

# Big Data

DOI 10.1007/s11576-014-0428-0

## Die Autoren

**Professor Vasant Dhar**  
Stern School of Business  
New York University  
44 West 4th Street  
New York, NY 10012  
USA  
[vdhar@stern.nyu.edu](mailto:vdhar@stern.nyu.edu)

**Professor Dr. Matthias Jarke** (✉)  
Information Systems  
RWTH Aachen University  
Ahornstraße 55  
52074 Aachen  
Deutschland  
und  
Fraunhofer FIT  
Schloss Birlinghoven  
53754 Sankt Augustin  
Deutschland  
[jarke@dbis.rwth-aachen.de](mailto:jarke@dbis.rwth-aachen.de)

**Dr. Jürgen Laartz**  
McKinsey & Company Inc.  
Budapester Str. 46  
10787 Berlin  
Deutschland  
[juergen.laartz@mckinsey.com](mailto:juergen.laartz@mckinsey.com)

Online publiziert: 2014-08-05

This article is also available in English via <http://www.springerlink.com> and <http://www.bise-journal.org>: Vasant Dhar, Matthias Jarke, Jürgen Laartz (2014) Big Data. Bus Inf Syst Eng. doi: [10.1007/s12599-014-0338-0](https://doi.org/10.1007/s12599-014-0338-0).

© Springer Fachmedien Wiesbaden 2014

Physiker und Astronomen waren mit die ersten, die mit massiven Datenmengen umgehen musste, um Antworten auf elementare Fragen über den Ursprung und Zustand des Universums zu finden. In den späten 1990er Jahren initiierte Jim Gray bei Microsoft Research ein ambitioniertes Astronomiedatenprojekt (Microsoft 2009), um damit seine These zu demonstrieren, dass datengetriebene Wissenschaft („data science“) gleichwertig neben die klassischen wissenschaftlichen Methoden der theoretischen Analyse, des Experiments und der Simulation treten werde. Heute findet diese damals gewagt erscheinende These allgemeine Anerkennung. Aber was bedeutet sie, worin unterscheidet sie sich von anderen Methoden und welche Fähigkeiten brauchen „data scientists“, um in einer mit Daten geradezu überschwemmten Welt produktiv zu sein?

Data-Science wird heute als die Wissenschaft von der systematischen Extraktion von Wissen aus Daten verstanden (Dhar 2013). Data-Science will dabei *handlungsleitendes* Wissen entdecken und extrahieren, also nicht nur *erklären*, was passiert, sondern Vorhersagen treffen und ggf. Entscheidungen vorbereiten oder gar automatisch fällen. Maschinen sollen die Fähigkeit gewinnen und zumindest menschlichen Entscheidungsträgern vermitteln, die richtigen Fragen an die Flut von Daten zu stellen, die heute von Menschen, Geräten und Sensoren erzeugt werden. „Big Data“ wird damit zum Kern der Data-Science.

Bereiche, in denen bereits heute großer Einfluss von Big Data absehbar ist, sind die Sozialwissenschaften, die Wirtschaft, das Gesundheitswesen, Unterhaltung und Sport. Nie zuvor waren wir sozial so stark vernetzt (und verfolgt) wie heute, nie zuvor hat Handel in der heutigen Form stattgefunden, bei der jeder Schritt potenziell für immer unlöslich aufbewahrt wird. Dieser digitale „Auspuff“ wächst exponentiell. Unser Leben verändert sich für immer durch die Rolle von Big Data als Maschine von Entdeckungen, Entscheidungen, Politiken, aber auch als Zerstörer eines Konzepts von Privatheit, das sich die Menschheit über Jahrhunderte erarbeitet hatte.

Ein kurzer Rückblick auf die Entstehungsgeschichte: In den 1990er Jahren wurden erstmals umfangreiche historische Unternehmensdaten in Data-Warehouses systematisch gesammelt und analysiert (Jarke et al. 2003). Im Finanzbereich ermöglichte die elektronische Verfügbarkeit feingranularer Handelsdaten erstmals automatischen Hochfrequenzhandel und erweiterte so die Datenanalyseforschung um Realzeit-Datenstromstudien. Suchmaschinen, Internethandel und soziale Medien ergänzten das Spektrum im neuen Jahrtausend schnell um Textanalyse, soziale Netzanalyse und heterogene Datenanalysen. Insgesamt hat die Digitalisierung der Geschäftsprozesse über zwei Jahrzehnte zu explodierenden Datenvolumina mit wesentlich größerer Variabilität und ständig steigender Velozität geführt. Zusätzlich zu diesen quantitativen und technischen Aspekten der „3V“ kristallisierten sich prädiktive Analyseverfahren zur Entscheidungsunterstützung oder gar Entscheidungsautomatisierung heraus.

Als Schlagwort nähert sich „Big Data“ sicherlich langsam seinem Höhepunkt im Gartnerzyklus (Dhar 2013); nichtsdestoweniger wird sich der inhaltliche Trend in den kommenden Jahren sogar noch beschleunigen. Das Internet der Dinge wird durch Milliarden neuer Sensoren, die Bilder, Audio und Video erfassen und in immer schnelleren Multicore-Prozessoren verarbeiten, Umfang und Geschwindigkeit der Datensammlung nochmals um Größenordnungen beschleunigen; das gleiche gilt für immer reichhaltigere Simulationsergebnisse im Ingenieursbereich, die verwaltet und analysiert werden müssen.

Die rasche Ausbreitung von Datenanalyseverfahren mit wachsender Qualität bei sinkenden Kosten ändert die Wirtschaftslandschaft grundlegend. Die Nachfrage nach Data-Scientists übersteigt deutlich das Angebot (Dhar 2013; Davenport und Patil 2012). Visualisierung, drahtlose Kommunikation und Cloud-Infrastrukturen ändern auch die Wettbewerbsstrukturen. Die Analysewerkzeuge helfen Unternehmen, Transparenz in ihre Geschäftsprozesse zu bringen, zu experimentieren, um Handlungsoptionen zu entdecken, ihre Variabilität und Performanz zu erhöhen, Kundengruppen zu segmentieren und neue Produkte und Dienstleistungen anzubieten. Ausgehend vom Vorbild von Pionieren wie Amazon oder Google lernen Firmen auch, mit Big Data Tests und Experimente durchzuführen, um ihre Produkte und Dienste zu verbessern.

Viele Unternehmen analysieren in Realzeit Blogs, Nachrichten und Twitter-Messages auf subtile Stimmungsänderungen hin, die Produkt- und Preisstrategien beeinflussen können.

Innovative Unternehmen nutzen zudem die Erfassung von immer mehr Daten auf operationaler Ebene, um aus Informationsprodukten zusätzliches Einkommen zu generieren. Zusammenfassend transformieren Big Data und seine speziellen Analysemethoden damit die Art und Weise des Tagesgeschäfts in fast jedem Unternehmen. Viele Unternehmen verändern aufgrund des neuen evidenzbasierten Ansatzes ihre Strategie weg von einer an ständiger Verbesserung und schlanker Organisation orientierten Zielsetzung hin zu einer experimentierfreudigen lernenden Organisation. Dieses Potenzial wird zwar weithin erkannt, aber organisatorische wie technologische Komplexitäten und der Wunsch nach Perfektion verzögern oft die Umsetzung.

In diesem Schwerpunktheft verfolgen wir eine interdisziplinäre Sichtweise des Themas Big Data. Viele der Herausforderungen zeigt der Übersichtsartikel „Big Data – Eine interdisziplinäre Chance für die Wirtschaftsinformatik“ auf. Er fasst eine breit angelegte BMBF-Studie aus Sicht der Disziplinen Wirtschaftsinformatik (Krcmar und Schermann, TU München), Datenmanagement (Markl und Hensen, TU Berlin) und Rechtswissenschaft (Hoeren, Bitter und Buchmüller, Universität Münster) zusammen und dürfte für viele Leser besonderen Wert durch seine klare Herausarbeitung von Big-Data-Herausforderungen im europäischen und speziell deutschen Rechtsrahmen bieten.

Nach mehreren Begutachtungs- und Überarbeitungsrunden wurden drei der eingereichten Papiere für das Schwerpunktheft akzeptiert, in denen das Thema Big Data aus drei betriebswirtschaftlichen Perspektiven betrachtet wird: betriebliche Entscheidungsstile, Umgang mit Unsicherheit und erforderliche Personalkompetenzen.

In „Big Data und Informationsverarbeitung in organisatorischen Entscheidungsprozessen – Eine multiple Fallstudie“ untersuchen Kowalczyk und Buxmann (TU Darmstadt) die Beziehungen zwischen dem Kontext von Entscheidungen, der Nutzung von Big Data und den grundlegenden Informationsmodellen der Entscheidungsfindung, somit einem der ältesten Themen der Wirtschaftsinformatik. Die Beobachtung, dass eine rein technische Anwendung von Big-Data-Methoden durch Informationsüberflutung, widersprüchliche Aussagen und mangelnde Akzeptanz klare Entscheidungen eher behindern kann, führt zur Empfehlung, gemischte Vorgehensweisen aus Data-Science und klassischen organisatorischen Entscheidungsverfahren je nach Kontext anzupassen.

Die zweite Arbeit, „Informationsunschärfe in Big Data – Erkenntnisse aus sozialen Medien in Stadtgebieten“ von Bendler, Wagner, Brandt und Neumann (Universität Freiburg), macht einen Vorschlag zur Reduktion der inhärenten Unsicherheit, die aus der wachsenden Vielzahl von Datenquellen mit jeweils einzelnen zweifelhafter Datenqualität in vielen Big-Data-Anwendungen resultiert. Die Autoren nutzen Twitter-Kurznachrichten aus dem Raum San Francisco als Beispiel, um den Nutzen von Raum-Zeit-Mustern zu demonstrieren, die durch maschinelles Lernen einer Kombination von Geodaten und sehr großen Nutzerzahlen entstehen. Grundidee ist dabei die automatische Erzeugung und Validierung kausaler Beziehungen zwischen einzelnen Kategorien geographischer Interessenpunkte und den typischen Nachrichtenmustern in ihrer Umgebung.

Der Mangel an qualifiziertem Fachpersonal kann sich als eines der größten Fortschrittshemmnisse bei Big Data erweisen. Aber welche Fähigkeiten sind eigentlich erforderlich, und wie unterscheiden sie sich von den traditionellen Anforderungen im etablierteren Bereich Business Intelligence? Debortoli, Müller und vom Brocke (Universität Liechtenstein) analysieren Stellenanzeigen, um diese Frage zu beantworten. Die Hauptbeobachtung in ihrem Papier „Vergleich von Kompetenzanforderungen an Business-Intelligence- und Big-Data-Spezialisten – Eine Text-Mining-Studie auf Basis von Stellenausschreibungen“ ist, dass Big-Data-Projekte offenbar bisher forschungs- und fachkräfteintensiver sind als die etablierten BI-Verfahren, die lediglich den Umgang mit Standardprodukten und gut bekannten Analyseverfahren erfordern. Zusätzlich bevorzugen Anwenderunternehmen im Bereich Big Data offenbar Open-Source-Methoden, während traditionelles BI meist mit proprietären Standardprodukten bekannter Player arbeitet.

Interviews mit zwei bekannten deutschen Forschern und Unternehmern vervollständigen das Schwerpunkthema. Der Kernphysiker und Unternehmensgründer Michael

Feindt (KIT und Blue Yonder) stellt eine radikale Lösung für den Umgang mit Unsicherheit und Entscheidungsstilen vor, indem er Anwendungsbereiche von Präskriptiver Analyse heraushebt, bei der Entscheidungen in digitalen Geschäftsanwendungen mit Big Data vollständig automatisiert werden. Forschungsmanager Stefan Wrobel (Fraunhofer IAIS und Universität Bonn) illustriert die Themen Interdisziplinarität und Fachkräftekompetenz am Beispiel der Fraunhofer-Allianz Big Data, der derzeit größten Organisation der angewandten Forschung zu diesem Thema in Europa.

## Danksagung

Die Herausgeber danken allen Autoren und Gutachtern der Beiträge zu diesem Heft, sowie auch dem Hauptherausgeber und seinem Team für die Anleitung und Unterstützung in der Vorbereitung.

## Literatur

- Davenport TH, Patil DJ (2012) Data scientist: the sexiest job of the 21st century. *Harv Bus Rev* 90(10):70–76
- Dhar V (2013) Data science and prediction. *Commun ACM* 56(12):64–73
- Jarke M, Lenzerini M, Vassiliou Y, Vassiliadis P (2003) *Fundamentals of data warehouses*, 2. Aufl. Springer, Heidelberg
- Microsoft (Hrsg) (2009) *The fourth paradigm: data-intensive scientific discovery*. Microsoft Research, Redmond