



Dr. Siegfried Timpf, März 2017

BIG DATA

Dossier „Digitalisierung“ – Teil 6 einer Mini-Serie im Rahmen der Kommission „Arbeit der Zukunft“

Über den Autor

Siegfried Timpf, Jg. 1958, Dr. rer. pol., ist Sozialökonom (HWP Hamburg) mit den Forschungsschwerpunkten Diskursanalyse, Ideengeschichte, Wissenssoziologie und Zeitpolitik. Aktuelle Forschungsinteressen sind vergleichende Feld- und Netzwerkanalyse, Geschichte und Zukunft der Digitalisierung, Entstehung und Entwicklung des zeitpolitischen Feldes.

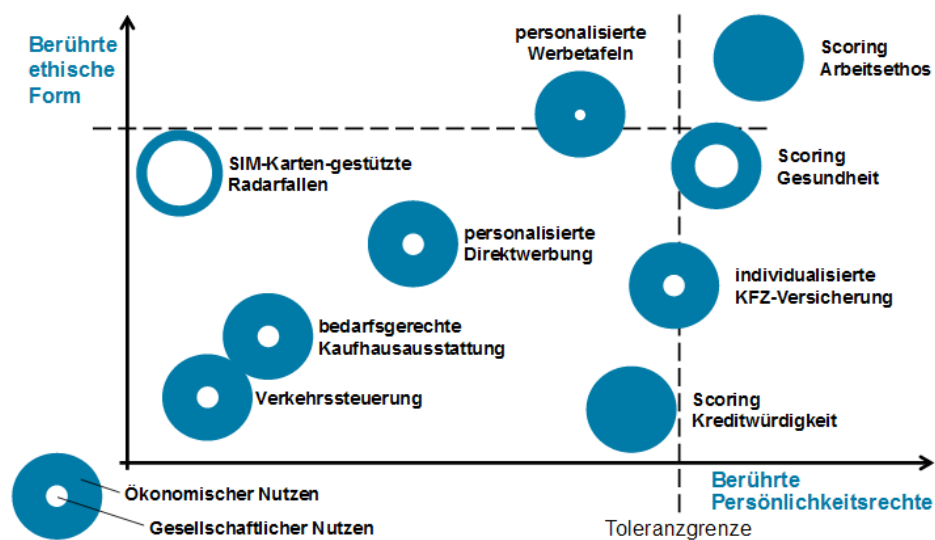
Big Data – „We don’t want the name. The name is noise.“

Die anekdotische Bemerkung, die hier zitiert wird (Hardy 2012), markiert einen Schnitt und ist zugleich Ausdruck einer Hybris. Für die ökonomische Nutzung war bisher der Name/die Identität und möglichst viele zusätzliche Daten der Ausgangspunkt, etwa für Marketingstrategien. Ungeordnete Daten erschienen als Rauschen. Basis für die Vorstellung von „Big Data“ ist, dass Daten in einer Menge anfallen, wie sie bisher unbekannt war, und dass es mittels Algorithmen möglich sei, aus diesen Daten ein Wissen zu generieren, das ökonomisch nutzbar gemacht werden kann. Verbreitet ist auf dieser Grundlage die Rede vom „Erdöl des 21. Jahrhunderts“. Es müssten Wege gefunden werden, diese Daten zu „fördern“ und zu „raffinieren“, „...also in nützliche Informationen und in Wissen zu verwandeln.“ (Helbing 2014, 4). Die Herausforderungen, die zu bewältigen sind, werden in der Regel als technisch lösbar definiert: die schiere Menge der zu bewältigenden Daten und ihre Auswertung in Echtzeit, die Unstrukturiertheit und Verschiedenartigkeit der Daten sowie das Problem, dass die anfallenden Daten unvollständig, nicht repräsentativ, fehlerhaft oder manipuliert seien (Helbing 2014, 4/5). Allein das ökonomische Potential von Open Data (also die Öffnung der Daten in der Verwaltung) wird unter Bezug auf Schätzungen von McKinsey (2013, 2) mit 3000 bis 5000 Mrd. US-Dollar weltweit angegeben. Für Deutschland hat die Konrad-Adenauer-Stiftung eine Studie (Dapp et al. 2016) erstellen lassen, in der für verschiedene Nutzungsszenarien von Open Data jährliche Summen von 12 – 131 Mrd. EUR jährlich angegeben sind.

Big Data wird mit dem künftigen Internet der Dinge in Verbindung gebracht. Während die mit digitalen Anwendungen bestückten Dinge Alltagsprozesse steuern, fallen Daten „nebenbei“ an. Die Frage ist dann, wer diese Daten erhält, bevorratet und auswertet. Doch bereits heute ist jenseits des „Rauschens“ unstrukturierter Datenmengen die konventionelle, strukturierte Datensammlung, etwa über Datenbroker wie Axciom (Offlinedaten) in Verbindung mit Facebook (Onlinedaten) bereits Realität. Der Zugriff auf Daten über Plattformen wie Google, Facebook, Apple, Amazon ist ein erster Schritt, folgen muss in der Logik ökonomischer Nutzung zwingend eine Auswertung über Algorithmen, um von relativ unstrukturierten Datenmengen zu ökonomisch verwertbaren Informationen zu gelangen. Die Wachstumsraten der

Datenmengen sind exponentiell: „...90% of the data now circulating on the internet was created less than two years ago (...). The Big Data sector has growth of 40% a year, seven times higher than that of the information data market’ (European Commission 2015).“ (Degryse 2016, 10).

Mayer-Schönberger (2015) vertritt die Auffassung, dass für den Erkenntnisgewinn primär die Daten und nicht der Algorithmus seien (so auch Brynjolfsson/McAfee 2012, 5). Kritisch wird dies deshalb beurteilt, weil im besten Fall induktiv gleichzeitig auftretende Merkmale in Verbindung gebracht werden, ohne dass die Kausalitäten klar werden (Scholz 2016a). Auch Helbing (2014) sieht ähnliche Risiken, zudem könne die naive Anwendung von Algorithmen zu falschen Schlussfolgerungen führen oder zu Fehlentscheiden, Diskriminierungen und Benachteiligungen. Es gibt inzwischen auch Erkenntnisse, die auf eine unterschiedliche Wirkung von Big Data in unterschiedlichen Milieus hindeuten (Jerome 2013). Eine Übersicht bietet Schröder (2016) zu verschiedenen Anwendungsbereichen von Big Data, wobei jeweils nach ökonomischem und gesellschaftlichem Nutzen unterschieden wird.



Schröder 2016, 8

Es hat den Anschein, als würden auch in Bezug auf Big Data die Anwendungsmöglichkeiten in einer Geschwindigkeit entwickelt, die vernünftige Überlegungen zu Zweckmäßigkeit und Nebenwirkungen weit hinter sich lassen. Nutzerinnen liefern über ihre Geräte bereits Daten über ihre Aktivitäten für das sogenannte

Nudging (Anstupsen). Auf der Basis gesammelter Daten werden individuelle Angebote möglich, bis hin zu einer dynamischen Preisbildung. Und – die Anonymisierung von Daten durch Weglassen des Namens ist mit der Fülle gesammelter Daten jederzeit umkehrbar ohne Information der betroffenen Person. Zu bedenken ist hier, dass Voraussetzungen für das Recht auf informationelle Selbstbestimmung geschaffen werden müssen und der Verarbeitungsprozess von Daten transparent und beeinflussbar gehalten wird. Wenn die Algorithmen¹ für die Abschätzung des privaten wie gesellschaftlichen Nutzens von ausschlaggebender Bedeutung sind, so sind sie entsprechend ihrer Bedeutung zu behandeln. Insbesondere betrifft dies Algorithmen, die dem Bereich des maschinellen Lernens zuzuordnen sind, etwa *Klassifikation* (Zusammenfassung ähnlicher Objekte unter einem Oberbegriff), *Clustering* (Gruppierung von Daten auf der Basis struktureller Ähnlichkeiten), *Regellernen* (aus expliziten Merkmalen werden implizite Regeln gelernt) und *Mustererkennung* (Muster werden aus bekannten Daten extrahiert und über neue Datenmengen gelegt) (Fraunhofer Institut 2015, 23f.).

¹ Algorithmen entstehen z. B. im Kontext der Übertragung menschlicher Handlungen und Denkvorgänge auf Maschinen. Dies gelingt, „*wenn sich Handlung und Denken mittels Abstraktion in eine Folge von „elementaren und rekombinierbaren Teilaktivitäten zerlegen lässt und dabei der Kontext so eingeschränkt werden kann, dass alle Entscheidungsalternativen, die im Handlungs- oder Denkprozess auftreten können, sich schon vorher vollständig bestimmen lassen.*“ (Trogemann 2010, 160). Menschliche Handlungen und Denkabläufe werden so sowohl entkörperlicht als auch dekontextualisiert und bilden Muster von Handlungsabläufen, die aus einfachen Elementen bestehen. Diese sind kombinierbar zu beliebig komplexen Artefakten. Aktuell werden interessante Fragen im Diskurs gestellt. Die Art, wie Algorithmen Prozesse ordnen, ist an die Existenz bestehender Muster orientiert, werden damit Interaktionsprozesse auf bekannte Muster beschränkt? Sind Algorithmen eigene Entitäten oder sind sie in einem Grad relational und assoziativ, der Zweifel an ihrer Eigenständigkeit aufkommen lassen? Wenn wir annehmen können, dass Algorithmen die Teile eines Codes sind, die Entscheidungen treffen, wie beeinflussen Algorithmen die Entscheidungsprozesse von Institutionen, Unternehmen und Administration? Wie beeinflussen Algorithmen Ergebnisse und Optionen? Und – wie sortieren, ordnen und prognostizieren Algorithmen, wie tragen sie damit bei zur Erfindung, zum Erhalt oder zur Verstärkung von Normen (Beer 2017, 5/6)? Darüber hinaus ist zu fragen, wie der Begriff des Algorithmus selbst unsere Erkenntnismöglichkeiten und Weltsicht reflektiert, welches Verständnis von Werten und kalkulierender Rationalität sich darin ausdrückt (Beer 2017, 7).

Auch Frey/Osborne haben sich mit der Frage befasst, was Algorithmen gegenwärtig leisten können. Sie stellen zunächst fest, dass Maschinen gegenwärtig nicht denken und begründen können, wie es Menschen vermögen. Menschliches Denken und Handlungsfähigkeit basiert auf einem impliziten Wissen über unsere Umwelt. Sie illustrieren am Beispiel der Entzifferung handschriftlicher Aufzeichnungen, dass zu ihrer Entschlüsselung intuitives Wissen über das Zusammenwirken von Schreibgerät und Papier, das Ausblenden von Unvollkommenheiten der Ausführung gehört, aber auch die Einschätzung der Bedeutung von Worten eines Tiefenwissens über die typische Struktur einer Sprache bedarf. Weiter nutzen wir Anhaltspunkte, die für die Wahrscheinlichkeit einer Deutung sprechen, wie etwa die Absichten des Autors und die Umstände, unter denen der Text entstand. Für die meisten dieser kognitiven Prozesse nehmen Frey/Osborne an, dass sie weit außerhalb dessen liegen, was Algorithmen derzeit reproduzieren können. Dies bedeute aber nicht, dass es ihnen unmöglich sei, menschliche Tätigkeiten auszuführen. Algorithmen, die dem Bereich des Maschinellen Lernens (vgl. oben) zuzuordnen seien, hätten in den späten 90ern und frühen 2000ern immerhin über 10% der Schecks in den USA eingelesen. Neuere Entwicklungen in der Automatisierung komplexer Aufgaben haben nach Beobachtung von Frey/Osborne die Lücke zwischen Algorithmen und menschlichem Wissen geschlossen, indem sie die zunehmende Verfügbarkeit von Big Data nutzen (Frey/Osborne 2015, 23). Sie beziehen sich hier auf die Google Algorithmen zur automatischen Übersetzung von Texten, die unter anderem eine große Menge von UN-Dokumenten nutzen, und in der Lage sind, kurze Phrasen zu identifizieren, die nicht nur äquivalent in andere Sprachen übersetzt werden, sondern auch substituiert werden, um die Übersetzung zu verbessern. Die Google-Algorithmen seien weit davon entfernt, die Tiefensemantik von Texten zu „verstehen“, doch seien sie für viele Anwendungen innerhalb des Big Data Ansatzes mehr als ausreichend. Frey/Osborne gehen offensichtlich davon aus, dass es eine Analogie zwischen „deep tacit knowledge“ und „Big Data“ gibt, die es ermöglichen soll, die Entwicklung von Algorithmen voranzutreiben.

Probleme in der Beurteilung von Big Data bestehen gegenwärtig bezüglich von Austauschverhältnissen, in denen die unentgeltliche Erbringung von Diensten an datenschutzrechtliche Einwilligungen gekoppelt ist.² Dies hat zur Folge, dass personenbezogene Daten in diesen Austauschverhältnissen die Funktion einer Währung annehmen mit der Folge einer Kommerzialisierung persönlicher Daten (Hermstrüwer 2016, 75).³ Rinne/Zimmermann (2016, 8) kommen angesichts von Positionen, in denen die Menge der gespeicherten Daten als das „wirklich innovative Gut der digitalen Revolution“ eingeschätzt werde, zu zwei Schlussfolgerungen. Zum einen bestünde dann der Grundkonflikt des „Informationskapitalismus“ zwischen Kapital und persönlichen Daten. Zweitens löse dieser Konflikt den alten Konflikt zwischen Kapital und Arbeit ab. Wenn diese Analogie stimme, so müsse dem neuen wie dem alten Konflikt mit geeigneten Regulierungsmaßnahmen begegnet werden. Dem gegenüber stehen Positionen, die von der pragmatischen Sichtweise ausgehen, dass „Privatsphäre“ ein Auslaufmodell sei und wir längst in der Zeit der „post-privacy“ angekommen seien (Heller 2011).

Morozov hingegen fordert ein rechtliches und sozioökonomisches Umfeld, in dem Akte des Teilens von Daten nicht manipulativ wirken können (Morozov 2015b, 7). Um zu verhindern, dass Bürger vom datenorientierten Kapitalismus überwältigt werden,

² Unter Bezug auf das Institut für Demoskopie Allensbach (2015) stellt das Vodafone Institut (2016, 21) fest, dass mobile InternetnutzerInnen kostenpflichtige und „kostenlose“ Anwendungen im Verhältnis von 1 : 3 nutzen, aber Vodafone stellt für den europäischen Bezugsrahmen fest, dass die NutzerInnen mehrheitlich eher bereit sind, für eine Anwendung zu zahlen, als die Verfügung über ihre Daten abzutreten.

³ Als Reaktion auf Unsicherheiten bezüglich der Nutzung von Big Data hat der BDI prüfen lassen, wie die Rechtslage sich darstellt (BDI/Noerr 2015). Danach greifen Big Data Analytics auf erfasste Daten zu, die unter bestimmten Gesichtspunkten (Kunden, Produktionsmaschinen) geordnet sind. Diese Datenbestände können auf „Muster, Schnittmengen, Parallelen und Entwicklungen“ hin analysiert werden. *„Die erhobenen Daten werden im Rahmen der Big Data Analyse neuen Fragestellungen und Auswertungen unterworfen, z.B. zum Zwecke der Prozessoptimierung, Produktentwicklung oder dem Marketing. Durch die Sichtung der Daten auf zunächst nicht sichtbare Muster und Strukturen gewinnen die Daten einen neuen Aussagegehalt. Der vorhandenen Datenmasse werden neue Informationen entnommen, so dass letztlich neue Daten geschaffen werden.“* (BDI/Noerr 2015, 19). Für solche Daten ist als Schlussfolgerung ein eindeutiges Verfügungsrecht konstruierbar.

schlägt er vor, die Daten in öffentlicher Hand zu behalten. *„Erforderlich ist eine stärkere Konsolidierung des öffentlichen Interesses bei allen Schichten – den Sensoren, Filtern und Profilen. Die Idee der Schaffung eines Gemeinguts der Daten mag zunächst abwegig erscheinen, aber es sollte wirklich so sein, dass jeder getätigte Klick auf eine App oder eine Seite, jede Interaktion mit einem intelligenten Thermostat oder intelligenten Auto, jede Bewegung in der Stadt dem Bürger zufallen sollte und nicht den Unternehmen, die diese Dienste anbieten. Ansonsten würde die Öffentlichkeit feststellen, dass sie die Kontrolle über dieses »Klick-Kapital« zusehends verliert.“* (Morozov 2015a, 33). Morozov schlägt konkret vor, eine Differenzierung in grundlegende und erweiterte Dienstleistungen vorzunehmen. Die grundlegenden Dienstleistungen würden unentgeltlich angeboten bei allgemeiner Zugänglichkeit, und erweiterte Dienstleistungen könnten einer ökonomischen Nutzung ohne Datenanhäufung offen stehen. Dies setzt voraus, dass Daten als öffentliches Gut reklamiert und einer Marktregulierung entzogen werden.

Hauptproblem ist die Betrachtung der Daten als Rohstoff. Es handelt sich jedoch um artifizielle Produkte, die im Zusammenspiel von physisch-sozialer Welt und digitalen Anwendungen entstehen. In ihre Konstruktion gehen viele Bedingungen und Gestaltungs- bzw. Verwertungsaspekte ein, die im Zweifel durch die automatische Datengenerierung⁴ über alltägliche Gebrauchsgegenstände verdeckt werden. Diese Unsichtbarkeit mitlaufender Datengenerierung und –übermittlung geht einher mit einer gleichzeitigen Ermöglichung (von Kontakten) und Begrenzung durch Algorithmen (Cohen 2012, 17). Fragwürdig wäre es, sich im Diskurs um die mögliche Regulierung von Big

⁴ Werden (2016) betont die Bedeutung der Geräte, die große Mengen an Daten während des Gebrauchs erzeugen und übermitteln, oft unter Wechsel des Rechtsraumes, der für die Beurteilung von Datensouveränität und Rechtskonformität dieser Datenströme bedeutend ist: *„Wichtig und wesentlich für Big Data sind „smarte Geräte“. Diesen Geräten kommt eine Schlüsselfunktion zu. Diese fungieren sinnhaft als intelligente Mess-Sensoren und sind massenhaft verbreitet. Jedes noch so kleine Datenschnipsel wird erfasst und versendet, je nach der Hersteller- und Anwenderkonfiguration.“* (Werden 2016, 38). Diese Prozesse basieren auch auf einer veränderten Konstruktion von Software, die nicht allein an der Funktionalität ausgerichtet ist (Werden 2016, 39).

Data auf diese „Automatismen“ als nicht hinterfragte Voraussetzung zu beziehen. Schweighofer (2016, 245) diskutiert die Möglichkeit der Eigentumskonstruktion an Daten für Nutzerinnen von Google, Facebook u. a., andere sehen die Gefahr, dass Menschen mit wenig Mitteln dann ihre Eigentumsrechte verkaufen und die Entstehung einer 2-Klassen-Datengesellschaft. Wer es sich leisten kann, verzichtet auf den Verkauf seiner Daten und genießt weiterhin „Privatsphäre“. Kognitiv wie ökonomisch besteht ein Problem in der Analogie zwischen Big Data und „deep tacit knowledge“, wenn auch die Fortschritte, wie sie von Frey/Osborne (2015) im Bereich der Übersetzungs-Algorithmen skizziert werden, beeindruckend sind.

Literatur

- BDI/Noerr LLP (2015): Digitalisierte Wirtschaft/Industrie 4.0. Ein Gutachten der Noerr LLP im Auftrag des BDI zur rechtlichen Situation, zum Handlungsbedarf und zu ersten Lösungsansätzen.
- Beer, D. (2017): The social power of algorithms. In: Information, Communication & Society, 20:1, S. 1-13.
- Brynjolfsson, E./McAfee, A. (2012): Big Data: The Management Revolution. In: Harvard Business Review, October 2012, S. 1-9.
- Cohen, J. E. (2012): Configuring the Networked Citizen. <http://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=1826&context=facpub> (Zugriff: 08.01.2017).
- Dapp, M. M./Balta, D./Palmetshofer, W./Krcmar, H. (2016): Open Data. The Benefits. Das volkswirtschaftliche Potential für Deutschland. Studie im Auftrag der Konrad-Adenauer-Stiftung. St. Augustin/Berlin: Konrad-Adenauer-Stiftung.
- Degrype, C. (2016): Digitalisation of the Economy and its Impact on Labour Markets. Etui: Working Paper 2016.02, Brüssel.
- Fraunhofer Institut für sichere Informationstechnologie (2015): Begleitpapier Bürgerdialog. Chancen durch Big Data und die Frage des Privatsphärenschutzes. Stuttgart: Fraunhofer Verlag.
- Frey, C. B./Osborne, M. A. (2015): Technology at work. The future of innovation and employment. (with contributions from Citi research). Citi GPS: Global Perspectives & Solutions/Oxford Martin School, http://www.oxfordmartin.ox.ac.uk/downloads/reports/Citi_GPS_Technology_Work.pdf (Zugriff: 03.01.2017).
- Hardy, Q. (2012): Big Data Era and Privacy. In: New York Times, 11.06.2012.
- Helbing, D. (2014): Big Data – Zauberstab und Rohstoff des 21. Jahrhunderts. In: Die Volkswirtschaft. Das Magazin für Wirtschaftspolitik 5, S. 4-6.
- Heller, C. (2011): Post-Privacy: Prima leben ohne Privatsphäre. München: C. H. Beck.
- Hermstrüwer, Y. (2016): Informationelle Selbstgefährdung. Tübingen: Mohr/Siebeck.
- Institut für Demoskopie Allensbach (2015): Allensbacher Computer- und Technik-Analyse – ACTA 2015. <http://www.ifd-allensbach.de/acta/> (Zugriff: 30.12.2016).
- Jerome, J.(2013): Buying and selling privacy. Big Data`s burdens and benefits. In: Stanford Law Review online 66, S. 47-53.
- Mayer-Schönberger, V. (2015): Was ist Big Data? Zur Beschleunigung des menschlichen Erkenntnisprozesses. In: In: Aus Politik und Zeitgeschichte, 65. Jahrgang, 11–12, S. 14-19.
- McKinsey/Global Institute (2013): Open data: Unlocking innovation and performance with liquid information. <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>.
- Morozov, E. (2015a): Digitale Technologie und menschliche Freiheit. In: Neue Gesellschaft/ Frankfurter Hefte 03/2015, S. 30-34.

- Rinne, U./Zimmermann, K. F. (2016): Die digitale Arbeitswelt von heute und morgen. In: In: Aus Politik und Zeitgeschichte, 66. Jahrgang, 18-19, S. 3-9.
- Scholz, C. (2016a): Smart durch Big Data? Plädoyer für eine kritische Betrachtung. In: Wirtschaftspsychologie Aktuell, Nr. 2/2016.
- Schweighofer, J. (2016): Zur Befreiung des Menschen von mühevoller Arbeit und Plage durch Maschinen, Roboter und Computer – Auswirkungen der Digitalisierung auf die Arbeitsmärkte. In: Wirtschaft und Gesellschaft, 42. Jg., Heft 2, S. 219-255.
- Schröder, L. (2016): Beschäftigtendatenschutz aus der Sicht der Gewerkschaften. Präsentation Berlin. TK IT Ver.di.
- Trogemann, G. (2010): Algorithmen im Alltag. In: Trogemann, G. (Hrsg.): Code und Material, Exkursionen ins Undingliche. Springer: Wien/New York, S. 159-185.
- Vodafone Institut für Gesellschaft und Kommunikation (2016): Big Data. Wann Menschen bereit sind, ihre Daten zu teilen. Eine europäische Studie. Berlin.
- Werden, S. (2016): Digitale Souveränität, ein Orientierungsversuch. In: Friedrichsen, M./Bisa, P.-J. (Hrsg.): Digitale Souveränität. Vertrauen in der Netzwerkgesellschaft, S. 35-52.