkultur und -technik wirklich an Tiefe gewinnt, oder ob durch die Platzhalterfunktion des Datenbegriffs für diverse Datenphänomene nicht der kritische Blick getrübt wird. Im schlimmsten Fall werden datenpositivistische Narrative weitergetragen, wenn der Gegenstand der Kritik nicht klar gefasst werden kann. Der Feminismus ist zu wichtig, als ihn an Daten-

³²² Loukissas, 2019, S. 8.

³²³ D'Ignazio und Klein, 2020.

³²⁴ Ebd., S. 10.

narrativen zu verlieren. Warum dann nicht eine klare Technologiekritik, in der Tradition von Sadie Plant formulieren?³²⁵

Zusammenfassend vermute ich, dass sich selbst in Diskursen der CDS die Datenvergessenheit als Teilaспект des Datenexzessionalismus zeigt. Daten sind in manchen der Beiträge omnipräsent, wobei die Gründe ihrer Existenz kaum besprochen werden. Zu welchem Grad und in welcher Intensität wäre eine weiterer Forschungsschwerpunkt, den ich an dieser Stelle offen halte möchte. Stattdessen möchte ich im Folgenden auf Ansätze verweisen, die sich aus den diskutierten Gründen bewusst von einem positivistisch besetzten Datenbegriff abgrenzen. Vielleicht braucht es eine radikale Abkehr von Begriff der Daten insgesamt?

4.2 Post-Data

Wenn selbst die Datenkritik Gefahr läuft, Daten zu affirmieren oder zumindest weiterhin zu mystifizieren, besteht eine Möglichkeit der Reaktion darin, den Datenbegriff an sich zu verabschieden. Solche Ansätze möchte im Folgenden als *post-data* fassen. Das Präfix post kommt aus dem Lateinischen und bedeutet soviel wie zeitlich nachgestellt. Eine Beispiel für diese Verwendung des Präfix ist die Postmoderne, die konzeptionell nach der Moderne verortet wird. Es wird aber auch benutzt, um einen Zustand nach einer bestimmten Erkenntnis zu beschreiben. So bezeichnet etwa *post-digital* den Versuch der Überwindung einer Analog-Digital-Opposition durch die Erkenntnis, dass sich Digitaltechnik gesellschaftlich umfassend etabliert hat und sich zudem durch eine zunehmende Unsichtbarkeit kennzeichnet.³²⁶

Dieser Versuch zeigt aber auch Schwächen durch seine Nähe zu affirmativen Konzeptionen des Digitalmythos, wie Jan Distelmeyer argumentiert.³²⁷ Auch im Kontext der Datenkritik spricht beispielsweise Bowker von post-positivistischen Ansätzen, die einen Datenpositivismus überwinden wollen.³²⁸ Analog zu Distelmeyer erachte ich meinen vorgeschlagenen Begriff *post-data* als problematisch, da er die konzeptionelle Nähe zur Datenaffir-

³²⁵ Plant, 1998.

³²⁶ Mehr zum post-digital-Diskurs bei Berry und Dieter, 2015.

³²⁷ Distelmeyer, 2013.

³²⁸ Kitchin, 2015.

mation braucht um als Gegenposition lesbar zu werden. Da ich aber in meiner Argumentation den Begriff nicht als eigenständige Kategorie, sondern nur als Unterscheidung zu Ansätzen, die im bisherigen Datenbegriff verbleiben, nutzen möchte, erscheint mir die begriffliche Unschärfe für die weitere Argumentation tragbar.

Ein konkretes Beispiel für einen post-data-Ansatz ist Johanna Druckers Neubestimmung des Datenbegriffs als „*capta*“ von 2011:

*„The concept of data as a given has to be rethought through a humanistic lens and characterized as *capta*, taken and constructed.“*
– Drucker, 2011.

Der Begriff „*capta*“ leitet sich aus dem lateinischen Pendant zu „data“ als das Gegebene ab und bedeutet als Deklination von „*capere*“ soviel wie das „das Gefangene“ oder „das Genommene“. Diese Gegenüberstellung ist keine Erfindung Druckers, sondern fiel beispielsweise H. E. Jensen schon in den 1950er Jahren auf:

*„It is an unfortunate accident of history that the term *datum* [...] rather than *captum* [...] should have come to symbolize the unit-phenomenon in science. For science deals, not with ‚that which has been given‘ by nature to the scientist, but with ‚that which has been taken‘ or selected from nature by the scientist in accordance with his purpose.“*
– H. E. Jensen, zitiert nach Kitchin, 2014a, S. 29.

Der Kern des Konzepts

sowohl bei Jensen als auch bei Drucker ist die Bewusstwerdung, dass Daten immer eine kulturelle Abstraktion sind und nie das betrachtete Phänomen selbst abbilden. Beiden folgen der Einsicht, dass der bisher benutzte Datenbegriff nicht ihre kulturelle Konstruktion abbilden kann oder diese gar verschleiert und daher durch den alternativen „*capta*“-Begriff ersetzt werden soll. „*Capta*“ akzentuiert vor allem den subjektiven und intentionalen Charakter von Datenprozessen. So werden essenzielle Eigenschaften von Daten hervorgehoben und zum Wesensmerkmal dieser epistemischen Struktur stilisiert. Andererseits bleiben so andere Eigenschaften, wie beispielsweise der Prozess der Abstraktion zur Datenbildung mit seinen reduktiven Aspekten, ungenannt.

Eine Abkehr vom narrativ aufgeladenen Datenbegriff im Sinne des *post-data* ist eine produktive Form der Datenkritik, da sie konzeptionelle Grenzen aufzeigen kann. Sie formuliert eine Gegenreaktion auf die Datenvergessenheit des Datenexzensionalismus, da sie auf strukturelle Merkmale von Daten verweist. Daher muss der post-data-Begriff auch gar nicht als alleinstehender Begriff im Diskurs bestehen, sondern erfüllt die Aufgabe eines Perspektivwechsels gerade durch seine Abhängigkeit vom bestehenden Datenbegriff. Aber vielleicht liegt die primäre Aufgabe von post-data-Ansätzen nicht im Finden des einen Begriffs, der Daten als Idee komplett ablösen kann. Worauf diese Ansätze vielmehr hinweisen, ist die primäre Herausforderung der Schaffung eines gesellschaftlichen Umgangs mit Daten über den Datenpositivismus hinaus und notfalls auch gegen etablierte (technozentristische) Datenkonzepte. Eine so konzipierte tiefergehende Datenkritik hilft bei der Neudefinition einer neuen Data Literacy. Die konkretere Formulierung des post-data-Ansatzes stellt einen offenen Forschungsansatz für mich dar. Obwohl im späteren Teil dieses Buches Vorschläge für alternative und erweiterte Konzeptionen gemacht werden, möchte ich die grundsätzliche Beschäftigung als anschlussfähig für weitere Forschungsbemühungen verstanden wissen. Anknüpfungspunkte finden sich dabei in jüngeren Stimmen der Technikphilosophie, so etwa bei Yuk Hui³²⁹ oder Luciana Parisi.³³⁰

³²⁹ Hui, 2016a.

³³⁰ Parisi, 2019.

5. Datenexzceptionalismus als Modell

Das Ziel dieses Kapitel war es zu überprüfen, inwiefern kulturelle Modelle die Vorstellungen, Praxis und Reflexion von Daten prägen. Ich führte dafür die Figur des Datenexzceptionalismus ein,



anhand derer es galt diese Prägung zu skizzieren. Die Kalibrierung des Begriffs erfolgte über drei thematische Felder hinweg. In einem ersten Schritt machte ich auf die Narrative bezüglich Daten aufmerksam, die eine klare Erwartungshaltung gegenüber ihrer Erscheinungsform, Verwertbarkeit und ihren Möglichkeiten ausdrücken. Im zweiten Schritt machte ich auf die verschiedenen Dimensionen der Datenaffinität aufmerksam. Dabei erfolgte die konkrete Bestimmung des Datenbegriffs und den grundsätzlichen Eigenschaften von Daten, vor allem als Korrektur zu rein technischen Definitionen von Datenstrukturen. Über den Digitaldiskurs konnte dann mithilfe Nassehis soziologischer Theorie der digitalen Gesellschaft einen Gegenentwurf zu Lesarten positioniert werden, die Daten als reine Folgephänomene von Digitaltechnik auffassen. Daten sind so verstanden tief in der gesellschaftliche Struktur spätestens seit der Moderne verankert und müssen dementsprechend auch anders theoretisch konzipiert werden. Datenkritische Diskurse geben Vorschläge dafür, wie solche nicht datenpositivistische Ideen aussehen könnten, und wurden daher in einem letzten Schritt besprochen. Jedoch wurde deutlich, dass sich selbst in der Datenkritik eine gewisse Datenvergessenheit einschreibt, die die gesellschaftlichen Dimensionen des Datenbegriff zugunsten einer Technologiekritik mystifiziert und damit Spuren des Datenexzceptionalismus weiterträgt. Zusammenfassend zeigt sich durch diese drei Schritte – Datennarrative, Datenaffinität und Datenkritik – ein klareres Bild vom Datenexzceptionalismus.

Abb. 30 „Design Follows Data“, Superdot, 2022.

Das Modell des Datenexzptionalismus zielt grundsätzlich auf ein Narrativ der Ausnahme gegenüber Daten. Als theoretische Figur ermöglicht er, Ideen der Besonderheit von Daten bis hin zu einer Datenhybris, einer Überlegenheit von Daten oder einem *Datenchauvinismus* offenzulegen. Es handelt sich um die Sichtbarmachung einer datenpositivistischen Perspektive, die das Denken über und mit den Daten vor allem rhetorisch modelliert. Beispielhaft steht dafür die Datengetriebenheit im Datenjournalismus, das Positionierungsproblem der Digitalisierung in den Digital Humanities und der Idee des Lernens von Daten, welches in den Gründungsmomenten der Datenwissenschaften propagiert wird. Durch den Datenexzptionalismus entsteht eine Vorstellung einer Sonderstellung von Daten, die sich aus sich selbst heraus begründet und sich somit einer Selbstreflexion entledigt. Es wird eine *Datenversessenheit* gepflegt, die subjektive Einwirkungen außen vor lässt. Das bestimmende Paradigma ist: die Daten sind da und verändern die Gesellschaft von Grund auf und diese Gelegenheit muss ebenso affirmativ wie kritisch genutzt werden.

„Der Mythos von der technischen, politischen, und gesellschaftlichen Zwangsläufigkeit ist ein wirksames Beruhigungsmittel für das Bewußtsein. Seine Funktion besteht darin, die Verantwortung jedem von den Schultern zu nehmen, der an ihn glaubt. Aber in Wirklichkeit gibt es handelnde Personen!“

– Weizenbaum, 1978, S. 317.

Der Datenexzptionalismus will einen kulturellen Wendepunkt konstruieren, der vor allem mit der gesellschaftlichen Etablierung von Digitaltechnik einhergeht. Es geht dabei viel mehr um die Folgen der vernetzten Computertechnologie und ihrer Datenphänomene, als um Fragen nach dem Ursprung und Wesen von Daten an sich. Ich verbinde mit dieser Figur eine Möglichkeit zur Sichtbarmachung dieses konstruierten Modells und der damit verbundenen Konsequenzen für einen Datendiskurs.

Zusammenfassend lässt sich der Datenexzptionalismus als eine intendierte Reduktion des Datenbegriffs definieren. Diese folgenreiche Simplifikation basiert vor allem auf zwei folgenreichen Gleichsetzungen. Zuerst die Idee, dass Daten immer in numerisch-technischer Form auftreten, womit alle anderen Datenformen negiert werden. Und die Vorstellung von Daten als exklusives Phänomen von Digitaltechnik, was die lange Traditionsgeschichte vom Datenbegriff ignoriert. Daten werden somit oft in einer tech-

nodeterministischen Perspektive im Zuge der Sekundärfolgen der Computertechnologie problematisiert. Mit Nassehis Theorie der digitalen Gesellschaft ließ sich aber verstehen, dass Datenstrukturen in der Gesellschaft weit eher vorgefunden wurden, als dass die Digitaltechnik sie hervorgebracht hätte. Verkürzt formuliert könnte man sagen, dass die Datengesellschaft nicht durch die steigende Quantität der Daten geprägt wird, sondern vielmehr durch die kulturelle Projektion des Datenexzptionalismus.

In dieser Deutung kennzeichnet den Datenexzptionalismus eine zweite Idee der ausgeprägten *Datenvergessenheit*. Obwohl Daten ins Zentrum dieser Beobachtung gestellt werden, interessiert sie sich abseits dieses Modells nicht dafür, warum es Daten überhaupt gibt. Der Datenexzptionalismus projiziert einzig Folgen und nicht Gründe der Beschäftigung mit Daten. Eine grundlegende Beschäftigung mit dem Datenbegriff bleibt aus. Eine solche Anordnung ist allerdings nicht nur dem Datenpositivismus vorbehalten. Wie im Kapitel angedeutet wurde, nehmen teilweise auch datenkritische Stimmen den Gegenstand ihrer Kritik als gegeben an und fallen somit in die Datenaffirmation. Gerade für ein kritisches Verständnis von Datenphänomenen muss der Vorsatz Data generell mit besonderer Vorsicht betrachtet werden. Die bisher vorgeschlagenen Gegenpositionen zum Datenexzptionalismus bestehen in der Erweiterung des Datenbegriffs (als grundlegende Struktur der modernen Gesellschaft), die Rückbesinnung auf bestehende Datenkonzepte (Intentionalität) oder gar die Verabschiedung des Datenbegriffs (post-data) an sich. Was alle drei Ansätze eint, ist die Formulierung einer Alternative zu einem starren Datenmodell. Es gilt, Daten nicht als passives Material zu konzipieren, sondern sie als aktiv projiziertes Motiv zu verstehen. So gilt insbesondere für Datengestalter:innen: Daten sind kein Material für Datenbilder, sondern bringen vielmehr Themenprojektionen mit sich, die es mit Bedacht zu behandeln gilt.

„Data without a model is just noise.“

– Anderson, 2008.

Die wohl wichtigste Einsicht, die aus diesem Kapitel und der Betrachtung des Datenexzptionalismus erfolgte, ist der Aspekt der Modellierung. Der Mehrwert von Daten liegt nicht in der ersehnten automatisierten Analyse von Mustern in ihnen, sondern vielsprechender zeigte sich die Reflektion der Modelle, die sie hervorbringen. Daten sind zugleich Spiegel und Produzent von sozialer Realität.³³¹ Daten sind so gesehen nicht Ursache gesellschaftlicher Asymmetrien, sondern eher ein Effekt einer bestimmten Vorstellung was mit den Daten zu tun ist. Der Datenexzptionalismus beschreibt dann nur ein bestimmtes Modell in einer datenpositivistischen Art und Weise zu verfahren. Die Fragen nach diesem Modell, also Fragen warum und wozu Daten genutzt werden, verspricht dann womöglich sogar mehr Erkenntniswert als die Analyse der Daten selbst. Was gebraucht wird, ist dieser Argumentation nach kein weiterer algorithmic, computational oder digital turn, sondern die Rückbesinnung auf die Ideen, Vorstellungen und Konzepte, kurz die Modellierung der Daten. Daten werden per Definition verstanden als Abstraktionen, nicht als Abbilder von Realität, sondern immer Projektionen eines Modells über diese Realität. Das Mangelhafte an den Daten ist nicht, dass sie reduziert in ihrer Kapazität sind, sondern, dass ihnen das Vertrauen zur Vollständigkeit von der Gesellschaft zugeschrieben wird. Ganz im Sinne Vilém Flussers³³² beschreibt es auch Thomas Bächle:

„Die Modellierung der Realität ist stets ein konstruktiver Akt, in dem die Repräsentation der Welt zu einer durch Code abzubildenden Tatsache wird. Abbildung der Realität erschafft Realität.“

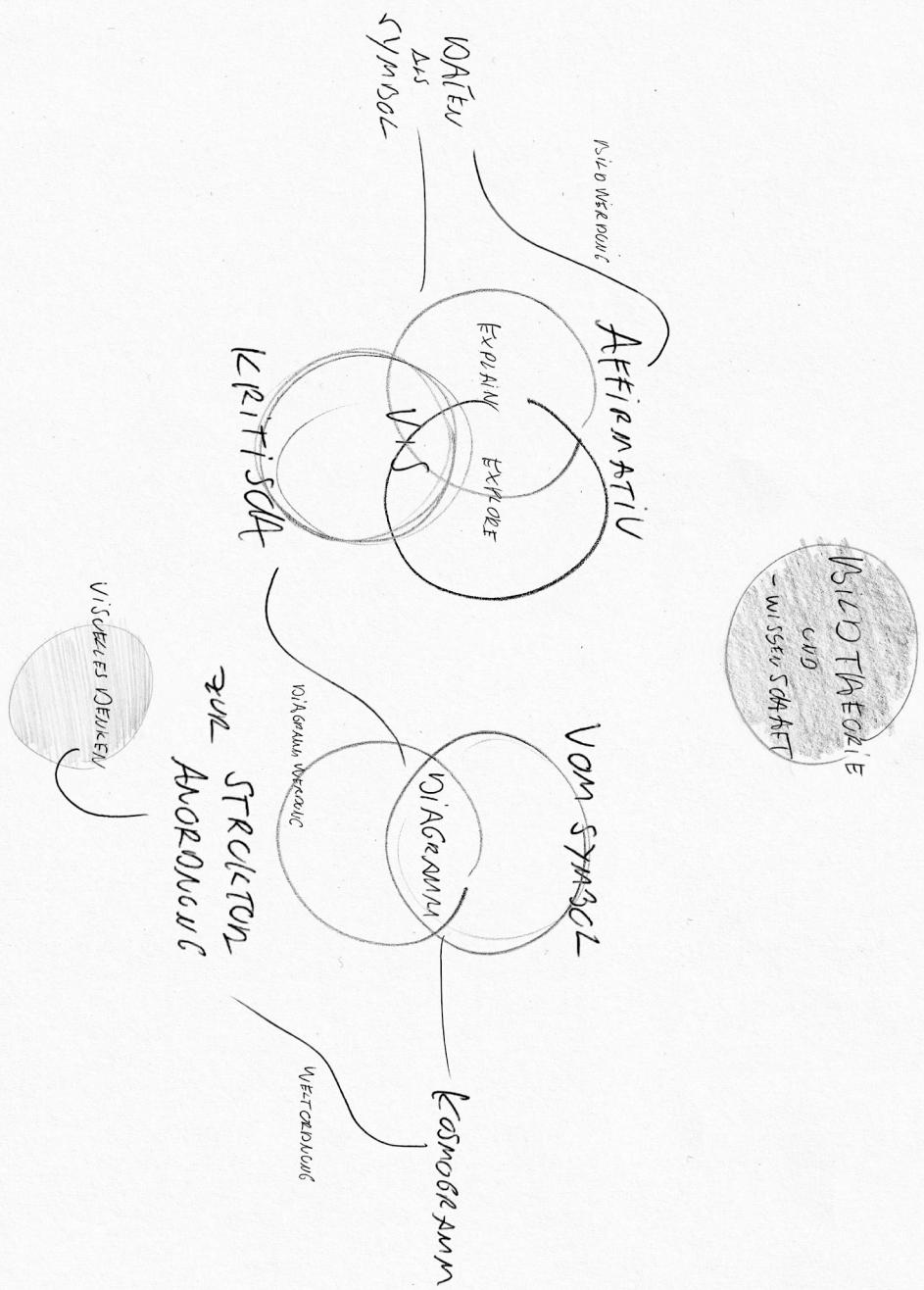
– Bächle, 2016, S. 91.

Der Datenexzptionalismus

ist ein ganz bestimmtes Modell mit Daten umzugehen. Dieses Modell wirkt sich auch stark auf die ästhetische Praxis und Theorie von Daten aus. Datenvisualisierungen sind so gesehen ein zeitgenössisches Phänomen, welches unter der simplifizierten Vorstellung von Daten als rein technisches und computerbezogenes Artefakt und Visualisierung als reine Repräsentationsidee leidet. Auf die hier angedeutete Abhängigkeitsbeziehung von Datenexzptionalismus und Datenvisualisierung möchte im nächsten Kapitel näher eingehen.

³³¹ Bächle, 2016, S. 149.

³³² Flusser, 1991.



KAPITEL 3

Visualisierte Modelle