

Banken, Basen, Reservoirs

Information Storage and Retrieval

Anfang der 1960er Jahre wurde die Speicherung und Abfrage großer Informati-onssammlungen in Computern auf den Begriff *Datenbank* gebracht. Ihre Konzeption ist eng mit der theoretischen Beschreibung, technischen Implementierung und kommerziellen Distribution von Computern verbunden, geht jedoch nicht gänzlich darin auf. Die Entwicklung von Datenbanken wurde durch die Hoffnung beför-dert, »dass alle relevanten internen oder externen, vergangenen oder zukünftigen, wirtschaftlichen oder menschlichen Informationen in einer einzigen Struktur untergebracht werden können« (Haigh 2007: 60). Dieser Wunsch bildet, obwohl un-eingelöst, das Imaginäre digitaler Datenbanken, das nicht nur die Speicherung aller möglichen Informationen, sondern auch deren universelle Verwendung möglich erscheinen lässt. Im Zentrum dieses Versprechens steht ein generalisierter, ab-strakter und reifizierter Informationsbegriff, der aus der Nachrichtentheorie und Kybernetik heraus entstand (vgl. Hayles 1999: 50ff.; Bowker 1994) und durch das sich neu formierende Forschungsgebiet des *Information Retrieval* auf den Bereich digitaler Datenbanken übertragen wurde.¹

Im Folgenden sollen die Konturen dieses Informationsbegriffs nachgezeichnet werden. Hierbei wird die These vertreten, dass Information keinesfalls ein ein-heitliches Konzept und damit auch keine einheitliche Größe darstellt, welche als uniforme Basis medientechnischer Innovationen und medialer Praktiken dient.

1 | Zwischen dem Aufkommen des abstrakten Informationsbegriffs und der Formu-lierung der Idee digitaler Datenbanken sowie ihrer technischen Realisierung lässt sich nicht nur eine Verbindungslinie knüpfen. So hat Thomas Haigh in einer Reihe von Publikationen die Genese der Datenbankidee mit Informationsutopien des Managementdiskurses der 1950er Jahre in Verbindung gebracht. Da es im Folgenden nicht darum geht, eine möglichst umfassende (Vor-)Geschichte der Datenbank zu schreiben, sei an dieser Stelle auf die sozial- und technikhistorischen Arbeiten von Haigh verwiesen (vgl. 2001a, 2001b, 2007, 2009).

Information fungiert vielmehr als *travelling concept*², das sich in verschiedenen Gebrauchskontexten auf unterschiedliche Weise konkretisiert und dabei disparate Bedeutungen annimmt. Vor diesem Hintergrund werden im Unterkapitel »Kommunikation mit Informationssammlungen« Modelle der Datenbankkommunikation diskutiert, die die theoretische Lücke zwischen dem abstrakten Informationsbegriff und den medialen Praktiken mit Datenbanken füllen. Die Spezifik der Datenbankkommunikation zeigt sich hierbei vor allem in Differenz zu dem von Claude Shannon (1976 [1948]) entwickelten nachrichtentechnischen Kommunikationsmodell. Wenn die Pluralität und Heterogenität von Information im Spannungsfeld von begrifflicher Abstraktion und technischer Konkretion offenkundig wird, bedarf es eines Vokabulars, welches erlaubt, unterschiedliche Formen von Information zu beschreiben. Diesem Problem widmet sich der abschließende dritte Teil »Daten und Information«.

INFORMATION: ZWISCHEN BEGRIFFLICHER ABSTRAKTION UND TECHNISCHER KONKRETION

Den Horizont digitaler Datenbanken bildet ein abstrakter Informationsbegriff, auf dem das Versprechen beruht, in Datenbanken *alle* Informationen versammeln, speichern und finden zu können. Obwohl die Utopie eines universellen und vollständigen Informationsreservoirs in der Praxis unerreicht bleibt, ist dieses Versprechen als das Imaginäre digitaler Datenbanken wirksam. Es ist der Motor der zeitgenössischen Sammellust, aber auch die Wurzel der Sorge, dass bereits alles in digitalen Datenbanken gefunden werden kann.³

2 | Den Begriff der *travelling concepts* hat Mieke Bal eingeführt, um auf die Transformationen hinzuweisen, denen die Bedeutung von Begriffen an verschiedenen Orten, zu verschiedenen Zeiten und in verschiedenen Kontexten unterliegt. Vor dem Hintergrund ihres Entwurfs der *Kulturanalyse* plädiert Bal für eine »eine ernsthafte Beschäftigung mit Begriffen« (Bal 2002: 9), bei der die unterschiedlichen Bedeutungsvarianten freigelegt werden, die sich in vermeintlich klar umrissenen Begriffen überlagern. Als Beispiel führt Bal den Begriff des Texts an: »In der AlltagsSprache verbreitet, in der Literaturwissenschaft selbstverständlich, in der Anthropologie metaphorisch gebraucht, in Kunstgeschichte wie Filmwissenschaft ambivalent zirkulierend und in der Musikwissenschaft vermieden, zieht dieser Begriff Auseinandersetzungen und Kontroversen auf sich, die wunderbar anregend sein können, wenn sie ›aufgearbeitet‹ werden. Doch wenn keine derartige Aufarbeitung stattfindet, können die gleichen Verwendungen des Begriffs ›Text‹ zur Quelle von Missverständnissen werden oder – schlimmer noch – zu unverstandener Parteinaahme verlocken« (Bal 2002: 12).

3 | In der digitalen Medienkultur überlagern sich das Streben nach mehr Informationen und die Angst vor zu viel Informationen. David Gugerli hat beide gegenläufigen

Das Reden über und der Umgang mit Information bewegt sich zwischen der Einheit eines abstrakten, aber unterbestimmten Informationsbegriffs und der Vielfalt informationeller Praktiken, an denen Menschen ebenso wie Technologien teilhaben. Als Ausgangspunkt für die Diskussion dieses Spannungsverhältnisses dient die von Calvin N. Mooers – einem der geistigen Väter des Information Retrieval – verfasste Spekulation über die Zukunft der Menschheit. Das Ziel der Ausführungen zu Mooers ist es, am Kreuzungspunkt von bibliothekarischer und technischer Informationsverarbeitung eine Perspektive auf den prekären Status des Informationsbegriffs freizulegen, der diesem seit den 1940er Jahren innewohnt.

Ein Jahr nach dem Ende des Zweiten Weltkrieges schrieb Mooers als damals 27-jähriger Student der Mathematik und Physik am *Massachusetts Institute of Technology* einen kurzen Text, den er mit *This is an inquiry into the future of mankind* betitelte.⁴ Sich dem schwerwiegenden Problem der Zukunft der Menschheit widmend, denkt Mooers über die Frage nach, ob Roboter⁵ in naher Zukunft alle Fähigkeiten von Menschen übertreffen und sie infolgedessen überflüssig machen werden. Auch wenn er nicht so weit gehen will, die bevorstehende Verdrängung von Menschen durch die neuen digitalen Technologien zu behaupten, so war Mooers davon überzeugt, dass Roboter imstande sein werden, eine Reihe von Aufgaben besser zu erfüllen als Menschen: »The true use of robots will be for jobs that are not fit for a human being« (Mooers 1946). Hierzu zählt er zum Beispiel die Vermittlung von Telefongesprächen, die damals bei Anrufen im lokalen Ortsnetz der USA bereits teilweise automatisiert war.⁶ Musste es für Mooers in Anbetracht dessen nur

Tendenzen pointiert zusammengefasst: »Wenn die Datenflut vorhandene Verarbeitungskapazitäten übersteigt oder die Datenknappheit einen im Dunkeln tappen lässt, dann werden Daten zum Problem« (Gugerli 2007b: 7).

4 | Das unpublizierte Manuskript, welches auf den 14. Dezember 1946 datiert ist, findet sich im Nachlass von Mooers im Charles Babbage Institute der University of Minnesota, Minneapolis.

5 | Heute wäre wahrscheinlich eher von Computern als von Robotern die Rede. Als Mooers seine Prognose verfasste, diente das Wort Computer jedoch noch nicht als allgemeiner Sammelbegriff für die programmierbare Maschine, sondern bezeichnete den Gebrauch digitaler Technologien als Rechenmaschinen und damit eine spezifische Verwendungsweise.

6 | Ein Verfahren zur automatischen Vermittlung von Telefongesprächen wurde bereits 1889 von Almon Brown Strowger entwickelt. Die Automatisierung der Telefonvermittlung war jedoch ein langwieriger Prozess (vgl. Fliehy 1994: 197ff.). Hiervon zeugt beispielsweise der Ende der 1920er Jahre von der *American Telephone & Telegraph Corporation* produzierte Lehrfilm »How to Use the Dial Phone« (1927), der die Telefonkunden über den richtigen Gebrauch von Selbstwahltelefonen instruierte. Die Möglichkeiten zur Direktwahl beschränkten sich zunächst auf Gespräche im örtlichen Telefonnetz. Das *Direct Distance Dialing*, also die automatische Vermittlung von Ferngesprächen, war in den USA erst ab 1951 möglich (Dempewolff 1951).

noch als eine Frage der Zeit erscheinen, bis auch bei Ferngesprächen kein menschlicher Vermittler mehr notwendig sein wird, um diese Verbindungen herzustellen, so sind die weiteren von ihm ersonnenen Einsatzmöglichkeiten von Computern weitaus spekulativer.⁷ Der Abbau von Kohle, die Reinigung von Abflussrohren, das Fliegen von Flugzeugen, aber auch das Führen internationaler Beziehungen, die Steuerung nationaler Finanzmärkte, die Entwicklung einer befriedigenden Theorie der Ökonomie, die Planung von Städten sowie die Erforschung und das Schreiben von Geschichte können, so Mooers, künftig von Computern erledigt werden (vgl. Mooers 1946).

Retrospektiv muss Mooers' Prognose als utopisch erscheinen. Seine Zukunft ist nicht unsere Gegenwart geworden. Vieles kam anders als erwartet und es zeigte sich, dass die Steuerung von Volkswirtschaften oder das Schreiben von Geschichte nicht von Computern vollbracht werden kann.⁸ Auch das Fliegen von Flugzeugen bedarf noch immer menschlichen Zutuns, gleichwohl die meisten Flugzeuge heute ohne den Einsatz elaborierter Computersysteme ebenfalls am Boden bleiben würden. Das Gleiche gilt wohl auch für die Ökonomie. Ein Leben, Arbeiten und Forschen ohne Computer oder Internet scheint heute kaum mehr vorstellbar.⁹

Mooers' Spekulation über die Zukunft der Menschheit ist jedoch nicht deshalb interessant und instruktiv, weil seine Vorhersagen zutreffend gewesen wären. Im Gegenteil, in bequemer historischer Distanz vermögen uns seine Prognosen wahrscheinlich kaum mehr als ein staunendes und zugleich beschämtes Lächeln zu

7 | Um seine Zukunftsprognose zu plausibilisieren, rekuriert Mooers auf das bereits technisch Mögliche. Hierin kommt ein typisches Muster technischer Utopien zum Vorschein. Das technisch Bekannte bildet den Hintergrund, vor dem sich die Visionen technischen Fortschritts abzeichnen. Deutlich tritt dies beispielsweise auch in Bill Gates »Information at Your Fingertips«-Vortrag aus dem Jahr 2005 zu Tage, in dem er unter anderem die flächendeckende Durchsetzung von Videotelefonie prognostiziert, welche als Erweiterung der gebräuchlichen Telefone, wie z.B. von Telefonzellen, präsentiert wird (vgl. Gates 1994: 30'40"). Doch anders als von Gates vorhergesagt, wurden sämtliche Telefonzellen nicht schlicht um diese Funktion erweitert. Im Gegenteil, mit dem Aufkommen von Mobiltelefonen wurden Telefonzellen tendenziell obsolet.

8 | Im Rahmen des 1971 initiierten Cybersyn-Projekts beauftragte die chilenische Regierung Stafford Beer, ein Computersystem zu entwickeln, welches in der Lage ist, die Wirtschaft Chiles zu lenken. Das Projekt fand 1973 durch den Militärputsch Pinochets ein jähes Ende (vgl. Pircher 2004: 90).

9 | Die Durchdringung aller Lebensbereiche mit digitalen Kommunikationstechnologien zeigt sich vielleicht am deutlichsten in den publizierten Erfahrungsberichten verschiedener Autoren und Publizisten, die im Selbstexperiment versucht haben, auf das Internet, Handy usw. zu verzichten (Röhlig 2008; Rühle 2010; Koch 2010). In deren Selbstversuchen tritt die Bedeutung dieser Technologien paradigmatisch zum Vorschein.

entlocken. Daher ist Mooers' Spekulation in erster Linie als Bild einer möglichen Zukunft ernst zu nehmen, in dem das zum Ausdruck kommt, was in den 1940er Jahren als realisierbar erachtet wurde. Zukunftsvisionen wie die von Mooers sind nicht an einer künftigen Gegenwart zu messen, sondern sind Spiegel der Zeit, in der sie entstanden sind.¹⁰

Was Mooers' Vision zu erkennen gibt ist – in einem gänzlich unemphatischen Sinn – der Zeitgeist, dem zufolge es damals möglich erschien, dass Computer in naher Zukunft in der Lage sein würden, unterschiedlichste Aufgaben zu erfüllen, die vom Kohleabbau bis zur Geschichtsschreibung reichen. Auf die Frage, worin dieser Glaube an die Computertechnik gründet, findet sich in Mooers' Text eine bemerkenswert konzise Antwort: Computer sind dazu prädestiniert, derart heterogene Funktionen zu erfüllen, weil es sich bei diesen gleichermaßen um Probleme der Informationsverarbeitung handelt:

»These are all tasks of assembling immensely complicated masses of varied information, bringing a system out of the disorder, and then constructing an engineering solution, i.e. the best solution in view of the facts and of the requirements.«
(Mooers 1946)

Die Überzeugung, dass Computer der Sammlung, Speicherung und Verarbeitung von Informationen dienen, gründet auf einem generalisierten, begrifflich abstrakten und ontologisch reifizierten Informationskonzept, mit dessen Herausbildung heute vor allem die Namen Claude E. Shannon und Norbert Wiener in Verbindung gebracht werden (vgl. Hayles 1999: 50ff.).

Bereits Ende der 1930er Jahre setzte sich Shannon mit diversen Problemen der technischen Nachrichtenübertragung auseinander, was 1948 schließlich in der Publikation der *Mathematical Theory of Communication* mündete.¹¹ Zwar standen

10 | Dies bringt Geoffrey Nunberg zum Ausdruck, wenn er schreibt: »Nothing betrays the spirit of an age so precisely as the way it represents the future« (Nunberg 1996: 103). Die Wahl des Worts *betray* ist bemerkenswert, da es doppeldeutig ist und sowohl den Verrat von etwas (wie z.B. Geheimnisverrat) als auch den Verrat an etwas (wie z.B. Betrug) meinen kann. Hieraus ergeben sich zwei unterschiedliche Lesarten. Spiegeln Zukunftsdarstellungen in zeitlich diachroner Perspektive wider, was zu einer Zeit als möglich erachtet worden ist, erscheinen sie in einer synchronen Betrachtungsweise als Symptome eines (Selbst-)Betrugs, der auf seine Ursachen, seine Ziele und seine Politik hin zu befragen ist.

11 | Wie Axel Roch in der 2010 erschienenen Monographie über das Leben und Werk Shannons zeigt, hat Shannon im Februar 1939 in einem Brief an seinen Mentor Vannevar Bush sein Interesse bekundet, sich mit den Grundlagen der Kommunikationstheorie auseinanderzusetzen (vgl. Roch 2010: 37). Zur Genese von Shannons Informationstheorie und ihren medientheoretischen Implikationen siehe auch Schüttelpelz 2002a, 2003.

während des Zweiten Weltkriegs die militärischen Probleme der Geheimkommunikation sowie der Fernsteuerung von Flugabwehrsystemen im Vordergrund von Shannons Arbeiten, nach 1945 übersetzte er seine Forschungsergebnisse jedoch zunehmend in den zivilen Kontext.¹² Die Frage der möglichst effizienten Geheimhaltung von Informationen im Zuge ihrer technischen Übertragung wich dem Problem der möglichst effizienten technischen Übertragung von Informationen in einem gegebenen Kommunikationskanal (vgl. Hagemeyer 1979: 430ff.). Im selben Jahr wie Shannons mathematische Kommunikationstheorie erschien auch Norbert Wieners *Kybernetik*, ein weiterer grundlegender und wirkmächtiger Text, in dem ein generalisiertes Informationskonzept propagiert wird. »Information ist Information«, schreibt Wiener darin, »weder Materie noch Energie« (Wiener 1968 [1948]: 166).

Dass der Student Mooers seine Spekulation über die Zukunft der Menschheit bereits zwei Jahre vor dem Erscheinen dieser späteren Gründungstexte der Informationstheorie verfasst hat, ist ein Indiz dafür, dass ein allgemeines und abstraktes Informationskonzept bereits verbreitet war.¹³ »Information lag«, um eine Formulierung von Thomas Haigh aufzugreifen, »in der Luft« (Haigh 2007: 60), als Mooers ans MIT kam, wo er eigenen Angaben zufolge einen Vortrag Shannons besuchte und dessen Informationstheorie kennengelernte (vgl. Corbitt 1993: 9).¹⁴

12 | Diese sowohl militärische als auch zivile Anschlussfähigkeit von Fortschritten auf dem Gebiet der Nachrichtentechnik und der Nachrichtentheorie ist, so Roch, charakteristisch für die Forschungen in den Bell Labs, an denen Shannon seit 1941 tätig war: »Die Bell Labs verwandelten seit 1940 Prinzipien der Kommunikation in Feuerleitung, seit 1945 hingegen Prinzipien der Feuerleitung in Kommunikation. Das Bell System forschte nicht binär oder digital, sondern dual, militärisch und zugleich – wenigstens potentiell – auch zivil« (Roch 2010: 154). Möglich war diese doppelte Ausrichtung der Forschungen, weil das militärische Streben nach Geheimkommunikation und die zivile Kommunikationsforschung sich demselben Problem zuwandten, gleichwohl unter verschiedenen Vorzeichen: »Aus Sicht des Empfängers sind Kryptographie und Kommunikation fast identisch: Aus einem mit Rauschen gemischtem Signal wird eine gesendete Botschaft ermittelt. [...] Aus der Sicht des Senders sind beide Verfahren verschieden: Kryptographie führt Rauschen als Verschlüsselung bereits im Sender ein, im Fall von Kommunikation stört Rauschen erst im Kanal« (Roch 2010: 105).

13 | Durch seine Tätigkeit am *Naval Ordnance Laboratory* zwischen 1941 und 1946, wo er an dem von John Vincent Atanasoff geleiteten, aber später unvollendet eingestellten Computerprojekt mitwirkte, war Mooers mit den aktuellen Entwicklungen im Bereich digitaler Computertechnologien bereits vor Aufnahme seines Masterstudiums vertraut (vgl. Mooers 2001).

14 | Aus Mooers' Nachlass lässt sich nicht genau rekonstruieren, ob seine Spekulation über die Zukunft der Menschheit direkt von Shannons Vortrag inspiriert wurde. Auch ein Interview mit Calvin und Charlotte Mooers aus dem Jahr 1993

Sowohl die Kybernetik als auch die nachrichtentechnische Kommunikationstheorie rücken Information ins Zentrum ihrer Betrachtungen, wobei diese als mess- und quantifizierbare Eigenschaft von Nachrichten begriffen wird.¹⁵ Hierin besteht das Neue der kybernetischen und nachrichtentechnischen Hinwendung zu und der Behandlung von Information. Denn der Informationsbegriff ist nicht erst in den 1940er Jahren gebräuchlich geworden, sondern findet sich spätestens seit dem 15. Jahrhundert in der englischen Sprache, wie Nunberg dargelegt hat (1996: 109). Der Gebrauch des Begriffs war jedoch an das Verb *informieren* gekoppelt und bezog sich stets auf eine Person, die durch etwas informiert wird. An die Stelle der von Nunberg als partikulare Bedeutungsvariante bezeichneten Gebrauchsform tritt in den 1940er Jahren zunehmend eine abstrakte Bedeutung, der zufolge es sich bei Information um eine Art von »intentional substance« (Nunberg 1996: 110) handelt, die in der Welt existiert. Auf Grundlage eines solchen abstrakten Verständnisses von Information wird nachvollziehbar, warum Mooers der Überzeugung war, dass so unterschiedliche Tätigkeiten wie das Fliegen von Flugzeugen, das Schreiben von Geschichtsbüchern oder die Formulierung von ökonomischen Theorien im Wesentlichen Informationsverarbeitungsprozesse sind.

gibt hierüber keinen Aufschluss (vgl. Corbitt 1993). Obwohl dies aus historischer Perspektive durchaus von Interesse wäre, ist die Frage für die hier entwickelte Argumentation unerheblich.

15 | Bereits in den 1920er und 30er Jahren haben Harry Nyquist und Ralph Hartley, die beide ebenso wie Shannon an den Bell Labs tätig waren, an Modellen und Methoden gearbeitet, die Übertragungskapazität eines Kanals zu bestimmen und Information zu quantifizieren. Hagemeyer vergleicht die Ansätze von Nyquist, Hartley und Shannon anhand der Position, die die drei Wissenschaftler in den Bell Labs innehatten: »Will man die wichtigsten Stadien einer Nachrichtentheorie, die in der Forschung des Bell Systems entstanden etwas einseitig durch die institutionelle Position ihrer Autoren, Nyquist (1924), Hartley (1928) und Shannon (1948) charakterisieren, so kann man mit einem Recht von der Nyquistischen Telegrafentheorie als einer ›Line‹ Theorie, dem Hartleyschen Informationskonzept als einer ›Staff‹ Theorie und der Shannonschen Informationstheorie als einer ›consulting‹ Theorie sprechen; die erste die methodisch klare Theorie eines eng begrenzten Sachgebietes und Problembereiches (Vergleich der Effizienz verschiedener Telegrafiersysteme), entstanden auf der untersten Ebene der Forschungsstruktur in einer klassischen Theorieabteilung; die zweite die methodisch vage Theorie eines umfassenden Gebietes (Vergleich der Effizienz aller Übertragungstechniken), entstanden in der dieses Gebiet leitenden Ebene des technischen Managements; die dritte die methodisch klare und strenge Theorie eines umfassenden Gebietes (Vergleich verschiedener Nachrichtenquellen und gestörter Übertragungssysteme allgemein), entstanden in einer Consulting-Abteilung (für die besondere Operationsbedingungen galten) mit eigener Forschungskompetenz« (Hagemeyer 1979: 94).

Paradigmatisch tritt die von Nunberg diagnostizierte Abstraktionsbewegung in Shannons und Wieners Informationstheorien zutage, in denen von dem Sinnzusammenhang abgesehen wird, in dem eine Information auftritt, d.h. bei der Quantifizierung von Information hat die Bedeutung von Information keine Bedeutung. Indem Shannon und Wiener in ihren jeweiligen Ansätzen systematisch von den in einer Nachricht enthaltenen Inhalten abstrahierten, war es ihnen möglich, Information als eine mess- und berechenbare Größe zu behandeln, deren Wert sich losgelöst von der Kommunikationssituation und den an der Kommunikation beteiligten Personen bestimmen lässt. Gerade hierin bestand, so N. Katherine Hayles, der Clou dieser Ansätze (vgl. 1999: 53). Macht es aus Sicht von Kommunikanten einen Unterschied, in welchem Kontext eine Nachricht wie zum Beispiel »Die Sonne scheint« geäußert wird, so wird dieser Nachricht in der Nachrichtentechnik und der Kybernetik unterschiedslos ein Wert zugewiesen, d.h. man sieht von den möglichen Unterschieden ab und fokussiert das Gemeinsame der Nachricht.

Die in Shannons nachrichtentheoretischem und Wieners kybernetischem Ansatz vollzogene Ablösung der Information von ihrem jeweiligen Kontext hat sich in der Folgezeit als ebenso produktiv wie problematisch erwiesen. Einerseits gewann Information die Konturen einer eigenständigen Entität, deren Eigenschaften empirisch messbar und mathematisch beschreib- und berechenbar sind, andererseits aber verlor man den konkreten Bezugsrahmen aus dem Blick, vor dessen Hintergrund sich die Informationstheorien entwickelt haben:

»Taken out of context, the definition allowed information to be conceptualized as if it were an entity that can flow unchanged between different material substrates, as when Moravec envisions the information contained in a brain being downloaded on a computer. Ironically, this reification of information is enacted through the same kind of decontextualization moves that the theory uses to define information as such. The theory decontextualizes information; Moavc decontextualizes the theory.« (Hayles 1999: 54)

Laut Hayles sind die Informationstheorien von Shannon und Wiener aus ihrem theoretischen und praktischen Kontexten gerissen worden. Symptomatisch für diese Dekontextualisierung ist die zunehmende Generalisierung des Informationsbegriffs, vor der Shannon in den 1950er Jahren noch gewarnt hat, als er zu bedenken gab, dass ein Informationsbegriff allein wohl kaum geeignet sei, um den vielfältigen Aspekten und Anwendungsgebieten des Forschungsfelds der Informationstheorie gerecht zu werden.¹⁶ Obwohl die von Shannon gleichermaßen konstatierte wie ein-

16 | In *The lattice theory of information* stellt Shannon heraus: »The word 'information' has been given many different meanings by various writers in the general field of information theory. It is likely that at least a number of these will prove sufficiently useful in certain applications to deserve further study and permanent recognition. It is hardly to be expected that a single concept of information would satisfactorily

geforderte Pluralität verschiedener Informationsbegriffe in der technischen Praxis und der wissenschaftlichen Forschung durchaus fortbesteht, wurde sie von einer diskursiv imaginierten Einheit von Information und des Informationsbegriffs überlagert, welche im Zentrum der von Geoffrey Bowker als *information mythology* bezeichneten Überzeugung steht, dass das Universum grundlegend aus Information besteht:

»In the story that we are looking at, ›information‹ can travel anywhere and be made up of anything. Sequences in a gene, energy levels in an atom, zeros and ones in a machine and signals from a satellite are all ›information‹ and are thus subject to the same laws. If everything is information, then a general statement about the nature of information is a general statement about the nature of the universe.« (Bowker 1994: 233).

Bowker, der sich für die narrativen und praktischen Bedingungen interessiert, unter denen eine solche Annahme wahr wird oder zumindest in einer bestimmten Periode als wahr erscheint, findet bereits in den Schriften von Charles Babbage Grundzüge der Informationsmythologie vorgezeichnet.¹⁷ Kennzeichnend ist für ihn dabei nicht der abstrakte Informationsbegriff, sondern das Verschwinden der infrastrukturellen Bedingungen bzw. Voraussetzungen hinter einer abstrakten Idee (von Wissen, Information, Determiniertheit oder Berechenbarkeit). Diese erscheint fortan als natürliches Prinzip oder universelle Gesetzmäßigkeit, welche das Universum, Gesellschaften, Menschen, Ökonomien etc. bestimmt.¹⁸ Auch wenn in Bowkers Studie dem Aufkommen des abstrakten Informationsbegriffs seit den 1940er

account for the numerous possible applications of this general field« (Shannon 1953: 105).

17 | Objektive Information existiert Bowker zufolge nicht an sich, sondern ist Effekt einer Objektivierung, für deren Bedingungen er sich interessiert: »Information mythology [...] describes an integral part of the economic process of ordering social and natural space and time so that ›objective‹ information can circulate freely. The global statement that everything is information is not a preordained fact about the world, it becomes a fact as and when we make it so« (Bowker 1994: 245).

18 | Bowker macht es sich zum Ziel, die in den Hintergrund gerückten infrastrukturellen Bedingungen in den Vordergrund zu rücken. Dieses Vorgehen bezeichnet er als infrastrukturelle Inversion. In *Sorting Things Out* beschreibt Bowker gemeinsam mit Susan Leigh Star diese Inversion als »struggle against the tendency of infrastructure to disappear (except when breaking down). It means learning to look closely at technologies and arrangements that, by design and by habit, tend to fade into the woodwork [...]. Infrastructural inversion means recognizing the depths of interdependence of technical networks and standards, on the one hand, and the real work of politics and knowledge production on the other. It foregrounds these normally invisible Lilliputian threads« (Bowker/Star 2000: 34).

Jahren kein zentraler Stellenwert beigemessen wird, zeigt sich in der seither zu beobachtenden Ablösung des Begriffs von seiner Theorie die für die Informationsmythologie charakteristische Struktur. Der Begriff beginnt, das konkrete Kommunikationsproblem zu überdecken, für dessen Behandlung er eingeführt wurde und Gültigkeit beansprucht.¹⁹

Vor dem Hintergrund dieser sicherlich nicht erschöpfenden Verortung des Diskurses über Information wird deutlich, was mit der Eingangs formulierten Charakterisierung von Information als generalisiert, abstrakt und reifiziert gemeint ist. Information ist generalisiert, weil alles Information ist (oder zumindest als solche betrachtet werden kann). Information ist abstrakt, da der Begriff erstens von der Bindung an das Geschehen des Informierens entkoppelt wird und sich zweitens als abstrakter Begriff von den Gebrauchskontexten löst, in denen er entstanden ist. Schließlich ist Information reifiziert, da Informationen als eigenständige Entitäten begriffen werden, die manipulierbar sind. Alle drei Momente sind nicht voneinander abzulösen, sie greifen vielmehr ineinander und bilden den Horizont, vor dem Information als Leitbegriff zur Beschreibung des Status quo dient, was sich in Komposita wie Informationsgesellschaft und Informationszeitalter zeigt.

Die konkreten Praktiken des Umgangs mit Informationen, die sich in einem Netzwerk von menschlichen und technischen Akteuren ereignen, stehen zu dem

19 | Es besteht kein Zweifel daran, dass sich Shannons Ansatz auch in anderen Kontexten als fruchtbar erwiesen hat. Falsch aber wäre es, hieraus die allgemeine Gültigkeit dieses oder irgendeines anderen Ansatzes abzuleiten. Zunächst ging es ihm um die Lösung eines Kommunikationsproblems, das er in seiner mathematischen Kommunikationstheorie auf denkbar klare Weise auf folgende nachrichtentheoretische Definition von Kommunikation gebracht hat: »Das grundlegende Problem der Kommunikation besteht darin, an einer Stelle entweder genau oder angennähert eine Nachricht wiederzugeben, die an einer anderen Stelle ausgewählt wurde« (Shannon 1976 [1948]: 41). Kommunikation stellt den Nachrichtentechniker demzufolge vor ein Replikationsproblem: Wie lässt sich die Replikation einer an Ort A geäußerten Nachricht an Ort B möglichst effizient gewährleisten? Die von Shannon als Antwort auf diese Frage entwickelte stochastische Beschreibung von Information vermag Auskunft darüber zu geben, wie viel Informationen in einem bestimmten Kanal übertragen werden können (vgl. Shannon 1976 [1948]: 46ff.). Zugleich aber stellt sie, wie Shannon einige Jahre später schreibt, kein Kriterium zur Verfügung, ob es sich bei zwei Nachrichten um dieselbe aktuelle Information handelt, da »two entirely different sources might produce information at the same rate (same H) but certainly they are not producing the same information« (Shannon 1953: 105). Um nicht nur die Übertragungskapazität potenzieller Information in einem Kanal zu berechnen, sondern die Eigenschaften aktualisierter Informationen mathematisch zu beschreiben, ist das ursprüngliche Konzept Shannons also nicht hinreichend.

abstrakten Informationsbegriff in einem gewissen Spannungsverhältnis.²⁰ Denn obwohl sich der Informationsbegriff diskursiv zunehmend verselbstständigte, indem seine Bedeutung von den Gebrauchskontexten abstrahiert wurde, situiert sich der Diskurs über Information im Kontext einer Vielzahl informationeller Praktiken, konkreter Problemzusammenhänge und technischer Informationssysteme, wie z.B. der Geheimhaltung von Kommunikation, der Optimierung der Kanalkapazität, der Raketenlenkung, der bibliothekarischen Versammlung und Verwaltung von Dokumenten, dem Management von Unternehmen, dem World Wide Web, Internetsuchmaschinen, medizinischen und Gendatenbanken etc.

In Informationssystemen realisiert sich Information nie als eine losgelöste, abstrakte Entität. Im Gegenteil, Informationssysteme bedingen, was als Information zur Erscheinung kommt, als solche adressiert, gesucht, gefunden und verarbeitet werden kann. In Anbetracht dessen stellt die abstrakte Vorstellung von Information ein uneingelöstes Versprechen oder eine Wunschkonstellation dar. Nur in konkreten informationellen Praktiken wird von Informationen Gebrauch gemacht, werden sie erzeugt, gespeichert und verarbeitet. Doch auch wenn das Versprechen abstrakter, reifizierter Information uneingelöst bleibt, besteht es als Imaginäres diskursiv fort, was die Entwicklung immer neuer Medientechnologien anspornt und zu neuartigen Problemlösungsansätzen, aber auch zu neuen Problemen führt. Computer sind, wie Hartmut Winkler in seiner Studie *Docuverse* eingewandt hat, nicht nur als konkrete Medientechnologien, sondern auch als Wunschmaschinen zu betrachten, an denen sich Wunschkonstellationen formieren und durch die sich Wünsche und Utopien perpetuieren (vgl. Winkler 1997a: 11ff.).

Deutlich wird das Wechselspiel zwischen abstraktem Informationsbegriff und partikularen Gebrauchskontexten, wenn man die späteren Arbeiten von Mooers betrachtet. Seine Spekulation über die künftigen Einsatzgebiete von Computern steht am Anfang seiner Karriere, die der Entwicklung von Informationstechnologien gewidmet sein sollte. Der Kontext, in dem sich die Arbeit von Mooers situiert, ist die Versammlung, Speicherung und Abfrage von bibliographischen Informationen. Wie Mooers Anfang der 1990er Jahre rekapitulierte, inspirierte ihn Shannon zu seiner Arbeit im Bereich des Information Retrieval (Corbitt 1993: 9), einem Forschungsgebiet, dem Mooers im März 1950 auf einer Tagung der Association for Computing Machinery (ACM) seinen Namen gab, und welches er im Herbst desselben Jahres auf dem *International Congress of Mathematicians* folgendermaßen definiert hat: »The problem of directing a user to stored information, some of

20 | Mit dieser Formulierung wird an die Akteur-Netzwerk-Theorie (ANT) Bruno Latours angeschlossen, der in seinen wissenschaftssoziologischen Studien immer wieder auf die Verflechtung von menschlichen und nicht-menschlichen Akteuren hingewiesen hat (vgl. Latour 2002: 111ff.). An anderer Stelle beschreibt Latour die ANT als Wissenschaft, welche eine Welt demonstriere, »die aus Verkettungen von Mittlern besteht, wo von jedem Punkte gesagt werden kann, daß er agiert« (Latour 2007b: 103f.).

which may be unknown to him, is the problem of ›information retrieval‹« (Mooers 1950a: 572).²¹ Obwohl es Mooers in dieser Definition nicht explizit macht, richtet sich das Information Retrieval nicht auf das Suchen und Finden von Informationen im Allgemeinen, sondern behandelt die Frage, wie mithilfe von Technologien das Auffinden von Informationen ermöglicht werden kann. Die Herausforderungen der bibliothekarischen Praxis bei der Ordnung und Katalogisierung von Dokumenten werden beim Information Retrieval in das ingenieurtechnische Problem des maschinellen Findens von Informationen übersetzt, wie Mooers einige Jahre später herausstellt: »When we speak of information retrieval [...], we are really thinking about the use of machines in information retrieval« (Mooers 1960: 229).²²

21 | In einem späteren Rückblick auf die Geschichte des Information Retrieval rekapituliert Mooers, dass er den Begriff des Information Retrieval erstmals in dem Artikel »The Theory of Digital Handling of Non-Numerical Information and its Implications to Machine Economics« gebrauchte, den er auf einer Konferenz der Association for Computing Machinery, die am 29. März 1950 an der Rutgers University stattfand, präsentierte (vgl. Mooers 1960: 229, Fn 1). Das Manuskript ist als *Zator Technical Bulletin*, Nummer 48, von Mooers im Anschluss selbst publiziert worden (Mooers 1950b).

22 | Greift Mooers' ursprünglicher Definitions vorschlag in dieser Hinsicht vielleicht nur zu kurz, erweist er sich aus einem medien- und techniktheoretischen Blickwinkel als problematisch, denn Information Retrieval wird als ein dezidiertes Problemlösungshandeln beschrieben. Zwar trifft zu, dass damals wie heute die Fülle der potenziell verfügbaren Informationen ein Problem darstellt, doch wurde und wird Information nicht nur als Last empfunden. Die Entwicklung neuer Techniken zur Verwaltung von Informationssammlungen wurde häufig auch durch eine Informationslust motiviert. Bereits Holleriths Erfindung des Lochkartenautomaten zur Speicherung und statistischen Auswertung von Daten (vgl. Hollerith 1889) und dessen Einsatz bei der US-amerikanischen Volkszählung 1890 zeugen von einem gestiegenen Verlangen nach Information: »The invention of the first punched-card system in the United States in the 1880s grew out of a public demand for more detailed census statistics« (Heide 2009: 253). Die von Heide als öffentliche Forderung nach statistischen Daten zur Bevölkerungszusammensetzung beschriebene Informationslust wird begleitet von einem Optimierungs- und Rationalisierungsstreben, welches in den Vereinigten Staaten von Amerika spätestens 1890 einzog und kurz vor Beginn des Ersten Weltkrieges auch Deutschland erreichte (vgl. Krajewski 2002: 167). Für die beschriebene Informationslust lassen sich auch eine Reihe von Beispielen aus der jüngeren Vergangenheit anführen, wie z.B. das *Human Genome Project*, das *Google Library Book Search*-Projekt sowie der riesige Datenmengen produzierende *Large Hadron Collider* am CERN in Genf. Insofern wäre es zu kurz gegriffen, die Entwicklung von Anwendungen zur computergestützten Informationsverarbeitung nur als Reaktion auf den drohenden Information Overload zu verstehen, auch wenn dieser diskursiv oft als Anlass angeführt wird,

Die Apparate, an die Mooers denkt, sind nicht gleichzusetzen mit Computern. Denn auch wenn er deren Einsatz zum Zweck des Findens und Abrufens von Informationen in Erwägung zieht, erscheinen Mooers die Ende der 1950er Jahre verfügbaren Computer noch nicht wirklich dazu geeignet, die Aufgaben des Information Retrieval effizient zu erfüllen: »As computing machines are now designed, they are not matched for the job of information retrieval [...] and the situation of using a computing machine for such purposes is comparable to using a bulldozer to crack peanuts« (Mooers 1960: 230). Gibt man Mooers' Bedenken über die Leistungsfähigkeit damaliger Computer eine medientheoretische Wendung, dann wird deutlich, dass der Computer einerseits als eine Technologie begriffen werden kann, die Apparaturen mit dem reifizierten Informationsbegriff verschränkt, er aber andererseits als Informationstechnologie selbst unterschiedliche Ausgestaltungen erfahren kann und für die Anforderungen des Information Retrieval erst angepasst werden muss. Wer Medien, wie zum Beispiel Wolfgang Ernst, vorrangig auf die durch sie realisierte Verbindung von »physikalischen Praktiken und logischen Operationen« (Ernst 2008b: 162) hin befragt und in Computern demzufolge eine »Hochzeit aus Physik und Logik« (Ernst 2008b: 173) vollzogen sieht, vermag das von Mooers konstatierte Ungenügen damaliger Computer nicht richtig zu bewerten: Computertechnologien und die mit ihnen realisierten Anwendungen sind das Resultat von Gestaltung und Design, welches sich weder auf die Gesetzmäßigkeiten der mathematischen Logik reduzieren lässt, noch auf einen Informationsbegriff, wie beispielsweise das stochastische Informationsmaß Shannons. Vielmehr bilden Computer einen gestaltbaren und medienhistorisch wandelbaren Rahmen, in dem unterschiedliche Formen von Information verarbeitet werden können. Gerade dann, wenn es um die medientheoretische Analyse von Datenbanken geht, gilt es dies anzuerkennen. Was es heißt, Informationen technisch zu sammeln, sie zu ordnen, zugänglich zu machen und zu verarbeiten, zeigt sich nur im Kontext partikularer Anwendungen. Zwischen dem abstrakten, aber relativ unbestimmten

wie z.B. in Bushs Entwurf der Memex (vgl. Bush 1945: 102). Neben Information Overload und Information Underload darf auch das »Luxus- und Spielbedürfnis« (Blumenberg 2009: 75) von Ingenieuren und Erfindern als wichtiger Motor technischer Entwicklungen nicht vernachlässigt werden. Ein solcher Spieltrieb wird beispielsweise dann wirksam, wenn es darum geht, neue Anwendungsmöglichkeiten für Medientechnologien zu ersinnen. Der Entwurf der Memex kann, wie Fairthorne vorgeschlagen hat, auch in diese Richtung gelesen werden: »Quite properly he [gemeint ist Vannevar Bush, M.B.] was concerned to get jobs for the machine, not machines for the job. Fundamental problems were untouched. The Memex conception, even if it had grown to engineering perfection, would have been useful only to an individual who could apply his own criteria of relevance to cumulatively stored micro-copy of world literature, and who, having read and digested all of this already, had marked it appropriately for retrieval« (Fairthorne 1961a: 132).

Informationsbegriff und konkreten Informationssystemen klafft eine theoretische Leerstelle, die praktisch immer wieder gefüllt werden muss.²³

Ungeachtet Mooers' berechtigter Bedenken gegenüber der Eignung der damals verfügbaren Computer für das Information Retrieval, traute man ihnen nicht nur in wissenschaftlichen und technischen Kontexten bereits einiges zu. In Büchern, Magazin- und Zeitungsartikeln sowie in Fernsehsendungen wurden die Möglichkeiten von Computern verhandelt und Überlegungen über die Zukunft des Menschen in computerisierten Gesellschaften angestellt. So wurde die Bevölkerung der USA seit den ausgehenden 1940er Jahren mittels populärer Massenmedien immer wieder mit den aktuellen Entwicklungen im Bereich der Computertechnologie vertraut gemacht und über die erhofften Leistungen künftiger Computersysteme informiert (Vgl. Malone 2002: 16f.).²⁴ Auch Hollywood nahm sich rasch des Themas an. Im Jahr 1957 erschien der Film *Desk Set* (Regie: Walter Lang), der sich als einer der ersten, wenn nicht sogar als *der* erste (Hollywood-)Film mit den Hoffnungen und Ängsten auseinandersetzte, die man dem »Electronic Brain« entgegenbrachte.²⁵

23 | Die Unterbestimmtheit des abstrakten Informationsbegriffs findet ihren Widerhall in dem unterminologischen Gebrauch des Begriffs in der Alltagssprache, den Konitzer konstatiert: »Wir gebrauchen den Ausdruck ‚Information‘ in der Alltagssprache gewöhnlich unterminologisch. Das gilt auch dort, wo er Spuren seiner Herkunft aus der mathematischen Informationstheorie an sich trägt. Wir bezeichnen durch ihn entweder einen Vorgang – die Tatsche, daß jemand etwas Neues erfährt, daß er eine Einsicht, ein Wissen gewinnt, daß ihm etwas mitgeteilt wird – oder dasjenige, was ihm mitgeteilt wird: die Information als den Inhalt der Mitteilung, als das Mitgeteilte. Würde man uns weiter darum bitten, den Vorgang der Benachrichtigung genauer zu beschreiben, so würden wir davon sprechen, daß etwas jemanden (einen Empfänger) erreicht oder von ihm verstanden wird. Und würde man uns danach fragen, was das ist, die Nachricht, so würden wir vermutlich zwei Kandidaten nennen, die aber nur gemeinsam auftreten können: den materiellen Träger der Nachricht, und das, was dieser irgendwie vermittelt, nämlich ein Wissen« (Konitzer 2006: 326).

24 | Diesbezüglich stellt Malone fest: »[T]he coverage of developments in the nascent computer industry in popular magazines and on television reached mass audiences that no researcher could. Over the course of the 1950s, newspapers and periodicals slowly increased their coverage of new technology developments and tracked installations of giant brains in various government and commercial settings« (Malone 2002: 16). Dominant war die Vorstellung vom Computer als elektronischem Gehirn, welche sich nicht zuletzt aus der konzeptuellen Verbindung von Computertechnologie und Kybernetik ergab. Doch auch Alan Turings Überlegungen zu maschineller Intelligenz und der von ihm vorgeschlagene Turing Test trugen zur Parallelisierung von Geist und Computer bei (vgl. Turing 1987 [1950]).

25 | Die vom Charles Babbage Institute herausgegebene Filmliste »Hollywood & Computers« nennt *Desk Set* als den ersten Film, in dem sich Hollywood Computern

Bemerkenswert ist jedoch nicht nur, dass der Film die möglichen Auswirkungen der Computerisierung auf die US-amerikanische Gesellschaft verhandelt, sondern der Gebrauchskontext, in den Computer gestellt werden.²⁶ Die Geschichte von *Desk Set* entspinnst sich um die Einführung eines Computersystems namens EMMERAC (Electromagnetic Memory and Arithmetical Research Calculator) im Reference & Research Department einer Fernsehanstalt. Hier werden Computer weniger als Rechen-, sondern vielmehr als Informationsmaschinen dargestellt, die das Suchen und Finden gespeicherter Informationen ermöglichen und durch deren Einsatz sich ein Übergang vom bibliothekarischen Wissen zum computergestützten Suchen vollzieht.

Miss Bunny Watson (Katherine Hepburn), die Leiterin des Reference & Research Departments, und ihre drei Mitarbeiterinnen Peg Costello (Joan Costello), Sylvia Blair (Dina Merrill) und Ruthie Saylor (Sue Randall) zeigen sich über die drohende Installation von EMMERAC in ihrer Abteilung enorm besorgt, da sie befürchten, von dem Computer ersetzt zu werden. Betreut wird die Einführung von dem geheimnisvollen Richard Sumner (Spencer Tracy), dem Erfinder von EMMERAC höchstpersönlich, der sich selbst als »methods engineer« bezeichnet und freilich von den Vorzügen des von ihm erbauten Computers überzeugt ist.

Im Verlauf des gesamten Films wird der Kampf Mensch gegen Maschine²⁷ inszeniert, wobei Miss Watson mit ihrem Wissen, ihrem Erinnerungs- und ihrem Assoziationsvermögen als Gegnerin von EMMERAC antritt. Im Zentrum steht die Angst, dass Computer, wenn nicht den Menschen insgesamt, so doch zumindest dessen Arbeits- und Leistungsvermögen überflüssig machen werden. Dies wird in einem Dialog zwischen Sumner und Watson deutlich, der sich bei einem gemeinsamen Mittagessen der beiden abspielt:

»Sumner: Did you ever see one of these electronic brains work? EMMARAC, for Example?

thematisch zuwandte (Charles Babbage Institute 2003). Obwohl der Film nicht als Klassiker in die Filmgeschichte eingegangen ist, genießt er Cheryl Knot Malone zufolge in Kreisen von Bibliotheks- und Informationswissenschaftlern heute noch Kultstatus (vgl. Malone 2002: 14).

26 | *Desk Set* vermag es nach Ansicht von Malone, dem heutigen Betrachter zu zeigen, wie der Computer in den 1950er Jahren konzeptualisiert wurde und wie Menschen auf die von ihnen befürchteten Folgen der Computerisierung reagierten (vgl. Malone 2002: 15): »Although it is fiction, in which the characters, computers, and corporate library are imaginary, it nevertheless can attest to the ways in which computers were perceived during the decade when they first became available for business (not just military and governmental) applications« (Malone 2002: 14).

27 | Genauer gesagt ist es sogar der Kampf »Frau gegen Maschine«. Auf die Rolle von Frauen in diesem Film soll an dieser Stelle jedoch nicht genauer eingegangen werden.

- Watson: Yeah, yeah. Just this morning, as a matter of fact, I saw a demonstration at IBM.
- Sumner: Oh. Did you see it translate Russian into Chinese?
- Watson: Yeah. Saw it do everything. Frightening. Gave me the feeling that maybe, just maybe people were a little outmoded.
- Sumner: Mmm. Wouldn't surprise me a bit if they stopped making them.« (Lang 1957: 35:48)

Mit ihrer Vermutung, dass Menschen angesichts der neuen Computertechnologie antiquiert und überflüssig erscheinen könnten, bringt Watson die Befürchtung auf den Punkt, die in den 1950er Jahren nicht nur die Protagonistinnen des Films, sondern auch die US-amerikanische Bevölkerung umtrieb. Indem Sumner die Befürchtung Watsons nicht entkräf tet, sondern sie ironisch mit dem Kommentar übersteigert, dass er sich nicht wundern würde, wenn man folglich aufhörte, Menschen zu machen, wird eine Opposition zwischen dem Techniker Sumner und der Person Miss Watson aufgebaut, die erst ganz am Ende des Films aufgelöst wird, als sich die beiden ineinander verlieben. Ziel des Mittagessens ist für Sumner jedoch nicht, mit Miss Watson über die möglichen Konsequenzen der Einführung von Computern zu diskutieren, sondern die Überlegenheit von EMMERAC gegenüber der menschlichen Auffassungsgabe und dem menschlichen Gedächtnis vorzuführen. So stellt Sumner Miss Watson allerhand Fragen, die ihr Gedächtnis ebenso herausfordern sollen wie ihre Fähigkeit, Muster zu erkennen und logische Schlüsse zu ziehen. Anders jedoch als von Sumner intendiert, vermag Watson bei diesem menschlichen Benchmark-Test alle noch so kniffligen Fragen zu parieren, sodass sich Sumner schließlich von ihrer messerscharfen Auffassungsgabe sichtlich beeindruckt zeigt. Bereits hier deutet sich an, dass EMMERAC wohl keine ernsthafte Konkurrenz für Miss Watson darstellen wird. Es dauert jedoch bis zum Ende des Films, bis endgültig klar wird, dass Computer die Arbeit gut ausgebildeter Bibliothekare und Wissenschaftler nicht zu ersetzen, sondern nur zu unterstützen vermögen.²⁸ Es wird deutlich herausgestellt, dass es nicht hinreicht, einen Computer zu installieren und ihn durch eine Technikerin bedienen zu lassen. Vielmehr bedarf es weiterhin der Expertise der Bibliothekarin, da nur sie in der Lage ist, Anfragen an EMMERAC zu richten, die auch zum gewünschten Ergebnis führen. Damit verschiebt sich jedoch ihr Aufgabenprofil bzw. ihre Rolle. Diese Rollenverschiebung weist in Richtung einer neuen Wissensform, für die der Übergang vom Bücherwissen zu einem Wissen des elektronischen Gehirns charakteristisch ist, welches selbst aber nur ein virtuelles Potenzial bildet, das erst im Prozess des kompetenten Suchens aktualisiert wird. War bis zu diesem Zeitpunkt die Bibliothekarin diejenige, an welche Anfragen gerichtet wurden und welche sie

28 | Eben dies war zu dieser Zeit auch die Message von IBM. Dass die Firma die Herstellung des Films unterstützte, vermag in Anbetracht dessen auch nicht zu überraschen (vgl. Malone 2002: 15).

prozessierte, ist sie unter den Bedingungen des elektronischen Gehirns nun zur Fragenden geworden, die die im Computer gespeicherten Informationen richtig zu durchsuchen weiß.²⁹ Die allwissende Bibliothekarin wird damit zur kompetent Suchenden. Der Computer jedoch ist das vermeintliche Gehirn, welches wie von Geisterhand Antworten auf die an ihn gerichteten Anfragen ausgibt.

Damit zeigt der Film nicht nur ein versöhnliches Ende, sondern zeichnet auch das Bild einer damals noch fernen Zukunft. In *Desk Set* kommt demzufolge weniger die technische Realität zum Vorschein, als vielmehr das Imaginäre der Informationsverarbeitung. Denn existierende Computer waren damals noch weit davon entfernt, das leisten zu können, was der Film als deren Leistungsvermögen ausstellte. Und auch heute sind wir manchen der im Film dargestellten Visionen erst wenige Schritte näher gekommen. So formuliert Miss Watson natürlichsprachliche Anfragen an den Computer, die dieser zu verarbeiten und zu beantworten weiß, und sofern eine Frage nicht präzise genug gestellt war, fordert EMMERAC sogar die Präzisierung der Anfrage ein. Kurz vor Ende des Films, als bereits klar ist, dass EMMERAC nicht installiert wurde, um die Bibliothekarinnen der Forschungsabteilung zu ersetzen, wird Miss Watson von einem Anrufer vor die schwierige Frage gestellt, wie schwer die Erde sei. Langsam in EMMERAC keine Konkurrenz, sondern ein Werkzeug sehend, tippt sie die Frage »What is the total weight of the earth?« in das Schreibmaschineninterface des Computers ein, woraufhin EMMERAC die Anfrage piepend verarbeitet und nach kurzer Zeit über den Drucker etwas ausgibt, das Sumner zum Lachen bringt, sodass Miss Watson nachfragt:

»Watson: What's the matter?
Sumner: It's asking you a question.
Watson: What's the question?
Sumner: With or without people?« (Lang 1957: 94:41)

Diese spitzfindige Reaktion von EMMERAC versöhnt Miss Watson mit der Maschine.³⁰ Die ausgestellte Raffinesse von Computern war und ist noch immer weitgehend Wunschvorstellung. Entscheidend aber ist, dass Computer im Film nicht als leistungsfähige Rechenmaschinen vorgeführt werden, sondern als interaktive Informationsmaschinen, die der Speicherung großer Informationsmengen

29 | Dies spiegelt sich in der Etymologie des Wortes Computer wider. In den 1920er und 30er Jahren wurden Frauen in Forschungseinrichtungen Computer genannt, die repetitive Rechenaufgaben erfüllten. Mit der Einführung elektronischer Rechenmaschinen ging diese Bezeichnung von den rechnenden Frauen auf die rechnende Maschine über (vgl. Light 1999; Grier 2005; Skinner 2006). Dass der EMMERAC-Computer auch weiblich konnotiert wird, indem Sumner diesen liebevoll Emmy nennt, unterstreicht dies.

30 | Die Überwindung ihrer Vorbehalte bringt Miss Watson zum Ausdruck, indem sie EMMERAC als »Good girl. Good girl.« hätschelt (Lang 1957: 94:49).

sowie der selektiven Ausgabe relevanter Informationen dienen. Den Horizont dieser filmischen Imagination bildet der abstrakte Informationsbegriff, der sich jedoch nicht nur in die antizipierten Einsatzmöglichkeiten von Computern einschreibt, sondern auch in die Entwicklung von partikularen Informationssystemen. Er wird als technisches Imaginäres operativ und begleitet die Entwicklung digitaler Datenbanken, wie auch Haigh in Bezug auf die Computerisierung der institutionellen Informationsverarbeitung festgestellt hat:

»Die Auffassung, Information sei eine allgemeine Größe, die in Maschinen aufbewahrt und durch sie verarbeitet wird, dient in all diesen Fällen dazu, die Differenz zwischen einem sehr breit gefassten menschlichen oder betriebswirtschaftlichen Informationsbegriff und den viel begrenzteren Fähigkeiten spezifischer automatisierter Systeme zu verwischen.« (Haigh 2007: 58f.)

Bemerkenswert ist dies nicht zuletzt, weil die begrenzte Leistungsfähigkeit tatsächlicher Informationssysteme häufig verdeckt wird, indem man die in diesen Systemen verwalteten Formen von Information universalisiert. Dies lässt sich in den Debatten um die Einsatzgebiete von DBMS ebenso aufzeigen wie in den neueren Diskussionen um Data Warehousing, Data Mining, Knowledge Management, Web 2.0, Semantic Web und Linked Data (vgl. Haigh 2007: 58). Derartige Universalisierungstendenzen oder »Unifizierungsphantasien« (Winkler 1997a: 55) sind Teil der digitalen Medienkultur und als solche auf ihre Motive, Mechanismen und Funktionen hin zu befragen. Zugleich gilt es, die vielfältigen und unterschiedlichen Formen der computertechnischen Verarbeitung von Informationen freizulegen. Zu beschreiben sind partikulare Software-Hardware-Konfigurationen, die zwar nicht vorschreiben, welche Informationen in einem Informationssystem verarbeitet werden, aber was im Kontext eines solchen Systems als Information adressiert werden kann.

Mit der Feststellung, dass sich der abstrakte Informationsbegriff nie in konkreten Technologien realisiert, weil er zu generell und zu unbestimmt ist, soll dessen historische Bedeutung nicht revidiert werden. Sie erfährt nur eine Korrektur, die der diskursiven Macht des abstrakten Informationsbegriffs als ein Imaginäres Rechnung trägt, welches gleichermaßen Wunsch und Versprechen ist. In ihm steckt immer zu viel und zu wenig zugleich. Zu wenig, weil er im Prozess der technischen Realisierung stets eine Transformation erfährt. Zu viel, weil die Realisierung von einem Überschuss, einem (uneingelösten) Versprechen begleitet wird, welches jedoch das technische Imaginäre speist. Dieses Imaginäre geht ebenso in die Entwicklung neuer Technologien ein wie es die Erwartungen gegenüber existierenden Technologien strukturiert. Im abstrakten, reifizierten Informationsbegriff verdichtet sich die Imagination universeller und vollständiger Informationsreservoirs.

Neben den mittels partikulärer Computeranwendungen realisierten informationellen Praktiken gilt es auch die Wünsche und Versprechen zu thematisieren, welche sich an digitalen Datenbanken re-aktualisieren, obwohl sie durch diese nicht verwirklicht werden. Es sind nicht nur die digitalen Medientechnologien,

welche die Informationsgesellschaft hervorbringen, sondern auch das Imaginäre der *information mythology*. Dieses nimmt Eingang in die Diskurse um die Informations- und Wissensgesellschaft und befördert utopische und dystopische Beschreibungen unserer medialen Gegenwart sowie Prognosen für die Zukunft. Diese Beschreibungen sind selten gänzlich fernab der Wirklichkeit, aber ebenso selten völlig zutreffend.³¹

KOMMUNIKATION MIT INFORMATIONSSAMMLUNGEN

Wenn das Aufkommen des generalisierten, abstrakten und reifizierten Informationsbegriffs das Resultat einer zunehmenden Dekontextualisierung des Begriffs von seinen Gebrauchskontexten ist, dann gilt es im Folgenden nach dem theoretischen und praktischen Kontext zu fragen, in dem die Datenbankidee auftaucht und in dem sich die Entwicklung digitaler Datenbanken vollzieht. Eine Perspektive hierauf eröffnet die Auseinandersetzung mit Modellen der Datenbankkommunikation, welche den kommunikativen Zusammenhang beschreiben, in denen sich Datenbanken situieren, und vor deren Hintergrund sich die Herausforderungen abzeichnen, auf die Datenbanktechnologien antworten.

Im ersten Schritt wird das Kommunikationsmodell nachgezeichnet, auf dem Mooers' Konzeption des Information Retrieval beruht. Hierbei werden die Differenzen zum nachrichtentechnischen Kommunikationsmodell Shannons und infolgedessen zu dessen Informationsbegriff offenkundig. Das Kommunikationsmodell des Information Retrieval mündet Ende der 1950er Jahre in ein dezidiertes Modell der Datenbankkommunikation. Ausgehend hiervon werden im zweiten Teil die Erwartungen, aber auch Herausforderungen des Findens von Informationen in der unsichtbaren Tiefe des Computers betrachtet. Abschließend wird in Rekurs auf Luhmanns systemtheoretisch fundierten Erfahrungsbericht über die Kommunikation mit seinem Zettelkasten die Frage diskutiert, inwiefern es gerecht fertigt ist, die Interaktion mit Datenbanken als Kommunikation zu verstehen.

Das Kommunikationsmodell des Information Retrieval

Das Information Retrieval widmet sich dem Problem des maschinen- respektive computergestützten Suchens und Findens von Informationen. Mooers' ursprünglicher Definition dieses Problem- und Forschungszusammenhangs liegt eine Vorstellung von Information zugrunde, die seinem eigenen Erinnerungsbericht zu-

31 | Winkler kommt am Ende von *Docuverse* zu einem ähnlichen Schluss, wenn er einerseits die Forderung »zu einer realistischen Untersuchung der Strukturen, die im Datenuniversum sich entwickelt haben« (Winkler 1997a: 338) formuliert und andererseits dafür plädiert, »die Wünsche als Wünsche ernster zu nehmen« (Winkler 1997a: 338).

folge an Shannons Informationsbegriff anknüpft und diesen auf den Bereich der bibliothekarischen Informationsverarbeitung anwendet. Im Zentrum des Information Retrieval steht für Mooers die Frage der Auswahl, welche seines Erachtens gleichermaßen den Kern von Shannons Informationstheorie bildet: »Well, the whole message of information theory turns on the matter of choice, and choice in comparison to other choices« (Corbitt 1993: 9).

Auch wenn hierdurch ein wichtiger Aspekt von Shannons Informationsbegriff berührt wird, verschiebt sich in Mooers' Aneignung die Ebene, auf der die Wahl stattfindet. Geht es auf der einen Seite um das nachrichtentechnische Problem der Übertragung von Informationen in einem Kanal und somit um die Selektionen, die der Äußerung von Nachrichten inhärent sind, widmet sich das Information Retrieval auf der anderen Seite der Herausforderung, aus einer Zahl von Nachrichten (oder genauer Dokumenten) diejenigen auszuwählen, die die gesuchten Informationen enthalten – selbst wenn der Suchende noch eine unspezifische Vorstellung von dem hat, was er sucht. Insofern erfährt der shannonsche Informationsbegriff durch seine Anwendung im Information Retrieval eine Transformation. Diese resultiert aus der Verschiebung des Kontexts, der bedeutsam bleibt, auch wenn sich der Informationsbegriff Shannons in Folge seiner Popularisierung zunehmend verselbstständigt hat und Information losgelöst von der Theorie als eigenständige Entität erscheint, die theoretisch beschrieben und praktisch verarbeitet werden kann.

Der Ansatz Shannons basiert auf einem Kommunikationsmodell, vor dessen Hintergrund er sich dem Problem der nachrichtentechnischen Informationsübermittlung stellt und es theoretisch löst. Wenn Mooers und mit ihm die Proponenten des Information Retrieval auf ein generalisiertes, abstraktes und vor allem reifiziertes Informationskonzept rekurrieren, dann geschieht dies nicht auf der Grundlage desselben Modells. Das Information Retrieval stützt sich auf ein eigenes Kommunikationsmodell, welches Mooers wie folgt beschreibt:

»In information retrieval, the addressee or receiver rather than the sender is the active party. Other differences are that communication is temporal from one epoch to a later epoch in time, though possibly at the same point in space; communication is in all cases unidirectional; the sender cannot know the particular message that will be of later use to the receiver and must send all possible messages; the message is digitally representable; a ›channel‹ is the physical document left in storage which contains the message; and there is no channel noise because all messages are presumed to be completely accessible to the receiver. The technical goal is finding in minimum time those messages of interest to the receiver, where the receiver has available a selective device with a finite digital scanning rate. (Mooers 1950a: 572)

Indem Mooers die Kommunikation des Information Retrieval in impliziter Abgrenzung zu Shannons Kommunikationsmodell expliziert, rückt er sich abermals in die Nähe zu dessen Informationstheorie. Dennoch wird in Mooers' Bearbeitung von Shannons Modell ein weitreichender konzeptioneller Wandel manifest, da

Kommunikation nicht mehr als Austausch zwischen Personen entworfen wird, sondern als ein mehr oder minder zielgerichteter Zugriff auf Informationen, die zuvor gespeichert wurden. Das nachrichtentechnische Problem der möglichst ungestörten Übertragung von Signalen wird übersetzt in die bibliothekarische Herausforderung, aus einer Menge vorhandener Dokumente diejenigen auszuwählen, die das Informationsbedürfnis eines Nutzers befriedigen. Information wird zu einem Potenzial, welches im Prozess des Information Retrieval aktualisiert werden muss. Im suchenden Zugriff auf ein Reservoir vorhandener Dokumente kommuniziert der Nutzer, indem er Informationen selektiert und durch diese überrascht wird. Aus diesem Informationspotenzial speist sich das Imaginäre der Datenbank als einem unerschöpflichen und umfassenden Informationsreservoir. Kevin Kelly hat dies 1984 in seinem Bericht zur Lage der Netzwerknation so formuliert:

»I've always wanted a World Brain. The kind that lives in science-fiction books, where you can ask it any question from any terminal and it beeps out the answer. [...] Now I'm getting my change. I've enrolled in a two-day training workshop in how to search Dialog. What we are courting is a memory that cradles all government documents, every telephone book in the country, several million magazines, the major newspapers, the entire patent office, a backlog of scientific reports, and encyclopedias (note plural).« (Kelly 1984: 40)

Bei *Dialog* handelt es sich um den ersten kommerziellen Datenbankservice, der seit 1972 die Suche nach Forschungsliteratur in ausgewählten thematischen Datenbanken erlaubte.³² Darin sieht Kelly die Vision eines World Brain erfüllt, die H.G. Wells in den 1930er Jahren formuliert hat (vgl. Wells 1971 [1938]). Rückblickend hat Roger K. Summit, der Entwickler von *Dialog*, seine Motivation zur Arbeit an dem System wie folgt beschrieben: »A common statement around Lockheed at the time was that it is usually easier, cheaper, and faster to redo scientific research than to determine whether it has been done previously« (Summit 2002). Dies sollte sich mit *Dialog* ändern: Das Finden des bereits Gewussten wurde leichter und effektiver als das erneute Erfinden.³³

Auch für Mooers ist diese Idee leitend. Er modelliert den suchenden Zugriff auf eine Informationssammlung als Kommunikation, bei der sich die Rollen von

32 | Die Entwicklung von *Dialog* und anderen Online-Datenbankservices zwischen 1963 und 1976 haben Bourne und Hahn (2003) ausführlich aufgearbeitet.

33 | In den Fragmenten über den Ideenlauf hat Josias Ludwig Gosch bereits im ausgehenden 18. Jahrhundert darauf insistiert, dass es leichter sei, Wissen zu erlernen, als es zu erfinden: »Fast unglaublich ist der Unterschied zwischen der Mühe, welche bei der Erfindung der Wahrheiten angewendet werden muß, und derjenigen, die bei der Erlernung derselben nötig ist, wenn sie bereits von andern erfunden sind« (Gosch 2006 [1789]: 118). Dies hat aber zur Voraussetzung, dass bekannt ist, was bereits erfunden wurde und damit auch erlernt werden kann.

Sender und Empfänger umkehren. Ist in Shannons Modell der Adressat einer Nachricht passiv, weil er die von einer Nachrichtenquelle oder einem Kommunikator an ihn gerichtete Nachricht schlicht empfängt, kommt ihm im Modell von Mooers eine aktive Rolle zu. Beim Information Retrieval ist es der Adressat, der entscheidet, welche Nachrichten für ihn von Bedeutung sind, d.h. welche Informationen er sucht. Da der Sender auf der anderen Seite nicht wissen kann, wonach der künftige Empfänger einmal suchen wird, besteht seine Aufgabe darin, alle möglichen Nachrichten auszusenden. Der Sender artikuliert streng genommen keine bestimmte Nachricht mehr, sondern fungiert als Sammler, der alle möglichen Informationen zusammenträgt, die potentiell einmal von Bedeutung sein könnten. Übertragen werden die Nachrichten in Form von Dokumenten, bei denen das Rauschen einer Störquelle, anders als in Shannons Konzeption, kein Problem mehr darstellt. Insofern verschwindet das Rauschen des Kanals in Mooers' Modell und wird allenfalls implizit auf Seiten des Suchenden wiedereingeführt, wenn dieser nicht nur Unbekanntes, sondern Unerwünschtes findet.³⁴ Damit sind die zentralen Elemente von Mooers' Kommunikationsmodell des Information Retrieval beschrieben, welches zwar in Anlehnung an Shannons Modell formuliert wurde, aber in zentralen Punkten mit ihm bricht. Infolgedessen kann das nachrichtentechnische Modell von Kommunikation nicht als alleinige Basis des medientheoretischen Denkens über Informationstechnologien dienen. Vielmehr gilt es, die Spezifik der Kommunikation mit Informationssammlungen anzuerkennen und danach zu fragen, welche Forschungs- und Untersuchungsperspektiven sich hieraus für die medientheoretische Analyse digitaler Datenbanken eröffnen.

Das Kommunikationsmodell des Information Retrieval besetzt die Leerstelle zwischen abstraktem Informationsbegriff und partikularen Informationstechnologien bzw. informationellen Praktiken. Es bildet den Hintergrund, vor dem sich die spezifischen Informationsprobleme abzeichnen, die es durch die Entwicklung von Information Retrieval-Technologien zu lösen gilt. Aus den unterschiedlichen Modellierungen von Kommunikation resultieren verschiedene Perspektiven auf Information. Hierauf hat Robert A. Fairthorne, der ebenfalls zu den Pionieren der Informationswissenschaft zählt, in dem 1954 publizierten Aufsatz *The Theory of Communication* hingewiesen:

»[T]he communication engineer has different interests. In particular he is not concerned with completed messages, but how to deal with bits of them during the course of communication. He cannot do much to alter the temporal order of

34 | Michel Serres hat die relative Unterscheidung von Signal und Rauschen im Anschluss an Shannon ins Zentrum seiner kommunikationsphilosophischen Theorie des Parasiten gerückt (Serres 1987). Sollte die Unterscheidung von unbekannten und unerwünschten Suchergebnissen ihre Entsprechung in derjenigen von Signal und Rauschen finden, dann ließe sich das Suchen und Finden von Informationen ebenso als parasitär beschreiben.

these bits, because this implies storage and storage implies delay, and delay is what communication engineers are paid to fight. We, on the other hand, deal with spatial collections of completed messages and, after recognition or identification, questions of their ordering and disordering predominate. Finally, the communication engineer has a much narrower semantics, which refers only to the statistical properties of the language of the message, not to its sense or seemliness.« (Fairthorne 1961 [1954]: 65)

In Übereinstimmung mit Mooers charakterisiert Fairthorne das Kommunikationsproblem, dem sich die Nachrichtentechnik zuwendet, als eines der Übertragung von Nachrichten im Raum.³⁵ Diese Form der Kommunikation ist zeitkritisch in einem technischen Sinn, da es darum geht, Signale in möglichst kurzer Zeit von einem Punkt A an einen Punkt B zu übertragen. Die Informationen, denen sich das Information Retrieval widmet, liegen hingegen gespeichert im Raum vor: »Library communication differs from telegraphic mainly in that all the messages have been sent already, and you have to pick out the right one to suit a query not known beforehand« (Fairthorne 1961 [1953]: 25). Der Fokus der bibliothekarischen Informationsverarbeitung liegt dementsprechend nicht auf der Übermittlung von Nachrichten, sondern auf deren Erfassung, Identifizierung, Ordnung und Umordnung. Schließlich weist Fairthorne auf einen weiteren zentralen Unterschied zwischen nachrichtentechnischer und bibliothekarischer Informationsverarbeitung hin. Während der Kommunikationsingenieur weitgehend oder ganz vom Bedeutungsgehalt und der Angemessenheit von Nachrichten absieht, interessieren sich Informationswissenschaftler genau hierfür. Unter dieser Prämisse sucht Fairthorne nach den Einsatzmöglichkeiten von Technologien zur automatisierten Verarbeitung bibliothekarischer Informationen, wobei er seinen Fokus auf diejenigen semantischen Aspekte lenkt, die nicht auf dem Erschließen von Informationen durch Menschen beruhen, sondern auf routinemäßigen Vorgängen. Entscheidend für Fairthornes Arbeit ist die Frage: »[H]ow far can we go in documentation, as in computing, by using ritual in place of understanding?« (Fairthorne 1961b: X). Entsprechend der dieser Herangehensweise zugrunde liegenden pragmatischen Maxime versucht Fairthorne nicht, eine Technologie zu entwickeln, die der menschlichen Interpretations- und Verständnisfähigkeit semantischer Informationen ebenbürtig ist. Sein Ziel ist bescheidener und beruht auf der Beobachtung, dass es im Bereich der bibliothekarischen Informationsverarbeitung eine Reihe repetitiver und routinemäßiger Vorgänge gibt, die ebenso gut von Computertechnologien wie von Menschen erfüllt werden können.

35 | Obwohl Fairthorne nicht explizit auf Mooers verweist, war ihm dessen Kommunikationsmodell des Information Retrieval bekannt. Beide lernten sich spätestens 1950 auf dem *International Congress of Mathematicians* kennen, wo Fairthorne dem Vortrag von Mooers beiwohnte (vgl. Fairthorne 1998).

Die Voraussetzung für die technische Automatisierung von »clerical activities« (Fairthorne 1961c: 94) in der bibliothekarischen Informationsverarbeitung ist, dass sich die semantischen Informationen, auf die das Information Retrieval abzielt, in syntaktische Informationen übersetzen lassen.³⁶ Laut Fairthorne befasst sich die Kommunikationstheorie Shannons mit letzterem Typus von Information und kann auf dieser Ebene für die Konzeptualisierung von Information Retrieval-Technologien fruchtbar gemacht werden:

»This kind of ›information‹ is what communication theory usually deals with. The function of telegraphy is to inform the receiver which particular message has been handed in at the transmitter. It is not concerned with any information that may be interpreted within the message itself. So long as the physical events presented at input are reproduced faithfully enough at output, all the information within the original will have been transmitted also. If not, language would not work.« (Fairthorne 1961 [1954]: 65)

Der technischen Verarbeitung semantischer Information auf dem Niveau syntaktischer Information geht die Übersetzung von Semantik in Syntax voraus. Indem Fairthorne auf den Unterschied zwischen den beiden Formen von Information hinweist, markiert er präzise die Leerstelle, die im Information Retrieval theoretisch und praktisch überbrückt werden muss: die Differenz zwischen den Konzepten einerseits und der Physik andererseits, eben zwischen Semantik und Syntax. Diese kann, so Fairthornes Ansatzpunkt, durch die Einführung eines Notationssystems überbrückt werden: »The bridge between the concepts and the physics of retrieval is the notation or system of marking the text« (Fairthorne 1961b: X).³⁷

Als Lösung schlägt er einen Übersetzungsprozess vor, bei dem Bedeutungen anhand vordefinierter Regeln in Markierungen übertragen werden. Fairthorne unterscheidet zwei Formen des Markierens von Dokumenten, die intrinsische Veränderung eines Dokuments und die dem Dokument äußerlich bleibende Anord-

36 | Die Tätigkeiten, die Fairthorne als *clerical* bezeichnet, sind routinemäßige Abläufe, welche sich in einem Protokoll formalisieren lassen: »By ›clerical‹ activities are denoted routines of observation, identification, and manipulation of marked material objects according to some protocol involving only observations on the marks, including those belonging and accessible to the observing device« (Fairthorne 1961c: 94).

37 | Wie Fairthorne darlegt, wurden Texte in Bibliotheken traditionellerweise auf zwei Weisen betrachtet: als Träger von Konzepten einerseits und als materielle Objekte andererseits. Beide Aspekte wurden seines Erachtens jedoch gesondert behandelt: »For some millennia librarians have had to deal with texts as carriers of concepts, and with texts as heavy objects with marks on. They have evolved efficient techniques and principles to cope with these aspects severally« (Fairthorne 1961b: IX).

nung von Dokumenten, welche er auch als »marking and ›parking« (Fairthorne 1961 [1953]: 95) bezeichnet: »[W]hen counting sheep, you must distinguish those already accounted for from those not. This can be done with a branding iron [...]. Or you can segregate the counted sheep in a fold« (Fairthorne 1961 [1953]: 95). Geläufiger sind heute die Bezeichnungen »deskriptives Markup« und »formale Strukturierung«. Beim Markup werden Informationen deskriptive Metainformationen hinzugefügt, welche als Kennzeichnungen fungieren, die über das Gekennzeichnete informieren. Demgegenüber erschließt sich die Bedeutung von formal strukturierten Informationen aus dem Platz, der ihnen innerhalb einer Struktur zugewiesen wurde.

Eine ähnliche Idee der Übersetzung steht auch im Zentrum von Mooers' Ansatz, die Probleme des Information Retrieval technisch zu lösen: »To avoid scanning all messages in entirety, each message is characterized by N independently operating digital descriptive terms (representing ideas) from a vocabulary V , and a selection is prescribed by a set of S terms« (Mooers 1950a: 573). Im Unterschied zu Fairthorne begründet Mooers die Notwendigkeit, Dokumente mit Deskriptoren zu versehen, mit einem ökonomischen Argument. Das Finden von Informationen werde effektiver, wenn nicht der Volltext aller verfügbaren Dokumente durchsucht werden müsse, sondern nur die beschreibenden Terme. Den für die Praxis der computer-technischen Informationsverarbeitung nicht unerheblichen Effizienzerwägungen von Mooers gibt Fairthorne hingegen eine informationstheoretische Wendung.³⁸ Ohne die Übersetzung von Semantik in Syntax können Technologien nicht zur bibliothekarischen Informationsverarbeitung im Allgemeinen und zum Suchen und Finden von Informationen im Besonderen eingesetzt werden. Dieses Postulat gilt es medientheoretisch ernst zu nehmen.

Einerseits kommt in Fairthornes Vorschlag die geläufige Überzeugung zum Ausdruck, dass Computer stets nur auf dem Niveau der Physik operieren und demzufolge lediglich syntaktische Merkmale von Zeichen interpretieren können. Andererseits bedeutet dies entgegen der oft vertretenen Meinung keinen Verzicht auf oder Abschied von Semantik. Im Gegenteil: Semantik kann in Syntax übersetzt und technisch in nicht-semantischen Routinen verarbeitet werden. Wenn Kulturtechniken wie von Sybille Krämer vorgeschlagen als Modi des nicht-interpretativen operativen Umgangs mit Zeichen definiert werden und damit die durch

38 | Wenn im Folgenden Fairthornes Position in den Vordergrund gerückt wird, soll dies nicht implizieren, dass die Effizienzerwägungen von Mooers nebensächlich sind. Im Gegenteil, derartige Erwägungen sind noch immer von großer Bedeutung. Heute ist zwar die Volltextsuche relativ unproblematisch, dafür hat das WWW neue Herausforderungen mit sich gebracht, welche die verfügbaren Informationssysteme an ihre technologischen Grenzen bringen. Infolgedessen müssen bei der Implementierung neuer Technologien weiterhin verschiedene Zielstellungen gegeneinander abgewogen werden. Ein Informationssystem, das alle Aufgaben gleich gut erledigen kann, ist eine Wunschvorstellung.

Kulturtechniken vollbrachte Entsemantisierungsleistung betont wird, dann können Information Retrieval-Systeme im Besonderen und Datenbanken im Allgemeinen als Kulturtechnologien des Speicherns und Findens begriffen werden (vgl. Krämer 2003b: 169f.).³⁹

Im Anschluss hieran ist es eine Aufgabe der medienkulturwissenschaftlichen Auseinandersetzung mit digitalen Datenbanken, die unterschiedlichen Formen und Strategien der Übersetzung von Bedeutung in syntaktische Strukturen sowie die verschiedenen Modi der computertechnischen Verarbeitung von Semantik als Syntax freizulegen und zu analysieren. Das Versehen von Dokumenten mit deskriptiven Markierungen stellt in diesem Zusammenhang nur eine Möglichkeit dar, wie Semantik jenseits semantischer Interpretationsprozesse verarbeitet werden kann. Die Speicherung von Daten in bestimmten Strukturen, wie z.B. in Tabellen, ist eine andere Form der Übersetzung semantischer Informationen in Syntax. Auch Algorithmen können eine solche Übersetzung vollziehen. Die Regeln, gemäß denen die Übertragung zwischen den beiden Ebenen vollzogen wird, sind dem Computer jedoch äußerlich und stellen zunächst nur Hypothesen dar, die einer experimentellen Überprüfung bedürfen, welche die Brauchbarkeit für Menschen und nicht ein wie auch immer geartetes Verstehen durch Computer zum Gradmesser nimmt. Hierauf weist Mooers hin, wenn er darauf insistiert, dass bei der Entwicklung von Information Retrieval-Systemen weniger die Maschinen, als vielmehr die menschlichen Nutzer zentral seien: »At all times, one should remember that the human customer who uses the information retrieval system is the one who must be served, and not the machine« (Mooers 1960: 229). Auch in dieser Hinsicht weist die Konzeptualisierung des Information Retrieval Parallelen zu Krämers Kulturtechnikkonzept auf. Kulturtechniken bewirken, so Krämer, nicht nur eine »Aufspaltung von ›Operation‹ bzw. ›Konstruktion‹ einerseits und ›Interpretation‹ andererseits« (Krämer 2003b: 169), sondern auch eine Entkopplung von Rezept- und Begründungswissen, d.h. »das Wissen, wie eine Aufgabe zu lösen ist, [trennt sich, M.B.] vom Wissen, warum diese Lösung ›funktioniert‹« (Krämer 2003b: 170). Mit anderen Worten: Die Begründung ist der Lösung nicht intrinsisch, weshalb externe Kriterien über den Erfolg oder Misserfolg von kulturtechnischen Operationen entscheiden und nicht die dem routinemäßigen Vorgehen inhärente Funktionslogik. Um mit semantischen Informationen in nicht-semantischen kulturtechnischen Routinen umzugehen, bedarf es demzufolge keines Verstehens von Bedeutungen. Das interpretative Verständnis des Gehalts von Informationen bleibt den Verfahrensweisen äußerlich: »[S]emantic knowledge is needed only at the very beginning of communication. Once translation rules have been established it becomes irrelevant« (Fairthorne 1961 [1953]: 24f.).

39 | An dieser Stelle wird das weite Verständnis von Datenbanken zugrunde gelegt, wie es im vorangegangenen Kapitel »Datenbank« (S. 120ff.) dargestellt wurde, d.h. *Datenbank = Daten + Zugriff*.

Festzuhalten bleibt, dass sich der Informationsbegriff vor dem Hintergrund verschiedener Kommunikationsmodelle ausdifferenziert. Bleiben diese Unterschiede bei Mooers noch implizit, so werden sie in Fairthornes Schriften zum Information Retrieval deutlich herausgearbeitet. Zugleich zeigt Fairthorne aber auch, dass die shannonsche Informationstheorie für das Information Retrieval nutzbar gemacht werden kann, sofern die Übersetzung von semantischer Information in syntaktische Information vollzogen ist.

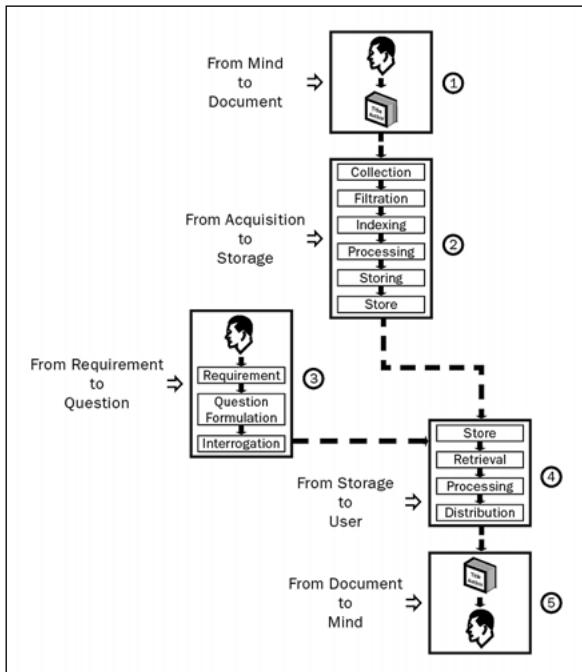
Data Banking: Vom Finden in einer Black Box

Dem Speicher als Ort der Sammlung misst Mooers in seinem Modell noch keine besondere Bedeutung zu, er wird als passiver Container begriffen, dessen reibungsloses, d.h. rauschfreies, Funktionieren vorausgesetzt wird.⁴⁰ An diese Stelle tritt Ende der 1950er Jahre die Datenbank, als im Rahmen eines vom *Office of Naval Research* finanzierten Forschungsprojekts bei der *Benson-Lehner Corporation* das Kommunikationsmodell des Information Retrieval zu einem Modell der Datenbankkommunikation erweitert wurde (vgl. Worsley et al. 1959). Im Mittelpunkt des Projekts stand eine prozessorientierte Analyse der Speicherung und Abfrage von Informationen, die im Abschlussbericht unter dem Oberbegriff *data banking* zusammengefasst werden: »Data Banking«, heißt es, »is the process of communicating between many conceivers to many receptors through a store. By providing mechanisms for decision-making the receptor selects resting information« (Worsley et al. 1959: 5f.). Das Hauptinteresse der Forschergruppe liegt wie auch bei Mooers auf den Informationspotenzialen, die sich auf Seiten des Nutzers im suchenden Zugriff aktualisieren können (vgl. Abb. 3). Jedoch erweitert sich das Problembewusstsein in Richtung der Speicherungsprozesse, denn bevor Informationen aus einer Daten-

40 | An der Rolle, die dem Rauschen in den beiden Kommunikationsmodellen beigemessen wird, zeichnet sich ein zentraler Unterschied zwischen diesen ab. Obwohl bereits Mooers' Wahl des Vortragstitels *Information Retrieval Viewed as Temporal Signalling* nahe legt, dass der zentrale Unterschied zwischen den beiden Ansätzen seines Erachtens darin besteht, dass es sich auf der einen Seite um Signalübertragung im Raum und auf der anderen Seite um Signalübertragung in der Zeit handelt, greift diese Gegenüberstellung zu kurz. Denn auch in Shannons Modell ist die Möglichkeit der asynchronen Kommunikation gegeben. Kommunikation über Zeit bedeutet hier das Aussenden einer bestimmten Nachricht, die verkörpert als respektive gespeichert in einem Dokument zu einem späteren Zeitpunkt einen Empfänger erreicht. Das mögliche Rauschen des Kanals ist dabei noch immer ein Problem. Erstens können unter den Bedingungen der (digital-)technischen Speicherung von Dokumenten Störungen bei deren Codierung und Dekodierung auftreten und zweitens ist jedes Dokument aufgrund der ihm eigenen Materialität anfällig für Zer-Störung. Mooers zieht diese Probleme in seinem Modell nicht in Betracht; oder genauer: Er setzt ihre Lösung voraus.

bank abgefragt werden können, müssen diese trivialerweise zuerst im Speicher abgelegt werden.

Abb. 3: Modell des Data Banking

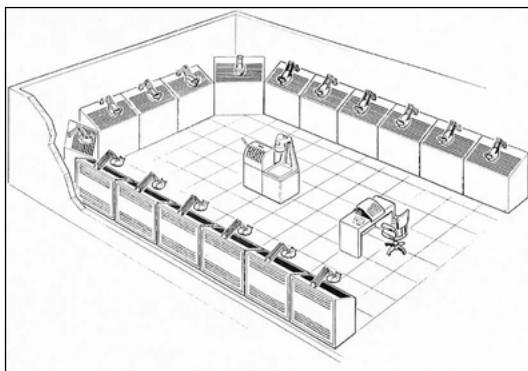


Quelle: Worsley et al. 1959: 30f.

An der funktionalen Schnittstelle von Speicherung und Abfrage setzt der im Rahmen des Projekts entwickelte Entwurf einer Datenbank an, der die in Punkt 2 und 4 analysierten Funktionalitäten, d.h. die Sammlung, Filterung, Indexierung, Verarbeitung, Speicherung, Abfrage und Distribution von Dokumenten, technisch, aber nicht computertechnisch umsetzt. Konzeptualisiert wurde diese Datenbank als Mikrofilmbibliothek, die bis zu einem gewissen Grad an Bushs Memex und vielleicht sogar noch mehr an Emanuel Goldbergs Vision der Bibliothek der Zukunft erinnert (vgl. Gesell 1926; Anonymus 1938; Buckland 2006). Der im Bericht eingeführte Begriff der Datenbank geht jedoch über diesen spezifischen Entwurf hinaus. Er dient als Oberbegriff für alle Formen von Informationssammlungen respektive Informationsspeichern, die sowohl Resultat als auch Ausgangspunkt für das *data banking* sind. Dementsprechend können auch Bibliotheken als Datenbanken bezeichnet werden (vgl. Worsley et al. 1959: 9).⁴¹

41 | In dieser Hinsicht ähnelt die im Rahmen des Forschungsprojekts bei der Benson-Lehner Corporation entwickelte Konzeption von Datenbanken der juristischen Defini-

Abb. 4: *The Data Bank*



Quelle: Worsley et al. 1959: 91

Der so entwickelte Entwurf einer Datenbank lässt sich als eine mediale Konfiguration begreifen, die der Speicherung und Abfrage von Dokumenten dient, welche in Form von Mikrofilmen hinterlegt sind und deren Inhalte anhand von Stichworten durchsucht werden können (vgl. Abb. 4).⁴² Entscheidend ist hieran, dass diese Datenbank es ihren Nutzer ermöglichen soll, automatisch im Informationsbestand zu suchen. Hierdurch wird die Datenbank zu einer Black Box, auf die apparativ zugegriffen werden kann. Diesen Aspekt rückt Charles Bachman wenige Jahre später in der Vordergrund seiner Konzeption von Datenbanken:

»Let us design a ›black box‹ with a number of input-output plugs. Each plug would have the ability to receive or transmit a record. In addition, this plug would be able to receive and cause the ›black box‹ to respond to Integrated Data Store commands. Let us build these boxes with a capacity to store one, two, or ten million records. The box would be extremely useful if only electric typewriters could be attached. A man could manually type a new record on his typewriter, then Put it into the data store. He could Modify or Delete records that are already there. Best of all, he could Get a record when he wanted it. Based upon the information gained from one record, he could ask for another record. He could continue until he had exactly the quantity of information he wanted. He would have every advantage that the marvelous little game, ›twenty questions‹ gives its player. Each question asked is based upon new knowledge, a new strategy, for solving the problem. Contrast our player of ›twenty questions‹ at a typewriter of an Integrated Data Store and the man with a three inch report printed two weeks ago with data that was *then* a week out of date. Which

tion der Europäischen Union, S. 128f. Der Datenbankbegriff ist weder an eine spezifische Technologie, noch an eine bestimmte Form der Verkörperung der Sammlung gebunden.

42 | Ob und in welcher Form der Entwurf je realisiert wurde, ist unbekannt.

will make the best decision? In our dynamic business there is nothing equivalent to Instant Information. With the ability to attach dozens of typewriters, this black box would become one of the most significant information processing tools available.« (Bachman 1962b: IIB-4-2)

Diese im Zusammenhang mit der Entwicklung des *Integrated Data Store* – einem Vorläufer späterer Datenbankmanagementsysteme – stehende Beschreibung entwirft Datenbankkommunikation als Interaktion mit einer Black Box, in die man mithilfe eines Ein- und Ausgabegeräts Datensätze eingeben, gegebenenfalls modifizieren und löschen sowie interaktiv abfragen kann. Mit den Anfang der 1960er Jahre verfügbaren Computertechnologien war insbesondere der interaktive Umgang mit digitalen Informationssammlungen nur schwer möglich. Zu dieser Zeit war die Stapelverarbeitung von Prozessen dominant, bei der eine Reihe von Aufgaben erst gesammelt und zu einem späteren Zeitpunkt gebündelt ausgeführt wird. Im interaktiven Zugriff werden die Befehle der Nutzer sofort ausgeführt, sodass die Ergebnisse neue Befehle nach sich ziehen können. Durch die instantane Verarbeitung von Befehlen entsteht der Eindruck direkter Interaktion mit dem Computer, wodurch sich nicht zuletzt die Weise des Zugriffs auf Information ändert. Bachman vergleicht die Informationssuche mit dem Spiel *Zwanzig Fragen*, bei dem ein Spieler den Namen einer Person herausfinden muss, indem er 20 Fragen stellt, die mit Ja oder Nein beantwortet werden können (vgl. Bachman 1962b: IIB-4-2). Im Verlauf des Spiels werden die Fragen sukzessive präzisiert, sodass der Spieler – vorausgesetzt er stellt die richtigen Fragen – am Ende weiß, welche Person er ist. Übertragen auf die Suche in Datenbanken bedeutet dies, dass das Informationsbedürfnis eines Nutzers nicht sofort befriedigt werden muss, sondern durch sukzessive Reformulierung und Präzisierung der Anfragen gestillt werden kann.⁴³ Hinzu kommt, dass die Abfrage von Informationen aus digitalen Datenbanken im Allgemeinen und der interaktive Zugriff im Besonderen höhere Aktualität verspricht, was speziell im wirtschaftlichen und administrativen Kontext von hoher Bedeutung ist. Denn wer über aktuellere Informationen verfügt, vermag, so das Diktum des Managementdiskurses seit Anfang der 1960er Jahre, bessere strategische Entscheidungen zu treffen sowie die Vorgänge im Unternehmen effizienter zu steuern (vgl. Haigh 2007: 59ff.).

Ohne auf Bachmanns Parallelisierung von interaktiven Datenbankabfragen mit dem Spiel *Zwanzig Fragen* zu rekurrieren, zieht David Gugerli denselben Vergleich, wenn er die dem gleichen Spielprinzip folgende Fernsehsendung *Was bin*

43 | Ein ähnliches Ziel verfolgte auch Edgar Codd mit seinem relationalen Datenbankentwurf, wie Gugerli herausgearbeitet hat: »Codd's Suchmaschine [...] sollte ergebnisoffener funktionieren. Die Nutzer seiner Suchmaschinen brauchten nicht zu wissen, wie der Datenraum aufgebaut, wie er strukturiert war und was er genau enthielt. Sie konnten die Datenbank als Black Box behandeln, an die sich auch Fragen richten ließen, deren Beantwortbarkeit bislang nicht getestet worden war« (Gugerli 2009: 72).

ich? als Suchmaschine interpretiert (vgl. Gugerli 2009: 19ff.). Bei seiner Analyse der zwischen 1955 und 1989 von Robert Lembke moderierten Spielshow stehen weniger die Potenziale des interaktiven Frage- und Antwortspiels für das Information Retrieval im Vordergrund, als die normierende und normalisierende Funktion, die durch eine Suchmaschine vom Typ *Was bin ich?* bewirkt wird. Normal ist, wenn man über einen eindeutig benenn- und zuweisbaren Beruf verfügt, der im Rahmen des Spiel erraten werden kann (vgl. Gugerli 2009: 21).⁴⁴ Obwohl es sich bei dem von Gugerli analysierten Beispiel um eine Fernsehshow handelt, lässt sich hieran ein wichtiger Aspekt der Kommunikation mit digitalen Datenbanken freilegen. Wird in *Was bin ich?* eine stabile Zuweisung von Person und Beruf hergestellt, geht es bei der Suche in digitalen Datenbanken um die Verknüpfung von Frage und Antwort. Normierung ist hierbei sowohl Strategie zur Lösung dieses Zuweisungsproblems als auch sein Effekt. Das gemeinsame Versprechen der Spielshow sowie der Datenbank ist, dass etwas gefunden werden kann; der Beruf einer Person einerseits und Informationen andererseits.

Der Zugriff eines Informationssuchenden auf die Datenbank operiert mit zwei Unbekannten: den potenziell auffindbaren Informationen und der internen Wissensordnung der Datenbank, die das Finden von Informationen zwar ermöglicht, aber das Suchen zugleich unter Regeln stellt.⁴⁵ Gefunden werden kann nur im Rahmen der durch die digitale Suchtechnologie zur Verfügung gestellten Möglichkeiten, weshalb es, wie Weingarten hervorgehoben hat, notwendig ist, dass der Nutzer »sein Informationsbedürfnis zunächst nach den Maßgaben der Systematik der Datenbank reformuliert« (Weingarten 1994: 169). Der interaktive Umgang mit einer digitalen Datenbank ermöglicht es dem Nutzer, seine Anfragen sukzessive der Logik der Black Box anzunähern. Verfügt der Informationssuchende über keinerlei Vorwissen im Umgang mit der digitalen Suchtechnologie, sind die Möglichkeiten der interaktiven Annäherung an die interne Datenbanklogik in der Regel jedoch sehr beschränkt. Hierauf hat Roger Summit hingewiesen:

»It occurred to me that we should be able to simply parse the plain text statement of a query and match those words against a database of textual citations, identify the

44 | Gugerli entwickelt die These, dass Lembkes Spielshow eine konservative Antwort auf »die dramatischen Veränderungen in der Berufswelt der sechziger und siebziger Jahre des 20. Jahrhunderts« (Gugerli 2009: 22) war, wobei die eindeutige Zuweisbarkeit eines Berufs zu einer Person als gesellschaftlicher Normalzustand vorgeführt wurde: »Sendung für Sendung, Gast um Gast sollte die doppelte Gewißheit darüber erzeugt werden, daß sich jeder ›normalen‹ Person eine Berufstätigkeit zuweisen ließ und daß sich dabei die Zeugen dieses Zuweisungsprozesses ihrer eigenen beruflichen Normalität versichern durften« (Gugerli 2009: 24).

45 | Die Wissensordnung der Datenbank wird durch die Regeln der Übersetzung von Bedeutung in syntaktische Strukturen sowie durch die hieran anschließenden Verarbeitungsroutinen bestimmt.

relevant items and then sort them according to word hit frequency [...]. The results of this process were disappointing. One of the main issues I recall had to do with the mystery of how to modify the query to obtain better results. Because the search and relevance algorithms within the search engine are unknown to the user, how to modify the query to improve the results is not apparent. We referred to this as black box searching and abandoned further work along these lines.« (Summit 2002)

Ist die Ordnungs- und Suchlogik der Datenbank (d.h. die Regeln der Übersetzung von Semantik in Syntax, die Prinzipien der Selektion sowie die Konventionen der Formulierung von Anfragen) dem Nutzer fremd, dann vermag dieser, so das von Summit beschriebene Bedenken, seine Suchanfragen nicht zu präzisieren, um bessere Suchresultate zu erzielen. Mit anderen Worten: Die Nutzer müssen den Umgang mit Datenbanken erst erlernen, damit sie in der Lage sind, in der Interaktion mit dem Suchinterface ihre Anfragen zu präzisieren. Dies gilt Summit zufolge insbesondere dann, wenn Suchmaschinen es zum Beispiel erlauben sollen, natürlichsprachlich formulierte Fragen an die Datenbank zu richten. Zu welchen Ergebnissen verschiedene Suchanfragen führen, ist dabei abhängig von ihrer algorithmischen Interpretation. Summits Einwand stellt die von Bachman vorgeschlagene Konzeptualisierung der Datenbank als Black Box nicht grundlegend in Frage, sondern erweitert sie kritisch um einen wichtigen Aspekt.⁴⁶ Denn ganz gleich wie kompetent der Suchende im Umgang mit der Datenbank ist, dem Erfolg der Suche bleibt ein gewisses Maß an Unsicherheit inhärent, da der Suchende seinen Erfolg normalerweise nur an den Ergebnissen seiner Anfrage bemessen kann und es nicht möglich ist, das Gefundene vor dem Hintergrund des potenziell Auffindbaren zu bewerten. Ob eine spezifische Information in der Datenbank nicht existiert oder nur nicht gefunden wurde, lässt sich anhand der Ergebnisse allein nicht entscheiden.⁴⁷

Strategien, um diese Unsicherheit zu kompensieren und die Zuweisung von Suchanfragen und Suchergebnissen zu stabilisieren, sind beispielsweise die Verwendung von kontrollierten Vokabularen, Klassifikationssystemen und Taxonomien sowie

46 | Mit ihren jeweiligen Verweisen auf die Black Box verbinden Summit und Bachman unterschiedliche Assoziationen. Dies wird in Summits Plädoyer für die interaktive Suche deutlich, das mit Bachmans Beschreibung der Datenbank als Black Box konform geht: »In my view of an interactive system, information retrieval should be thought of as a process, not as a probe (as is the case with batch systems). With the exception of simple, explicit searches, the searcher is neither completely aware of what is contained in the database, nor confident of just which words to use in the query to elicit a desired response. Because of this, there needs to be a high degree of interaction between the searcher and the database to gain the desired outcome« (Summit 2002).

47 | Die Diskussion der Datenbank als Black Box wird im Kapitel »Techno-Logik« (S. 259ff.) erneut aufgegriffen. Relevant wird hierbei die Unterscheidung zwischen dem Datenbestand als Black Box und der Suche als Black Box.

die Entwicklung formaler Anfragesprachen, die das computergestützte Finden von Informationen erleichtern, wenn nicht sogar ermöglichen sollen. Wie der Linguist Rüdiger Weingarten Anfang der 1990er Jahre festgestellt hat, kann die Herausbildung rigider Ordnungsstrukturen als Antwort auf diejenigen Probleme verstanden werden, die aus der Entkopplung der Kommunikationspartner in der Datenbankkommunikation resultieren. Für Weingarten stellt die Datenbank ein Medium dar, welches »nicht auf ein gemeinsames Zeigfeld der Kommunikationspartner zurückgreifen« (Weingarten 1994: 168) kann. Zudem sind dem Bereitsteller von Informationen in Datenbanken der Pfad und der Ausschnitt unbekannt, den der Suchende durch die Datenbank nimmt bzw. den er von der Datenbank wahrnimmt. Dies hat zur Konsequenz, dass »die ansonsten üblichen textdeiktischen Mittel [...], aber auch makrostrukturelle Verfahren der Kohärenzbildung, wie z.B. narrative Strukturen« (Weingarten 1994: 168) nicht mehr funktionieren.⁴⁸

Obwohl die Anwendung kontrollierter Vokabulare zur thematischen Beschreibung von medialen Konstellationen, ihre Einordnung in ein rigides Klassifikationssystem sowie die formale Strukturierung von Information durchaus ein geeignetes Mittel sein können, um mit die Herausforderungen der Datenbankkommunikation technisch umzugehen, erweisen sich diese Strategien dann als problematisch, wenn sie zwischen unterschiedlichen *Communities of Practice* vermitteln sollen, in denen die gleichen Informationen Verschiedenes bedeuten und in denen auf unterschiedliche Weise mit Informationen umgegangen wird. An der Schnittstelle zwischen unterschiedlichen Gemeinschaften können *Boundary Objects* vermitteln, wie Geoffrey Bowker und Susan Leigh Star in *Sorting Things Out* dargelegt haben⁴⁹:

»Boundary objects are those objects that both inhabit several communities of practice and satisfy the informational requirements of each of them. Boundary objects are thus both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use and become strongly structured in individual-site use.« (Bowker/Star 2000: 297)

Boundary Objects ermöglichen die Vermittlung zwischen verschiedenen Praxisgemeinschaften und ihren jeweiligen Klassifikationen sowie Bedeutungen, da sie einerseits relativ abstrakt sind, andererseits aber den Anforderungen verschiedener

48 | Siehe hierzu auch Weingartens linguistische Analyse des Datenbankabfragen zugrunde liegenden Dialogmusters (vgl. Weingarten 1988).

49 | Entwickelt wurde das Konzept der Boundary Objects von Star und James R. Griesemer in einer Studie zu den unterschiedlichen Praktiken von professionellen Wissenschaftlern, Amateurforschern und Verwaltern sowie den Vermittlungen zwischen diesen Communities of Practice im *Museum of Vertebrate Zoology* der University of California, Berkeley (Star/Griesemer 1989).

Gemeinschaften angepasst werden können.⁵⁰ Im Vergleich dazu sind die standardisierten Vokabulare und Klassifikationen, mit denen viele digitale Informationssysteme operieren, sehr unflexibel.⁵¹ Deshalb machen sich Bowker und Star für den Entwurf und die Entwicklung von *Boundary Infrastructures* stark, welche die »differing constitution of information objects within the diverse communities of practice that share a given infrastructure« (Bowker/Star 2000: 314) berücksichtigen. Wie genau ein solches Informationssystem aussehen könnte und wie in diesem die technischen Prozesse der Informationsverarbeitung an die diversen menschlichen und sozialen Formen des Umgangs mit Informationen angeschlossen werden sollen, lassen die Autoren offen. Wichtiger als ein fertiger Entwurf für eine solche Infrastruktur ist jedoch, dass der Vision eines einheitlichen, universellen Informationssystems eine Alternative zur Seite gestellt wurde: die Idee eines fragmentierten, zwischen den partikularen informationellen Praktiken verschiedener Gemeinschaften vermittelnden Informationssystems. Die Kontingenz des Entwurfs von Informationssystemen wird hierdurch deutlich und erlaubt es, die Politiken digitaler Datenbanken zu hinterfragen sowie die in Datenbanken eingeschriebenen Machtstrukturen kritisch zu analysieren.⁵² Hierbei ist nicht nur von Interesse, welche Informationen auf welche Weise in einer Datenbank gespeichert werden, sondern auch, an wen sich die Datenbank richtet, wie die Zuweisung von Suchanfragen und Suchergebnissen stabilisiert wird und in welche informationellen Praktiken sie eingebunden ist.

Selbst die in diesem Kapitel bisher thematisierte Suche nach dem Bekannten, d.h. nach vorhandenen Informationen, stellt nur eine, wenn auch wichtige Form der medialen Datenbankpraxis dar. Die Datenbank – verstanden als konkrete Sammlung von Informationen – dient hierbei als Reservoir, Fundus oder Ressource zur Befriedigung künftiger Informationsbedürfnisse. Sie ist ein Informationspotenzial, welches auf abgelegten Informationen basiert, die unsichtbar abgespeichert durch Suchanfragen selektiert und auf der Benutzeroberfläche zugänglich gemacht werden können.

In anderer Hinsicht erscheint die Datenbank als Ressource, die potenziell mehr Informationen bereithält als in ihr abgespeichert wurden. Die Datenbank ist hierbei

50 | Star und Griesemer unterscheiden in ihrem Aufsatz aus dem Jahr 1989 vier Typen von Boundary Objects: Sammlungen (*repositories*), Idealtypen (*ideal types*), übereinstimmende Grenzen (*coincident boundaries*) und standardisierte Formulare (*standardized forms*) (vgl. Star/Griesemer 1989: 410f.). Diese Taxonomie erhebt jedoch keinen Anspruch auf Vollständigkeit.

51 | Dies trifft gemeinhin auch auf algorithmische Verfahren der Auswertung von Information zu. Algorithmen beschränken die Interpretationsmöglichkeiten auf vordefinierte Regeln, wie in Bezug auf Nakes Unterscheidung von Oberfläche und Unterfläche bereits festgestellt wurde, S. 96ff.

52 | Kritik wird hier zunächst im buchstäblichen Sinn von *krinein* als Aufgabe begriffen, Unterschiede zu markieren.

kein Bestand, in dem bereits Bekanntes entdeckt werden kann, sondern Basis für neues Wissen. Aus der Kombination und Rekombination von Informationen können neue Informationen entstehen, die es erlauben, Terroranschläge vorauszusehen und vorzubeugen, ökonomische Entwicklungen in Echtzeit nachzuvollziehen, den Ausbruch von Krankheitsepidemien zeitnah zu erkennen, Verbrechen zu bekämpfen, Freunde und Lebenspartner zu finden etc.⁵³ Im Großen und im Kleinen birgt die findige Kombination und Rekombination vorhandener Informationen die Möglichkeit, neue Informationen zu erhalten. Insofern können digitale Datenbanken unter Umständen etwas wissen lassen, was so noch nicht gewusst, was allenfalls latent und rein virtuell als potentielle Information vorhanden war. Dies ist die Signatur des Spiels vollständiger Information, welches nach Ansicht von Lyotard charakteristisch für das postmoderne Wissen ist, welches sich in Datenbanken materialisiert:

»Solange man es mit einem Spiel unvollständiger Information zu tun hat, kommt der Vorteil dem zu, der über Wissen verfügt und sich einen Zusatz an Information verschaffen kann. [...] Aber in Spielen mit vollständiger Information kann die höchste Performativität per hypothesis nicht im Erwerb einer solchen Ergänzung bestehen. Sie ergibt sich aus einer neuen Anordnung von Daten, die eben einen ›Spielzug‹ darstellen. Diese neue Anordnung wird meist durch die Verknüpfung von Datenreihen erreicht, die bis dahin für unabhängig gehalten wurden. Man kann diese Fähigkeit, zusammen zu artikulieren, was nicht zusammen war, als Phantasie (*imagination*) bezeichnen.« (Lyotard 2009 [1979]: 126)

In dem von Lyotard beschriebenen Spiel vollständiger Information geht es nicht mehr darum, das verfügbare Wissen durch das Hinzufügen neuer Informationen zu erweitern, sondern aus dem Vorhandenen etwas Neues zu schöpfen. Paradigmatisch kommt dieses Bestreben beispielsweise im Data Mining zum Vorschein, dessen Ziel Cios et al. wie folgt beschreiben: »The aim of data mining is to *make sense of large amounts of mostly unsupervised data, in some domain*« (Cios 2010: 3). Das Data Mining, welches auch als »*knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing*« (Cios 2010: 10) bekannt sei, macht es sich zur Aufgabe, in einer großen Menge vorhandener aber weithin unstrukturierter Daten sinnvolle Zusammenhänge zu entdecken. Dabei werden unter anderem Klassifikations-, Regressions- und Clustering-Verfahren angewendet, um Muster zu entdecken, die so der Interpretation durch einen Analysten zugänglich werden.⁵⁴

53 | Angesprochen ist hier der Aspekt der *Knowledge Discovery* in Datenbanken. Hierauf wird im Kapitel »Phänomeno-Logik« (S. 307ff.) näher eingegangen.

54 | Eine besondere Form der Mustererkennung ist die negative Rasterfahndung, welche Gugerli zufolge Ende der 1970er Jahre beim Bundeskriminalamt angewandt wurde, um Mitglieder der RAF ausfindig zu machen. Diese Terroristen »zeichneten fahndungstechnisch durch jene fatale Merkmalslosigkeit aus, die auch das

Luhmanns Zettelkasten: Datenbankkommunikation systemtheoretisch betrachtet

Auf der Idee, nicht nur eigentlich schon Bekanntes, sondern Neues oder zumindest Überraschendes aus einer Datenbank zu schöpfen, basiert auch Luhmanns systemtheoretische Beschreibung der Kommunikation mit seinem Zettelkasten (vgl. Luhmann 1981). In Luhmanns Erfahrungsbericht findet sich das derzeit vielleicht bekannteste Modell der Kommunikation mit einer Informationssammlung. Dies verdankt sich der Berühmtheit des Zettelkastens sowie seines Betreibers. Für die hier verfolgte Argumentation sind Luhmanns Ausführungen von Interesse, da er sich mit der Frage auseinandersetzt, ob man in Bezug auf die Interaktion mit einer Zettelsammlung überhaupt von Kommunikation sprechen darf.⁵⁵ Damit adressiert Luhmann ein Problem, das in den bisher skizzierten Modellierungen der Datenbankkommunikation stets außen vor geblieben ist. Mooers diskutiert beispielsweise nicht, inwiefern es gerechtfertigt ist, das Information Retrieval als einen Kommunikationsprozess zu beschreiben. Daher könnte der Einwand erhoben werden, dass es als Modell unzulänglich ist, um Kommunikation zu beschreiben. Bei Mooers' Modell des Information Retrieval handelt es sich hingegen im gleichen Sinne um ein Kommunikationsmodell wie bei Shannons mathematischer Theorie der Kommunikation. Beide Ansätze sind ungenügend, um das gesamte Spektrum kommunikativen Handelns zu beschreiben, da es fokussierte Modellierungen kommunikativer Situationen vor dem Hintergrund spezifischer Gebrauchs- und Problemkontexte sind. Die Bedingungen erfolgreicher Kommunikation, die ihnen eingeschrieben sind, lassen sich demzufolge als notwendige Voraussetzungen für das Zustandekommen von bestimmten Formen von Kommunikation begreifen, stellen aber keine hinreichenden Bedingungen für Kommunikation dar.

Luhmanns Frage, ob und unter welchen Bedingungen die Interaktion mit einer Informationssammlung Kommunikation darstellt, wird durch die diskutierten Modelle noch nicht beantwortet. Voraussetzung für das Zustandekommen von Kommunikation im Sinne der Systemtheorie, d.h. für die Etablierung eines Kommunikationssystems, ist die Einheit der dreifachen Selektion von Information, Mit-

Schreckgespenst der bürgerlichen Kriminalistik, die Figur des Gauners, geprägt hatte – sie waren nicht nur als Biedermann und Gattinnen verkleidet, sie fuhren inzwischen auch Wagen, die als Dubletten gar nicht gesucht werden konnten, und verfügten über konspirative Wohnungen, deren ostentative Normalität nur schwerlich zu charakterisieren war. [...] Das BKA konnte also nicht nach einem bestimmten Eintrag in seinen verschiedenen Dateien suchen, sondern mußte eine bedeutungsschwere Lücke im Adressraum seiner Dateien orten« (Gugerli 2009: 63).

55 | Der Bericht ist noch in einer weiteren Hinsicht bemerkenswert. Bei Luhmann findet sich das Modell der Datenbankkommunikation praktisch im Zettelkasten, dessen Funktionsweise der Autor systemtheoretisch reflektiert.

teilung und Verstehen (vgl. Luhmann 1999: 203).⁵⁶ In Bezug auf eine mögliche Kommunikation mit Zettelkästen erscheint Luhmann einzig der Aspekt der Selektion von Information erklärbungsbedürftig.⁵⁷ Die Unterscheidungen von Information, Mitteilung und Verstehen sind innersystemisch und müssen nicht auf beiden Seiten der Kommunikation vollzogen werden. Wichtig ist nur, dass der Betreiber des Zettelkastens diese Unterscheidungen trifft und so den »Zettelkasten als Kommunikationspartner« (Luhmann 1981: 222) begreift. Voraussetzung hierfür ist, »daß die Partner sich wechselseitig überraschen können« (Luhmann 1981: 222). Infolgedessen kann der Zettelkasten in seinem Gegenüber Informationen erzeugen. Hierbei ist entscheidend, dass Information im Rahmen der luhmannschen Systemtheorie nicht als eine reifizierte Entität begriffen wird, die gespeichert und abgerufen werden kann, sondern als der innersystemische Effekt einer »Selektion aus einem (bekannten oder unbekannten) Repertoire von Möglichkeiten« (Luhmann 1999: 195).⁵⁸

56 | Konitzer erläutert in seiner Rekonstruktion von Luhmanns systemtheoretischen Kommunikationskonzept die drei Selektionen mit folgendem Beispiel: »Stellt man sich einen Mann vor, der sehr eilig seines Weges geht, so kann man sein Verhalten als Anzeichen dafür nehmen, daß er dringend irgendwohin will – aber er teilt das eben nicht unbedingt mit. Geht er aber schnell, und ein anderer versteht die Eile als Demonstration von Unansprechbarkeit, so sind in seinem Verstehen schon die ersten beiden Differenzen, die von *Information* und *Mitteilung*, gegeben. Die drei Unterscheidungen sind also in diesem Beispiel: Einmal, daß er so und so geht – das ist die *Information*. Dann: daß er nicht ›einfach so‹ geht, sondern daß diese *Handlung* ›etwas besagt‹, daß sich also *Information* und *Mitteilung* voneinander unterscheiden lassen. Und schließlich: daß einer da ist, der diesen Unterschied beobachtet, und vice versa« (Konitzer 2006: 325).

57 | Die Aspekte von Mitteilung und Verstehen werden in *Kommunikation mit Zettelkästen* nicht erwähnt. Luhmann diskutiert einzig das Vermögen des Zettelkastens, verstanden als Alter Ego im Ego, Information zu erzeugen (Luhmann 1981: 222).

58 | Es ist anzumerken, dass sich Luhmann bei der Formulierung seines Informationsbegriffs explizit auf die Informationstheorie Shannons beruft (vgl. Luhmann 1999: 195). Dies verbindet er jedoch mit einer grundlegenden Kritik an dem Konzept der Übertragung, auf dem Shannons Kommunikationsmodell fußt: »Die Übertragungsmetapher ist unbrauchbar, weil sie zuviel Ontologie impliziert. Sie suggeriert, dass der Absender etwas übergibt, was der Empfänger erhält. [...] Die Übertragungsmetapher legt das Wesentliche der Kommunikation in den Akt der Übertragung, in die Mitteilung. Sie lenkt die Aufmerksamkeit und die Geschicklichkeitsanforderungen auf den Mitteilenden. [...] Ferner übertreibt die Metapher die Identität dessen, was ›übertragen‹ wird« (Luhmann 1999: 193f.). Indem Luhmann auf Grundlage dieser Kritik ein alternatives Kommunikationsmodell entwickelt, verschiebt sich der Ort der Information. Wird sie in Shannons Modell der

Luhmann ist es möglich, mit seinem Zettelkasten in Kommunikation zu treten, weil er als Autor respektive Betreiber von ihm überrascht werden kann. Dies ist dann der Fall, wenn der Zettelkasten über ausreichend Eigenkomplexität verfügt. Ist dies nicht gegeben, dient er nur als Nachrichtencontainer, aber nicht als Kommunikationspartner: »Der Zettelkasten braucht einige Jahre, um genügend kritische Masse zu gewinnen. Bis dahin arbeitet er nur als Behälter, aus dem man das herausholt, was man hineingetan hat« (Luhmann 1981: 226). Doch durch Zeit allein entwickelt kein Zettelkasten Komplexität. Diese entsteht vielmehr durch die interne Funktionslogik, gemäß der der Zettelkasten betrieben wird. Von zentraler Bedeutung ist für Luhmann die Aufgabe der thematischen Ordnung der Notizen, die durch die feste Adressierung der Zettel ersetzt wird:

»Für das Innere des Zettelkastens, für das Arrangement der Notizen, für sein geistiges Leben ist entscheidend, daß man sich gegen eine systematische Ordnung nach Themen und Unterthemen und statt dessen für eine feste Stellordnung entscheidet. [...] Diese Strukturentscheidung ist diejenige Reduktion der Komplexität möglicher Arrangements, die den Aufbau hoher Komplexität im Zettelkasten und damit seine Kommunikationsfähigkeit erst ermöglicht.« (Luhmann 1981: 223f.)

Durch die Entscheidung für eine unthematische Stellordnung wird zunächst der Frage aus dem Weg gegangen, wo neue Zettel einzuordnen sind. Dies stellt in thematisch geführten Zettelkästen insbesondere immer dann ein Problem dar, wenn einem Zettel mehr als nur ein thematischer Ort zugewiesen werden kann. Erlaubt das Fehlen einer thematischen Ordnung im Zettelkasten einerseits dessen beliebige Erweiterbarkeit, wird der thematische Zugriff auf den Kasten hierdurch erschwert, was es, so Luhmann, notwendig macht, ein Register zu führen. Schließlich ermöglicht die feste Adressierung den Aufbau einer internen Verweisstruktur, wodurch im Zettelkasten ein komplexes Netzwerk thematischer Querverweise bzw. Links zwischen einzelnen Zetteln entstehen kann (vgl. Luhmann 1981: 224).⁵⁹

Es gilt abschließend hervorzuheben, dass es sich hierbei um die Beschreibung einer analogen Datenbank handelt. Denn die Struktur des Zettelkastens lässt sich,

Nachricht zugeschrieben, ist Information im systemtheoretischen Verständnis ein »innersystemisches Ereignis« (Luhmann 1981: 222). Dass dies eine nicht unerhebliche Transformation von Shannons Informationskonzept darstellt, macht Luhmann jedoch nicht deutlich.

59 | Der Verweisstruktur zwischen den einzelnen Zetteln wird von Luhmann ein zentraler Stellenwert beigemessen: »Jede Notiz ist nur ein Element, das seine Qualität erst aus dem Netz der Verweisungen und Rückverweisungen im System enthält. Eine Notiz, die an dieses Netz nicht angeschlossen ist, geht im Zettelkasten verloren, wird vom Zettelkasten vergessen. Ihre Wiederentdeckung ist auf Zufälle angewiesen und auch darauf, daß dieser Wiederfund im Moment des Vorfalls zufällig etwas besagt« (Luhmann 1981: 225).

wie Luhmann selbst unterstrichen hat, keinesfalls generalisieren. Es handelt sich vielmehr um ein Beispiel dafür, wie ein Zettelkasten angelegt sein kann, »damit er entsprechende kommunikative Kompetenz erwirbt« (Luhmann 1981: 223). Demzufolge lassen sich aus der Beschreibung des Zettelkastens keine generellen Aussagen über eine technische Funktionslogik von Datenbanken ableiten.

DATEN UND INFORMATION: BEGRIFFSKLÄRUNG

In den bisherigen Kapiteln blieb der Begriff *Information* im philosophisch-terminologischen Sinne unthematisch.⁶⁰ Anstatt diesen zu definieren, wurde in einer dem Wunsch nach terminologischer Präzision entgegenlaufenden Bewegung gezeigt, dass sich mit dem Aufkommen der sogenannten Informationstechnologien und der Entwicklung von digitalen Datenbanken die Bedeutungen des Informationsbegriffs vervielfachten. Informationen sind nicht nur das, was durch eine mediale Konfiguration vermittelt wird, sondern auf einer konzeptuellen Ebene auch ihr Effekt. Was Informationen sind erschließt sich dem medienanalytischen Blick bei der Betrachtung partikularer informationeller Praktiken, wie zum Beispiel bei der Analyse von konkreten Informationssystemen.⁶¹ Eine solche Perspektive rückt die Pluralität und Heterogenität all dessen in den Vordergrund, was gleichermaßen als Information bezeichnet wird. Zugleich wurde das Verhältnis von Daten zu Information bislang nicht diskutiert und beide Begriffe weitgehend undifferenziert gebraucht.

In Anbetracht dessen wird die medienphilosophische Reflexion über den Informationsbegriff nicht obsolet, sondern geradezu vordringlich. Es bedarf eines Vokabulars, um die Differenzen zwischen den unterschiedlichen in Informationssystemen operativ werdenden Formen von Information zu fassen und die Begriffe *Daten* und *Information* zueinander ins Verhältnis zu setzen. Wertvolle Anknüpfungspunkte finden sich in der vom italienischen Philosophen Luciano

60 | In seiner Diskussion der husserlischen Phänomenologie hat Eugen Fink 1957 die Unterscheidung zwischen *thematischen* und *operativen* Begriffen eingeführt. Während thematische Begriffe Gegenstand des Denkens sind und reflektiert werden, sind die unthematisch bleibenden operativen Begriffe die blinden Flecken einer Philosophie. Es sind diejenigen Begriffe, die die philosophische Reflexion ermöglichen, ohne selbst Gegenstand einer solchen Reflexion zu werden (vgl. Fink 1957: 324f.). Folgt man dieser Unterscheidung Finks, dann wurden die Begriffe Daten und Information bisher operativ gebraucht.

61 | Datenbanken im engen Sinn von Informationen, die mit DBMS verwaltetet werden, und im weiten Sinn von digitalen Informationssammlungen sind nur Aspekte hiervon. Der algorithmischen Generierung von Informationen in Simulationen sind beispielsweise andere Konzepte von Daten und Information inhärent.

Floridi propagierten *Philosophy of Information*.⁶² Eine vordringliche Aufgabe der philosophischen Hinwendung zu Information ist Floridi zufolge die Reflexion der konzeptuellen, kulturellen und ethischen Grundlagen der sich rasant entwickelnden Informationsgesellschaften, wodurch zu einem besseren Verständnis des Informationszeitalters beigetragen werden soll.⁶³ Darüber hinaus erhebt Floridi den weiterreichenden Anspruch, zentrale philosophische Fragen informationstheoretisch zu reformulieren und hierdurch lösen zu können.⁶⁴

Bevor auf Floridis Informationsphilosophie näher eingegangen werden soll, gilt es auf eine maßgebliche Schwäche seines Ansatzes hinzuweisen. Kennzeichnend für Floridis Zugang ist die nahezu vollständige Vernachlässigung der medialen Bedingtheit informationeller Praktiken. Wie aus einer seiner wenigen Äußerungen zu Medien deutlich wird, wird Information von ihm medienunabhängig gedacht:

»We all know that the actual *format*, *medium*, and *language* in which semantic information is encoded is often irrelevant and hence disregardable. So it is trivial to acknowledge that the same semantic information may be analogue or digital, printed on paper or viewed on a screen, in English or in some other language, expressed in words or pictures.« (Floridi 2011: 86)

62 | Seit Ende der 1990er Jahre hat sich Floridi in zahlreichen Publikationen unterschiedlichen Aspekten der Philosophie der Information zugewendet. Die Ergebnisse seiner Forschungen sind in der 2011 erschienenen Monographie *The Philosophy of Information* zusammengefasst.

63 | Um die Dringlichkeit einer philosophischen Auseinandersetzung mit den Herausforderungen der Informationsgesellschaft zu unterstreichen, bemüht Floridi das Bild eines Baums ohne Wurzeln: »The information society is like a tree that has grown its far-reaching branches much more widely, hastily, and chaotically than its conceptual, ethical, and cultural roots. [...] As a consequence, today, any advanced information society faces the pressing task of equipping itself with a viable philosophy of information. Applying the previous analogy, while technology keeps growing bottom-up, it is high time we start digging deeper, top-down, in order to expand our understanding of our information age, of its nature, of its less visible implications, and of its impact on human and environmental welfare and thus give ourselves a chance to anticipate difficulties, identify opportunities, and resolve problems« (Floridi 2010: 8).

64 | Der doppelte Anspruch, den Floridi mit der Philosophie der Information einzulösen versucht, spiegelt sich in seiner Definition des Forschungsfelds wider: »The philosophy of information (PI) is the philosophical field concerned with (a) the critical investigation of the conceptual nature and basic principles of information, including its dynamics, utilization, and sciences; and (b) the elaboration and application of information-theoretic and computational methodologies to philosophical problems« (Floridi 2011: 14).

Nach Ansicht Floridis kann der Aspekt des Medialen vernachlässigt werden, weshalb es wenig verwunderlich ist, dass er in seinen Texten weder auf die vor allem im deutschsprachigen Raum virulente medienphilosophische Debatte Bezug nimmt, noch einen Begriff des Mediums ausarbeitet. Medien stehen für Floridi in einer Reihe mit Sprachen und Formaten, die gleichermaßen nicht berücksichtigt werden müssen.⁶⁵ Das Argument, mit dem er dies zu legitimieren versucht, ist die Indifferenz von Information gegenüber medialen, formalen und sprachlichen Differenzen, d.h. dieselbe Information kann in unterschiedlichen Medien auf verschiedene Weise verkörpert werden. Obwohl stets auf eine materielle Verkörperung angewiesen, ist Information, so Floridi, losgelöst von dieser zu betrachten (vgl. Floridi 2011: 90f.). Vor dem Hintergrund dieser Überlegung wird Floridis Prognose verständlich, dass der Bedeutungszuwachs, den Information in Informationsgesellschaften erfährt, zu einem Bedeutungsverlust der Physis (Material, Physik) führt (vgl. Floridi 2011: 8).⁶⁶

Während die Medienunabhängigkeit für Floridi selbstevident erscheint und deshalb keiner weiteren Erläuterung bedarf, hinterfragt die Medienforschung diese Annahme, wie in Wiesings phänomenologischer Erörterung des Medienbegriffs besonders offenkundig wird.⁶⁷ Als Werkzeuge zur Trennung von Genesis und Geltung ermöglichen Medien artifizielle Selbigekeit. Die Selbigekeit medialer Konstellationen ist folglich Effekt von Medien. Welche Unterschiede zwischen medialen Konstellationen einen Unterschied machen und welche nicht lässt sich nicht losgelöst von ihrer medialen Konfiguration beantworten. Infolgedessen können die in Medien verkörperten Informationen nicht losgelöst von ihrer medialen Verfasstheit betrachtet werden. Dies gilt insbesondere auch dann, wenn es, wie Floridi in der eingangs zitierten Passage behauptet, aus informationstheoretischer Sicht keinen Unterschied macht, wie schriftsprachlich verfasste Informationen verkörpert sind (analog oder digital) bzw. rezipiert werden (auf Papier oder am Bildschirm).⁶⁸ Es mag sich hierbei um dieselben Informationen handeln, aber *dass* es

65 | Über den Medienbegriff Floridis lässt sich nur spekulieren, doch die Gegenüberstellung von Medien, Sprachen und Formaten legt die Vermutung nahe, dass er Medien als Träger von Information begreift.

66 | Floridi folgt dem Paradigma der Kommunikationsforschung, die, wie Schüttelpelz in *»Get the Message Through«* (2002b) dargelegt hat, die Gemeinsamkeiten der Kommunikation jenseits aller medialen Unterschiede in den Vordergrund rückt. Demgegenüber treten Medien als genuiner Untersuchungsgegenstand in den Vordergrund, wenn man den Blick von der »Isomorphie und Kongruenz aller Kommunikation und Kommunikationstechniken« (Schüttelpelz 2002b: 64) ab- und den Unterschieden zwischen diesen zuwendet.

67 | An dieser Stelle soll Wiesings Medienbegriff nicht erneut dargelegt werden. Vielmehr sei auf die ausführliche Diskussion von dessen Position im Kapitel »Medium« (S. 44ff.) verwiesen.

68 | Es gilt anzumerken, dass diese These nicht nur medientheoretisch problematisch ist, sondern auch in den Debatten zur Informationsphilosophie umstritten ist.

dieselben Informationen sind, wird bedingt durch die in und mit Medien vollzogene Trennung von Genesis und Geltung. Information ist Medien gegenüber nicht autonom, sondern verdankt ihre Autonomie den Medien, weshalb sich unter den Bedingungen medialer Umbrüche die Kriterien der Selbigekeit von Informationen verändern und verschieben können.

Medientheorie eröffnet hierbei den Blick auf die Kontingenz der Unterscheidung zwischen Demselben und Unterschiedlichem.⁶⁹ So hat Matthew Kirschenbaum beispielsweise auf die Herausforderungen hingewiesen, welche neue Formen elektronischer Literatur für die historischen Textwissenschaften darstellen (vgl. Kirschenbaum 2008: 27f.). Was in archivarischer und philologischer Hinsicht denselben digitalen Text darstellt, ist seinen Ausführungen zufolge bisher keineswegs ausgemacht. Verfällt man einem Bildschirmessentialismus, dann läuft man Gefahr, digitale Literatur am Modell des Buchs zu behandeln und sie auf dieses Modell zu reduzieren.⁷⁰ Dagegen spricht sich Kirschenbaum dezidiert aus, indem er in exemplarischen, zwischen der Bildschirmoberfläche und unterschiedlichen Codeebenen hin- und herwechselnden Analysen digitaler Medienobjekte Bedeutungsebenen freilegt, die dem ausschließlich auf dem Bildschirm verharrenden Blick verborgen bleiben.⁷¹

Dass Medien, wie von Floridi konstatiert, für die philosophische Analyse von Information keinen Unterschied machen, ist damit grundlegend infrage zu stellen. Dennoch erweist sich das von Floridi vorgeschlagene informationstheoretische Vokabular für die medientheoretische Analyse von Datenbanken als anschlussfähig. Insbesondere erlaubt es seine Unterscheidung verschiedener Typen von Information, unterschiedliche Niveaus freizulegen, auf denen Computer mit Daten respektive Informationen operieren. Diesem Aspekt wird sich das folgende Teilkapitel widmen. Darauf aufbauend werden unterschiedliche Bedeutungsvarianten

Eine Gegenposition zu Floridi nimmt beispielsweise Cornelius ein, der sich für die Einbeziehung sozialer Aspekte bei der Konstruktion von Bedeutung sowie für die systematische Betrachtung der Empfängerseite von Informationen stark macht (vgl. Cornelius 2004: 385). Diesem Einwand zufolge sind situative Faktoren der Rezeption, wie z.B. die durch die Art der Verkörperung von Information bedingte Rezeptionssituation, nicht zu vernachlässigen.

69 | Welche Informationen beispielsweise eine Fotografie enthält und ob diese mit den Informationen übereinstimmen, die eine andere Fotografie beinhaltet (oder nicht), wird auf Grundlage anderer Kriterien entschieden als bei Büchern.

70 | Geprägt wurde der Begriff des *screen essentialism* von Nick Montfort, der mit diesem kritisiert, dass viele Literaturwissenschaftler bei ihrer Auseinandersetzung mit elektronischer Literatur allein die Oberfläche thematisieren (vgl. Montfort 2004).

71 | In *Mechanisms* analysiert Kirschenbaum (2008) das Spiel *Mystery House* aus dem Jahr 1980 sowie das gemeinsame Kunstprojekt des Verlegers Kevin Begos Jr., des Künstlers Dennis Ashbaugh und des Science Fiction-Autors William Gibson, das diese 1992 unter dem Titel *Agrippa: A Book of the Dead* veröffentlichten.

des Datenbegriffs rekonstruiert und deren jeweiliges Verhältnis zu Information diskutiert.⁷² Abschließend wird das Augenmerk auf die informationelle Umwelt digitaler Informationen gelegt.

Information als, über und für Realität

Es sei, so Bernd-Olaf Küppers, nicht untypisch, »daß bei der Grundlegung einer neuen Theorie zunächst immer von unscharfen, eher intuitiv gebildeten Begriffen ausgegangen wird und diese dann in einem Iterationsprozeß zwischen Begriffs- und Theorienbildung sukzessiv verschärft werden« (Küppers 1999: 32). Ein Versuch, der Komplexität des Informationsbegriffs beizukommen, ist die Anwendung der semiotischen Dreiteilung in Syntax, Semantik und Pragmatik auf den Informationsbegriff. Diese auf Charles William Morris (vgl. 1979 [1938]: 24) zurückgehende Unterscheidung von drei Zeichenebenen stellt nach Ansicht von Holger Lyre die »wohl wichtigste Charakterisierung des Informationsbegriffs« (Lyre 2002: 16) dar, die es erlaubt, zwischen Informationseinheiten sowie deren Relationen zueinander (Syntax), deren Bedeutung (Semantik) und ihrer Wirkung (Pragmatik) analytisch zu unterscheiden.⁷³ Hierdurch werde es möglich, spezielle Aspekte von Information zu fokussieren. Für eine »Gesamtcharakterisierung« (Lyre 2002: 16) des Informationsbegriffs greift dies jedoch zu kurz. Werden nur einzelne Ebenen betrachtet, führt dies zu einer verkürzten Sichtweise auf Information, was dem von Lyre erhobenen Anspruch auf Vollständigkeit zuwider laufen würde (vgl. Lyre 2002: 17). Im Streben nach einer vollständigen Analyse des Informationsbegriffs kommt der Glaube an einen Einheitsbegriff der Information zum Ausdruck. Dieser Glaube liegt auch Küppers' Diagnose zugrunde, der in der Unschärfe des Informationsbegriffs ein bloßes Durchgangsstadium hin zu einem theoretisch fundierten einheitlichen Verständnis von Information sieht.

Floridi zeigt sich gegenüber der Möglichkeit, einen Einheitsbegriff von Information zu formulieren, skeptisch. Zwar steht seines Erachtens eine detaillierte philosophische Erörterung der Frage, ob eine »Grand Unified Theory« (Floridi 2011: 33) der Information möglich ist, noch aus, aber dennoch erscheint es ihm wahrscheinlicher, »that we are [...] facing a network of logically interdependent, but mutually irreducible, concepts« (Floridi 2011: 33). Folglich stellt er sich gegen reduktionistische Ansätze, die danach streben, das Wesen von Information freizulegen. Diese Grundeinstellung zieht sich wie ein roter Faden durch das Werk

72 | Bedeutsam ist dies weniger, weil der Informationsbegriff, wie Rafael Capurro herausgestellt hat, in der Informatik »zunächst allgemein und undefiniert als Synonym von Daten gebraucht« (Capurro 1978: 232) wurde, sondern weil beide Begriffe in Bezug auf Computer weiterhin mehrdeutig gebraucht werden.

73 | Die Unterscheidung zwischen Syntaktik, Semantik und Pragmatik wurde erstmals von Warren Weaver (vgl. 1976 [1949]: 35) auf Information angewandt; siehe hierzu Ott (2004: 32).

Floridis. Sie drückt sich in dem fortwährenden Versuch aus, Unterscheidungen zwischen Perspektiven auf und Typen von Information einzuführen. Hierdurch erkennt Floridi nicht nur die Heterogenität unterschiedlicher Gebrauchsweisen des Begriffs an, sondern auch die Legitimität verschiedener theoretischer Zugänge zu Information. Dies spiegelt sich in seinem Vorschlag wider, Informationstheorien danach zu unterscheiden, auf welchen Typ von Information sie sich beziehen. Er grenzt drei Arten von Information voneinander ab, die sich durch ein je unterschiedliches Verhältnis zur Realität auszeichnen: Information *als* Realität, Information *über* Realität sowie Information *für* Realität (vgl. Floridi 2011: 30).⁷⁴

Theorien, die Information *als* Realität thematisieren, interessieren sich für den Formgehalt oder die Struktur von Dingen, wie z.B. Stühle oder Signalfolgen. Diese Perspektive bewegt sich nahe an der Bedeutung des lateinischen Wortes *informare* (formen, gestalten, herausbilden, darstellen, bilden), von dem der Terminus Information abgeleitet ist.⁷⁵ Die Realität beinhaltet diesem Verständnis zufolge Information, welche es zu beschreiben gilt. Auch die syntaktischen Strukturen sprachlicher Äußerungen lassen sich in dieser Hinsicht als Information begreifen. Insofern ist Shannons *Mathematische Theorie der Kommunikation* eine Theorie der Information *als* Realität. Die Betrachtung von Information auf diesem Niveau erlaubt es Shannon, von der Bedeutung von Nachrichten gänzlich abzusehen.

Die Bedeutungsdimension tritt in denjenigen Ansätzen in den Vordergrund, die Information als Information *über* Realität behandeln. Hierbei tritt der Form- respektive Strukturaspekt in den Hintergrund. Zwar setzt Information *über* Realität eine Verkörperung als mediale Konstellation voraus, die gleichsam als Information *als* Realität behandelt werden kann, aber sie lässt sich nicht auf diese zurückführen. Information *über* Realität bringt in der Realität etwas in eine Form, um über etwas anderes zu informieren. Damit rückt nicht nur die Ebene der Bedeutung ins Zentrum des theoretischen Interesses, sondern auch die Frage danach, ob die Sinn-dimension ausreicht, um Information *über* Realität hinreichend zu charakterisieren. Floridi zweifelt dies an, da es seines Erachtens nicht hinreichend ist, Information *über* Realität als wohlgeformte und bedeutungstragende Daten zu definieren. Er plädiert daher für die Erweiterung der Definition um ein Wahrheitskriterium. Nicht jeder sinnvolle Satz enthält, gemäß seiner Kernthese, Informationen *über* Realität, sondern nur solche, die auch wahr sind (vgl. Floridi 2011: 80ff.).⁷⁶ An dieser Stelle

74 | Bereits 1999 hat Albert Borgmann in *Holding on to Reality* diese Unterscheidung eingeführt, auf den Floridi jedoch keinen Bezug nimmt (vgl. Borgmann 1999: 1ff.).

75 | In seiner Studie zum Informationsbegriff arbeitet Capurro die ideengeschichtlichen Bezüge des lateinischen Verbs *informare* sowie der hiervon abgeleiteten Substantivierung *informatio* zu den griechischen Begriffen *typos*, *morphe*, *eidos* und *idea* heraus (vgl. Capurro 1978: 17ff.).

76 | Fiktionen sind im Anschluss an Floridi nicht sinnlos oder gar unsinnig, sondern nur nicht informativ.

ist es unerheblich, ob Floridis These einer kritischen Erörterung standhält.⁷⁷ Es ist vielmehr bedeutsam, dass die Thematisierung von Information *über* Realität zu anderen theoretischen Fragestellungen führt, als die Betrachtung von Information *als* Realität.

Bei Information *für* Realität tritt schließlich der Wirkungsaspekt in den Vordergrund. Es ist Information, welche die Realität verändert oder eine neue Realität hervorbringt. Auch Information *für* Realität bedarf der Verkörperung von Information *als* Realität, jedoch informiert diese nicht vorrangig *über* etwas, sondern *zu* etwas. Hierbei ist es unwichtig, ob diese Information an Menschen, CPUs oder Zellen gerichtet ist. Information *für* Realität dient als Bauanleitung oder als Programm, welches abgearbeitet werden muss, um zu einem bestimmten Ergebnis zu gelangen. Der Bedeutungsaspekt geht somit im Wirkungsaspekt auf, da die Bedeutung von Information *für* Realität ihre Wirkung ist.

An der Darstellung der drei Typen von Information respektive Informationstheorien, die Floridi unterscheidet, ist deutlich geworden, dass sie weitreichende Parallelen zur Binnendifferenzierung von Information entlang der drei Zeichenebenen Syntax, Semantik und Pragmatik aufweisen. Diese Ähnlichkeiten sind jedoch nur vordergründig und dürfen nicht über den tiefgreifenden konzeptionellen Wandel hinwegtäuschen, für den sich Floridi stark macht. Syntax, Semantik und Pragmatik dienen Floridi gerade nicht dazu, um unterschiedliche Aspekte freizulegen, die Informationen stets innwohnen, sondern der Differenzierung verschiedener Typen von Information, in denen sich die Frage nach Syntax, Semantik und Pragmatik auf eine je eigene Weise stellt und jeweils unterschiedlich zu beantworten ist. Damit tritt die Pluralität unserer informationellen Umwelt in den Vordergrund und die Frage nach der Vermittlung und der Übersetzung zwischen unterschiedlichen Typen von Information. Wer sich beispielsweise mit Information *über* Realität befasst, ist notgedrungen auch mit Fragen der Syntaktik beschäftigt. Die Antworten hierauf sind jedoch andere als bei der Auseinandersetzung mit Information *als* Realität. Das bedeutet nichts anderes, als dass es einen irreduziblen Unterschied zwischen Shannons Informationstheorie und beispielsweise einer Grammatik der deutschen Sprache gibt. Ebenso wird mit dieser Unterscheidung nicht angezweifelt, dass Information *über* Realität eine praktische Wirkung haben kann. Die Konsequenzen, die Informationen *über* Realität zeitigen, sind diesen nicht inhärent und infolgedessen kann »[d]er pragmatische Informationsgehalt eines Zeichens [...] für zwei Empfänger ganz unterschiedlich sein, während der semantische annähernd identisch ist« (Ott 2004: 33).

Die Anerkennung der Pluralität und Heterogenität von Information, informationellen Praktiken und Informationstheorien entlastet philosophisch von der Notwendigkeit, die Aspekte der Syntax, Semantik und Pragmatik einheitlich

77 | Es ist umstritten, inwiefern das von Floridi vorgebrachte Wahrheitskriterium für die Konzeptualisierung semantischer Information zentral ist; siehe hierzu exemplarisch Pimiero (2009) und Cornelius (2004).

in einem Konzept von Information aufeinander zu beziehen. Ein solcher Versuch birgt nicht zuletzt die Gefahr, dass nur ein Aspekt in den Vordergrund gerückt wird und die anderen als sekundäre Effekte betrachtet werden.⁷⁸ Zugleich macht es dieser Zugang notwendig, informationelle Praktiken daraufhin zu befragen, auf welche Weise mit welcher Art von Information umgegangen wird. Dies erweist sich als grundlegender Vorteil der externen Unterscheidung von Informationstypen für eine medientheoretische Analyse von Datenbanken, da es hierdurch möglich wird, Unterschiede überall dort sichtbar zu machen, wo nur scheinbar unterschiedslos mit Information operiert wird. Infolgedessen gilt es, sich von der Vorstellung zu verabschieden, dass digitale Medienobjekte Information in einem singulären Sinn beinhalten oder verkörpern. In ihnen überlagern sich unterschiedliche Typen von Information, so dass jede mediale Konstellation stets Information im Plural ist bzw. beinhaltet. Betrachtet man die informationellen Praktiken im Kontext digitaler Medientechnologien, werden hierdurch die Spannungen offenkundig, die überall dort entstehen, wo mit derselben medialen Konstellation auf verschiedenen Ebenen als Information umgegangen wird, sowie dort, wo Vermittlungen zwischen den unterschiedlichen Ebenen vollzogen werden.

Daten zwischen Information als, über und für Realität

Die Heterogenität von Information zeigt sich auch in dem mehrdeutigen Gebrauch des Begriffs Daten und dem unbestimmten Verhältnis von Daten zu Information. Werden die Begriffe auf der einen Seite in Kontrast zueinander gestellt, erweisen sich Daten in einer anderen Perspektive als Information *sui generis*. Ziel der Auseinandersetzung mit dem Datenbegriff ist nicht, dessen Vieldeutigkeit aufzulösen und durch einen exakt definierten Begriff zu ersetzen. Vielmehr soll die Ambiguität des Begriffs dargelegt werden, die sich in der digitalen Medienpraxis im Allgemeinen und der Datenbankpraxis im Besonderen zeigt. Mindestens fünf Bedeutungsvarianten des Begriffs *Daten* lassen sich unterscheiden, die in der digitalen Medienkultur bedeutsam sind.

Daten als Voraussetzung respektive Vorstufe von Information

In wissenschafts- und informationstheoretischen Erörterungen werden Daten häufig als Vorstufe von Information betrachtet und somit in Differenz zu dieser gestellt. Eine solche Position vertritt beispielsweise Floridi, für den Daten noch keine Information (*über* Realität) konstituieren, sondern eine notwendige Voraus-

78 | Diesem Problem entkommt auch Lyre nicht. Um die Ebene der Semantik theoretisch handhabbar zu machen, argumentiert er dafür, sie mit der Pragmatik in einem Konzept von Semantopragmatik zu verschränken (vgl. Lyre 2002: 20). Hierdurch wird Semantik jedoch auf Pragmatik zurückgeführt.

setzung für Information sind.⁷⁹ Diese Auffassung leitet sich aus dem von ihm propagierten Datenbegriff ab, den er etymologisch auf das griechische Wort diafora& (diaphora) zurückführt, welches ins Deutsche übersetzt *Verschiedenheit* bedeutet. Die typischerweise von der lateinischen Wortherkunft des Begriffs abgeleitete Definition von Daten als etwas *Gegebenem* erfährt hierdurch eine differenzlogische Reformulierung: »A datum is a putative fact regarding some difference or lack of uniformity within some context« (Floridi 2005b).⁸⁰ An diese Definition anschließend lassen sich, Floridi zufolge, drei Ebenen unterscheiden, auf denen von Daten gesprochen werden kann (vgl. Floridi 2011: 85):

1. Unterschiede *de re*: Daten sind proto-epistemische Unterschiede in der Realität.
2. Unterschiede *de signo*: Daten sind Unterschiede zwischen mindestens zwei physikalischen Zuständen, d.h. Signalen.
3. Unterschiede *de dicto*: Daten sind Unterschiede zwischen mindestens zwei Symbolen.

Diese Typen von Daten differieren hinsichtlich der jeweils in Betracht gezogenen Unterschiede. Während sich Signale beispielsweise als physikalische Unterschiede manifestieren, kann von diesen materiellen Differenzen bei Symbolen abgesehen werden, da sich ein und dasselbe Symbol auf verschiedene Weisen verkörpern lässt. Floridis Differenzierung verschiedener Arten von Daten richtet sich demzufolge auf die unterschiedlichen Gegebenheitsweisen des als Unterschied Gegebenen. Insofern lassen sich Daten als Information *als* Realität begreifen.⁸¹ Wenn Floridi darauf insistiert, dass Daten keine Information sind, so ist dies vor dem Hintergrund seines Interesses für eine philosophische Analyse von Information *über* Realität zu betrachten. Der Hinweis auf die Verschiedenheit von Daten und Information markiert einen Unterschied, der als Unterschied zwischen Information *als* und *über* Realität beschrieben werden kann. Dies erklärt, warum Daten auf der einen Seite von Information unterschieden werden und auf der anderen Seite Informationen sind, wenn auch nicht auf die gleiche Weise. Die Abgrenzung von Daten gegen-

79 | Auch in der Informationswissenschaft ist eine solche Sicht verbreitet. Beispielsweise definiert Morgenroth Daten in seiner Monographie zum kontextbasierten Information Retrieval als Vorstufe von Wissen und Information (vgl. Morgenroth 2006: 5f.).

80 | Batesons berühmter Definition zufolge ist Information »ein Unterschied, der einen Unterschied macht« (Bateson 1985b: 582). Stellt man Daten in Differenz zu Information, dann müssten diese folglich als Unterschiede verstanden werden, die keinen Unterschied machen.

81 | Auch wenn Daten Information *als* Realität sind, darf ihr relationaler Charakter nicht unterschlagen werden. Insofern warnt Floridi vor der zunehmenden Substanzialisierung von Daten und damit einhergehend auch von Informationen (1999, 2005a).

über Information ist relativ und zeigt an, dass es unterschiedliche Niveaus gibt, auf denen an ein und dieselbe mediale Konstellation als Information angeschlossen werden kann. Auf etwas Ähnliches weist auch Ernst hin, wenn er feststellt: »Daten, Information und Wissen unterscheiden sich durch die Apparaturen ihrer Verarbeitung, durch die Institutionen ihrer hermeneutischen Aufladung und durch die Fähigkeit zur Selektion« (Ernst 2002: 179). Während Ernst mit dieser Bemerkung einerseits die Relativität der Unterscheidung zwischen Daten und Information (und Wissen) anerkennt, identifiziert er sie andererseits mit den Agenturen der Verarbeitung. Damit kommt ein weiterer Aspekt ins Spiel, der dem Verständnis von Daten als digitaler Repräsentation von Information zugrunde liegt.

Daten als Repräsentation von Information

In der Fachsprache der Informatik werden Daten als binäre Repräsentationen von Information begriffen. Dieses Begriffsverständnis kommt in der DIN-Norm 44 300 zum Ausdruck, in der Daten definiert sind als »Gebilde aus Zeichen oder kontinuierliche Funktionen, die aufgrund bekannter oder unterstellter Abmachungen Information darstellen, vorrangig zum Zweck der Verarbeitung oder als deren Ergebnis« (DIN 1989: 149).⁸² Daten sind keine Vorform oder Voraussetzung für Information, sondern Information in einem anderen Aggregatzustand; sie sind codierte Information und stehen demzufolge in einem Abbildungsverhältnis zu ihr.⁸³ Zweck der Unterscheidung zwischen Daten und Information ist die Differenzierung unterschiedlicher Agenturen der Informationsverarbeitung: Während Menschen in der Lage sind, mit Informationen umzugehen, vermögen Computer lediglich, Daten zu verarbeiten. Dementsprechend wird das Lemma *Daten* im *Lexikon der Informatik* von Fischer und Hofer definiert als »alles, was sich in einer für die Datenverarbeitungsanlage, den Computer, erkennbaren Weise

82 | Ein Zeichen ist in diesem Kontext definiert als »[e]in Element (als Typ) aus einer zur Darstellung von Information vereinbarten endlichen Menge von Objekten [...], auch jedes Abbild (als Exemplar) eines solchen Elements« (DIN 1989: 149). Bemerkenswert hieran ist, dass Zeichen zunächst ausschließlich als Typen beschrieben werden, von denen Exemplare (Tokens) in Gebilden oder Konfigurationen instanziert werden, die man dann Daten nennt. Eine Bedeutung im Sinne einer Zeichenrelation von Signifikant und Signifikat besitzen Zeichen entsprechend der gegebenen Definition nicht. Diese kommt der Norm zufolge erst bei Symbolen ins Spiel.

83 | Neben der Definition von Daten findet sich in der DIN-Norm auch eine Definition von Nachricht, die der Datendefinition nahezu gleicht: Eine Nachricht ist ein »Gebilde aus Zeichen oder kontinuierliche Funktionen, die aufgrund bekannter oder unterstellter Abmachungen Information darstellen und die zum Zweck der Weitergabe als zusammengehörig angesehen und deshalb als Einheit betrachtet werden« (DIN 1989: 149). Einziger Unterschied in den Definitionen ist die funktionale Bestimmung der Zeichengebilde. Während Daten der Verarbeitung durch den Computer dienen, sind Nachrichten zur Weitergabe bestimmt.

codieren, speichern und verarbeiten lässt, also abstrahierte und ›computergerecht‹ aufbereitete Informationen« (Fischer/Hofer 2008).

In dieser Deutung des Datenbegriffs kommt der Sachverhalt zum Ausdruck, dass Computer nicht unmittelbar mit natürlichsprachlich verfassten Texten, Bildern oder Filmen operieren, sondern immer nur vermittels binär codierter Repräsentationen. Die Reduktion des digitalen Codes auf die elementare Unterscheidung zwischen zwei Zuständen im Bit hat den Vorteil, dass die Unterscheidung gleichermaßen semantisch (wahr/falsch), logisch-mathematisch (1/0) sowie materiell (z.B. hohe/niedrige Spannung in einem Schaltkreis) dargestellt werden kann (vgl. Floridi 2010: 28f.). Auf Grundlage dessen sei es, so Floridi, möglich, technische Apparate zu bauen, »that can recognize bits physically, behave logically on the basis of such recognition, and therefore manipulate data in ways which we find meaningful« (Floridi 2010: 29).

Das Verständnis von Daten als Repräsentationen von Information trägt der Tatsche Rechnung, dass Computer technische Artefakte sind, die Informationen nur in binär codierter Form verarbeiten können. In dieser Hinsicht unterscheidet sich die technische Datenverarbeitung von der kognitiven Informationsverarbeitung, wobei letztere weithin unbestimmt bleibt. Dies stellt solange kein Problem dar, wie der Datenbegriff der Informatik nur herangezogen wird, um zu beschreiben, womit Computer materiell und logisch operieren. Als problematisch erweist es sich jedoch dann, wenn dieses Begriffsverständnis mit der Konzeption von Daten kurzgeschlossen wird, die aus der Betrachtung von Daten als Voraussetzung und Vorstufe von Information folgt. Dies führt zu der falschen Annahme, dass Computer Daten ausschließlich auf dem Niveau von Information *als* Realität verarbeiten. Falsch ist diese Annahme nicht erst, seitdem man sich in der Informatik »mit der steigenden Leistungsfähigkeit der Rechner hin zur Semantik und Pragmatik« (Ott 2004: 184) orientierte, wie Ott in seiner Auseinandersetzung mit dem Informationsbegriff der Informatik nachgezeichnet hat. Die Entwicklung des Computers während und nach dem Zweiten Weltkrieg kann im Gegenteil als Versuch verstanden werden, computertechnisch mit Information *über* Realität zu operieren, hierdurch Zugriff auf Realität zu erlangen und diese zu steuern. Bereits in der Frühzeit des Computers war Information *über* Realität von zentraler Bedeutung, welche zumeist in Form numerischer Daten vorlag. Die Bedeutungsvariante von Daten, auf die hierbei rekurriert wird, ist vor allem in den quantitativen Wissenschaften gebräuchlich und unterscheidet sich sowohl von Floridis diaphorischer Definition als auch der Bestimmung von Daten als digitaler Repräsentation von Information.

Daten als (numerische) Information über Realität

Ein Kapitel seiner phänomenologischen Studie zum Sammeln widmet Manfred Sommer dem Datensammeln, wobei er sich der grundlegenden Frage widmet, ob und unter welchen Bedingungen es überhaupt möglich sei, Daten zu sammeln (vgl. Sommer 2002a: 392ff.). In seiner Annäherung an dieses Problem macht er sich für eine tentative Definition von Daten als »zahlenmäßig bestimmte Fakten« (Sommer

2002a: 399) stark. Diese sind nicht von Information *über* Realität verschieden, sondern stellen einen bestimmten Typus ebenjener dar. Charakteristisch für Daten ist seines Erachtens, dass sie quantitative Informationen über Realität bereitstellen: »Ein Faktum ist ein Datum genau dann, wenn folgende Bedingungen erfüllt sind: (a) Der Satz, der auf das Faktum referiert, hat die Form $f(x,y)$, und (b) y steht für einen Zahlenwert« (Sommer 2002a: 403). Floridis These, dass Wahrheit ein zentrales Kriterium für Information (*über* Realität) darstellt, spiegelt sich in Sommers Feststellung wider, dass »ein Sachverhalt genau dann eine Tatsache sei, wenn ein wahrer Satz auf ihn referiert[t]« (Sommer 2002a: 394). Daten sind diesem Begriffsverständnis zufolge nicht unmittelbar gegeben, sondern stellen eine bestimmte Form der Bezugnahme auf Realität dar, die auf Messung beruht. In eine ähnliche Richtung weist die Definition von Daten, die sich in der *Encyclopedia of Statistical Sciences* findet:

»A datum is defined to be a fact (numerical or otherwise) from which a conclusion may be drawn such as, for example, that none of four apples have the same number of worms. A datum contains information whereas a number, adjective, or other form of description may not.« (Federer 1982: 269)

Anders als in Sommers Definition werden Daten hier jedoch nicht auf numerische Information beschränkt, sondern mit faktischen Informationen gleichgesetzt. Hierdurch erfährt der Datenbegriff Sommers zwar eine Erweiterung, jedoch keine grundlegende Umdeutung. Daten werden verstanden als faktische Information *über* Realität und sind somit unterschieden von Fiktionen respektive symbolischen Probehandlungen.⁸⁴

Ganz gleich, ob man die engere oder weitere Bedeutungsvariante zugrunde legt, in Bezug auf Computer ist in der digitalen Medienkultur häufig auch von Daten in diesem Sinn die Rede. Seit einigen Jahren wird nicht nur von einem drohenden Information Overload gesprochen, sondern auch eine *data deluge* diagnostiziert. Diese Beobachtung einer Datensintflut bezieht sich nicht nur auf binäre Repräsentationen von Informationen, sondern vor allem auch auf die zunehmende Verfügbarkeit quantitativer Daten über Realität, die mit Computern erfasst werden und gespeichert vorliegen (vgl. Borgman 2007: 6f.). Auch im Begriff der Datenbank schwingt die Bedeutungsvariante von Daten als (numerische) Information *über* Realität mit, welche beispielsweise im Kontext der institutionellen Informationsverarbeitung und des Managements von zentraler Bedeutung ist. Hier werden Datenbanken nicht in erster Linie als Sammlungen digitaler Repräsentationen von

84 | In der Forschungs- und Medienpraxis lässt sich eine Vielzahl disparater Formen von Daten als faktische Information *über* Realität unterscheiden. Dies hat, wie Carlson und Anderson darlegen, auch Konsequenzen für die Versammlung, Archivierung und Nutzbarmachung solcher Daten mittels digitaler Medien (vgl. Carlson/Anderson 2007).

Informationen begriffen, sondern als Sammlungen quantitativer ökonomischer Daten, die mithilfe des Computers verwaltet und verarbeitet werden können.

Daten als Inhalt computertechnischer Operationen

Eine weitere in der digitalen Medienkultur wichtige Bedeutungsnuance des Datenbegriffs kommt in der Unterscheidung zwischen Daten und Programmen (Algorithmen, Quellcode, etc.) zum Ausdruck. Die Gegenüberstellung von Daten einerseits und Befehlen andererseits stützt sich auf den in der Mathematik virulenten Gebrauch des Begriffs. Daten werden hier als die zur Lösung einer Aufgabe vorgegebenen Größen begriffen. Die Unterscheidung zwischen faktischen und fiktiven Daten, die dem Verständnis von Daten als (numerischen) Informationen über Realität zugrunde liegt, ist hierbei irrelevant. Daten sind vielmehr die Ausgangszustände, die gemäß vorgegebener Regeln manipuliert werden, sowie das Ergebnis dieser Manipulation.

Strukturell ist diese Konzeption von Daten im Entwurf der Turingmaschine angelegt. Indem Turing die Regeln der Manipulation von Symbolen als Maschine bezeichnet und diese vom Band unterscheidet, welches die Maschine »versorgt« (Turing 1987 [1937]: 20), führt er die Differenz zwischen Programm und Daten ein. Die auf dem Band gespeicherten Symbole sind Gegenstand der Operationen der Maschine, sie sind das, was gemäß den Regeln der Turingmaschine verarbeitet wird, und sind somit gewissermaßen der Inhalt computertechnischer Operationen. Obwohl mit der Einführung der Von-Neumann-Computerarchitektur die strukturelle Differenz zwischen Daten und Programmen aufgehoben wurde, lebt sie in der Computerpraxis fort und zeigt sich im Gebrauch des Datenbegriffs.⁸⁵ Als Gegenbegriff zu Programm, Algorithmus und Befehl sind die Daten der Ausgangs- und Zielpunkt bzw. Basis und Resultat ihrer computertechnischen Verarbeitung.⁸⁶ Auf diesem Verständnis gründet die problematische, dass Daten als Inhalte computertechnischer Verfahren medientheoretisch vernachlässigt werden können. Dass das Gegenteil der Fall ist, wurde bereits dargelegt. Doch auch wenn sich die Gegenüberstellung von Daten und Programmen in dieser Hinsicht als suggestiv erweist, markiert sie einen wichtigen, nicht zu vernachlässigenden Unterschied. Die Instruktionen, die den Computer steuern, sind nicht mit den Informationen zu wechseln, auf die Nutzer mittels des Computers Zugriff erhalten. Auch wenn die Differenz zwischen Daten und Befehlen allenfalls relativ ist und vom Standpunkt der Beobachtung bzw. dem spezifischen Gebrauch des Computers abhängt, ist die hierdurch getroffene Unterscheidung nicht zuletzt für die medientheoretische

85 | Die Von-Neumann-Architektur sieht einen gemeinsamen Speicher für Daten und Programme vor (vgl. Von Neumann 1987 [1945]: 19f.).

86 | Auf der Gegenüberstellung von Daten und Programmen basiert Lustis Definition von Daten respektive Datenbanksystemen: »Werte, die sich auf Rechnern darstellen lassen, nennt man Daten. Programme, die Massendaten verwalten, heißen Datenbanksysteme« (Lusti 1997: 3).

Analyse des Computers bedeutsam. Die Bedeutungsvariante von Daten als Inhalt computertechnischer Operationen weist zudem gewisse Parallelen zu dem Verständnis von Daten als binärer Repräsentation von Information auf. Beide sind extensional weitgehend deckungsgleich. Sie differieren jedoch hinsichtlich ihrer Intensionen, die sich in den Begriffen zeigen, von denen der Datenbegriff jeweils unterschieden wird. Während computerlesbare Daten einerseits zu Information in Differenz gesetzt werden, unterscheiden sich Daten andererseits von Programmen, die ihre Verarbeitung steuern.

Daten als digitaler Text

In der terminologisch weitesten und umgangssprachlich vielleicht verbreitetsten Bedeutungsvariante werden Daten als digitaler Text verstanden. Der Begriff Daten bezeichnet hierbei die Gesamtheit der binär codierten, maschinenlesbaren Inschriften und dient damit als Sammelbegriff für all das, was auf digitalen Datenträgern gespeichert vorliegt, d.h. sowohl Daten im vorgenannten Sinn als auch Programme.

Sommer erkennt in diesem Begriffsgebrauch eine metaphorische Ausweitung der von ihm beschriebenen Datenkonzeption. Die Tatsachen, auf die mit Daten im Sinn von zahlenmäßig verfassten Fakten Bezug genommen wird, basieren auf Messungen und können nur mittels Sätzen ausgedrückt werden. Computerlesbare digitale Daten sind nach Ansicht Sommers entsprechend zu beschreiben. Denn Computer messen und verarbeiten Unterschiede (das Vorhandensein oder Nichtvorhandensein von Löchern auf Lochkarten, Spannungsunterschiede, etc.), die sich als eine »Serie von Sätzen« (Sommer 2002a: 416) der Form $f(x,y)$ verstehen lassen, welche folgende drei Bedingungen erfüllen: »(a) f steht für ‚ist belegt mit‘; (b) x referiert auf einen Speicherplatz; (c) y steht für den Wert 0 oder den Wert 1« (Sommer 2002a: 414). Die Mittelbarkeit, die digitalen Daten diesem Verständnis zufolge innewohnt, wird durch die Identifikation von Daten mit deren materieller Realisation verdeckt.⁸⁷ Daten können infolgedessen, »was sie als Sachverhalte gar nicht können dürften: Sie befinden sich an einem Ort, sie brauchen Platz, sie werden durch den Raum transportiert, sie gehen verloren oder werden gelöscht« (Sommer 2002a: 419).⁸⁸

87 | Das Sammeln von Daten beruht nach Ansicht Sommers auf einer doppelten Mittelbarkeit, die er wie folgt beschreibt: »Was wir zusammentragen und abspeichern, hat also bei den elektronisch verarbeitbaren Daten – wie bei allen Sachverhalten – eine zweifache Mittelbarkeit: Erstens: statt der Daten nehmen wir Sätze; zweitens: statt der Sätze nehmen wir Zeichen. Wir Computerbenutzer haben uns jedoch eine Redeweise angewöhnt, in welcher diese Mittelbarkeit zum Verschwinden gebracht wird durch metaphorischen Sprachgebrauch.« (Sommer 2002a: 418)

88 | Gesammelt werden können nach Ansicht von Sommer nur Dinge. Will man immaterielle Gegenstände, wie z.B. Romane sammeln, so ist dies nur mittelbar möglich, indem »der Sinn versinnlicht und verdinglicht« (Sommer 2002a: 105)

Wenn Sommer auf die Mittelbarkeit und damit auf die Konstruertheit von Daten hinweist, rekurriert er implizit auf das Begriffsverständnis der neuzeitlichen Wissenschaften, welches, wie Rheinberger herausgestellt hat, der etymologischen Bedeutung des Datenbegriffs entgegenläuft:

»Dem lateinischen Wortsinn nach sind die Daten das Gegebene, die Fakten hingegen das Gemachte. Nun ist es verblüffend, dass im Sprachgebrauch der neuzeitlichen Wissenschaften sich die Bedeutung der beiden Begriffe genau umgedreht hat: Daten werden als etwas von Instrumenten gemachtes und von Prozeduren hervorgebrachtes angesehen, während Fakten als etwas gegebenes angenommen und unterstellt werden.« (Rheinberger 2007: 117)

Diese Umdeutung des Begriffs erfährt unter den Bedingungen digitaler Medientechnologien eine erneute Umkehrung, die sich in dem Verständnis von Daten als digitalem Text widerspiegelt. Hierbei sind Daten das Gegebene, mit dem einerseits Computer und andererseits Menschen mittels Computern operieren. Dies zeige sich, so Rheinberger, auch in der Wissenschaftspraxis, die »von der hypothesenleiteten zur datengeleiteten Forschung« (Rheinberger 2007: 123f.) übergeht.⁸⁹ Am Datenbegriff tritt damit der Schriftcharakter sowie der Aspekt ihrer Verkörperung respektive Speicherung in den Vordergrund, weshalb Rheinberger in der Bedeutungsverschiebung keine metaphorische Ausweitung des Datenbegriffs erkennt, sondern eine Rückkehr zu dessen etymologischer Wurzel.

Die widersprüchlichen Deutungen von Sommer und Rheinberger sind symptomatisch für den Gebrauch des Datenbegriffs in der digitalen Medienkultur. Sie resultieren aus der Inanspruchnahme verschiedener Begriffsverständnisse als primärer Bedeutungsvariante und dem Absehen von den je anderen Gebrauchsweisen des Begriffs. Durch diese Interpretationen wird verdeckt, dass Daten nicht eindeutig bestimmt sind und sie demzufolge gleichermaßen als Gemachtes und als Gegebenes erscheinen und gebraucht werden. Deutlich wird dies am Begriff der Datenbank, in den die unterschiedlichen Verständnisse von Daten eingehen. In der Datenbank als Speicher, Ressource und Basis überkreuzen sich die verschiedenen Verständnisse von Daten als digitaler Text, als Übersetzung bzw. Vorstufe von Information, als spezifische Form von Information und als Inhalt technischer Operationen. Die in der Datenbank enthaltenen Daten sind zugleich Konstrukt und Gegebenes. Sie sind Resultat der Versammlung, Übersetzung, Formatierung und Inschriftion von Information sowie Ressource für Auswertung, Interpretation, Kombination und Rekombination von Daten zu Information. Datenbanken versammeln Daten, die

wird. Daten sind demzufolge nur dann unmittelbar sammelbar, wenn sie mit den materiellen Inschriften gleichgesetzt werden. Numerische Informationen über Realität sind hingegen nicht unmittelbar sammelbar, da sie keine Dinge, sondern Sätze bezeichnen.

⁸⁹ | Siehe hierzu auch S. 311ff.

nie bloß gegeben sind, machen sie aber als Gegebenes verfügbar und damit zum Reservoir vorhandener Information sowie zur Basis für neue Information. Daher sind Datenbanken zugleich Produkt (Gemachtes) von und Ausgangspunkt (Gegebenes) für informationelle Praktiken, sie bewegen sich in dieser Hinsicht zwischen medialer Konstellation und medialer Konfiguration.

Die informationelle Umwelt digitaler Informationen

Will man die mediale Logik digitaler Datenbanken verstehen, dann gilt es nicht nur die unterschiedlichen Formen von Informationen respektive Daten zu unterscheiden, die in Datenbanken enthalten sind, sondern auch danach zu fragen, wie diese verwaltet und verarbeitet werden. Zu bedenken ist dabei, dass die algorithmische Auswertung gespeicherter Informationen potenziell zu neuen Informationen führen kann, dass im Umgang mit Computern ständig neue (Gebrauchs-)Informationen erzeugt werden und dass die Verwaltung sowie Verarbeitung von Informationen in und mittels digitaler Medientechnologien häufig auf Zusatzinformationen angewiesen ist. Diese bilden die informationelle Umwelt digitaler Informationen.

Aus den Informationen, die auf digitalen Datenträgern gespeichert vorliegen, können mithilfe von Algorithmen vielfältige Informationen abgeleitet werden. Beispielsweise lässt sich das Alter einer Person nach einfachen Regeln berechnen, wenn ihr Geburtsdatum bekannt ist. Diese nicht unmittelbar in Datenbanken gespeicherten Informationen lassen sich als virtuelle Information begreifen. Bereits 1974 haben Folnius et al. darauf hingewiesen, dass die alleinige Betrachtung physisch verkörperter Informationen beim Entwurf von Datenbanksystemen zu kurz greift, denn »physically recorded data is only one point within a spectrum of ways to obtain information« (Folnius et al. 1974: 3). Um virtuelle Information handelt es sich ihrer Ansicht nach dann, wenn gespeicherte Daten algorithmisch ausgewertet werden. Damit besetzen virtuelle Informationen einen Zwischenraum, zwischen gespeicherten Informationen einerseits und reinen Programmen andererseits. An dieser Stelle setzt auch das Data Mining an, welches darauf abzielt, durch mathematische Verfahren der Mustererkennung neue Informationen zu entdecken (vgl. Cios 2010: 255ff.). Erkenntnisse werden hierbei nicht aus einzelnen in der Datenbank enthaltenen Informationen abgeleitet, sondern aus der Analyse des gesamten Informationsbestandes. Unter dem Stichwort »Data finds Data« beschreiben Jeff Jonas und Lisa Sokol (2009) einen weiteren Typus virtueller Informationen, der die impliziten Verknüpfungen zwischen einzelnen Datensätzen betrifft. Als Beispiel nennen die Autoren das Informationssystem eines Kasinobetreibers. Werden dem System die persönlichen Daten eines entlarvten Trickbetrügers hinzugefügt, gilt es mögliche Verbindungen zum Personal zu beachten, um auszuschließen, dass der Betrüger einen Komplizen unter den Angestellten hat. Findet sich in der Personaldatenbank ein Eintrag mit derselben Adresse oder der gleichen Telefonnummer, dann besteht zwischen dem Betrüger und einem der Beschäftigten eine Verbindung, die aus den Daten abgeleitet werden kann (vgl. Jonas/Sokol 2009: 108f.).

Eine weitere nicht zu vernachlässigende Form von Information stellen die Nutzungs- und Gebrauchsinformationen dar. Im Umgang mit Computern sowie beim Prozessieren des Computers entstehen Informationen, die gespeichert und ausgewertet werden können, d.h. es lassen sich nicht nur Informationen aus Informationen ableiten, der Gebrauch von digitalen Informationen allein ist bereits potenziell informativ. Der Erfolg des Web 2.0 basiert in erheblichem Maß auf der Speicherung und Auswertung von Nutzungsinformationen. Dienste wie das Internet-radio last.fm sammeln Informationen über die Hörgewohnheiten ihrer Nutzer, um ihnen auf Grundlage des so bestimmten Geschmacks neue Musik vorzuschlagen. Mit der Erhebung der Hörgewohnheiten einer Vielzahl unterschiedlicher Nutzer wird es möglich, diese in ein Nähe- und Distanzverhältnis zueinander zu setzen. Gefunden werden können hierdurch andere Nutzer, die einen ähnlichen Musikgeschmack haben, sowie neue Musik, die dem eigenen Geschmack ähnlich ist.

Die dritte und für die medientheoretische Betrachtung von Datenbanken grundlegendste Form von Information in der Umwelt digitaler Informationen gerät in den Blick, wenn man Informationssysteme auf die Bedingung ihres Funktionierens hin fragt. In Datenbanken werden Informationen zum Zweck ihres späteren Gebrauchs gespeichert, wie zum Beispiel ihrer Lektüre, ihrer algorithmischen Auswertung, ihrem Abgleich oder ihrer generischen Wiederverwendung. Insbesondere dann, wenn die in Datenbanken gespeicherten Informationen nicht nur auf dem Niveau von Realität verarbeitet werden sollen, sondern als Informationen *über* Realität, bedarf es zusätzlicher Informationen, sogenannter Metainformationen, die über die Bedeutung der Primärinformationen informieren. Diese Informationen über Informationen ermöglichen es Computern »Daten zu «erkennen» und zu finden, zu verschieben und zu extrahieren, mit anderen Daten zu verbinden« (Manovich 2005: 29). Durch die Explikation impliziter Bedeutungen zum Zweck der computertechnischen Verarbeitung von Informationen gerät nach Ansicht Weinbergers jedoch »die empfindliche Ökologie des Impliziten und Expliziten« (Weinberger 2008: 184) ins Wanken. Seines Erachtens lässt sich das Implizite nie vollständig darlegen; es gibt stets einen Überschuss des Impliziten, weshalb sich Metainformationen immer als defizitär erweisen, um die potenziellen Bedeutungen von Informationen vollständig zu erfassen oder zu beschreiben.

Auch wenn der Explikation des Impliziten enge Grenzen gesteckt sind, kommt Metainformationen bei der Verwaltung und Verarbeitung digitaler Informationen ein zentraler Stellenwert zu, weshalb sie medientheoretisch ernst zu nehmen sind. Es gilt nach den Strategien der Explikation des Impliziten zu fragen und zu eruieren, wie Metainformationen in Computern operativ werden. Im Folgenden Kapitel zur technischen Logik digitaler Datenbanken werden diese Fragen einer eingehenden Betrachtung unterzogen.

