

„Big Data“ als Erwartungsraum

von Jan-Felix Schrape

Utopien und Dystopien zu Massendaten seit 1960

Seit die britische Wochenzeitung *The Economist* digitalen Datenfluten 2010 eine große Sonderausgabe gewidmet hat,¹ genießt der Begriff „Big Data“ eine sehr hohe öffentliche Aufmerksamkeit. Bereits in den Jahrzehnten davor hat sich allerdings eine facettenreiche Debatte um die Computerisierung der Gesellschaft entwickelt und in der aktuellen Diskussion spiegeln sich viele der utopischen bzw. dystopischen Prophetien und Prognosen wider, die seit den 1960er-Jahren mit Massendaten verknüpft wurden. Diese Erwartungsdynamiken möchte ich im Folgenden nachzeichnen, um anschließend die Rolle der Sozialwissenschaften in diesem Diskurs zu beleuchten.

Der Mythos der Maschine

Auch wenn der Begriff „Big Data“ selbst erst Ende der 1990er-Jahre entstanden ist, unternahmen Schriftsteller wie Philip K. Dick in *The Minority Report* (1956) schon in der Frühzeit des digitalen Computers² literarische Annäherungsversuche an die zunehmende Informatisierung des Lebens und die Ambivalenzen der damit einhergehenden Datenansammlung. Nur wenig später entwarf der Medientheoretiker Marshall McLuhan das erste sozialwissenschaftlich informierte Zukunftsszenario einer elektronisch vernetzten Gesellschaft, die von totaler wechselseitiger Überwachung geprägt sein sollte: „The new electronic interdependence recreates the world in the image of a global village [...]. And as our senses have gone outside us, Big Brother goes inside.“ Den aufkommenden Medientechnologien schrieb er vielfältige Auswirkungen auf die menschlichen Wahrnehmungs- und Organisationsweisen zu, deren Bewertung aber je nach Beobachtungsperspektive variere. Insofern ging es McLuhan zunächst einmal darum, überhaupt ein Bewusstsein für technikinduzierten Wandel zu schaffen.³

Der Jurist Arthur R. Miller nahm 1967 eine weitaus eindeutigeren Einordnung vor. Er vermutete, dass die positiven Effekte der Computertechnologien den Blick auf deren negativen Folgen verstellen könnten. Vor allem anderen warnte er vor einem schleichenden Verlust der Privatsphäre: In früheren Zeiten sei eine umfassende Anhäufung personenbezogener Daten schlicht unmöglich gewesen, nun aber wachse das individuelle Datendossier kontinuierlich an. „Our success or failure in life ultimately may turn on what other people decide to put into our files and on the

programmer's ability, or inability, to evaluate, process, and interrelate information."⁴ Das, so befürchtete der Techniksoziologe Lewis Mumford in seinem Buch *Mythos der Maschine* (1967/1970), könnte letztlich „nicht nur die Invasion der Privatsphäre bedeuten, sondern die totale Zerstörung der menschlichen Autonomie [...]“.⁵

Der Journalist Ben H. Bagdikian hingegen sah 1971 die Gefahr der elektronischen Medien vorrangig in der „Überschwemmung des Individuums mit Informationsfluten frei Haus“: Sie ermutigten „eher zur Reaktion auf aktuell wichtige Ereignisse als auf große Trends“ und erweckten nur „die Illusion umfassenden Wissens“.⁶ Ähnlich argumentierte der Futurologe Alvin Toffler, der auf die Gefahr der kognitiven Überstimulation durch neue Technologien hinwies.⁷ Während Toffler dieses Risiko des *information overload* auf stetig dezentralere Strukturen der Medienlandschaft zurückführte, warnten einige Informatikpioniere freilich bereits in dieser Zeit vor dem anderen Extrem – vor zunehmend zentralisierten Systemen, deren Kontrolle einigen wenigen Anbietern überlassen würde.⁸

Gleichzeitig verbreitete sich ab Mitte der 1970er-Jahre eine Vielzahl positiver Visionen zur informatisierten Gesellschaft, die sich verdichtet in einem *Time*-Artikel von 1978 wiederfinden, der sich nicht zuletzt mit dem Wandel des Alltagslebens durch den Computer auseinandersetzte: „Paper clutter will disappear as home information management systems take over from memo pads, notebooks, files, bills and the kitchen bulletin board. [...] The computer might appear to be a dehumanizing factor, but the opposite is in fact true. It is already leading the consumer society away from the mass-produced homogeneity of the assembly line [...]. Far from George Orwell's gloomy vision of *Nineteen Eighty-Four*, the computer revolution is stimulating intellects, liberating limbs and propelling mankind to a higher order of existence.“⁹

Die Informatisierung des Alltags

Hierzulande wurden die neuen Formen elektronischer Vernetzung ab den frühen 1980er-Jahren in der Lebenswelt der Rezipienten sichtbar. Medieninnovationen wie die Bildkassette, der Bildschirmtext und das Kabelfernsehen weckten in den Sozialwissenschaften rasch weitreichende Veränderungserwartungen; sie wurden von der Bevölkerung anfänglich indes eher zurückhaltend aufgenommen.¹⁰

Derweil konkretisierte sich die Diskussion über die beschleunigte Produktion von Daten sowie deren Management auf der Basis empirischer Studien und mündete 1986 in Hal Beckers idiomatische Frage: „Can users really absorb data at today's rates? Tomorrow's?“¹¹ Der Informatiker Peter Denning sah einen Ausweg aus dieser Überlastungssituation in Maschinen „that can recognize or predict patterns in data without understanding the meaning of the patterns“.¹² Und Tim Berners-Lee

entwickelte ab 1989 sein Konzept des World Wide Web, wobei er zunächst kein globales Informationssystem erschaffen, sondern ein typisches Problem großer Organisationen lösen wollte: Deren Dokumentationssysteme waren in der Regel hierarchisch angelegt und konnten den stetig komplexeren Verweisen zwischen Personen, Projekten und Dokumenten immer weniger gerecht werden.¹³

Die Schwierigkeit, die Berners-Lee ursprünglich angehen wollte, war also die der Verknüpfung und Strukturierung von Daten – und mit dem World Wide Web als dezentralem Organisationsprinzip und leicht zugänglichem ‚Interface‘ des Internets schuf er zugleich die Grundlage für eine erneute Beschleunigung der Datenproduktion. Ebendiese Wechselwirkung zwischen wachsenden Datenmassen und erweiterten Möglichkeiten, Informationen auszutauschen, beschrieb der IT-Unternehmer John Mashey 1998 als „Big Data and the Next Wave of InfraStress“.¹⁴ Ab 1996 war es günstiger, Daten digital statt auf Papier zu speichern, der Datenverkehr im Internet wuchs um über 100 Prozent per anno und erstmals in der Geschichte wurde das Gros der Daten durch individuelle Nutzer generiert.¹⁵

Vor diesem Hintergrund wurde das World Wide Web einerseits schnell als freies Medium gepriesen, das eine „Verschiebung der Intelligenz vom Sender zum Empfänger“¹⁶ bewirke, da sich nun „die Rollentrennung von Kommunikator und Rezipient auflöst“.¹⁷ Andererseits monierte der polnische Autor Stanisław Lem, das Netz öffne seine Tore einem jeden, „der betrügen, Daten stehlen und Geheimnisse aushorchen will“.¹⁸ Und der Medienkritiker Neil Postman betonte, dass nicht mehr die schiere Verbreitung von Daten das Hauptproblem der Moderne sei, sondern die Frage, „wie man Information in Wissen verwandelt und wie Wissen in Erkenntnis“.¹⁹

Web 2.0: „Data is the next Intel Inside“

Die Protagonisten dieser eher allgemein gehaltenen Debatte um die Chancen und Risiken des neuen Mediums erinnerte der IT-Entrepreneur Tim O’Reilly in seinem Artikel „What is Web 2.0“ (2005) wieder an die eigentliche Herausforderung der Digitalisierung: „Database management is a core competency of Web 2.0 companies [...]. This fact leads to a key question: Who owns the data? [...] Data is indeed the Intel Inside of these applications. In many cases [...] there may be an opportunity for an Intel Inside style play, with a single source for the data.“²⁰ Ein Hauptmerkmal des ‚Web 2.0‘ sah O’Reilly in seinem ursprünglich internetökonomisch angelegten Essay insofern in einer so noch nie dagewesenen Zentralstellung von Daten in der Geschäftswelt sowie den damit verbundenen Fragen nach ihrer Kontrolle und Auswertung.

Bald darauf avancierte das ‚Web 2.0‘ allerdings zum Synonym für eine erneute allumfassende Aufbruchsstimmung um das Netz und der öffentliche Diskurs entfernte sich zügig von dem eigentlichen Topos des Datenmanagements. Dabei lassen sich im Wesentlichen drei interagierende Veränderungserwartungen

unterscheiden, die auf einen Rückbau eingespielter sozialer Differenzierungen und Machtasymmetrien hinauslaufen: (1) das Aufbrechen starrer Rollenverteilungen zwischen Produzenten und Konsumenten bzw. „the implosion of production and consumption“²¹; (2) ein sukzessiver Relevanzverlust massenmedialer Nachrichtenverteilung (*one-to-many*) gegenüber nutzerzentrierten Austauschprozessen im Social Web (*many-to-many*); sowie (3) eine technikinduzierte Demokratisierung politischer Entscheidungsprozesse.

Nach der ersten öffentlichen Aufregung um die neuen Möglichkeiten im ‚Web 2.0‘ rückten ab 2010 indes auch wieder dystopische Thesen in den Blickpunkt. Der Publizist Evgeny Morozov warnte vor der „Falle der Self-Empowerment-Diskurse“, die er als eine „ideologische List“ beschrieb, um „Unternehmensinteressen [zu] verschleiern“.²² Der Soziologe Zygmunt Bauman vermutete, die meisten Nutzer seien sich der Überwachung im Internet durchaus bewusst, allerdings werde heute „die alte Angst vor Entdeckung von der Freude darüber abgelöst, dass immer jemand da ist, der einen wahrnimmt. [...] Das Versprechen erhöhter Sichtbarkeit [...] kommt dem ersehnten Beweis gesellschaftlicher Anerkennung nahe [...]“²³ Und der iranische Blogger Hossein Derakhshan stellte 2015 nach sechsjähriger haftbedingter Netzabstinenz fest: „The rich, diverse, free web [...] is dying. [...] We seem to have gone from a non-linear mode of communication [...] toward a linear one, with centralization and hierarchies.“²⁴

‚Big Data‘ als Chiffre und Erwartungsraum

Parallel zu diesem übergreifenden gesellschaftsphilosophischen Diskurs hat in den vergangenen Jahren die Debatte um große Datenmengen im engeren Sinne wieder Fahrt aufgenommen. 2010 erschien die eingangs erwähnte Ausgabe der Zeitschrift *The Economist*, 2011 beschrieb das Marktforschungsunternehmen *Gartner* ‚Big Data‘ als die größte ökonomische Herausforderung unserer Tage und 2014 prognostizierte die *International Data Corporation*, dass das digitale Datenuniversum von ca. 0,8 Zettabyte (ZB) im Jahr 2009 auf 44 ZB im Jahr 2020 anwachsen werde. 2013 lag dieser Wert bei 4,4 ZB, wovon rund 3 ZB durch die Nutzer selbst generiert wurden (zumeist auf unternehmensgeführten Plattformen).²⁵

Anders als in den Dekaden davor rückt nun ‚Big Data‘ als begriffliches Kondensat für Massendaten sowie deren Kontrolle und Auswertung explizit in den Fokus der öffentlichen Diskussion um die Digitalisierung. Und ähnlich wie im Falle des Labels ‚Web 2.0‘, das die kommunikationserleichternden Eigenschaften der Onlinetechnologien betont hat, legt der Begriff ‚Big Data‘ als kommunikativer Bezugspunkt und Erwartungsraum heute in nahezu allen gesellschaftlichen Bereichen erneut den Eindruck eines disruptiven Umbruchs nahe, auf den unmittelbar reagiert werden muss. Die Einschätzungen reichen dabei von einem „neuen Versprechen der Allwissenheit“ über eine ubiquitäre Vernetzung der Welt und Vorhersagbarkeit menschlichen Verhaltens bis hin zu dem Orwell’schen

Albtraum einer Totalüberwachung durch den Staat oder führende IT-Konzerne.²⁶

Dabei wird ‚Big Data‘ als Chiffre für verschiedenartige soziotechnische Trends genutzt.²⁷ Versicherungen erproben Preismodelle auf der Grundlage von *self-tracking*, die eine gerechtere Beitragsberechnung versprechen, aber zugleich die Gefahr einer Entsolidarisierung mit sich bringen. In der Verbrechensbekämpfung sollen auf der Basis aggregierter Bewegungsdaten Delikte und Ausschreitungen vorhergesagt werden. ‚Big Data Marketing‘ stellt ein vollpersonalisiertes Einkaufserlebnis in Aussicht. In der Medizin findet die Auswertung von Massendaten nicht mehr nur im individuellen Gesundheitsmanagement, sondern überdies in der Früherkennung von Epidemien oder in der Katastrophenhilfe Anwendung. Und auch in anderen Forschungsbereichen führt die Verfügbarkeit wachsender Datenmengen zur Erprobung neuer Erhebungs- und Auswertungsmethoden sowie neuartiger Darstellungsweisen.

Am offenkundigsten zeigt sich die Zwiespältigkeit von ‚Big Data‘ im öffentlichen Diskurs bislang auf dem Feld der mobilen Endgeräte. Smartphones, Tablets und Smartwatches haben den medialen Erfahrungsraum individueller Nutzer erneut deutlich erweitert und erleichtern deren Alltag durch massendatenbasierte Echtzeitanwendungen. Gleichzeitig führt die Anbieterkonzentration auf diesem Markt aber zu einer medienhistorisch bis dato singulären Bündelung privatwirtschaftlicher Verfügungsmacht über Interaktionsdaten, mit der letztlich eine Privatisierung des Schutzes persönlicher Daten einhergeht. Die damit verknüpfte Ambivalenz kommt nirgends besser zum Ausdruck als in einem Zitat von Eric Schmidt aus dem Jahr 2010, der heute Executive Chairman der aus *Google Inc.* hervorgegangenen *Alphabet Holding* ist: „The more information we have about you [...] we can improve the quality of our searches. [...] We don't need you to type at all, 'cause we know where you are - with your permission. We know where you've been - with your permission. We can more or less guess what you're thinking about.“²⁸

Evangelisten, Apokalyptiker, Integrierte – und Sozialwissenschaftler

Insgesamt spiegeln sich im aktuellen Diskurs um die gesellschaftlichen Folgen von ‚Big Data‘ viele der Erwartungen wider, die bereits die Debatten um Massendaten seit den 1960er-Jahren beherrscht haben – von der Gefahr eines *information overload* und der Illusion umfassenden Wissens über die Angst vor einer Invasion der Privatsphäre oder einer schleichenden Entdemokratisierung bis hin zu vielfältigen positiven Visionen, die entlang von Stichworten wie Dezentralisierung, Personalisierung, Transparenz oder *Human Enhancement* mit der Informatisierung der Gesellschaft verknüpft werden. Und ganz ähnlich wie im Falle früherer technologischer Umbrüche treten auch im gegenwärtigen Diskurs die bekannten Figuren des berauschten Evangelisten, des kritischen Apokalyptikers und des

bereits nutzungspraktisch Integrierten auf, für den etwa das Thema Datenschutz längst das Signum einer vergangenen Epoche ist.

Angesichts dieser regelmäßig wiederkehrenden Erwartungen, kann es Soziologinnen und Soziologen nun aber gerade nicht darum gehen, als Advokaten der einen oder der anderen Seite aufzutreten oder oft direkt in der San Francisco Bay Area geschöpfte Schlagworte wie ‚Big Data‘ oder ‚Web 2.0‘ ungefiltert als „quasi-soziologische Fachbegriffe in die wissenschaftliche Diskussion“ zu übernehmen, „in der sie dann aber weiterhin nur von ihrer im Praxisfeld begründeten Evidenz zehren.“²⁹ Vielmehr sollte neben der Diskussion der methodischen Fallstricke und Potenziale von ‚Big Data‘³⁰ eine der Hauptaufgaben der Soziologie darin bestehen, die gesellschaftlichen Effekte der entsprechenden utopischen bzw. dystopischen Zukunftserwartungen zu eruieren und generalisierbare Muster sowie Kontinuitäten herauszuarbeiten.

Denn obgleich bislang keine der vorgestellten Prophetien und Prognosen in ihrer Radikalität empirisch eingelöst worden sind, erfüllen Sie in ihren jeweiligen Anwendungskontexten zentrale kommunikative Funktionen: Sie erzeugen öffentliche Aufmerksamkeit für neue technologische Möglichkeiten sowie die damit verbundenen Chancen und Risiken, kanalisieren den Diskurs in eine bestimmte Richtung und dienen nicht zuletzt als Legitimationsgrundlage in wirtschaftlichen oder politischen Entscheidungsprozessen.³¹ Insofern bieten die aktuellen Erwartungen um ‚Big Data‘ sozialwissenschaftlichen Beobachtern durchaus Orientierung – zwar nicht darüber, was die Zukunft bringen wird, aber zweifellos darüber, welche Veränderungshoffnungen, Unsicherheiten und Konflikte den gegenwärtigen gesellschaftlichen Diskurs prägen.

Dieser Beitrag ist Teil eines Themenschwerpunkts zu Big Data. Weitere Texte finden Sie hier.

Fußnoten

¹ The Economist o.V., The Data Deluge, Special Report, 27.2.2010.

² Der erste in Serie hergestellte digitale Computer UNIVAC I kam 1951 in den USA auf den Markt, kostete bis zu 1,5 Mio. US-Dollar und war deutlich größer als ein handelsüblicher Kleiderschrank.

³ Marshall McLuhan, The Gutenberg Galaxy, Toronto 1962, S. 31. In einem Interview aus dem Jahr 1972 gab McLuhan konkretisierend zu Protokoll: „The biggest job in the world will be espionage. [...] Espionage at the speed of light will become the biggest business in the world.“ Zit. nach: Peter C. Newman, The Lost Marshall McLuhan Tapes, in: In: Maclean’s Magazine, 16.7.2013, S. 48–51.

⁴ Arthur R. Miller, The National Data Center and Personal Privacy, in: The Atlantic 11 (1967), S. 53–57, hier: S. 54.

- 5[□] Lewis Mumford, *Mythos der Maschine*, Frankfurt 1977 [1967].
- 6[□] Ben Haig Bagdikian, *The Information Machines*, New York 1971. Zit. nach: O.V., „Die Elektronen haben keine Moral“, in: *Der Spiegel* 17 (1972), S. 158-164, hier: S. 164.
- 7[□] Alvin Toffler, *Future Shock*, New York 1970.
- 8[□] Vgl. Karl Steinbuch, *Massenkommunikation der Zukunft*, in: *Umschau in Wissenschaft und Technik* (Hrsg.), *Forschung '72*. Frankfurt 1971, S. 201-211.
- 9[□] O.V., *Living: Pushbutton Power*, in: *TIME*, 20.2.1978, S. 46-49, hier: S. 46.
- 10[□] Vgl. Jan-Felix Schrape, *Wiederkehrende Erwartungen an neue Medien*, Boizenburg 2012.
- 11[□] Hal B. Becker, *Can users really absorb data at today's rates? Tomorrow's?*, in: *Data Communications* 15.8 (1986), S. 177-193. Vgl. ferner Ithiel De Sola Pool / Hiroshi Inose / Nozomu Takasaki / Roger Hurwitz, *Communications Flows. A Census in the United States and Japan*, Tokio 1984.
- 12[□] Peter Denning, *Saving all the Bits*, in: *American Scientist* 78 (1990), S. 402-405, S. 402.
- 13[□] Tim Berners-Lee, *Information Management: A Proposal*, Arbeitspapier, Bern 1989.
- 14[□] John R. Mashey, *Big Data and the Next Wave of InfraStress*. Präsentation, 25.4.1998. Mit „InfraStress“ beschrieb Mashey die Überlastung technischer und organisationaler Infrastrukturen durch die anwachsenden Datenfluten.□
- 15[□] Vgl. Robert Morris / Brian Truskowski, *The Evolution of Storage Systems*, in: *IBM Systems Journal* 42.2 (2003), S. 205-217. Sowie: Peter Lyman / Hal R. Varian, *Reprint: How Much Information?*, in: *Journal of Electronic Publishing* 6.2 (2000).
- 16[□] Nicholas Negroponte, *Being Digital*, New York 1995, S. 29.
- 17[□] Joachim Höflich, *Technisch vermittelte interpersonale Kommunikation*, Opladen 1996, S. 13.
- 18[□] Stanisław Lem, *Zu Tode informiert*, in: *Der Spiegel* 11 (1996), S. 108-109, hier: S. 108.
- 19[□] Neil Postman, *Die zweite Aufklärung*, Berlin 1999, S. 124.
- 20[□] Tim O'Reilly, *What Is Web 2.0*, in: *O'Reilly Network*, 30.9.2005.
- 21[□] George Ritzer / Nathan Jurgenson, *Production, Consumption, Prosumption. The Nature of Capitalism in the Age of the Digital 'Prosumer'*, in: *Journal of Consumer Culture* 10.1 (2010), S. 13-36, hier: S. 19.
- 22[□] Evgeny Morozov, *Back to the Roots: Cyberspace als öffentlicher Raum*, in: *Blätter für deutsche und internationale Politik* 9 (2011), S. 114-120, hier: S. 118.
- 23[□] Zygmunt Bauman, *Das Ende der Anonymität*, in: *Blätter für deutsche und internationale Politik* 10 (2013), S. 51-62, hier: S. 54f.
- 24[□] Hossein Derakhshan, *The Web We Have to Save*, in: *Matter*, 14.7.2015.
- 25[□] Gartner Inc., *Pattern-Based Strategy: Getting Value from Big Data*, Stamford 2011. Sowie: EMC/IDC, *The Digital Universe of Opportunities*, Framingham 2014.
- 26[□] Vgl. Heinrich Geiselberger / Tobias Moorstedt (Hrsg.), *Big Data: Das neue Versprechen der Allwissenheit*, Berlin 2013. Ramón Reichert (Hrsg.), *Big Data:*

Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie, Bielefeld 2014.

27[□] Vgl. bereits: Danah Boyd / Kate Crawford, Critical Questions for Big Data, in: Information, Communication & Society 15.5 (2012), S. 662-679.

28[□] Google Inc. Eric Schmidt at Washington Ideas Forum 2010, Videodokument, 15. Minute.

29[□] Florian Süssenguth, Die Organisation des digitalen Wandels, in: Ders. (Hrsg), Die Gesellschaft der Daten, Bielefeld 2015, S. 93-121, hier: S. 99.

30[□] Vgl. Jochen Mayerl, Bedeutet ‚Big Data‘ das Ende der sozialwissenschaftlichen Methodenforschung?, in: Soziopolis, 29.11.2015.

31[□] Vgl. dazu Sascha Dickel / Jan-Felix Schrape, Dezentralisierung, Demokratisierung, Emanzipation. Zur Architektur des digitalen Technikutopismus, in: Leviathan 43.3 (2015), S. 442-463.