



Big Data

Das neue Versprechen
der Allwissenheit

edition unseld
SV

edition unseld Sonderdruck

Redaktion: Heinrich Geiselberger und Tobias Moorstedt

Die Ausspähprogramme PRISM und Tempora, die durch den Whistleblower Edward Snowden bekannt wurden, sind nur die Spitze des Datenbergs. Auskunftfeien, Internetunternehmen und Einzelhandelsketten sammeln schon seit Jahren Informationen über die Bürger. Möglich wird all das durch Big Data, also die Möglichkeit, ungeahnte Mengen an Daten zu erheben, zu speichern und zu analysieren. Big Data, darin sind sich Analysten und Wissenschaftler einig, könnte die Art und Weise, wie wir Entscheidungen treffen, konsumieren, Wahlkämpfe führen, forschen und leben, dramatisch verändern. Manche fühlen sich dabei an George Orwells *1984* oder Steven Spielbergs Film *Minority Report* erinnert, andere sehen darin eine digitale Glaskugel, die gigantische Wachstumschancen bietet und neue Forschungswelten eröffnet.

Dieser Band bietet einen Überblick über die vielfältigen Anwendungsbereiche und Akteure; renommierte Praktiker, Soziologen und Technikexperten beleuchten in ihren Essays das Thema Big Data aus unterschiedlichen Perspektiven. Mit Beiträgen von Chris Anderson (*The Long Tail*), Dirk Baecker, Peter Glaser, Michael Hagner und Dirk Helbing, Bruno Latour, Frank Schirrmacher, Thilo Weichert und anderen.

Big Data

Das neue Versprechen
der Allwissenheit

Suhrkamp

Die edition unseld wird unterstützt durch eine Partnerschaft
mit dem Nachrichtenportal *Spiegel Online*. www.spiegel.de

Umschlagfoto: Chris Clor/Blend Images/Corbis

Erste Auflage 2013

edition unseld

Originalausgabe

© Suhrkamp Verlag Berlin 2013

Alle Rechte vorbehalten, insbesondere das der Übersetzung,
des öffentlichen Vortrags sowie der Übertragung
durch Rundfunk und Fernsehen, auch einzelner Teile.

Kein Teil des Werkes darf in irgendeiner Form
(durch Photographie, Mikrofilm oder andere Verfahren)
ohne schriftliche Genehmigung des Verlages reproduziert
oder unter Verwendung elektronischer Systeme
verarbeitet, vervielfältigt oder verbreitet werden.

Druck: Druckhaus Nomos, Sinzheim

Umschlaggestaltung: Nina Vöge und Alexander Stubić

Printed in Germany

ISBN 978-3-518-06453-5

Inhalt

Vorwort	7
1. Anwendungsbereiche und Akteure	21
Barbara Junge: Wer hat meine Daten?	23
Tobias Moorstedt: Obamas Datenakrobaten	35
Jakob Schrenk: Ist Fußball etwa doch Mathematik? Ein Gespräch mit dem Datenscout Jannis Scheibe	55
Michael Moorstedt: Erscanne dich selbst!	67
Jörg Häntzschel: Die Datenbergwerker	76
Tobias Moorstedt: WWWissenschaft. Ein Gespräch mit Cameron Marlow, dem Haussoziologen von Facebook	90
Gerhard Lauer: Die digitale Vermessung der Kultur. Geisteswissenschaften als Digital Humanities	99
2. Essays	117
Bruno Latour: Achtung: Ihre Phantasie hinterlässt digitale Spuren!	119
Chris Anderson: Das Ende der Theorie. Die Datenschwemme macht wissenschaftliche Methoden obsolet	124
Thilo Weichert: Big Data. Eine Herausforderung für den Datenschutz	131
Dietmar Offenhuber/Carlo Ratti: Drei Mythen über Smart Cities und Big Data	149
Dirk Baecker: Metadaten. Eine Annäherung an Big Data.	156
danah boyd/Kate Crawford: Big Data als kulturelles, technologisches und wissenschaftliches Phänomen. Sechs Provokationen	187

David Weinberger: Die digitale Glaskugel	219
Michael Hagner und Dirk Helbing: Technologiegetriebene Gesellschaft oder sozial orientierte Technologie? Ein Gespräch	238
Frank Schirrmacher: Der verwettete Mensch	273
Peter Glaser: Erinnerung der Zukunft	281
Glossar	295
Über die Trägerinnen und Träger	306
Textnachweise	308

Vorwort

In dem Kinofilm *Moneyball* (2011), der auf dem gleichnamigen Bestseller des amerikanischen Wirtschaftsjournalisten Michael Lewis aus dem Jahr 2003 basiert, gibt es eine Reihe von Szenen, die wunderbar illustrieren, worum es bei dem opaken, vielversprechenden, vielleicht auch gefährlichen Begriff »Big Data«, der derzeit die gesellschaftliche Debatte bestimmt, eigentlich geht.

Das finanziell angeschlagene Baseball-Team Oakland Athletics hat seine drei besten Spieler verloren und sucht nun verzweifelt nach Ersatz. Die Zeit drängt. Manager Billy Beane, dargestellt von Brad Pitt, hat seine Scouts zusammengetrommelt, um ihnen den Ernst der Lage klarzumachen und zu erfahren, welche Athleten sie für die kommende Saison verpflichten wollen. Und warum. Da sitzen sie nun, ältere Herren in legeren Freizeitklamotten, und man bekommt einen Eindruck davon, wie Entscheidungsfindungsprozesse in traditionellen Institutionen oft ablaufen: »Ich mag diese Jungs, die ein paar Haare am Arsch haben.« »Gutes Gesicht.« »Energisches Kinn.« »Der Ball explodiert an seinem Schläger.« »Er hat die Schubkraft einer Rakete.« »Er hat ne hässliche Freundin.« »Was willst du damit sagen?« »Hässliche Freundin bedeutet: kein Selbstvertrauen.«

Billy Beane verdreht die Augen und ahmt mit der linken Hand das schnappende Maul eines Krokodils nach. »Blablabla« – soll das wohl heißen. Kurz darauf lernt er bei Verhandlungen mit einem anderen Club Peter kennen, der an einer Ostküstenuniversität Wirtschaft studiert und in seinem Leben mehr Zeit vor dem Computer als auf dem Sportplatz verbracht hat. Peter hat Übergewicht, trägt eine Brille und ein viel zu warmes Konfirmanden-Jackett. Im testosterongetriebenen Baseball-Betrieb

fällt Peter auf, weil er sich nicht für die Freundinnen der Spieler interessiert und schon gar nicht dafür, ob ein Spieler auch genug Haare am Arsch hat.

Billy Beane verpflichtet Peter als Assistenten. Um die revolutionäre Herangehensweise des Computer-Nerds zu illustrieren, zeigt *Moneyball*-Regisseur Bennett Miller endlose Aufnahmen von Excel-Tabellen. Peter hat Informationen über 20 000 Baseball-Spieler in einer Datenbank zusammengefasst. »Darin sind alle Daten enthalten, um die Entwicklung von Spielern zu prognostizieren«, erklärt er seinem neuen Chef. »Okay«, antwortet Billy Beane, und diesem »Okaaaaay« hört man an, dass er nicht mehr 100-prozentig mitkommt. »Es geht darum, alle Informationen auf eine Zahl zu reduzieren«, fährt Peter fort. Mithilfe statistischer Formeln will er billige Spieler finden, die von Scouts aus Fleisch und Blut bislang übersehen wurden. »Wenn wir die Daten richtig gewichten, können wir in ihnen Qualitäten entdecken, die niemand sonst in ihnen sieht.« Peter nennt diese Spieler »hässliche Entlein«, Wesen, deren versteckte Schönheit nur der Computer erkennt. Tatsächlich identifiziert Peter eine Reihe solcher »hässlichen Entlein«, doch als Billy Beane den etablierten Scouts die Liste präsentiert, stößt er auf erbitterten Widerstand. Die Brandrede des alten Chef-Scouts bringt die Herausforderung gut auf den Punkt, vor die das Big-Data-Prinzip heute nicht nur Sportvereine stellt, sondern auch Krankenkassen, Geheimdienste, Stadtplaner, Werber, Germanisten, Ministerialbeamte und Astrophysiker:

»Die gesamte Major League, und die Fans, wären mehr als glücklich, wenn sie dich und Google Boy den Wölfen zum Fraß vorwerfen könnten, wenn ihr nicht endlich zur Vernunft kommt. Man stellt ein Baseball-Team nicht an einem Computer zusammen, Billy. Baseball ist mehr als Zahlen. Es ist keine Wissenschaft. Wenn es so wäre, dann könnte jeder

unseren Job machen. Die wissen nicht, was wir wissen. Die haben nicht unsere Erfahrung und nicht unsere Intuition. Billy, du hast da drin einen Jungen, der in Yale Wirtschaft studiert hat, und hier hast du einen Scout mit 29 Jahren Baseball-Erfahrung. Du hörst auf den Falschen. Es gibt Unwägbarkeiten, die nur ein Baseball-Kenner versteht. Willst du das, was unsereins seit 150 Jahren tut, was du selbst getan hast, etwa mit Füßen treten?»

In dieser Szene prallen zwei Entscheidungs- und Weltwahrnehmungskulturen aufeinander: Augenschein vs. Algorithmen, Gedächtnis vs. Datenbanken, Bauchgefühl vs. Statistik, Erfahrung vs. Innovation. Man könnte natürlich – gerade als fußballfixierter Europäer – die internen Konflikte eines amerikanischen Baseball-Teams als peripheres Phänomen abtun. Aber schon die Tatsache, dass die Oakland Athletics (Lewis' Buch basiert auf einer wahren Begebenheit) in der Saison 2002 als erste Mannschaft in der über hundertjährigen Geschichte des American-League-Baseball zwanzig Spiele in Folge gewannen, ist ein Hinweis darauf, dass *Moneyball* mehr erzählt als nur die klassische Underdog-Story, die Hollywood so sehr liebt.

Denn: Daten zu sammeln, Software nach Mustern suchen zu lassen und anschließend aus den Ergebnissen die richtigen Schlüsse zu ziehen, das ist, verkürzt gesagt, Big Data. Dieser Begriff beschreibt interessanterweise weniger eine Methode oder ein konkretes Produkt als vielmehr eine Problematik bzw. Herausforderungslandschaft, in der sich Konzerne, Forscher, Politiker und Individuen aktuell neu orientieren müssen, ob sie es wollen oder nicht. Big Data, heißt es in der Internet-Enzyklopädie Wikipedia, sind alle Datenansammlungen, die so groß und komplex sind, dass es schwierig bis unmöglich wird, sie mithilfe traditioneller Datenbanken und Software-Werkzeuge zu verarbeiten. Der Begriff »Big Data«, schreiben Kenneth

Cukier und Viktor Mayer-Schönberger in ihrer Monographie zum Thema (Untertitel: *Eine Revolution, die die Art und Weise verändern wird, wie wir leben, arbeiten und denken*), »bezieht sich auf alle Dinge, die man auf einer höheren Ebene tun kann, auf einer niedrigeren Ebene jedoch nicht« (2013, S. 6).

So weit, so vage. (Dass man, sobald die Prozesse, die im 21. Jahrhundert in Datenbanken und Supercomputern ablaufen, mit sprachlichen Mitteln beschreiben soll, offenbar sofort an Grenzen stößt, ist nur ein weiterer Beleg dafür, wie komplex und fremd uns das Phänomen – noch – ist.) Nehmen wir einmal ein Beispiel, das nichts mit Baseball zu tun hat: Die Stadtverwaltung von Dublin hat 2013 ein Projekt gestartet, mit dem man verhindern will, dass sich die 1,3 Millionen Einwohner der irischen Metropole auf den Straßen permanent in die Quere kommen. Dazu sammeln die Behörden mithilfe der IBM-Software »Infosphere Streams« zunächst Daten: die Verkaufszahlen der Fahrkartenautomaten, Abgaswerte, die Twitter- und Facebook-Posts der Reisenden, Aufzeichnungen von Straßensensoren und Videokameras sowie die GPS-Daten der mehr als 1000 Busse, die jeden Tag in Dublin unterwegs sind. Die ganz unterschiedlichen Daten gleicht man dann mit Verkehrsdaten und Stauberichten aus der Vergangenheit ab. »So merkt man in Echtzeit, wo ein Problem besteht oder entstehen könnte, und kann darauf reagieren«, meint ein Verantwortlicher (zitiert nach Bertolucci 2013). Dabei kommen gelegentlich auch seltsame Phänomene zum Vorschein, etwa dass Busse, die um 18 Uhr in der Rushhour losfahren, oft Busse überholen, die schon um 17.30 Uhr gestartet sind.

Das Beispiel aus dem urbanen Alltag illustriert, was für das Phänomen Big Data charakteristisch ist: Die Stadtverwaltung arbeitet mit Datenmassen, die vor wenigen Jahren, als es längst noch nicht so viele Sensoren gab, die Speicherkapazitäten deutlich geringer und die Rechner noch langsamer waren, nicht ge-

sammelt, verwaltet und analysiert hätten werden können; sie bezieht die Daten – in Echtzeit – von unterschiedlichen Quellen und aus vielen Datenbanken; und sie ist regelmäßig mit Ergebnissen konfrontiert, die sie zwar nicht so ganz versteht, aber dennoch verwenden kann (vgl. zum Themenkomplex Smart Cities auch die Beiträge von Jörg Häntzschel sowie Dietmar Offenhuber und Carlo Ratti in diesem Band).

Big Data ist der neueste Trendbegriff aus dem Silicon Valley – nach E-Commerce, Web 2.0, Social Media, Smartphones und Mobile Entertainment. »Big Data verändert das Wirtschaften an sich, und der Effekt ist enorm«, schwärmt James Manyika, der Chef des MCKinsey Global Institute, des Thinktanks der bekannten Unternehmensberatung (zit. n. Fischermann/Hamann 2013). Matthias Schrader von der Digitalagentur SinnerSchrader bezeichnet Daten als »das Erdöl der Zukunft« (Schrader 2011). Die *New York Times* lässt den Informatiker Jon Kleiberg zu Wort kommen: »Der Begriff selbst mag vage sein, aber er bezieht sich auf etwas sehr Reales. Big Data ist ein Werbeslogan für einen Prozess, der das Potenzial hat, alles zu verändern.« (Lohr 2012)

Zurück ins Kino: Im Original erschien Michael Lewis' Bestseller *Moneyball* im Jahr 2003, der Begriff »Big Data« kommt darin trotz der vielen Ziffern, Leistungsdaten und Tabellen ebenso wenig vor wie in Steven Spielbergs Film *Minority Report* aus dem Jahr 2002, in dem Precrime, eine Abteilung der Washingtoner Polizei der Zukunft, schwerwiegende Straftaten vorhersieht und ahndet, bevor sie stattgefunden haben. Im Film schwärmt ein Werbespot: »Stellen Sie sich eine Welt ohne Morde vor. [...] [V]or wenigen Jahren sah es so aus, als könne nur ein Wunder das Blutbad stoppen. Aber statt einem Wunder haben wir drei Wunder bekommen: die Präkognitiven.« Die »Precogs« sind durchsichtige, sanfte Wesen, die die Zukunft voraussehen können – allerdings verlassen sie sich bei

der präemptiven Polizeiarbeit nicht auf Daten, Algorithmen und Rechnerleistung, sondern sie gelangen auf mystische Art und Weise zu ihren Prophezeiungen. Am Ende entsteht so eine Bürokratie, die sich nicht mit dem beschäftigt, was war oder ist, sondern mit dem, was sein wird. Die Hauptfigur des Films, Chief John Anderton (Tom Cruise), wird schließlich selbst wegen eines »zukünftigen Mordes« gejagt.

Es mag seltsam erscheinen, zehn Jahre alte Bücher und Filme zur Erklärung der Gegenwart heranzuziehen, aber in gewisser Weise haben uns *Moneyball* und *Minority Report* subkutan auf das vorbereitet, was wir nun erleben. Die Werke markieren zugleich die Pole, zwischen denen sich die Diskussion über das Thema Big Data heute abspielt: *Moneyball* bringt den Optimismus der Technologie-Branche an der amerikanischen Westküste zum Ausdruck, deren Big Player Google, Microsoft und Facebook, aber auch andere, uns versprechen, dass mithilfe des Internets und immer leistungsstärkerer Rechner alle Probleme gelöst und alle Fragen beantwortet werden können – Partnersuche, Wettervorhersage, Vermögensbildung, Innovation, Erkenntnis. Der Publizist Evgeny Morozov hat für diese Geisteshaltung des unbedingten Glaubens an technische Lösungen das schöne Wort »Solutionismus« erfunden (Morozov 2013). Der Film *Minority Report*, der auf einer Kurzgeschichte des Science-Fiction-Autors Philip K. Dick basiert, steht hingegen für die dystopisch-orwellischen Aspekte, die 2013 angesichts der Enthüllungen des Whistleblowers Edward Snowden in den Vordergrund gerückt sind: Überwachung, biometrische Ortung, totale Kontrolle.

Vielleicht muss man selbst Big-Data-Methoden anwenden, um dem Begriff »Big Data« näher zu kommen. Auf der Website des Internetgiganten Google, der so etwas ist wie die Mutter der Daten-Alchemie, da er in dem Bestreben, »das gesamte Wissen der Menschheit abzubilden« (Sergey Brin), mittlerweile so gut wie alles digitalisiert und in verwertbare Daten übersetzt, auf

der Website von Google also kann man sich mit einer Anwendung namens Trends anzeigen lassen, wie oft Nutzer in einem bestimmten Zeitraum nach einem Begriff gesucht haben. Gibt man »Big Data« in diese historische Suchen-Suchmaschine ein, erhält man für die Jahre 2002 bis 2011 eine Kurve, die eigentlich eine Gerade ist. Obwohl in dieser Zeit die technisch-konzeptionellen Grundlagen für Big Data gelegt wurden, wurde der Begriff selbst weder beworben noch benutzt oder bezweifelt. Die nuller Jahre können insofern als Latenzphase des Phänomens gelten. Der Graph zeigt keinen Ausschlag, als Google 2004 das Projekt Google Books startete, das nicht nur als gigantisches Backup des globalen Gedächtnisses dient, sondern auf der Grundlage der gescannten Wörter und Sätze eine Big-Data-Übersetzungssoftware entwickelt; er bewegt sich nicht, als die Technologie-Avantgardisten des US-Magazins *Wired* dem Thema im Juni 2008 ein Sonderheft widmeten und das Ende der Theorie ausriefen (der Aufmacher von Chris Anderson, der auch in diesem Band abgedruckt ist, versprüht noch die typische solutionistische Euphorie des Silicon Valley); kein Ausschlag auch, als Barack Obamas Wahlkampfteam im Herbst desselben Jahres den Wählern mittels der Algorithmen zu Leibe rückte oder als Google 2009 mittels Google Flu Trends die Ausbreitung der Schweinegrippe vorhersagte – ohne Bluttests, allein durch die Analyse der Suchanfragen der Nutzer. Im Februar 2011 demütigte IBMs Supercomputer Watson dann seine menschlichen Gegner im TV-Quiz »Jeopardy«. Es stellte sich heraus, dass Computer zwar nicht intelligent sind, aber dank endlos tiefer Datenbanken und guter Algorithmen besser Bescheid wissen. Und? Nichts!

Erst Anfang 2012 schießt die Big-Data-Kurve dann steil nach oben. Soft- und Hardwarekonzerne wie Intel, SAP, IBM oder Cisco versprechen mit Initiativen wie »Big Insight« und allen möglichen Begriffen, denen das Adjektiv »Smart« vorangestellt

ist, die Lösung von Menschheitsproblemen wie Verkehrsstaus oder Energieknappheit. Wirtschaftsredaktionen berichten schwärmerisch über Milliardenumsätze und Wachstumsraten im zweistelligen Bereich (vgl. Fischermann/Hamann 2013). Im August 2012 konstatierte Steve Lohr von der *New York Times*: »In diesem Jahr hat Big Data den Durchbruch geschafft – als Idee, als Wort und – ja, auch das – als Marketinginstrument. Big Data hat die Nische der Technologieexperten verlassen und den Mainstream erreicht.« (Lohr 2012)

Wir haben so lange unbemerkt (und unbenannt) Big-Data-Services in Anspruch genommen, haben uns von Google die richtigen Webseiten suchen lassen, von Amazon die richtigen Bücher und von Facebook die richtigen Freunde, dass es fast verwunderlich erscheint, wie verwundert die – deutsche – Zivilgesellschaft im Frühsommer 2013 auf die Enthüllungen Edward Snowdens reagierte. Die Fakten sind bekannt: Im Rahmen des PRISM-Programms hat die NSA bereits Hunderte Millionen Datensätze aus der ganzen Welt (E-Mails, Webabfragen, Kreditkartenbewegungen usw. usf.) gesammelt; die Geheimdienstleute versprechen (durchaus im solutionistischen Westküsten-tonfall) mehr Effizienz und Zielgenauigkeit bei der Suche nach Anschlagzielen und verdächtigen Individuen. Plötzlich ging es mitten im Jahr 2013 wieder um 1984 (vgl. dazu den Essay von Frank Schirrmacher in diesem Band).

Angesichts der gigantischen, datengetriebenen Rasterfahndung, die die Geheimdienste der »Vereinigten Daten« (Staub 2013) betreiben, kann man sowohl von einer Überwachungs- als auch von einer Berechnungsgesellschaft sprechen. Das Unbehagen in der Bevölkerung rührt nicht allein daher, dass die Datenspione möglicherweise Kenntnis von den Empfängern und Inhalten unserer E-Mail etc. erlangen, sondern auch daher, dass die Kriterien und Suchwörter der Mustererkennungsprogramme im Dunkeln liegen. Darf ich, fragt man sich, noch in

die USA einreisen, wenn ich radikalen Umweltorganisationen wie Earth First! oder den Sea Shepherds auf Twitter »folge«? Wenn ich schon einmal in Pakistan Urlaub gemacht oder bei Amazon ein paar Sekunden auf Seiten mit Werken von Hakim Bey oder Bakunin verweilt habe? Sollte ich, fragt man sich dann im nächsten Schritt, solche Sachen lieber lassen, wenn ich in Zukunft vielleicht einmal nach New York zum Shoppen fliegen will? Ein Satz, den das Bundesverfassungsgericht bereits vor 30 Jahren, also in der Steinzeit, datentechnologisch gesehen, formuliert hat, wirkt plötzlich wie eine düstere Prophezeiung: »Wer damit rechnet«, heißt es im Volkszählungsurteil aus dem Jahr 1983, »dass etwa die Teilnahme an einer Versammlung oder einer Bürgerinitiative behördlich registriert wird und dass ihm dadurch Risiken entstehen können, wird möglicherweise auf eine Ausübung seiner entsprechenden Grundrechte (Art. 8, 9 GG) verzichten.« (BVerfGE 65, 43)

Aber sind das überhaupt die richtigen Fragen bzw. Sorgen? Politische Parteien sammeln Unmengen von Daten über Wähler, suchen nach Mustern und schicken anschließend maßgeschneiderte Mails. Fußballmanager verlassen sich nicht länger allein auf ihr Bauchgefühl und Best-of-Videos, sondern heuern Datenscouts an, die Unmengen von Daten über Spieler sammeln, nach Mustern durchsuchen und anschließend Transferkandidaten präsentieren (vgl. dazu das Gespräch zwischen Jakob Schrenk und Jannis Scheibe in diesem Band). Der Online-Buchhändler Amazon sammelt Unmengen von Daten über seine Kunden, durchsucht sie nach Mustern und destilliert daraus Kaufempfehlungen. Literaturwissenschaftler digitalisieren Unmengen von Büchern, durchsuchen sie nach Mustern und schreiben auf dieser Grundlage Aufsätze (vgl. dazu den Beitrag von Gerhard Lauer). Ja, ganz normale Menschen sammeln Unmengen von Daten über sich selbst, durchsuchen sie nach Mustern und gehen dann vielleicht doch lieber noch eine Run-

de joggen, als sich eine weitere Nacht unruhig im Bett hin- und herzuwälzen (vgl. dazu den Beitrag von Michael Moorstedt). Im Prinzip sind also alle möglichen Akteure auf allen möglichen Ebenen, allen möglichen Geschäftsfeldern und in allen möglichen Funktionssystemen der Gesellschaft damit beschäftigt, Unmengen von Daten zu sammeln, sie nach Mustern zu durchsuchen und daraus irgendwelche Schlüsse zu ziehen (vgl. dazu den Beitrag von Dirk Baecker). Müssten wir nicht an der Zeitgenossenschaft von Geheimdiensten wie der NSA oder dem BND zweifeln, wenn sie nicht auch auf die Idee gekommen wären, genau das zu tun?

Wir erleben jedenfalls einige »funktionale Turbulenzen« (Dirk Baecker), und was Big Data noch so alles durcheinanderwirbeln und ob es wirklich alles verändern wird, bleibt abzuwarten. Als erstaunlichen Zwischenstand können wir einstweilen festhalten, dass zwei Dinge, die lange Zeit als banal, angestaubt und passé galten, plötzlich eine interessante Umwertung erfahren. Als Erstes wäre da die positivistische Planungseuphorie der sechziger Jahre zu nennen (vgl. Wiedemann/Seibel 2013). Vor fünfzig Jahren war man noch überzeugt, die Funktionsweise der Gesellschaft mithilfe kybernetischer Modelle durchschauen und zum Wohle aller feinjustieren zu können. Dieser optimistischen Haltung (man könnte glatt von einem Solutionismus *avant la lettre* sprechen) kam dann in der neoliberalen Postmoderne zunächst der Glaube an die Durchschaubarkeit und anschließend das Vertrauen in den Staat als kompetenten Steuermann abhanden. Heute favorisieren Wissenschaftler wie der deutsche Physiker Dirk Helbing, der an der Eidgenössischen Technischen Hochschule in Zürich Soziologie lehrt, gigantische digitale Plattformen, die offen und durch Partizipation vieler Teilnehmer charakterisiert sind (vgl. dazu den Beitrag von David Weinberger sowie das Gespräch zwischen Michael Hagner und Dirk Helbing).

Die zweite interessante Umwertung betrifft die Sozialfigur des Statistikers, der über Jahrzehnte als langweiliger Typ mit Ärmelschonern, Rechenschieber und Stift hinter dem Ohr galt. Kein Wunder, dass Franz Kafka, der bei der Arbeiter-Unfall-Versicherungs-Anstalt (AUVA) zwar nicht unmittelbar für die Statistik zuständig war – dafür gab es den Ingenieur Alois Gütling –, es aber dennoch mit Unfallhäufigkeiten und Gefahrenklassenschemata zu tun hatte (vgl. Alt 2005, S.175), so schwermütig war. Gini- und Korrelationskoeffizienten, Schwedenschlüssel und Permutationen – gab es in Oberstufe und Grundstudium irgendetwas noch Undurchschaubareres und Faderes? (Wer aufpasste, hatte freilich signifikant bessere Chancen beim Kartenspielen.) Mit Big Data hingegen wird der Statistiker nun neuerdings zum Action Hero des Informationszeitalters. Hal Varian, der Chefvolkswirt von Google, sagte dazu schon 2009 (man kann sich das bei Youtube ansehen):

»Wenn linke Schuhe richtig billig sind, und Sie haben ein Monopol auf rechte Schuhe, dann sind Sie in einer ziemlich guten Position. Was ist derzeit allgegenwärtig und billig? Daten. Was ist knapp und teuer? Das Talent, das man braucht, um diese Daten zu analysieren und sie dazu zu bringen, ihre Geschichte zu erzählen. [...] Deswegen sage ich den Leuten: In den zehner Jahren wird ein Job wirklich sexy sein: der des Statistikers.«

Als erster »Quant«, der es tatsächlich zu so etwas wie popkulturellem Sexappeal gebracht hat, kann wohl der Amerikaner Nate Silver gelten, der mit Baseball-Analysen anfing und weltberühmt wurde, als er 2008 für 49 von 50 Staaten den Ausgang der amerikanischen Präsidentschaftswahl richtig vorher sagte und 2012 bei Obamas zweitem Wahlsieg erneut richtig lag. Die Laufbahn als Statistik-Nerd verspricht dabei nicht nur

ein ziemlich passables Einkommen: Wenn alle Wissenschaften sich in angewandte Mathematik oder Informatik verwandeln (Bio-Informatik, Digital Humanities, Computational Social Sciences etc.), werden die Zahlenakrobaten zu dem, was früher einmal die Philosophen und Historiker, die Quantenphysiker, Soziologen und zuletzt die Neurowissenschaftler waren: Chefwelterklärer.

Big Data wird, je nach Standpunkt der Kommentatoren, einen Paradigmenwechsel einläuten oder auf dem Friedhof der Modetrends enden. Vorerst scheint es jedenfalls lohnenswert, sich mit dem Thema in seiner ganzen Breite und Tiefe zu befassen. Dieses Buch besteht daher aus zwei Teilen: Im ersten Teil werden in eher journalistischen Beiträgen und Interviews Akteure vorgestellt und die Felder abgesteckt, auf denen sich Big Data derzeit ereignet. Dabei geht es weniger um eine repräsentative Auswahl von Beispielen oder gar darum, das Phänomen komplett abzubilden. Es handelt sich eher um Probebohrungen, die veranschaulichen, wie Big Data funktioniert und was man damit machen kann. Im Glossar am Ende des Bandes finden sich weitere Anwendungsbeispiele sowie einige Grundbegriffe.

Im zweiten Teil des Buches haben wir Texte versammelt, in denen darüber nachgedacht wird, was das Prinzip »Big Data« mittelfristig bedeuten könnte, wie es sich zu der Art und Weise verhält, wie wir in der Vergangenheit gelebt, gearbeitet, gedacht, geforscht und entschieden haben, und welche Herausforderungen sich angesichts der großen Datenflut aus wissenschaftlicher, rechtlicher und politischer Perspektive stellen. Ein solcher Band kann natürlich dem Anspruch der Nerds und Quants nicht genügen, der, wie es in *Moneyball* heißt, darin besteht, alle Informationen zu sammeln und auf eine Zahl, auf eine eindeutige Aussage zu reduzieren. Vermutlich ist das allerdings im Angesicht von Big Data gar nicht möglich. »Wir können nicht länger

so tun«, schreibt Pat Helland, einer der führenden Datenbank-Experten bei Microsoft, in seinem Aufsatz »If you have too much data, then ›good enough‹ is good enough«, »als würden wir in einer klinisch sauberen Welt leben.« (Helland 2011) Wir müssten uns daher daran gewöhnen, dass wir nicht länger alle Informationen berücksichtigen können. In diesem Sinne funktioniert dieser Band wie eine Big-Data-Studie: Er bringt unterschiedliche Quellen zusammen, zeigt Zusammenhänge auf, die man zuvor vielleicht nicht gesehen hat, deutet Richtungen an und ist damit hoffentlich »good enough« – gut genug.

Literatur

- Alt, Peter-André (2005), *Franz Kafka. Der ewige Sohn*, München: C. H. Beck.
- Bertolucci, Jeff (2013), »Dublin points Big Data tech at traffic jams«, in: *InformationWeek* (20. Mai), online verfügbar unter: {<http://www.informationweek.com/big-data/news/big-data-analytics/dublin-points-big-data-tech-at-traffic-j/240155213>} (Stand August 2013).
- Fawcett, John (2013), »The new finance era: From HFT to Big Data«, auf: *Wall Street & Technology* (7. August), online verfügbar unter: {<http://www.wallstreetandtech.com/electronic-trading/the-new-finance-era-from-hft-to-big-data/240159602>} (Stand August 2013).
- Fischermann, Thomas/Götz Hamann (2013), »Wer hebt das Datengold?«, in: *Die Zeit* (6. Januar), online verfügbar unter: {<http://www.zeit.de/2013/02/Big-Data>} (Stand August 2013).
- Helland, Pat (2011), »If you have too much data, then ›good enough‹ is good enough«, online verfügbar unter: {<http://queue.acm.org/detail.cfm?id=1988603>} (Stand August 2013).
- Lohr, Steve (2013), »The origins of ›Big Data‹. An etymological detective story«, in: *New York Times* (1. Februar), online verfügbar unter: {http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/?_r=0&pagewanted=print} (Stand August 2013).
- Lohr, Steve (2012), »How big data became so big«, in: *The New York*