

Das Big Data Upgrade Neue Dimensionen der Datennutzung – Chancen und Risiken

Thomas Petzold

Derzeit werden täglich 2,5 Exabytes an Daten produziert. Ein Exabyte entspricht einer Eins, gefolgt von 18 Nullen. Darunter können wir uns gerade mal vorstellen, dass es sich um eine überwältigende Menge an Daten handelt. Datenanalysten ziehen gern den Vergleich heran, der auf den früheren Chief Executive Officer von Google, Eric Schmidt, zurückgeht: Die Summe an produzierten Informationen vom Beginn der Menschheit bis Mitte der 2000er betrug fünf Exabytes. Im Vergleich stellt sich das folgendermaßen dar: Heute werden alle zwei Tage so viele Informationen erzeugt wie zuvor in der gesamten Menschheitsgeschichte von Omo (der Wiege der Menschheit in Afrika) bis Euro (der Einführung der Gemeinschaftswährung).

Nun behauptet keiner ernsthaft, dass die schiere Produktion einer solch unvorstellbaren Menge an Informationen per se eine Errungenschaft ist. Das wird sie erst dann, wenn die dahinterliegenden Daten sinnvoll genutzt werden. Genau das ist der Fall beim Big Data Upgrade, der Umwandlung großer Datenmengen in neue Lösungen. Nach derzeitigen Erkenntnissen wurde der Terminus Big Data erstmals im Jahr 2000 von einem Forscher der Universität von Pennsylvania beschrieben. Karriere machte Big Data aber erst einige Jahre später, nachdem der Begriff erstmals 2007 in diversen Blogs erwähnt wurde. In Wikipedia tauchte der Eintrag im November 2009 auf. Dort wurde er zunächst wieder von Editoren gelöscht. Die Begründung war, dass „big“ lediglich ein Adjektiv aus dem Wörterbuch sei, das mit vielen anderen Wörtern verwendet werden kann, und die Kombination mit „data“ deshalb weder neu noch bemerkenswert genug sei für einen eigenständigen Eintrag. Der Begriff tauchte dann erst wieder Anfang 2010 in Wikipedia auf, zu einem Zeitpunkt, als die Popularität von Big Data bereits sprunghaft anstieg. Seitdem wird der enorme Zuwachs an Daten immer stärker in der öffentlichen Wahrnehmung unter dem Begriff Big Data diskutiert.

Woher kommen eigentlich all diese Daten? Daten entstehen überall: Sensoren erfassen Wetterinformationen, GPS-Signale werden genutzt, um das Restaurant um die Ecke zu empfehlen, neue Twitter-Nachrichten werden versandt, Fotos mit Freunden über Flickr geteilt, Unternehmen verarbeiten Informationen über eingekaufte Produkte von Kunden. Die Digitalisierung aller gesellschaftlichen Bereiche sowie die rasante Verbreitung und Nutzung von mobilen Endgeräten, Internetanwendungen und Sensoren produzieren immer größer und komplexer werdende Datenmengen. Da stellt sich die Frage, wie wird diese enorme Datenmenge eigentlich beherrschbar und welchen Nutzen bringt sie?

Big Data ist ein brisantes Thema in der Wirtschaft, weil diese sich signifikante Wertsteigerungen verspricht. Datenlawinen bestehen in Unternehmen derzeit vor allem in Geschäftstransaktionen. Die größte britische Supermarktkette Tesco nutzt diese zum Beispiel so: Während der Kunde an der Kasse steht, werden die Angaben der von ihm erworbenen Artikel an die Londoner Zentrale übermittelt, wo sie mit der Datenbank aller Tesco-Einkäufe abgeglichen werden. Auf Basis dieses Vergleichs werden Artikel identifiziert, die andere Kunden am häufigsten zusammen mit jenen Artikeln des an der Kasse stehenden Kunden gekauft haben. Die identifizierten Artikel werden daraufhin mit entsprechenden Rabatten auf die Rückseite des Kassensbons gedruckt, den der Kunde erhält. In wenigen Sekunden werden hier also Milliarden Datensätze nach relevanten Informationen durchforstet – durchaus mit Erfolg, denn Tesco-Kunden nutzen seit Einführung dieser Big Data Lösung die Rabattgutscheine 20 Prozent häufiger als zuvor.

Summary: More than ever before contemporary societies are characterized by the huge amounts of data being transferred. Authorities, companies, academia and other stakeholders refer to Big Data when discussing the importance of large and complex datasets and developing possible solutions for their use. Big Data promises to be the next frontier of innovation for institutions and individuals, yet it also offers possibilities to predict and influence human behavior with ever-greater precision.

Kurz gefasst: Das Kreislaufsystem gegenwärtiger Gesellschaften ist mehr denn je geprägt von Datenströmen. Unter dem Schlagwort *Big Data* versuchen Behörden, Unternehmen, Wissenschaft und andere Akteure die Bedeutung dieser Datenströme zu entschlüsseln und Lösungen für deren Nutzung zu erarbeiten. Big Data stellt ein beträchtliches Innovationspotenzial für Institutionen und Individuen dar, bietet aber auch Möglichkeiten, menschliches Verhalten mit immer größerer Präzision vorherzusagen und zu beeinflussen.



Thomas Petzold ist seit 2011 wissenschaftlicher Mitarbeiter in der Abteilung Kulturelle Quellen von Neuheit und Research Fellow am Centre for Excellence for Creative Industries and Innovation (Brisbane, Australien). Der promovierte Medien- und Internetforscher befasst sich mit Innovationskapazitäten im Bereich Medien und Kommunikation, vor allem mit der Vernetzung rechenbetonter und sozialer Innovation. [Foto: Udo Borchert]

thomas.petzold@wzb.eu

Literatur

Barabási, Albert-László: *Bursts – The Hidden Pattern Behind Everything We Do*. New York: Dutton 2010.

Bollier, David: *The Promise and Peril of Big Data*. Washington, DC: The Aspen Institute 2010.

Boyd, Danah/Crawford, Kate: *Six Provocations for Big Data*. Oxford Internet Institute: September 21, 2011.

Latour, Bruno: „Tarde's Idea of Quantification“. In: Mattei Candea (Ed.): *The Social after Gabriel Tarde – Debates and Assessments*. London: Routledge 2009.

Lazer, David et al.: „Computational Social Science“. In: *Science*, Vol. 323, No. 5915, S. 721–723.

Intelligente Analysemethoden und Nutzungskonzepte enormer Datenmengen sind also längst nicht mehr nur ein Alleinstellungsmerkmal erfolgreicher Internetunternehmen wie Amazon, Facebook oder Google. Ob es die Verarbeitung von Milliarden Flugverspätungsdaten bei einer Fluggesellschaft ist oder die Früherkennung von Kreditkartenbetrug bei einer Bank, ob Fakten zu tagesaktuellen Nachrichten visualisiert oder das individuelle Erbgut auf genetisch bedingte Krankheiten untersucht wird, ob sich fahrerlose Autos durch den dichten Berufsverkehr navigieren oder das Gesicht eines Straftäters in einer Menschenmasse identifiziert werden muss – überall werden brauchbare Lösungen gesucht, die in die Routinen des täglichen Lebens eingearbeitet werden können.

Um die bei Unternehmen, Behörden und anderen Institutionen entstehenden Datenfluten in nützliche Alltagsanwendungen zu übersetzen, sind innovative Analysemethoden und Werkzeuge gefragt. Die erforderliche technische und personelle Infrastruktur ist derzeit im Entstehen. Einrichtungen und Fachkräfte, die ausgeprägte statistisch-analytische Fähigkeiten mit fundierten Kenntnissen in maschinellem Lernen mit strategisch-visionären Eigenschaften bezüglich der Nutzung großer Datenmengen verbinden, können eine wichtige Rolle in vielen gesellschaftlichen Bereichen spielen. Denn sie können Antworten auf Fragen finden, die bislang keiner gestellt hat. Und sie sollen verhindern, dass die falschen Fragen gestellt werden, vorschnelle Entscheidungen getroffen oder sogar aussichtslose Initiativen gestartet werden, die von vornherein gar nicht die erhofften Resultate erbringen können. Um diese Institutionen, Initiativen und Individuen entwickelt sich eine Innovationskultur, die es erlaubt, durch Echtzeitanalyse großer Datenmengen die Gegenwart und nahe Zukunft zu beeinflussen.

Und genau das stellt die wohl größte gesellschaftliche Herausforderung um die Nutzung von Big Data dar. Weil das System der intelligenten Datenanalyse es ermöglicht, menschliches Verhalten vorherzusagen und zu beeinflussen, wird das Vertrauen in die Sammlung und Analyse von Daten ganz entscheidend davon abhängen, unter welcher Risikoabwägung Unternehmen und Behörden dieses System nutzen. Anwender sind skeptisch, was die Herausgabe ihrer persönlichen Daten betrifft, sie setzen diese in der digitalisierten Welt aber auch immer wieder dort ein, wo sie sich interessante Angebote versprechen. Für Unternehmen, Behörden und ganze Branchen heißt das, dass Bürgerinnen und Bürger das Sammeln von Nutzerdaten grundsätzlich mit Argwohn beobachten. Erst wenn Daten in nützliche, alltagsadäquate Serviceangebote übersetzt werden, kann das Vertrauen schaffen, also wenn die intelligente Datenanalyse hilfreich für Bürger auch jenseits kommerzieller Interessen wird. Verfügen beispielsweise Finanzbehörden über bessere statistische Daten und Analysemethoden, um Steuerdelikte aufzuklären, kann effizienter nach Steuersündern gefahndet und gleichzeitig die Quote der zu Unrecht behelligten Bürger reduziert werden. Sollte die datengesteuerte Früherkennung von Kreditkartenbetrug allerdings dazu führen, dass die Kreditkarte gesperrt wird, wenn man erstmals ein Land besucht und dort am Geldautomaten abheben will, dann behindert intelligente Datenanalyse den Alltag.

Deshalb stellt sich hinsichtlich Big Data vor allem die Frage, wie dessen Innovationspotenzial in den nächsten Jahren ausgestaltet wird. Neben den technischen Aspekten gibt es enorme soziale und innovationskulturelle Herausforderungen. Big Data ist ein Phänomen, das einerseits verspricht, massive Datenströme in raffinierte, smarte und nützliche Angebote umzuwandeln. Andererseits muss es sich aber zeigen, wie dabei die institutionellen und individuellen Nutzen und Risiken abgewogen werden.