



Marcus Burkhardt

DIGITALE DATENBANKEN

Eine Medientheorie
im Zeitalter von Big Data

[transcript] Digitale Gesellschaft

Aus:

Marcus Burkhardt

Digitale Datenbanken

Eine Medientheorie im Zeitalter von Big Data

Juli 2015, 390 Seiten, kart., zahlr. z.T. farb. Abb., 34,99 €, ISBN 978-3-8376-3028-2

Wir leben längst nicht mehr nur im Informationszeitalter, sondern in der Ära von Big Data. In dieser steht die Datenbank gleichzeitig für die riesigen Erkenntnispotenziale von Informationssammlungen wie für die bedrohlichen Informationsexzesse der digitalen Medienkultur. Zudem bezeichnet der Begriff konkrete Technologien und Verfahren der Sammlung und Bereitstellung von digitalen Informationen.

Zwischen diesen sehr unterschiedlichen Auffassungen gilt es, Datenbanken medientheoretisch zu verorten. Marcus Burkhardt zeichnet die Geschichte der Datenbanken nach und fragt, wie technische Verfahren der Verwaltung digitaler Informationen bedingen, was auf welche Weise *in* Datenbanken gefunden und *durch* sie gewusst werden kann.

Marcus Burkhardt (Dr. phil.) arbeitet am Centre for Digital Cultures der Leuphana Universität Lüneburg und leitet dort das Hybrid Publishing Lab.

Weitere Informationen und Bestellung unter:

www.transcript-verlag.de/978-3-8376-3028-2

Inhalt

Einleitung | 7

Medium: Zwischen Konstellationen und Konfigurationen | 21

Die Frage nach den Medien | 22

Wann sind Medien? | 33

Über Medien reden: Medienepistemologie | 58

Computer: Zwischen Oberfläche und Tiefe | 73

Phänomeno-Technische Konfigurationen | 75

Spielräume der computertechnischen Informationsvermittlung | 95

Datenbank: Zwischen digitalen Sammlungen und Sammlungstechnologien | 117

Was sind Datenbanken? | 121

Datenbanklogiken: Zur Datenbank als symbolischer Form | 131

Gegen die Datenbank als Prinzip: Mikrologiken der digitalen Datenhaltung | 145

Banken, Basen, Reservoirs: Information Storage and Retrieval | 149

Information: Zwischen begrifflicher Abstraktion und technischer Konkretion | 150

Kommunikation mit Informationssammlungen | 167

Daten und Information: Begriffsklärung | 187

Techno-Logik: Apparaturen, Architekturen, Verfahren | 205

Direct Access: Zur Festplatte als Herausforderung digitaler Datenbanken | 206

Datenbankmodelle: Architekturen für Datenunabhängigkeit | 221

Data + Access: Datenmodelle und Algorithmen | 242

Phänomeno-Logik: Mediale Praktiken mit Datenbanken | 283

Die Datenbank als latente Infrastruktur | 287

Orientierung im Vielen I: Das Eine finden | 296

Orientierung im Vielen II: Das Viele auswerten | 307

Schluss: Plädoyer für eine kritische Datenbankkultur | 327

Dekonstruktion der Datenbankfiktion | 333

Erinnern und Vergessen: Digitale Archive | 336

Fetischisierung des Quantitativen | 341

Literatur | 349

Einleitung

More, more, more! Das digitale Medienzeitalter ist das der Information. Mehr noch: Es ist ein Zeitalter der Informationsexzesse. Seit 2007 prognostiziert die *International Data Corporation* (IDC) regelmäßig eine immer rasantere Expansion des digitalen Informationsuniversums.¹ Wurden damals noch schätzungsweise 161 Exabyte, d.h. 161 Milliarden Gigabyte, digitaler Informationen kreiert, erfasst oder repliziert (vgl. IDC 2007: 3), waren es 2010 bereits ca. 1.200 Exabytes (vgl. IDC 2010: 1). Ein Ende dieser Expansion ist nicht in Sicht. Im Jahr 2020 werden, so die aktuellen Prognosen, 40.000 Exabyte digitale Informationen erzeugt (vgl. IDC 2012).² Auch wenn es sich hierbei zum größten Teil um flüchtige Daten handelt, die nicht permanent gespeichert werden, drängt sich die Frage auf, welche Effekte die scheinbar grenzenlose Verfügbarkeit von Informationen zeitigen wird.³ Wie ändert sich das

1 | Daten werden im Folgenden nicht strikt von Informationen unterschieden, sondern als eine Form der Materialisierung sowie ein Typus von Information begriffen. Im Kapitel »Banken, Basen, Reservoirs« (S. 187ff.) werden die Begriffe *Daten* und *Information* näher bestimmt. Hierbei wird deutlich, dass selbst der scheinbar eindeutige Begriff der Daten in der digitalen Medienkultur auf unterschiedliche Weise gebraucht wird.

2 | Die IDC-Studien betrachten hauptsächlich die zunehmende Akkumulation und Zirkulation von Daten. So geht es den Autoren darum, das Gesamtaufkommen digitaler Daten pro Jahr zu bestimmen. Daher werden nicht nur originäre Informationen oder Daten einbezogen, sondern auch all jene Daten, die anfallen, wenn Filme, Fernsehsendungen, Bilder, Texte, Audiodateien usw. kopiert und versandt werden (vgl. IDC 2007, 2008). Eine andere Perspektive haben die 2000 und 2003 an der Universität Berkeley angefertigten Studien zur Menge der jährlich erzeugten Informationen eingenommen (vgl. Lyman et al. 2000; Lyman et al. 2003). Im Zentrum der Studien stand die Frage wie viel »original content« (Lyman et al. 2000) jährlich erstellt wird. Zudem beschränkte man sich nicht auf den Bereich digitaler Informationen, sondern untersuchte, wie viele Informationen jährlich insgesamt erzeugt werden.

3 | Die Gesamtmenge der jährlich aufgezeichneten, erzeugten oder replizierten Daten überstieg dem IDC zufolge bereits 2007 die Speicherkapazität der verfügbaren Datenträger. Dabei werden in die Schätzungen des Datenaufkommens z.B. auch

menschliche Selbst- und Weltverhältnis, das gesellschaftliche und kulturelle Leben, unser Gedächtnis oder unser Wissen?

Antworten hierauf stehen noch weitgehend aus, denn das seit Jahrzehnten diagnostizierte Informationszeitalter ist noch immer im Werden begriffen.⁴ Gesichert ist nur, dass mit der fortschreitenden Digitalisierung weitere tiefgreifende gesellschaftliche und kulturelle Transformationen einhergehen werden. Jüngste Vorboten kündigen diesen Wandel an. Die im Frühsommer 2013 öffentlich gewordenen Spionageprogramme *Prism* und *Tempora* (vgl. Greenwald/MacAskill 2013; MacAskill et al. 2013a; MacAskill et al. 2013b), welche Orwells Überwachungs dystopie zu überbieten scheinen, sind nur ein Beispiel hierfür. Das Ende dieser Entwicklungen ist nicht abzusehen.

Wenn das fortwährende Werden des digitalen Medienzeitalters im Folgenden medienkulturwissenschaftlich in den Blick genommen werden soll, dann erfordert dies eine Untersuchungsperspektive, die sich der Vielgestaltigkeit der technischen Welt ebenso wie der Heterogenität von medialen Praxen stellt und diese medientheoretisch reflektiert.⁵ Den Ausgangspunkt hierfür bildet die medienhistorische Betrachtung der Diskurse über die Chancen und Herausforderungen der computer-technischen Informationsverarbeitung sowie der Genese digitaler Technologien. Die Hinwendung zur Diskurs- und Technikgeschichte ist dabei nicht Selbstzweck. Vielmehr vermag der Blick in die Geschichte des Computers, wie beispielsweise Wendy Chun (2011) in *Programmed Visions* vor Augen geführt hat, zu einem besseren theoretischen Verständnis unserer medialen Gegenwart beizutragen.

In den vergangenen Jahren dienten vor allem Netze und Netzwerke, technische wie soziale Netzwerke, sowie Algorithmen als Chiffren dieses medialen Transformationsprozesses.⁶ Doch mit dem Rumoren über die Sammelwut von Internetdienstleistern, Firmen und Staaten, den drohenden *Information Overload* und *Big Data* gerät zunehmend auch die Datenbank als spezifische, technische Infrastruktur und universelle Metapher digitaler Informationssammlungen in den Fokus medien-

digitale TV-Signale oder übertragene Webseitendaten mit einkalkuliert, die beim Empfänger allenfalls temporär gespeichert werden (vgl. IDC 2007: 3f.).

4 | Die Beschreibung unserer Epoche als Informationszeitalter kam spätestens in den 1970er Jahren auf. Angesichts der anhaltenden Dynamik gesellschaftlicher, politischer, ökonomischer und kultureller Transformationsprozesse ist die Entwicklung hin zum Informationszeitalter keineswegs abgeschlossen.

5 | Die Aufzählung der Disziplinen erhebt keinen Anspruch auf Vollständigkeit. Jenseits der Wissenschaften trägt auch die Kunst zu einem besseren Verständnis der gesellschaftlichen und kulturellen Transformationen im Informationszeitalter bei.

6 | Siehe hierzu exemplarisch den von Oliver Leistert und Theo Röhle herausgegebenen Sammelband *Generation Facebook* (2011) sowie *Die stille Revolution* von Mercedes Bunz (2012).

und kulturwissenschaftlicher Betrachtungen.⁷ Der notorische Stichwortgeber des aufkeimenden Datenbankdiskurses ist Lev Manovich, der bereits Ende der 1990er Jahre in einem vielbeachteten Aufsatz die Datenbank als symbolische Form der digitalen Medienkultur beschrieben hat.⁸ Durch die Datenbank wird seines Erachtens die Erzählung als dominante Form des kulturellen Ausdrucks abgelöst:

»After the novel, and subsequently cinema, privileged the narrative as the key form of cultural expression of the modern age, the computer age introduces its correlate – the database. Many new media objects do not tell stories; they do not have a beginning or end; in fact, they do not have any development, thematically, formally, or otherwise that would organize their elements into a sentence. Instead, they are collections of individual items, with every item possessing the same significance as any other.« (Manovich 2001: 218)

Für Manovich ist die Datenbank sowohl Symptom als auch Potenzial. So konstatiert er einerseits eine zunehmende Popularität der Datenbankform, für die die prinzipielle Unabgeschlossenheit, offene Anschlussfähigkeit und ungeordnete Multiperspektivität von Informationssammlungen charakteristisch sei (vgl. Manovich 2001: 219f.). Andererseits erblickt Manovich in Datenbanken die Chance, neue (medienkünstlerische) Ausdrucksformen und Gestaltungsprozesse zu erkunden (vgl. Manovich 2001: 227; Manovich/Kratky 2005).⁹

Manovichs holzschnittartige Gegenüberstellung von Erzählform und Datenbankform hat sich als ungemein populär erwiesen, wie die große Zahl an Publikationen zeigt, die an dessen Ansatz anschließen (vgl. exemplarisch Vesna 2007a; Böhme et al. 2012; Folsom 2007). Dennoch greift Manovichs Ansatz zu kurz, da die vielfältigen und teilweise gegenläufigen Transformationen unserer zeitgenössischen Medienkultur durch diese »easy binary« (McGann 2007: 1589) nicht adäquat beschrieben werden können. Vor allem die konkreten medientechnischen Verfahren der Versammlung, Verwaltung und Verarbeitung digitaler

7 | Die Diagnose eines Information Overload ist weder neu noch spezifisch für den Bereich digitaler Medientechnologien. Wie Markus Krajewski anmerkt, setzte bereits kurz nach der Erfindung des Buchdrucks ein Gefühl von Überwältigung durch die wachsende Zahl verfügbarer Bücher ein, welches in der Metapher der *Bücherflut* zum Ausdruck gebracht wurde (vgl. Krajewski 2002: 16). Ähnliche nautische Metaphern sind heute noch immer verbreitet, wie zum Beispiel im Begriff der *data deluge*, d.h. der Datensintflut, die im digitalen Zeitalter über uns hereinzubrechen droht (vgl. Borgman 2007: 6f.; Bell et al. 2009).

8 | Die Position Manovichs wird im zweiten Teil des Kapitels »Datenbank« ausführlich rekonstruiert und diskutiert.

9 | Unter dem Titel *Cultural Analytics* thematisiert Manovich in jüngster Vergangenheit auch verstärkt die Potenziale datenbankbasierter Forschungsmethoden für die Kulturwissenschaft; siehe hierzu exemplarisch Manovich (2009).

Informationen in Datenbanken bleiben bei Manovich weitgehend außen vor. Infolgedessen verschwindet die heterogene Vielgestaltigkeit der digitalen Datenbankkultur hinter der vermeintlichen Einheit der Datenbank als symbolischer Form.

Auch im öffentlichen Diskurs steht der informatische Begriff der Datenbank eher im Hintergrund. Das Wort *Datenbank* fungiert vielmehr als Projektionsfläche für die vielfältigen, heterogenen und scheinbar grenzenlosen Möglichkeiten der Verzeichnung, Zirkulation, Präsentation, Selektion und Auswertung von Informationen in Computern. Hierbei wird die Realität der computertechnischen Informationsverarbeitung von einem Imaginären überlagert. Stephan Porombka zufolge entfaltet sich an Datenbanken die »Phantasie vom virtuell vollständigen Gedächtnis« (Porombka 1998: 318) – alle möglichen Informationen scheinen in digitalen Datenbanken vorhanden und verfügbar zu sein. Auch wenn die Einlösung dieser Vollständigkeitsutopie stets ausbleiben wird, ist ihre Imagination ungemein verführerisch und suggestiv. Eben dies hat Jorge Luis Borges (1992 [1941]) in *Die Bibliothek von Babel* vor Augen geführt. In der kurzen Erzählung beschreibt Borges das Gedankenexperiment einer vollständigen Bibliothek, die nicht nur alle jemals verfassten, sondern alle möglichen Bücher enthält.¹⁰ Wenn jedes Buch, wie Borges beschreibt, 410 Seiten mit 40 Zeilen je Seite und 80 Zeichen je Zeile umfasst und das Alphabet aus 25 Zeichen (22 Buchstaben, Punkt, Komma und Leerzeichen) besteht (vgl. Borges 1992 [1941]: 68f.), dann ist die Zahl der Bücher in der Bibliothek zwar endlich aber dennoch unfassbar groß wie der Mathematiker William Goldbloom Bloch herausstellt: »*The number of books in the Library, although easily notated, is unimaginable*« (Bloch 2008: 22). Die totale Bibliothek enthält $25^{(410 \times 40 \times 80)} = 25^{1.312.000} \approx 1,956 \times 10^{1.834.097}$ Bücher.¹¹

Die Vollständigkeit der Bibliothek von Babel mag »ein überwältigendes Glücksgefühl« (Borges 1992 [1941]: 71) verursachen. Doch eigentlich sind ihre erhabene Größe und deren Berechenbarkeit nebensächlich, wenn nicht sogar trügerisch. Das Gedankenexperiment führt die Idee der Bibliothek ad absurdum. Gerade weil die totale Bibliothek jedes mögliche Buch enthält, finden sich in ihr keine tatsächlichen Bücher mehr. Jeder mögliche Autor¹² hat unter allen möglichen Titeln jeden möglichen Text verfasst sowie jeden möglichen Standpunkt und auch immer dessen Gegenteil vertreten. Die Universalbibliothek ist daher eine »widersprüchliche Bibliothek, deren vertikale Einöden aus Büchern unaufhörlich Gefahr laufen, sich in andere zu verwandeln, die alles bestätigen, leugnen und verwirren wie eine wahn-

10 | In dem 1939 verfassten Essay *Die totale Bibliothek* führt Borges die Idee der Universalbibliothek auf Gustav Theodor Fechner zurück, jedoch ohne eine genaue Quelle zu nennen (vgl. Borges 2003 [1939]: 165). Literarisch aufgearbeitet wurde diese Idee erstmals 1904 von Kurd Laßwitz in *Die Universalbibliothek* (1998 [1904]).

11 | Für eine Erläuterung dieser Berechnung siehe Bloch (2008: 11ff.).

12 | Der folgende Text macht weitgehend vom generischen Maskulinum Gebrauch. Sofern nicht genauer spezifiziert, sind stets Personen beider Geschlechter sowie Menschen, die sich außerhalb des binären Geschlechtersystems verorten, gemeint.

sinnige Gottheit« (Borges 2003 [1939]: 169). Mit der Universalisierung der Bibliothek löst sich die Spezifität des Buchs, verstanden als eine Sammlung von Aussagen, in der Kombinatorik des Möglichen auf.¹³ Wenn alles Mögliche geschrieben steht, ist eigentlich nichts mehr gesagt. Zudem wird auch das Suchen und Finden von Büchern an ihre paradoxe Grenze geführt. Existiert jede mögliche Kombination der 1.312.000 Zeichen, die ein Buch beinhaltet, muss der komplette Text bereits bekannt sein, um gezielt nach einem bestimmten Buch suchen zu können. Es kann, wie Bloch belegt, keinen Katalog geben, der die Suche in der Bibliothek erleichtern würde: »The Library is its own catalogue. Any other catalogue is unthinkable« (Bloch 2008: 39). Infolgedessen kommt der Akt des Suchens nach einem bestimmten Buch dem Schreiben des Buchs gleich.¹⁴

Die scheinbare Utopie der Vollständigkeit entpuppt sich in der Bibliothek von Babel als Dystopie. Es ist daher bemerkenswert wie hartnäckig die Signatur des Utopischen an der Vision einer totalen Bibliothek und in ähnlicher Weise auch an der Idee eines vollständigen Informationsbestands haftet.¹⁵ Das wahrscheinlich bekannteste Beispiel hierfür ist die von Vannevar Bush (1945) in *As We May Think* formulierte Idee der *Memex* – einer Art Bibliothek im Schreibtisch –, die rasch Popularität erlangte und als Vision die Entwicklung der digitalen Medienkultur maßgeblich mitprägte.¹⁶ Heute gilt Bushs *Memex*-Vision als einer der wichtigsten

13 | Das Spiel der kombinatorischen Möglichkeiten ist in der Bibliothek von Babel kein bloßes Potenzial, sondern aktualisiert sich in der Gesamtheit aller Bücher. In dieser Hinsicht unterscheidet sich die totale Bibliothek von kombinatorischen Maschinen, wie der von Jonathan Swift in *Gullivers Reisen* beschriebenen Maschine, welche nahezu selbstständig Bücher schreiben soll. Ein Professor der Akademie von Lagado beschreibt Gulliver den Zweck der von ihm erfundenen Maschine wie folgt: »Jedermann wisse, wie mühselig die übliche Methode sei, Kunst und Wissen zu erlangen, während durch seine Erfindung selbst die ungebildetste Person zu einem vernünftigen Preis und mit geringem körperlichen Einsatz Bücher in Philosophie, Poesie, Politik, Recht, Mathematik und Theologie schreiben könne, ohne die geringste Hilfe durch Begabung oder Lernen« (Swift 1987: 237).

14 | In Anlehnung an die Kurzgeschichte *Von der Strenge der Wissenschaft*, in der Borges die Idee einer Landkarte im Maßstab 1:1 vorführt, gelangt der Technikhistoriker David Gugerli zu einer ähnlichen Einschätzung: »Die vollkommene Suchmaschine ist so leistungsfähig wie die Karte im Maßstab 1:1. Sie kann alles suchen und findet eben deshalb nichts« (Gugerli 2006: 154).

15 | Eine wichtige Ausnahme stellen die Bedenken gegenüber Überwachung und Kontrolle durch staatliche und nichtstaatliche Organisationen dar, welche in Orwells Überwachungsdystopie zum Ausdruck kommen.

16 | Das von Paul Otlet und Henri La Fontaine Anfang des 20. Jahrhunderts begründete Mundaneum beruht auch auf der Idee einer universellen Bibliothek (vgl. Rayward 1975; Christolova 2012). Zur Geschichte der Idee der Weltbibliothek siehe auch Haber (2007).

Vorläufer von Hypertext und digitalen Literatur- und Informationssammlungen. Der schreibtsichgroße *Memory Expander* könne Bush zufolge als unerschöpflicher Informationsbestand fungieren, der seinen Nutzern den flexiblen Umgang mit vielfältigen Ressourcen ermöglicht.¹⁷ In jüngerer Vergangenheit hat Bill Gates – Mitbegründer und Vordenker von Microsoft – die Vision des technisch gestützten Zugriffs auf nahezu alle Informationen auf den Slogan *Information at Your Fingertips* gebracht. Als Mantra der digitalen Medienkultur führt *Information at Your Fingertips* die Vision des ungehinderten und vollständigen Zugriffs auf Information durch Computer vor Augen. Erstmals hat Gates diese Idee im Herbst 1990 im Rahmen seiner *Keynote* auf der *Computer Dealer's Exhibition* (COMDEX)¹⁸ postuliert.

Zu dieser Zeit standen Desktop-Computer bereits in vielen Büros und Haushalten. Jedoch mangelte es der Hardware wie der Software noch an Nutzerfreundlichkeit, wie Gates damals kritisch diagnostiziert. Hinzu kam, dass viele Nutzer nur sehr eingeschränkte Vorstellungen von den Potenzialen digitaler Computertechnologien hatten, die kaum über Textverarbeitung und Tabellenkalkulation hinausreichten (vgl. Gates 1990: 12:22). Vor diesem Hintergrund formuliert Gates seine Vision der digitalen Zukunft, für die der Slogan *Information at Your Fingertips* entsteht. Computer sollen ihm zufolge künftig als Informationsmaschinen dienen, indem sie Nutzern den Zugriff auf Information ermöglichen: »When somebody comes in to do their job they ought to sit down at their PC and see the information that's important to them. If they want more detail, they ought to just point and click and that detail should come up on the screen for them« (Gates 1990: 14:17). Gates zeichnet vom Computer jedoch nicht nur das Bild einer Informationsmaschine, mit der man auf diese oder jene Information zugreifen kann. Er formuliert vielmehr die Vision einer allgemeinen und vollständigen, d.h. universellen Informationstechnologie, die es ihren Nutzern ermöglicht, alle Informationen zu erhalten, die sie begehren. Kaum mehr als wenige Klicks entfernt, entfaltet sich das gesamte Weltwissen: »When I say information here I mean it in a very broad sense. I mean all the information that somebody might be interested in, including information they can't even get today« (Gates 1990: 14:05). Vier Jahre später aktualisiert Gates diese Vision in einer zweiten *Information at Your Fingertips*-Keynote auf der COMDEX in Las Vegas. Während er zunächst noch die Möglichkeiten eines weltweiten netzwerkbasierten Informationsaustausches und -zugriffs vernachlässigte, war das Internet nun integraler Bestandteil seiner Vision der digitalen Medienkultur. In ihren Grundzügen änderte sich die Vision jedoch nicht: Computer sollten zu universalen Informationsmaschinen

17 | Auf Bushs visionäre Beschreibung der Memex wird im Kapitel »Computer« (S. 103ff.) näher eingegangen.

18 | Die zwischen 1979 und 2003 jährlich im Herbst in Las Vegas veranstaltete COMDEX war eine der weltgrößten Computermessen. Seine erste Keynote hielt Bill Gates bereits 1983 auf dieser Messe.

werden, mit denen man vermeintlich alles finden und scheinbar unmittelbar mit Information interagieren kann (vgl. Gates 1994).

Zwei Dekaden später wissen wir, dass weniger Gates' Firma Microsoft, als vielmehr Google uns dieser Vision näher gebracht hat. Das Wort *googeln* hat Eingang in die deutsche Sprache gefunden und ist heute synonym für das Suchen im Internet, auch wenn Google es erfolgreich zu unterbinden wusste, dass der Duden diese allgemeine Bedeutungsvariante erwähnt.¹⁹ Mit Google scheint sich die Vision von Gates erfüllt zu haben – die Suchmaschine gibt Orientierung in der unübersichtlichen Informationsvielfalt des World Wide Web (WWW) und gewährt ihren Nutzern Zugriff auf vermeintlich grenzenlose Informationsressourcen: Nur wenige Suchanfragen entfernt liegt die Gesamtheit des Weltwissens. Wie die medien- und kulturwissenschaftliche Debatte über Websuchmaschinen in den vergangenen Jahren gezeigt hat, trägt dieser Eindruck jedoch (vgl. Becker/Stalder 2009a; Röhle 2010; Vaidhyanathan 2011). Der Suchmaschine sind trotz ihrer Größe und trotz ihrer enormen Leistungsfähigkeit Grenzen gesetzt. Google ist weder vollständig noch erlaubt es den neutralen Zugriff auf Informationen.²⁰ Ein Beispiel soll dies verdeutlichen: Als »Datenbank« des WWW ist die Suchmaschine Google paradoxerweise weitgehend blind gegenüber Informationen, die in Datenbanken gespeichert sind, d.h. von Informationen, auf die nicht als Webseite zugegriffen werden kann, sondern die vermittelt Webseiten abgefragt werden müssen (vgl. Madhavan et al. 2008).

Hieran zeigt sich jedoch nicht nur die Begrenztheit von Google, sondern auch das merkwürdige Doppelleben des Worts *Datenbank*. Dient es einerseits als Bezeichnung für jedwede Form digitaler Informationssammlungen, verweist es andererseits auf spezifische Technologien der Verwaltung von Informationssammlungen, sogenannte Datenbankmanagementsysteme (DBMS), deren Geschichte bis in die 60er und 70er Jahre des letzten Jahrhunderts zurückreicht. Zugleich macht das Beispiel deutlich, dass die Techniken und Praktiken der Versammlung, Verwaltung und Verarbeitung von Informationen Einfluss darauf haben, welche Informationen auf welche Weise gesucht, präsentiert und ausgewertet werden können. So selbstverständlich und unspektakulär diese Beobachtung zunächst zu sein scheint, so weitreichend sind ihre Konsequenzen für diejenige medientheoretische Tradition, die Shannons nachrichtentechnischen Informationsbegriff als theoretisches und praktisches Fundament der computertechnischen Informationsverarbeitung er-

19 | Auf Bitten von Google definiert der Duden seit der 2006 erschienenen 24. Auflage »googeln« als Suche im Internet mit Google. In einer breiteren Bedeutungsvariante wurde das Wort jedoch bereits 2004 in der 23. Auflage des Duden erstmals erwähnt (vgl. Heine 2005).

20 | Zur Funktionslogik von Websuchmaschinen im Allgemeinen und Google im Besonderen siehe S. 260ff. im Kapitel »Techno-Logik«.

achtet.²¹ Der vermeintlichen Einheit der Information steht in der medialen Praxis die Vielfalt partikularer Formen von Informationen gegenüber. Hiervon zeugt die konzeptuelle und technische Entwicklung von Datenbanken im engen und weiten Sinn. Denn im Kontext verschiedener Softwareanwendungen prozessieren Computer digitale Daten nicht unterschiedslos als Information im nachrichtentechnischen Sinne. Es werden vielmehr Unterschiede eingeführt, um digitale Daten als Information verwalten und verarbeiten zu können.

Mit der Einsicht in die Vielfalt der digitalen Informationsverarbeitung gilt es, sich von einem weiteren Diktum technischer Medientheorien zu verabschieden: der Irrelevanz der Semantik. Zwar mag Shannon Recht damit haben, dass der Aspekt der Bedeutung bei der Lösung des technischen Problems der Nachrichtenübermittlung zu vernachlässigen ist (vgl. Shannon 1976 [1948]: 41), aber im Kontext der digitalen Informationsverwaltung erweist sich die Frage der technischen Handhabung von Bedeutung geradezu als zentral – auch und gerade weil Computer den Sinn von Informationen nicht verstehen. Dementsprechend können digitale Datenbanken als Technologien der nicht-semantischen Verarbeitung von Bedeutung betrachtet werden.²² Hierauf beruht nicht zuletzt ihre Anschlussfähigkeit an vielfältige Gebrauchskontexte. Will man die mediale Eigenlogik von digitalen Datenbanken freilegen, gilt es die unterschiedlichen Modi der Herstellung »computer-lesbare[r] Signifikanz« (Becker/Stalder 2009b: 8) in den Blick zu nehmen. Daher wird im Folgenden keine medienkulturtheoretische Makroperspektive auf Datenbanken eingenommen, sondern auf unterschiedlichen Ebenen nach den medialen Mikrologiken der digitalen Datenhaltung gefragt.

Den Ausgangspunkt für die Auseinandersetzung mit digitalen Datenbanken bildet die Beobachtung, dass den vielfältigen, heterogenen und zum Teil gegenläufigen Praktiken der Versammlung, Verwaltung, Selektion und Auswertung von digitalen Informationen in der heutigen Medienkultur ein zentraler Stellenwert zukommt. Um diese medientheoretisch zu beschreiben und zu analysieren, gilt es, wie Manfred Faßler treffend herausgestellt hat, eine »beobachtende Sprachfähigkeit« (Faßler 2002: 21) über Datenbanken zu entwickeln, welche es ermöglicht, die vieltätige Realität digitaler Informationsverarbeitungspraxen sowie das Imaginäre digitaler Datenbanken zu analysieren. Hierbei besteht die Herausforderung darin,

21 | Gemeint ist die Tradition technischer Medientheorien in Folge Friedrich Kittlers, die heute unter anderem unter der Bezeichnung Medienarchäologie firmieren. Ein wichtiger Proponent ist Wolfgang Ernst, der Information im Anschluss an Shannon als »sub-semantische[n] Effekt der Hardware von Kommunikation« (Ernst 2002: 173) verstanden wissen will.

22 | Siehe hierzu vor allem den Vergleich des nachrichtentechnischen Kommunikationsmodells mit dem Modell des Information Retrieval im Kapitel »Banken, Basen, Reservoirs« (S. 167ff.) sowie die Unterscheidung verschiedener Modi der computertechnischen Verarbeitung und Zuschreibung von Bedeutung im dritten Teil des Kapitels »Techno-Logik« (S. 242ff.).

dass sich Datenbanken einer Einordnung in etablierte Raster medientheoretischer Reflexion entziehen. So erscheint die Datenbank auf der einen Seite als Metamedium, das der bloßen Versammlung vorhandener Informationen dient. Auf der anderen Seite erweist sich die Datenbank zugleich als Medium, aus dem neue Informationen geschöpft werden können und das vielfältige Möglichkeiten zur Formbildung eröffnet. Insofern drängt sich die Frage auf, wie die Medialität digitaler Datenbanken zu bestimmen und zu beschreiben ist. Diese Frage ist jedoch bereits auf der Ebene der medientheoretischen Betrachtung des Computers problematisch.

Als zweckoffen programmierbare Maschine widerstrebt der Computer einfachen Zuschreibungen, wie Georg Christoph Tholen (2002) in *Die Zäsur der Medien* argumentiert hat. Durch Programme wird der Computer hingegen auf Zwecke festgelegt, die dessen Gebrauch strukturieren. Im Spannungsfeld von zweckoffener Programmierbarkeit und zweckhaftem Gebrauch oszillieren Computer daher unentschieden zwischen Universalmedium und partikularem Medium. Doch auch auf der Ebene des gebrauchenden Umgangs von spezifischen Computeranwendungen eröffnen Computer vielfältige Möglichkeiten. So kann nahezu alles, was mit Computern getan wird, auf unterschiedliche Weise realisiert werden. Dieser doppelten Kontingenz Rechnung zu tragen, ohne zugleich die konkreten medialen Praxen mit Computern aus dem Auge zu verlieren, ist eine der zentralen Schwierigkeiten der medientheoretischen Beschreibung des Computers.

Die folgende Annäherung an digitale Datenbanken erfolgt aufgrund der skizzierten Herausforderungen im medientheoretischen Dreischritt vom Medium über den Computer zur Datenbank. Im ersten Schritt wird im Kapitel »Medium« die aktuelle Debatte über den Medienbegriff und die eng damit zusammenhängende Frage der Medialität der Medien rekonstruiert. Als zentrales Problem erweist sich dabei die relativ geringe (Trenn-)Schärfe des Medienbegriffs, die im medienphilosophischen Kontext insbesondere von Matthias Vogel (2001, 2003) und Lambert Wiesing (2005b) kritisiert wurde. Mit dieser Problemstellung sind zwei Fragen verbunden: Wie können Medien *erstens* von Nichtmedien unterschieden werden, und auf Grundlage welcher Kriterien ist es *zweitens* möglich, einzelne Medien sinnvoll zu unterscheiden? Beide Fragen entfalten sich an Zuschreibungen der Form »X ist ein Medium«. Anstelle einer Lösung wird eine teilweise Auflösung des Problems vorgeschlagen, mit der ein Wechsel der Perspektive auf den Forschungsgegenstand der Medienwissenschaft einhergeht. Während es durchaus sinnvoll ist, danach zu fragen, was den Bereich des Medialen von dem des Außermedialen unterscheidet, greifen alle Unterscheidungsversuche von Einzelmedien zu kurz. Infolgedessen erweist es sich stets als problematisch, wenn etwas als Medium charakterisiert wird. Hierauf weist Dieter Mersch (2008, 2010) im Kontext seiner negativen Medientheorie hin.

Es wird daher vorgeschlagen, Medien begrifflich als das Andere des Kommunikationsprozesses zu fassen, welches jedoch nicht positiv als *Medium X* oder *Medium Y* identifiziert werden kann. Die Untersuchungsgegenstände der Medienwissenschaften sind infolgedessen streng genommen keine Medien, sondern viel-

mehr historisch wandelbare mediale Konfigurationen, die sich in unterschiedlichen Hinsichten auf signifikante Weise verändern können. Die geläufige Bezeichnung Medium verweist insofern auf nichts anderes als temporär relativ stabile, mediale Konfigurationen, wie z.B. die Malerei, den Buchdruck oder das Fernsehen, welche sich anhand begrifflicher Kriterien jedoch nicht trennscharf voneinander unterscheiden lassen. Vermieden wird hierdurch zum einen die Gefahr von Essentialismen. Zum anderen wird es möglich, Verschiebungen medialer Konfigurationen auf der Mikroebene in den Blick zu nehmen. So lassen sich Großrechenanlagen, Desktop-Computer, Laptops und Smartphones zwar alle als Computer thematisieren. Dennoch sind sie es nicht in gleicher Weise.

Auf der Grundlage des im Kapitel »Medium« entwickelten begrifflichen Instrumentariums widmet sich das darauf folgende Kapitel der Frage nach der Medialität des Computers. Einen wichtigen Anknüpfungspunkt hierfür stellen Niklas Luhmanns bislang zu wenig beachteten Ausführungen zum Medium Computer in *Die Gesellschaft der Gesellschaft* dar.²³ Luhmann zufolge ist eine rein technische Analyse des Computers, die ausschließlich die in der Tiefe der Maschine ablaufenden Prozesse thematisiert, ebenso wenig zureichend wie Analysen, die sich auf die an den Bildschirmoberflächen erscheinenden Phänomene beschränken. Die Medialität des Computers sei vielmehr im Rahmen einer medialen Topologie zwischen Oberfläche und Tiefe zu verorten. Da Luhmann keine ausgearbeitete Medientheorie des Computers entwickelt hat, können dessen Bemerkungen allenfalls zum Ausgangspunkt genommen werden, um die im Computer realisierten Oberfläche/Tiefe-Verhältnisse genauer zu analysieren. Hierbei wird zum einen der Aspekt von Software beleuchtet und zum anderen die Frage nach den Daten. Hierauf aufbauend werden im letzten Teil des Kapitels »Computer« Spielräume der computertechnischen Informationsvermittlung diskutiert. Im Anschluss an Frieder Nakes (2001, 2005) Unterscheidung der Oberfläche und Unterfläche digitaler Medienobjekte wird der Fokus auf das Wechselverhältnis von menschlichen und technischen Formen der Verarbeitung von Informationen gelegt. Darüber hinaus wird das komplexe Wechselspiel von Einem und Vielem, d.h. von digitalem Medienobjekt und Sammlungen digitaler Medienobjekte im Computer betrachtet. Dies leitet zum Thema Datenbanken über, dem sich im weiteren Verlauf der Untersuchung aus unterschiedlichen Perspektiven angenähert wird.

Das Kapitel »Datenbank« widmet sich zunächst der Öffnung dieses medienwissenschaftlichen Forschungsfeldes. Anhand der Frage, was Datenbanken eigentlich sind, werden die unscharfen Grenzen des Gegenstandsbereichs freigelegt, welche aus der Ambiguität des Datenbankbegriffs resultieren. Datenbanken situieren sich zwischen allgemeinem Sammlungsbegriff und spezifischen Sammlungstechnologien. Die Spannungen, welche aus dieser Mehrdeutigkeit resultieren, werden

23 | Die Potenziale des neuen Mediums Computer diskutiert Luhmann in *Die Gesellschaft der Gesellschaft* auf knapp zehn Seiten unter der Überschrift »Elektronische Medien« (vgl. Luhmann 1998: 302ff.).

jedoch nicht aufgelöst, sondern zum Ausgangspunkt der weiteren Annäherung an dieses medienwissenschaftliche Untersuchungsfeld genommen. Kartiert werden die Dimensionen sowie die Ambivalenzen, die dem Datenbankbegriff seit dessen Auftauchen Ende der 1950er Jahre inhärent sind. Im Anschluss wird die These Manovichs verhandelt, dem zufolge die Datenbank eine symbolische Form der digitalen Medienkultur konstituiert. Manovichs Text stellt einen, wenn nicht sogar *den* zentralen Referenzpunkt der derzeitigen medienwissenschaftlichen Debatte über Datenbanken dar. Jedoch erweist sich dessen Position als problematisch, da er verschiedene Aspekte digitaler Datenbanken unsystematisch nebeneinanderstellt und diese in dem weitgehend unthematisch bleibenden Begriff der symbolischen Form vermengt. Infolgedessen werden die heterogenen Praktiken der Versammlung *von* und des Umgangs *mit* digitalen Informationen in einem vermeintlich einheitlichen Verständnis von Datenbanken homogenisiert. Der auf diese Weise entgrenzte Datenbankbegriff verliert jegliches Differenzierungsvermögen. Alternativ zur Einnahme einer solchen Makroperspektive schlage ich vor, auf der Mikroebene die medialen Konfigurationen zu betrachten, in der Datenbanken auftauchen bzw. die durch digitale Datenbanktechnologien konstituiert werden. Leitend ist dabei die These, dass der Datenbankbegriff auf keine einheitliche Form verweist, sondern vielfältige Technologien der Informationsverwaltung im Computer sowie heterogene Praktiken mit digitalen Informationssammlungen bezeichnet, deren Eigenheiten es medientheoretisch zu beschreiben und zu verstehen gilt.

Bevor die konkreten Techniken und Praktiken der Versammlung und Verwaltung von Informationen in Datenbanken betrachtet werden, ist der historische Kontext zu beleuchten, in dem sich die Datenbankidee in den 1950er und 60er Jahren herausgebildet hat. Im Hintergrund der Entwicklung von Datenbanken steht das Aufkommen eines abstrakten, generalisierten und reifizierten Informationskonzepts. Die Genese dieses Konzepts, welches sich in den 1940er Jahren aus der Nachrichtentheorie und der Kybernetik entwickelt hat, wird im ersten Teil des Kapitels »Banken, Basen, Reservoirs« in Anlehnung an Katherine Hayles (1999) und Geoffrey Bowker (1994) nachgezeichnet. Information erweist sich dabei nicht als uniforme Größe, mit der Computer stets auf die gleiche Weise operieren. Der abstrakte Informationsbegriff ist vielmehr ein *travelling concept*, das sich in verschiedenen Gebrauchskontexten auf unterschiedliche Weise konkretisiert.²⁴ Dabei bildet die abstrakte Vorstellung von Information die Basis für das Imaginäre digitaler Datenbanken, welches die realen Möglichkeiten digitaler Informationsverwaltungstechnologien einerseits zwar stets übersteigt, als *Wunschkonstellation* die mediale Praxis mit Datenbanken aber andererseits immer begleitet.²⁵ Im zweiten Schritt wird anhand verschiedener Modelle der Kommunikation mit Informations-

24 | Zum Begriff des *travelling concept* siehe Bal (2002).

25 | Auf die Notwendigkeit, digitale Medientechnologien auf die Wünsche und Utopien hin zu befragen, die ihnen eingeschrieben sind bzw. die sie produzieren, hat Hartmut Winkler in *Docuverse* hingewiesen (1997a).

sammlungen diskutiert, wie sich das abstrakte Informationskonzept im Kontext digitaler Datenbanken konkretisiert. Diese Modelle spannen, so die These, den Problemkontext auf, in dem sich die Entwicklung partikularer Datenbanktechnologien vollzogen hat. Um dies zu verdeutlichen, wird zunächst das Kommunikationsmodell des *Information Retrieval* rekonstruiert, welches Ende der 1950er Jahre in ein dezidiertes Modell der Datenbankkommunikation kulminierte. Die aufgezeigte Vielgestaltigkeit von Information macht es abschließend notwendig, den Gebrauch der Begriffe Daten und Information zu problematisieren. Auch hier soll es nicht das Ziel sein, Definitionen vorzuschlagen. Vielmehr gilt es, ein Vokabular zu entwickeln, um die verschiedenen Formen und unterschiedlichen Weisen zu differenzieren, auf denen Daten und Information in Computern operativ werden.

In den darauffolgenden Kapiteln »Techno-Logik« und »Phänomeno-Logik« werden die technischen Logiken *von* und medialen Praxen *mit* Datenbanken thematisiert. Auf der Ebene der Technik werden Apparaturen, Architekturen und Verfahren der Verwaltung von Informationen im Computer betrachtet und auf ihre mediale Eigenlogik hin untersucht. Die Entwicklung von Datenbankmanagementsystemen ist eng mit der Einführung der Massenspeichertechnologie *Festplatte* im Jahr 1956 verwoben, welche den wahlfreien Zugriff auf den Datenträger ermöglicht. Der Übergang von seriellen zu wahlfreien Zugriffsformen auf persistente Sekundärspeicher lässt einerseits neue Anwendungs- und Einsatzgebiete von Computern möglich erscheinen. Andererseits erweist sich die Festplatte zugleich als Herausforderung, die in dem Problem kulminiert, wie Informationen im Speicher automatisch adressiert und damit vom Computer verwaltet werden können. Diese Herausforderung wurde Ende der 1960er Jahre auf den Begriff der Datenunabhängigkeit gebracht. Hierdurch wird das Ziel bezeichnet, den Gebrauch von Datenbankinformationen an den Oberflächen vielfältiger Anwendungsprogramme weitgehend von der Verwaltung des Informationsbestandes in der unsichtbaren Tiefe des Computers zu entkoppeln. Zur Lösung dieses Problems wurde zunächst eine Zwei-Ebenen-Architektur von DBMS vorgeschlagen, auf deren Grundlage die *ANSI/X3/SPARC Study Group on Data Base Management Systems* schließlich eine Drei-Ebenen-Architektur entworfen hat, die als Metamodell der Informationsmodellierung in Datenbanken fungiert. Zwischen der internen Speicherlogik von Informationen im Computer und den externen Gebrauchslogiken wird eine Ebene der konzeptuellen Beschreibung von Information eingezogen, welche die Vermittlung zwischen der Tiefe des Informationsspeichers und den Oberflächen von Anwendungsprogrammen ermöglicht. Hierdurch wird die Ebene bestimmt, auf der Informationsmodelle in Datenbanken operativ werden. Seit nunmehr 30 Jahren ist das relationale Datenmodell *de facto* der Standard der Modellierung und Verwaltung von Informationssammlungen in Datenbanken. Es handelt sich um ein spezifisches Verfahren der nicht-semantischen Verarbeitung von Bedeutung im Computer, das es aufgrund seiner langjährigen Dominanz zu verstehen und zu beschreiben gilt. Im Kontext von Websuchmaschinen und des Semantic Web haben sich in der jüngeren Vergangenheit alternative Verfahren der computertechnischen Verarbeitung von

Bedeutung im doppelten Wortsinn von Sinn und Relevanz herausgebildet, deren technische Eigenlogiken abschließend diskutiert werden.

Auf dieser medientechnischen Betrachtung aufbauend werden im Kapitel »Phänomeno-Logik« unterschiedliche mediale Praxen mit Datenbanken beleuchtet. Dabei werden drei Erscheinungsformen digitaler Datenbanken differenziert, die jeweils unterschiedliche Fragen und Probleme aufwerfen. Die Datenbank als latente Infrastruktur bleibt weitgehend im Verborgenen, bedingt jedoch den Umgang mit medialen Konstellationen im Computer. Dies zeigt sich beispielsweise an datenbankbasierten *Content Management Systemen*, welche durch Templates die flexible Gestaltung und Umgestaltung von Webseiten ermöglichen. Auf einer zweiten Ebene erscheinen Datenbanken als Informationssammlungen, welche das Finden des Einen im Vielen einer Informationssammlung ermöglichen, wie z.B. in Websuchmaschinen oder digitalen Bibliothekskatalogen. Die Query stellt hierbei den zentralen Modus des Suchens und Findens von Informationen in Datenbanken dar. Jenseits der Homogenität der Suchinterfaces gilt es jedoch die Unterschiede zwischen verschiedenen Anfrage-logiken freizulegen. Neben dem Finden des Einen ist die Auswertung des Vielen in der zeitgenössischen Medienkultur von großer Bedeutung. Hierfür steht das derzeit vieldiskutierte Schlagwort Big Data. Obwohl es sich hierbei fraglos um ein *Buzzword* handelt, sind die Möglichkeiten und Grenzen des Versprechens in Datenbanken, nicht nur Bekanntes wiederfinden, sondern Neues entdecken zu können, medientheoretisch zu analysieren. Neben mathematischen Verfahren des *Data Mining* wird insbesondere Visualisierungen das Potenzial zugeschrieben, Zusammenhänge in ansonsten unanschaulichen Informationssammlungen aufzufinden und hieraus neue Erkenntnisse abzuleiten. Dieses Vermögen von Informationsvisualisierungen wird anknüpfend an Peirces (1960, 1976) Überlegungen zur Diagrammatik diskutiert. Dabei zeigt sich, dass das Erkenntnis-potenzial der *Information Visualization* in einem gewissen Spannungs-verhältnis zum Potenzial visueller Darstellungsformen steht, Einsichten auf leicht verständliche und evidente Weise zu kommunizieren.

In der Vielfältigkeit und Vielgestaltigkeit der medialen Praktiken, die auf Datenbanken beruhen, zeigt sich deren zentrale Bedeutung für die digitale Medienkultur. Zugleich wird die offene Anschlussfähigkeit digitaler Datenbanken offenkundig, welche der Bestimmung *einer* medialen Logik der Datenbank zuwiderläuft. Die unternommene Betrachtung von Mikrologiken der digitalen Datenhaltung führt daher nicht zu einer einheitlichen Medientheorie digitaler Datenbanken. Vielmehr werden unterschiedliche, zum Teil gegenläufige Dimensionen, Aspekte und Tendenzen der Versammlung und Verwaltung von Informationen in Computern freigelegt. In dieser Hinsicht möchte das folgende Buch einen Beitrag zu einem medientheoretischen Verständnis von Datenbanken leisten und Perspektiven für eine kritische Auseinandersetzung mit der sich rasant fortentwickelnden Datenbankkultur eröffnen.