



Sachstand

Big Data unter Berücksichtigung der Situation in der Volksrepublik China

Big Data unter Berücksichtigung der Situation in der Volksrepublik China

Aktenzeichen: WD 10 - 3000 - 068/17
Abschluss der Arbeit: 12. Januar 2018
Fachbereich: WD 10: Kultur, Medien und Sport

Die Wissenschaftlichen Dienste des Deutschen Bundestages unterstützen die Mitglieder des Deutschen Bundestages bei ihrer mandatsbezogenen Tätigkeit. Ihre Arbeiten geben nicht die Auffassung des Deutschen Bundestages, eines seiner Organe oder der Bundestagsverwaltung wieder. Vielmehr liegen sie in der fachlichen Verantwortung der Verfasserinnen und Verfasser sowie der Fachbereichsleitung. Arbeiten der Wissenschaftlichen Dienste geben nur den zum Zeitpunkt der Erstellung des Textes aktuellen Stand wieder und stellen eine individuelle Auftragsarbeit für einen Abgeordneten des Bundestages dar. Die Arbeiten können der Geheimschutzordnung des Bundestages unterliegende, geschützte oder andere nicht zur Veröffentlichung geeignete Informationen enthalten. Eine beabsichtigte Weitergabe oder Veröffentlichung ist vorab dem jeweiligen Fachbereich anzuzeigen und nur mit Angabe der Quelle zulässig. Der Fachbereich berät über die dabei zu berücksichtigenden Fragen.

Inhaltsverzeichnis

1.	Vorbemerkung	4
2.	Definition „Big Data“	4
3.	Einordnung	5
4.	Das „Sozialkredit-System“ – Big Data in der Volksrepublik China	6

1. Vorbemerkung

Dieser Sachstand enthält eine Definition sowie Einordnung des Begriffes „Big Data“. Hierfür wurde auf eine Ausarbeitung der Wissenschaftlichen Dienste (WD 10) zurückgegriffen, die sich bereits mit wesentlichen Aspekten des Themas beschäftigt und in Kapitel 3 wiedergegeben ist.

Darüber hinaus wird die Thematik im Zusammenhang mit der Situation in der Volksrepublik China eingeordnet.

2. Definition „Big Data“

Die „Brockhaus Enzyklopädie“ definiert Big Data wie folgt¹:

Big Data [-'deitə, englisch], Bezeichnung für Datenmengen, die so groß sind (typischerweise im Bereich Peta- [10¹⁵] bis Exabyte [10¹⁸]), sich so schnell ändern und in ihrem Format so stark variieren, dass sie nicht mehr in vertretbarer Zeit mit üblicher Software (allen voran Datenbankmanagementsystemen) verarbeitet werden können. Benötigt wird v. a. Software, die die Daten zur parallelen Abarbeitung auf Rechner-Cluster verteilt und die Ergebnisse zusammenführt. Man spricht dabei zwar auch von Datenbanksoftware, es handelt sich aber nicht mehr um relationale Datenbanken, sondern um sogenannte NoSQL-Datenbanken.

Beispiele für Big Data sind die Experimentdaten, die an den Detektoren des LHC anfallen (Grid-Computing), die geschäftlichen Rechenoperationen des Internethändlers Amazon, die täglich entstehenden Daten in sozialen Netzwerken (wie Facebook oder Twitter) oder die Daten, die von nationalen Geheimdiensten (wie der National Security Agency der USA) gesammelt werden. Big Data werden zunehmend zu einem wichtigen Wirtschaftsbereich: 2012 hatte der Big Data-Markt weltweit einen Umsatz von rd. 11,8 Mrd. US-Dollar, er soll bis 2017 auf etwa 50 Mrd. steigen. In keinem anderen Sektor des digitalen Universums wird ein ähnlich großes Wachstum vorhergesagt wie bei den unstrukturierten Daten, die v.a. von sozialen Netzwerken, dem E-Mail-Verkehr sowie der Onlinewerbung im Internet und auf mobilen Kommunikationsgeräten erzeugt werden. Vom sinnvollen Umgang mit Big Data versprechen sich Politik und Wirtschaft u.a. eine neue, solidere Basis für Entscheidungsfindungen.

1 Brockhaus: Big Data, URL: <https://deutscher-bundestag.brockhaus.de/enzyklopaedie/big-data> (Zugriff: 28.12.2017).

3. Einordnung

Spätestens seit den Enthüllungen um internationale Datenspionage und den Fall „Edward Snowden“ im Jahr 2013 hat sich der Blick auch auf die unter dem Stichwort „Big Data“ bekannt gewordenen neuen Möglichkeiten im Umgang mit großen Datenmengen gerichtet. Dabei geht es nicht um eine einzelne neue Technologie. Vielmehr bezeichnet Big Data ein Bündel neu entwickelter Methoden und Technologien, die die Erfassung, Speicherung und Analyse eines großen und beliebig erweiterbaren Volumens unterschiedlich strukturierter Daten ermöglicht. Für die IT-Branche wie auch die Anwender in Wirtschaft, Wissenschaft oder öffentlicher Verwaltung ist Big Data daher zum großen Innovationsthema der Informationstechnik geworden.

Daten sind heute im Wesentlichen durch drei Charakteristika gekennzeichnet, die ihren englischen Bezeichnungen zufolge als die „drei Vs“ bezeichnet werden. Dabei handelt es sich zum einen um die Datenmenge (*Volume*), die durch die fortschreitende Digitalisierung praktisch aller Bereiche des modernen Lebens in unvorstellbar großen Quantitäten produziert wird und sich etwa alle zwei Jahre verdoppelt. So wurden Schätzungen zufolge in diesem Jahr (2013) bereits über 2 Trilliarden Bytes an Daten weltweit gespeichert – was auf iPads gespeichert und gestapelt eine 21.000 km lange Mauer ergäbe. Ein weiteres Charakteristikum heutigen Datenverkehrs ist seine Geschwindigkeit (*Velocity*): Während früher Daten in bestimmten Abständen anfielen, die es erlaubten, sie nach und nach zu verarbeiten, ist man heute aufgrund von Vernetzung und elektronischer Kommunikation dem Datenfluss ununterbrochen ausgesetzt. Um sie nutzen zu können, müssen die einlaufenden Informationen immer schneller oder sogar in „Echtzeit“ aufgenommen und analysiert werden. Das dritte wichtige Merkmal ist die unterschiedliche Beschaffenheit (*Variety*) der heute in so vielfältig und komplex strukturierten Quellen wie z.B. sozialen Netzwerken, Fotos, Videos, MP3-Dateien, Blogs, Suchmaschinen, Tweets, Emails, Internet-Telefonie, Musikstreaming oder Sensoren „intelligenter Geräte“ vorkommenden Daten. Besonders interessant z.B. für Werbung, Marketing oder auch Wahlkämpfe sind dabei subjektive Äußerungen in Text- oder Wortbeiträgen aller Art, die Stimmungen oder Meinungen ausdrücken. Um letztere maschinenlesbar zu machen, werden Programme benötigt, die wertende Aussagen über Produkte, Marken u.ä. oder sogar Emotionen erkennbar machen, was technisch besonders herausfordernd ist.

Die wirtschaftliche Bedeutung von Daten wird inzwischen als so groß angesehen, dass sie neben Arbeitskraft, Ressourcen und Kapital als „vierter Produktionsfaktor“ angesehen werden. Denn der Wert von Erkenntnissen, die durch Auswertung vorhandener Daten gewonnen werden können, gilt als potentiell gewaltig. So versprechen sich Unternehmen unter anderem verbesserte Marketingmethoden oder auch neue Produktentwicklungen durch genauere Informationen über das Informations- und Konsumverhalten ihrer Kunden sowie Kostenersparnisse durch optimierte Logistikprozesse. Auch für die öffentliche Verwaltung sind die neuen Möglichkeiten interessant, wie erste Erfahrungen zeigen – z.B. beim Verkehrsmanagement in Stockholm. Dort konnten durch die Integration von Wetter- und Verkehrsdaten (Unfall- und Staumeldungen, Videos usw.), Verkehrsaufkommen und Emissionen um 20% und Fahrzeiten um 50% gesenkt werden. Auch der amerikanische Präsident Obama baute bei seinem Wahlkampf 2012 auf Big Data und beschäftigte in seinem Team fast 50 Datenanalytiker. Ihnen gelang es, mit Hilfe detaillierter Datenanalyse aus vielen Quellen die Wahlkampagne erheblich zu effektivieren, indem sie auf die Bundesstaaten und Zielgruppen konzentriert wurde, die – mit den für sie in Inhalt und Form jeweils passenden Botschaften – am ehesten überzeugt werden könnten. Auch die wissenschaftliche Forschung baut zunehmend auf die neuen Methoden der Datenanalyse. So stützen erste Erfahrungen mit Big Data-Anwendungen auf medizinischem Gebiet die Vision einer nicht mehr reaktiven, sondern präventiven und personalisierten Medizin, die durch die genaue Kenntnis individueller Risikofaktoren, subjektiver Befindlichkeiten und möglicher Nebenwirkungen verabreichter Medikamente möglich werden würde. Nach Schätzungen des McKinsey Global Institute wären durch den Einsatz von Big Data allein im

US-amerikanischen Gesundheitswesen Effizienz- und Qualitätssteigerungen im Wert von ca. 222 Mrd. Euro und für den gesamten öffentlichen Sektor in Europa von jährlich 250 Mrd. Euro möglich.

Das Besondere bei Big Data-Analysen ist vor allem die neue Qualität der Ergebnisse aus der Kombination bisher nicht aufeinander bezogener Daten. In der Regel sind dies Bestandsdaten, die zu 85% bislang technisch nicht ausgewertet werden konnten. Zu den technischen Voraussetzungen für Big Data-Analysen gehören vor allem die zwei Neuentwicklungen *MapReduce* und *Hadoop*. Letzteres ist eine Open Source-Software und Plattform, die aus einem Forschungsprojekt der Firma Yahoo hervorging und mittlerweile faktisch als Standard-Anwendung im Big Data-Bereich gilt. *Hadoop* ermöglicht es, schnell und dezentral große Datenmengen zu speichern und parallel zu bearbeiten. Dies wird durch ein Verteilsystem von Datenspeichern erreicht, durch das jeder Nutzer mit Netzanschluss große Datenmengen auf Gruppen oder Cluster von Rechnern verteilen und später wieder schnell auf sie zugreifen kann. Die eigentliche mathematische Analyse der Daten erfolgt dann durch den ursprünglich von Google entwickelten Algorithmus *MapReduce*, der sehr große Datenmengen parallel bearbeiten kann, indem sie zerlegt und auf zahlreichen Rechnern verteilt werden. Damit wird auch deutlich, dass Big Data-Analysen aufgrund der großen Datenmengen in der Regel nicht ohne dezentrale Speicherorte (*Clouds*) möglich sind. Auf der Basis der beiden frei erhältlichen zentralen Elemente von Big Data-Technologien wurden zwischenzeitlich diverse Erweiterungen und Werkzeuge entwickelt, die auch als externe Software-Dienstleistung angeboten werden, so dass die Nutzung finanziell und organisatorisch einer größeren Anzahl von Unternehmen möglich wird.

Neben den unbestritten großen Potentialen von Big Data für Wirtschaft, Wissenschaft und Gesellschaft werden in der zunehmend intensiver geführten Debatte über die neuen Möglichkeiten auch kritische Stimmen laut. Denn gerade die Nutzung der für Big Data besonders interessanten personenbezogenen Daten kollidiert mit zentralen europäischen datenschutzrechtlichen Prinzipien, wie dem Recht auf informationelle Selbstbestimmung, dem Schutz personenbezogener Daten und der Zweckbindung von erhobenen Daten, kodifiziert in der Europäischen Grundrechtecharta und dem Bundesdatenschutzgesetz. Auch eine Pseudonymisierung oder Anonymisierung von Daten ist hier nur von begrenztem Nutzen, weil die für Big Data typische Kombination vieler Datensätze häufig eine De-Anonymisierung ermöglicht. Einige Beobachter richten zudem den Blick auf die möglichen Auswirkungen auf unser wissenschaftliches Weltbild, in dem die Ergründung und die Wichtigkeit kausaler Zusammenhänge nun zunehmend durch statistische Korrelationen abgelöst werden könnte. Und schließlich bleibt zu fragen, wo in einer Welt, in der Entscheidungen zunehmend von datenverarbeitenden Maschinen dominiert werden, die menschliche Urteilsfähigkeit oder auch Intuition ihren Platz finden kann. Denn diese könnte manchmal auch nahelegen, bei bestimmten Entscheidungen eben gerade nicht der Datenlage zu folgen.

4. Das „Sozialkredit-System“ – Big Data in der Volksrepublik China

Auch in der Volksrepublik China spielt Big Data eine wichtige Rolle, wobei im Hinblick auf die vorstehende Fragestellung insbesondere personenbezogene Daten eine große Bedeutung haben. Derzeit wird dort ein umfassendes sogenanntes Sozialkredit-System auf der Basis von Big Data aufgebaut, das von einzelnen Autoren auch als „Gesellschaftliches Bonitätssystem“² bezeichnet

2 Vgl. Meissner, Mirjam: Chinas Gesellschaftliches Bonitätssystem. Marktregulierung mit Hilfe von Big Data hat weitreichende Folgen für Unternehmen in China; merics – Mercator Institute for China Studies; Berlin 2017.

wird. Dieses Online-Rating bzw. Scoring-System greift auf verschiedene Datenbanken zu, die beispielsweise Personendaten zur Kreditwürdigkeit, Strafregisterangaben und Informationen zum sozialen und politischen Verhalten von Unternehmen, Personen und Organisationen, wie z.B. Nichtregierungsorganisationen, beinhalten. Je nach Verhalten wird jeder Bürger dazu auf seinem Punkte-Konto bewertet, das seine Reputation ausweist.³

Der Aufbau des Systems wurde 2014 von der chinesischen Regierung beschlossen, bis 2020 soll es als Basisversion auf nationaler Ebene in Betrieb sein. Bis dahin sollen alle privaten und staatlichen Datenbanken in China miteinander verbunden sein. Ziel der Kommunistischen Partei ist es, jegliches Verhalten seiner Bürger zu erfassen, zu bewerten, zu belohnen oder zu sanktionieren, um die Ordnung des Marktes und die Ordnung in der Gesellschaft im Sinne des Leitbildes von Chinas Staatspräsident Xi Jinping, der „harmonischen Gesellschaft“, sicherzustellen.⁴

Im Rahmen diese Initiative gibt es derzeit mehrere kommunale Pilotprojekte. So hat seit 2014 die chinesische Küstenstadt Rongcheng damit begonnen, ein Sozialkredit-System einzuführen. Dort sammeln zahlreiche Stellen der öffentlichen Verwaltung und andere in das System eingebundene Institutionen Daten zum Familienstand, Strafregister, Verkehrsdelikten, der Kredithistorie, Informationen der Finanzbehörden und Sozialkassen oder auch Informationen aus Mobilfunk-Verträgen bei den staatlichen Telekommunikationsunternehmen. „Die Datenmenge, die der Staat über seine Bürger zusammen trägt, ist riesig.“⁵ Bereits heute werden für Vergehen wie Schuldensäumigkeit, „Schwarzfahren“ und anderen finanzielle Delikte staatliche Sanktionen in Form von Verboten für die Nutzung des Schnellzug- oder Flugverkehrs verhängt. Diese Strafe wurde 2016 rund 6,7 Millionen Mal verhängt. In das Sozialkredit-System fließen nun jedoch auch das gesellschaftliche und moralische Verhalten der Bürger in die Bewertung mit ein. „Alle Informationen sollen perspektivisch ein großes Ganzes ergeben. Der gläserne Bürger, über den alles bekannt ist.“⁶

Laut einer Mercator-Studie hat das geplante landesweite Sozialkredit-System zudem eine wirtschaftliche Komponente: Im projektierten Jahr der Fertigstellung 2020 sollen auch große Kon-

3 Vgl. State Council of the People's Republic of China: Planning Outline for the Construction of a Social Credit System (2014-2020); June 14 2014, GF No. (2014)21; in: China Copyright and Media. The law and policy of media in China – edited by Rogier Creemers; 14.06.2014, Online-Eintrag aktualisiert: 25.04.2015; URL: <https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/> (Zugriff: 11.01.2017). Dabei ist darauf hinzuweisen, dass bereits seit dem Jahr 2007 Planungen zum Aufbau eines Sozialkredit-Systems von offizieller Seite verlautbart worden sind. *Quelle*: State Council General Office: Some Opinions concerning the Construction of a Social Credit System; 23 March 2007, GBF No. (2007)17; in: Ebenda; 23.03.2007, Online-Eintrag aktualisiert: 26.03.2016; URL: <https://chinacopyrightandmedia.wordpress.com/2007/03/23/state-council-general-office-some-opinions-concerning-the-construction-of-a-social-credit-system/> (Zugriff: 11.01.2017).

4 Dorloff, Axel: Chinas Sozialkredit-System. Auf dem Weg in die IT-Diktatur; in: Deutschlandfunk Kultur (Hrsg.): Weltzeit; 05.09.2017; URL: http://www.deutschlandfunkkultur.de/chinas-sozialkredit-system-auf-dem-weg-in-die-it-diktatur.979.de.html?dram%3Aarticle_id=395126 (Zugriff: 11.02.2017).

5 Ebenda.

6 Ebenda.

zerne in das System integriert werden. Kernstück ist hierbei die massenhafte Erhebung und Bewertung von allgemeinen Unternehmensdaten, Informationen über die Einhaltung staatlicher Vorschriften und selbst von Echtzeitdaten zum Unternehmensverhalten. Letztlich soll das Sozialkredit-System Daten von allen staatlichen Ebenen, Branchenverbänden und gewerblichen Ratingagenturen auf einer von der Zentralregierung kontrollierten Plattform zusammenführen.⁷

Die Studie prognostiziert: „Angesichts der extrem schnellen Digitalisierung in China erscheint das zukünftige Potenzial einer Datenerhebung durch Echtzeitüberwachung nahezu unbegrenzt: das Gesellschaftliche Bonitätssystem könnte über Internethandelsplattformen Daten über die Vertrauenswürdigkeit von Unternehmen bei Onlinegeschäften erhalten, zum Beispiel hinsichtlich Zahlungsverhalten, Lieferung, Produktqualität und Kundenzufriedenheit. [...] Die Anwendung von Echtzeit-Überwachungssystemen wird bereits in Pilotprojekten getestet. Sollten sich diese als erfolgreich erweisen, kann das Gesellschaftliche Bonitätssystem [Sozialkredit-System, *Anm. des Autors*] unverzüglich und automatisch auf das Fehlverhalten eines Unternehmens reagieren: eine niedrigere Bewertung und damit verbundene Strafen könnten innerhalb von Sekunden erfolgen, als Reaktion auf einen erheblichen Zahlungsverzug im Internethandel, auf einen plötzlichen und über das zulässige Maß hinausgehenden Anstieg des Energieverbrauchs oder auf eine bestimmte Anzahl von Verkehrsbußgeldern eines Transportunternehmens.“⁸

Erst durch die neuen technischen Möglichkeiten von Big Data ist es der chinesischen Regierung möglich, ein umfassendes digitales System der Kontrolle von Markt und Gesellschaft einzuführen. Im Falle einer erfolgreichen landesweiten Einführung des Sozialkredit-Systems im Jahr 2020 haben Chinas Behörden damit auch im Internetzeitalter eine effektive Möglichkeit der Kontrolle, um insbesondere systemkritische Blogger, Autoren und Aktivisten bei abweichenden Meinungen zensieren und unterdrücken zu können.

7 Meissner: Chinas Gesellschaftliches Bonitätssystem..., a.a.O., S. 4 f.

8 Ebenda.