

Big Data

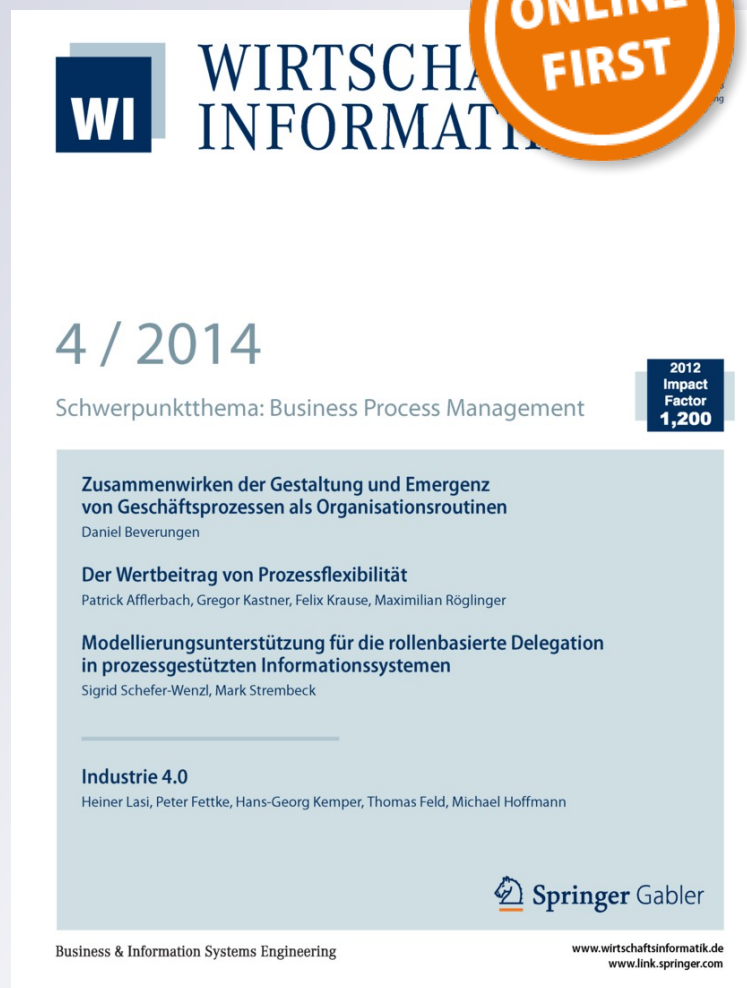
**Michael Schermann, Holmer Hensen,
Christoph Buchmüller, Till Bitter,
Helmut Krcmar, Volker Markl &
Thomas Hoeren**

WIRTSCHAFTSINFORMATIK

ISSN 0937-6429

Wirtschaftsinf

DOI 10.1007/s11576-014-0434-2



Your article is protected by copyright and all rights are held exclusively by Springer Fachmedien Wiesbaden. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Big Data

Eine interdisziplinäre Chance für die Wirtschaftsinformatik

Mit "Big Data" werden Technologien beschrieben, die nicht weniger als die Erfüllung eines der Kernziele der Wirtschaftsinformatik versprechen: die richtigen Informationen dem richtigen Adressaten zur richtigen Zeit in der richtigen Menge am richtigen Ort und in der erforderlichen Qualität bereitzustellen. Am Beispiel des Phänomens „Big-Data-Hybris“ werden technische, wirtschaftliche und rechtliche Voraussetzungen zur Erfüllung dieses Versprechens diskutiert. Auf Grund ihres interdisziplinären Selbstverständnisses ist die Wirtschaftsinformatik ideal positioniert, um Big Data kritisch zu begleiten und Erkenntnisse für die Erklärung und Gestaltung innovativer Informationssysteme in Wirtschaft und Verwaltung zu nutzen – unabhängig davon, ob sich Big Data nun tatsächlich als eine disruptive Technologie erweist oder doch nur eine flüchtige Modeerscheinung ist.

DOI 10.1007/s11576-014-0434-2

Die Autoren

Dr. Michael Schermann (✉)
Prof. Dr. Helmut Krcmar
 Lehrstuhl für Wirtschaftsinformatik
 (I17)
 Technische Universität München
 Boltzmannstraße 3
 85748 Garching
 Deutschland
michael.schermann@in.tum.de
krcmar@in.tum.de

Dr. Holmer Hemsén
Prof. Dr. Volker Markl
 Fachgebiet Datenbanksysteme und
 Informationsmanagement (DIMA)
 Technische Universität Berlin
 Einsteinufer 17
 10587 Berlin
 Deutschland
holmer.hemsén@tu-berlin.de
volker.markl@tu-berlin.de

Christoph Buchmüller
Till Bitter
Prof. Dr. Thomas Hoeren
 Institut für Informations-,
 Telekommunikations-
 und Medienrecht
 Westfälische Wilhelms-Universität
 Münster
 Leonardo-Campus 9
 48149 Münster
 Deutschland
christophbuchmueller@uni-muenster.de
till.bitter@uni-muenster.de
hoeren@uni-muenster.de

Eingegangen: 2014-04-01
 Angenommen: 2014-05-20
 Angenommen durch die Herausgeber des Schwerpunktthemas.

This article is also available in English via <http://www.springerlink.com> and <http://www.bise-journal.org>: Schermann M, Hemsén H, Buchmüller C, Bitter T, Krcmar H, Markl V, Hoeren T (2014) Big Data. An Interdisciplinary Opportunity for Information Systems Research. *Bus Inf Syst Eng*. doi: 10.1007/s12599-014-0345-1.

© Springer Fachmedien Wiesbaden 2014

1 Einleitung

Der Begriff „Big Data“ fasst technologische Entwicklungen im Bereich der Datenhaltung und -verarbeitung zusammen, welche es ermöglichen, immer größere Datenmengen unterschiedlichster Formate zu integrieren und diese in immer kürzerer Zeit zu verarbeiten (Chen et al. 2012; Lycett 2013). Big Data bietet so die Chance, die exponentiell steigenden Datenvolumina, welche durch die zunehmende Ubiquität von Informations- und Kommunikationstechnologie hervorgerufen werden, weiterhin beherrschen und vor allem wertschöpfend einsetzen zu können (McAfee und Brynjolfsson 2012). Glaubt man Herstellern, Forschungsprojekten und ersten Anwendungsbeispielen, verspricht Big Data nicht weniger als die richtigen Informationen dem richtigen Adressaten zur richtigen Zeit in der richtigen Menge am richtigen Ort und in der erforderlichen Qualität bereitzustellen zu können (Lycett 2013; Krcmar 2009).

Es ist daher nicht verwunderlich, dass Big Data zu einem der Modebegriffe der jüngsten Vergangenheit geworden ist und als neue disruptive Technologie diskutiert wird (Eberspächer und Wohlmuth 2013). Demgegenüber stehen allerdings auch kritische Stimmen, die den Innovationsgrad von Big Data anzweifeln und darin vor allem die Bemühungen von Technologieunternehmen sehen, neue Produkte und Dienstleistungen zu verkaufen (Buhl et al. 2013).

Steininger et al. (2009) folgend, behaupten wir aber, dass es eine Chance für die Wirtschaftsinformatik ist, sich mit Big Data auseinanderzusetzen. Auf Grund der rasanten technologischen Entwicklung von Big Data nehmen wir in diesem Beitrag Abstand davon, einen vollständigen Überblick zur Diskussion von Big Data zu geben. Stattdessen beleuchten wir ein konkretes Phänomen: Big-Data-Hybris, d.h., die organisatorische Selbstüberschätzung auf Grund des Einsatzes von Big Data (Lazer et al. 2014). Anhand dieses Phänomens zeigen wir, wie Big Data eine Chance für die Wirtschaftsinformatik sein kann, unabhängig davon ob Big Data nun tatsächlich eine disruptive Technologie oder doch nur eine flüchtige Modeerscheinung ist.

Der Beitrag ist wie folgt gegliedert. Zunächst stellen wir das Phänomen der Big-Data-Hybris vor. Anschließend beleuchten wir dieses Phänomen aus technologischer, rechtlicher und wirtschaftlicher Perspektive. Aus dieser Betrachtung stellen wir ein Plädoyer für den Umgang mit Big Data als Chance für die Wirtschaftsinformatik zur Diskussion.

2 Das Phänomen der Big-Data-Hybris

Typischerweise wird Big Data immer im Kontext völlig neuer Produkte und Dienstleistungen sowie innovativen betrieblichen Anwendungssystemen diskutiert (Eberspächer und Wohlmuth 2013; Schroeck et al. 2012). Kern dieser Beispiele sind immer die erheblichen Chancen aus der Verarbeitung von riesigen Datenvolumina unterschiedlichster Formate in immer kürzeren Zeiträumen (The Economist Intelligence Unit 2014). Dahinter steht die Annahme, dass die Verarbeitung von mehr Daten zu besseren Entscheidungen in Wirtschaft, Verwaltung, Politik und sogar im Privatleben führen. So argumentiert beispielsweise McAfee (2013), dass bisherige Entscheidungsprozesse auf Grund ihrer menschlichen Komponenten (z. B. aufgrund von Intuition oder „Bauchgefühl“) fehlerbehaftet sind und die Chance von Big Data darin liegt, datengestützte Entscheidungsprozesse ohne menschliche Interpretation etablieren zu können.

Ein besonders plastisches Beispiel dafür stellt das Projekt Google Flu Trends, eine Plattform zur Vorhersage von Grippewellen, dar (Ginsberg et al. 2009). Google hat historische Daten zu Meldungen von Grippefällen des amerikanischen Zentrums für Krankheitskontrolle mit eigenen Daten zu Häufigkeit, Zeitpunkten und Ort von Suchanfragen korreliert. Google identifizierte etwa 45 Suchbegriffe als Vorhersagekriterien für Grippewellen mit dem Ziel, in Echtzeit die Entstehung und Verbreitung von Grippewellen zeigen und vorhersagen zu können (Ginsberg et al. 2009).

Jüngste Studien zeigen aber, dass das Vertrauen in die Fähigkeit, schnell große heterogene Datenmengen zu verarbeiten, zum Teil noch verfrüht ist. Auch hier dient Google Flu Trends als plastisches Beispiel. So zeigen zum Beispiel Lazer et al. (2014), dass die Vorhersagen von Google Flu Trends teilweise nicht zutreffend sind. So hat Google Flu Trends zum Teil das Ausmaß von Grippewellen um 50 Prozent überschätzt und nichtsaisonale Grippeausbrüche gar nicht erkannt (Lazer et al. 2014).

Als Grund dafür wird die unbewusste Selbstüberschätzung auf Grund von Big Data („Big-Data-Hybris“) angeführt. Werden traditionelle Entscheidungsverfahren durch Big Data ersetzt, ohne jedoch Herausforderungen wie Konstruktvalidität, Reliabilität der Daten und den Kontext der Daten zu berücksichtigen, kann dies zu einer Selbstüberschätzung und damit zu potenziell dramatischen Fehlentscheidungen führen.

Lazer et al. (2014) führen das Phänomen der Big-Data-Hybris vor allem darauf zurück, dass bei den viel diskutierten Paradebeispielen von Big Data, Daten (Suchbegriffe, Inhalte von sozialen Netzwerken, etc.) zugrunde liegen, deren Qualität schwer einschätzbar ist. Auch wird oftmals ein Schwerpunkt auf die Verarbeitung großer, aber homogener Datenmengen gelegt (Suchanfragen im Fall von Google Flu Trends). Die wesentlich schwierigere und semantisch komplexere Integration unterschiedlichster Datenquellen wird hingegen in den meisten Fällen auf Grund des Aufwands zurückgestellt.

Ebenso basieren viele Pilotbeispiele auf öffentlich verfügbaren Daten aus sozialen Netzwerken (z. B. Google, Facebook oder

Twitter). Hier ist aber zu berücksichtigen, dass diese Unternehmen ihre Algorithmen ständig weiterentwickeln. Es stellt sich daher die Frage, inwieweit Daten über die verschiedenen Versionen der Algorithmen hinweg überhaupt verglichen werden können (Lazer et al. 2014).

Wir argumentieren, dass der systematische Umgang mit Big Data und damit die Vermeidung von Big-Data-Hybris einen interdisziplinären Ansatz benötigt, der technologische, rechtliche und wirtschaftliche Fragestellungen kombiniert. Nur so kann Big Data effektiv für die Gestaltung von innovativen Informationssystemen in Wirtschaft und Verwaltung genutzt werden.

3 Big Data: eine interdisziplinäre Analyse¹

Im Kern der Diskussion um Big Data steht die Chance, die richtigen Informationen dem richtigen Adressaten, zur richtigen Zeit in der richtigen Menge am richtigen Ort und in der erforderlichen Qualität bereitstellen zu können (The Economist Intelligence Unit 2014; Krcmar 2009). Im Folgenden beleuchten wir technologische, rechtliche und wirtschaftliche Voraussetzungen für die Realisierung dieser Chance.

3.1 Die technologische Voraussetzung: nutzerfreundliche Transparenz durch deklarative Systeme für Big Data

Die technologische Voraussetzung besteht darin, effektive Systeme für Big Data bereitzustellen, die es ermöglichen, Datenanalysen sicher und effektiv einem breiten Anwenderkreis zu öffnen.

Big Data erhöht die Komplexität der Methoden und Verfahren zur Analyse von Daten. Dazu werden neue Verfahren zur Erfassung, Integration und Analyse von Big Data benötigt. Hier stehen die Entwicklung und Anwendung komplexer statistischer Algorithmen, Verfahren des maschinellen Lernens, der linearen Algebra, Signalverarbeitung sowie des Data-Minings, Text-Minings, Graph-Minings, Video-Minings und der visuellen Analyse im Mittelpunkt des Interesses. Solche Verfahren gehen über die klassischen Operationen der relationalen Al-

¹Die Ergebnisse und Erkenntnisse in diesem Abschnitt basieren auf einer Studie, welche die Autoren im Auftrag des Bundesministeriums für Wirtschaft und Energie im Jahr 2013 durchgeführt haben. Diese Studie kann unter folgendem Link abgerufen werden: http://www.dima.tu-berlin.de/menue/forschung/big_data_management_report/.

gebra hinaus, welche in heutigen Datenbanksystemen realisiert sind (Klein et al. 2013).

Zentrales Ziel dabei ist es, dass entscheidende Methoden und Verfahren nicht nur durch einige wenige Spezialisten, sondern durch eine breite Schicht von Nutzern verwendet werden können. Eine breite Nutzung von Big Data erfordert jedoch, dass die Analyse von Daten keine Kenntnisse in der Systemprogrammierung bzw. Parallelverarbeitung erfordern darf, wie es bei derzeitigen Anwendungen der Fall ist.

Es wird daher an der Entwicklung von deklarativen Sprachen für die Spezifikation, Optimierung und Parallelisierung von komplexen Datenanalysen gearbeitet, die über die klassischen Möglichkeiten bestehender Sprachen, wie z. B. SQL, XPath und XQuery für XML Daten oder SparQL für RDF-Daten hinausgehen (American National Standards Institute 1992; Berglund et al. 2010; Boag et al. 2011; Prud'hommeaux und Seaborne 2008). Eine deklarative Spezifikation von Datenanalysen bedeutet, dass die Beschreibung des Analyseproblems von seiner Ausführung entkoppelt wird.

Für Big Data gibt es derzeit noch keine deklarative Spezifikation von Datenanalyseprogrammen, die auf beliebigen Rechnersystemen skalieren und datenunabhängig hinsichtlich der Speicherung und statistischen Eigenschaften automatisch optimiert und parallelisiert werden (Agrawal et al. 2012). Die Spezifikation einer solchen deklarativen Sprache für Big Data erfordert neben den klassischen Operationen der relationalen Algebra auch die Verwendung von komplexen benutzerdefinierten Funktionen sowie die Spezifikation von iterativen Algorithmen. Erst dadurch können beispielsweise Verfahren der linearen Algebra, des maschinellen Lernens, der Sprach-, Video- und Signalverarbeitung in die deklarative Beschreibung von Big-Data-Analysen integriert werden.

Für die Spezifikation und Verarbeitung von benutzerdefinierten Funktionen über Datensätze oder Gruppen von Datensätzen wurde das map/reduce-Paradigma für Big Data eingeführt (Dean und Ghemawat 2004). Dieses funktionale Programmiermodell besteht aus zwei Funktionen zweiter Ordnung, wobei beliebige Funktionen erster Ord-

nung zur Transformation oder Selektion von Datensätzen (map) sowie zur Aggregation (reduce) verwendet werden können. Das populäre Big-Data-System Hadoop implementiert dieses Programmiermodell und parallelisiert Analysen automatisch auf großen Rechenclustern (White 2012). Dieser Ansatz ist jedoch nur ein erster Schritt auf dem Weg zu einer deklarativen Sprache für Big Data, da beispielsweise keine automatische Optimierung und Parallelisierung komplexer, iterativer Datenanalysealgorithmen unterstützt werden. Datenanalysen bleiben damit nur einem eingeschränkten Nutzerkreis von Experten vorbehalten, die gleichzeitig Programmierkenntnisse und Kenntnisse in Datenanalyse und maschinellem Lernen besitzen.

Darüber hinaus liegt Big Data in unterschiedlichsten Qualitätsabstufungen vor. Geopositionierungsdaten haben beispielsweise eine gewisse Unschärfe auf Grund der jeweils verfügbaren Anzahl an Satelliten. Nutzer müssen daher die Herkunft der Daten und ihre Qualität zur Abschätzung der Korrektheit von Analyseergebnissen beurteilen, angemessene Stichproben ziehen und sinnvolle Modellannahmen treffen können. Ebenso kommen viel häufiger probabilistische Modelle zum Einsatz, deren Einschränkungen und Grenzen in die Entscheidungsunterstützung mit einfließen müssen.

Bei der Anwendung entsprechend komplexer Analyseverfahren ist ein iteratives Vorgehen zielführend, um die richtigen Parametereinstellungen zu ermitteln. Daher ist es notwendig, die verschiedenen Schritte der Datenanalyse transparent für den Nutzer zu gestalten, zum Beispiel mit Hilfe von interaktiven, oftmals visuell unterstützten, Werkzeugen („Visual Analytics“). Diese Werkzeuge unterstützen eine interaktive Modellierung von Annahmen, Rahmenbedingungen und Analyseinteressen, die iterativ durch entsprechende Teilanalysen unterstützt wird. Eine effektive Visualisierung kann dabei helfen, eine schnelle und genaue Bewertung der Ergebnisqualität in Abhängigkeit der Parametereinstellungen vorzunehmen (Thomas und Cook 2005).

3.2 Die rechtliche Voraussetzung: Verantwortung und Rechtssicherheit für Wertschöpfung mit Big Data²

Die rechtliche Voraussetzung besteht darin, effektive Rahmenbedingungen vorzugeben, die den Umgang mit Big Data aus gesellschaftlicher Sicht kanalisieren. Grundsätzlich wird Big Data vor allem aus der Perspektive des Datenschutzes diskutiert. Obwohl dies aus gesellschaftlicher Sicht äußerst relevant ist, zeigt das Phänomen der Big-Data-Hybris aber auch, dass urheberrechtliche, eigentumsrechtliche und haftungsrechtliche Überlegungen eine zentrale Rolle für den Einsatz von Big Data spielen müssen.

Das Urheberrecht schützt geistige Schöpfungen, die über eine ausreichende Gestaltungshöhe verfügen (§ 2 Abs. 2 UrhG). Daten erfüllen aber regelmäßig nicht die Voraussetzungen des § 2 Abs. 2 UrhG, weil sie entweder nur grundsätzlich freizuhaltende Tatsachen enthalten oder ihnen die erforderliche Individualität fehlt (Dreier und Schulze 2013, § 20 Rn. 130).

Trotzdem muss das Urheberrechtsgesetz beim Umgang mit Daten beachtet werden. Daher kann der sogenannte Sui-generis-Schutz für Datenbanken (§§ 87a UrhG) greifen. Darunter wird ein Schutz für das Investitionsrisiko der Erstellung einer Datenbank verstanden (Wandtke und Bullinger 2009, 87a Rn. 55 ff.; Dreier und Schulze 2013, § 87a Rn. 14; Gaster 1999, Rn. 476). Darüber hinaus können Urheberrechte bei der Auswertung nutzergenerierter Inhalte Dritter in sozialen Netzwerken dem Datenumgang entgegenstehen. Solchen Daten kann beispielsweise ein Schutz als Sprachwerk (§ 2 Abs. 1 Nr. 1 UrhG), Lichtbildwerk (§ 2 Abs. 1 Nr. 5 UrhG) oder Lichtbild (§ 72 UrhG) zukommen (Solmecke und Wahlers 2012).

Die Erlangung entsprechender Nutzungsrechte nach §§ 31 ff. UrhG liegt dabei zahlreichen praktischen. In digitalen Ökosystemen wird fast zwangsläufig in zahllose fremde Rechte eingegriffen, sodass eine erhebliche Anzahl vertraglicher Vereinbarungen getroffen werden müsste (Klass 2013). In der Praxis ist dies kaum möglich. Hier können beispielsweise „Fair-use-Klauseln“ nach US-Vorbild normiert werden, welche für den deutschen Rechtsraum bislang jedoch noch nicht existieren.

²Diese Fragen werden im Folgenden aus der Perspektive des deutschen Rechtsraums untersucht. Es ist aber naheliegend, dass in einer globalisierten Informationsverarbeitung international geltende Regelungen notwendig sind.

Darüber hinaus ist nach heutigem Stand der Rechtswissenschaft noch unklar, ob ein Eigentumsrecht an Daten bestehen kann. Eigentumsfähigkeit setzt voraus, dass Daten nach der Rechtsordnung einem Rechtssubjekt zugeordnet werden können. Die Rechtsprechung geht im Allgemeinen davon aus, dass ein Datenträger mit den darauf abgelegten Daten eigentumsfähig ist (OLG Karlsruhe 1996; LG Konstanz 1996). Auf Grund vernetzter Datenbestände ist aber der Rückgriff auf das Eigentum am Datenträger in der Praxis wenig hilfreich (Meier und Wehlau 1998).

Vielversprechend erscheint zur Klärung der Eigentumsfrage der Rückgriff auf das Strafrecht. So schützt beispielsweise § 303a StGB explizit Daten. Die Zuordnung des Schutzgutes zu einem Rechtssubjekt erfolgt anhand des „Skripturaktes“, also nach dem technischen Herstellungsprozess. Eine Übertragung dieser Konzeption auf das Zivilrecht ermöglicht eine eindeutige Zuordnung und damit die Möglichkeit der Etablierung eines „Dateneigentums“.

Obwohl eine solche eigentumsrechtliche Zuordnung von Daten zunächst konstruiert erscheinen mag, stellen Daten zunehmend einen Wert dar, der beispielsweise für die Bewertung von Unternehmen sowie bei insolvenzrechtlichen Fragestellungen an Bedeutung gewinnt. Würde zudem ein absolut wirkendes „Dateneigentum“ bejaht, so könnten unabhängig vom Datenschutzrecht oder Urheberrecht Ansprüche durch den „Dateneigentümer“ geltend gemacht werden, die den Umgang mit Datensätzen beeinflussen würden.

Fehlerhafte Daten oder fehlerhafte Analysen können gravierende Auswirkungen haben. Eine entsprechende Haftung eines Anbieters von Big Data setzt grundsätzlich ein Fehlverhalten voraus, beispielsweise durch die Übermittlung fehlerhafter Daten.

Im Falle fehlerhafter Datenübermittlung zwischen einem Anbieter und seinem Kunden kommt ein Schadensersatzanspruch wegen Pflichtverletzung eines Schuldverhältnisses gem. § 280 Abs. 1 BGB in Betracht. Unabhängig von einer vertraglichen Beziehung zum Schädiger kommt für den Geschädigten auch ein deliktischer Schadensersatzanspruch in Betracht, wenn fehlerhafte Daten zur Verletzung eines Rechtsgutes im Sinne des § 823 Abs. 1 BGB führen. Allerdings ist dies regelmäßig nur dann der

Fall, wenn Verkehrssicherungspflichten auf Seiten des Anbieters verletzt werden.

Welche Anforderungen an diese Pflichten im Bereich Big Data gestellt werden, ist bislang nicht geklärt. Eine Differenzierung bietet sich mit Blick auf den Bezug der Daten an. Sollen diese Grundlage für eine nachfolgende Entscheidung sein, die unmittelbare Auswirkungen auf Leib und Leben von Menschen hat, können hohe Anforderungen an die Verkehrssicherungspflichten des Datenübersmitters gestellt werden. Bestehen hingegen nur Auswirkungen auf geringwertigere Rechtsgüter, würden strenge Verkehrssicherungspflichten ein nicht zu unterschätzendes Hemmnis für die Unternehmen darstellen.

Im Ergebnis hängt das Haftungsrisiko also erneut von den Umständen des Einzelfalles ab, wobei im Regelfall von nur geringen Verkehrssicherungspflichten ausgegangen werden kann (Reese 1994). Selbst wenn eine Verkehrssicherungspflicht vorläge, müsste ein schuldhafter Verstoß dagegen vom Geschädigten nachgewiesen werden, um mit seinem Anspruch Erfolg zu haben. Zugutekommen könnten dem Geschädigten in einem solchen Fall die Grundsätze der Produzentenhaftung gem. § 823 Abs. 1 BGB. Hierbei handelt es sich um eine Verschuldenshaftung mit Beweiserleichterungen für den Kläger (Lehmann 1992). Weil die Produzentenhaftung ursprünglich für Industrieprodukte entwickelt wurde, ist jedoch unklar, ob ihre Grundsätze auch für Big Data Anwendung finden können (Meyer 1997).

Auf Grund der heterogenen Datenquellen unterschiedlicher Qualität und des zunehmenden Einsatzes von probabilistischen Analyseverfahren stellt sich insgesamt die Frage, wann fehlerhafte Daten vorliegen. Erste Hinweise auf Kriterien dafür werden in den USA im Kontext des „Federal Data Quality Act“ diskutiert (Office of Information and Regulatory Affairs 2002). In diesen Richtlinien werden Kriterien wie Nützlichkeit, Vollständigkeit, Objektivität, Nachvollziehbarkeit und Klarheit genannt (Gasser 2003). Es zeigt sich damit, dass allgemeingültige Kriterien nicht absehbar sind. Vielmehr muss im Einzelfall entschieden werden, ob Daten „gut“ sind und somit eine Haftung mangels Fehlinformation ausgeschlossen ist.

3.3 Die wirtschaftliche Voraussetzung: neue Wertschöpfungsnetzwerke für Big Data

Obwohl der Buchwert von Facebook zum Zeitpunkt des Börsenganges nur ca. USD 6 Milliarden war, hatte Facebook am Tag des ersten Börsenganges einen Marktwert von ca. USD 104 Milliarden erreicht. Unabhängig von den konkreten Marktgegebenheiten wird damit deutlich, dass Daten an sich zu einem zentralen Baustein in neuen Geschäftsmodellen geworden sind (Mandel 2012; Mayer-Schönberger und Cukier 2013). Daher etablieren sich neue Wertschöpfungsnetzwerke rund um Big Data.

Bisher waren Daten eher Nebenprodukte der unterschiedlichsten Anwendungssysteme im Kontext von Geschäftsprozessen. Zunehmend haben Daten im Kontext von Big Data einen eigenständigen Wert. Durch Big Data können Daten zu Etablierung von Wettbewerbsvorteilen genutzt werden (Manyika et al. 2011; Mayer-Schönberger und Cukier 2013). Insbesondere durch die Integration unterschiedlicher Datenquellen sind Daten nicht mehr zweckgebunden. Ebenso entscheidet die innovative Nutzung über den zu erreichenden Wert, sodass die Daten, z. B. Inhalte sozialer Netzwerke, auch durch konkurrierende Marktakteure verwendet werden können (Mandel 2012; Mayer-Schönberger und Cukier 2013).

Organisationen gehen dabei unterschiedlich mit Daten und ihrem potenziellen Wertbeitrag um (Manyika et al. 2011). Einige Marktakteure, wie zum Beispiel Banken, Telekommunikationsdienstleister oder Krankenkassen erzeugen und sammeln Daten für andere Zwecke. Diese können nun mit Hilfe von Big-Data-Technologie auch für neue Geschäftsmodelle nutzbar gemacht werden (Mayer-Schönberger und Cukier 2013).

Drei zentrale Fähigkeiten bilden die Grundlage für neue Geschäftsmodelle mit Big Data. Erstens müssen die Akteure in der Lage sein, verschiedene Datenquellen syntaktisch, semantisch und pragmatisch kombinieren zu können. Zweitens ist es notwendig, mit anderen Akteuren zu kooperieren um die notwendigen Daten verarbeiten zu können. Drittens, ist es wichtig, Daten so zu erheben, dass sie in unterschiedlichsten Kontexten verwendet werden können. Innerhalb der unterschiedlichen Kontexte haben Daten somit auch einen unterschiedlichen Wert. Ein Beispiel hierfür sind Google Maps

und Google Street View, die als Grundlage für die geographische Kontextualisierung anderer Datenquellen dienen (Mayer-Schönberger und Cukier 2013; Manyika et al. 2011).

Während bisher die Prozesse der Datenaufbereitung, -analyse und -verwendung vor allem unternehmensinterne Aktivitäten waren, zeichnen sich durch Big Data bereits Konturen eines Wertschöpfungsnetzwerks mit spezialisierten Akteuren ab (Gustafson und Fink 2013). Unternehmen in diesem Wertschöpfungsnetzwerk etablieren Wettbewerbsvorteile durch Spezialisierung auf zentrale Kernkompetenzen und bilden mit anderen Dienstleistern Wertschöpfungsketten.

Dabei können die folgenden zentralen Rollen in Big-Data-Wertschöpfungsnetzwerken identifiziert werden (Mayer-Schönberger und Cukier 2013; Gerhardt et al. 2012):

- Datensammler sind dabei Akteure, die sich auf die Sammlung von Daten fokussieren und entsprechende Kontrolle über den Zugang solcher Daten etablieren können. Oftmals nutzen Datensammler ihre Daten auch für eigene Zwecke, stellen sie aber darüber hinaus anderen Akteuren im Wertschöpfungsnetzwerk zur Verfügung.
- Technologiehersteller sind Akteure im Markt, die notwendige Hardware und Software bereitstellen. Zum Teil stellen sie auch entsprechende Methoden und Verfahren zur Aufbereitung, Integration und Analyse zur Verfügung. Ein wesentliches Merkmal von Technologieherstellern ist, dass sie standardisierte Produkte und Dienstleistungen anbieten (Manyika et al. 2011).
- Spezialisten hingegen sind Dienstleister, die ihre Kompetenzen in der kundenspezifischen Datenaufbereitung, -integration, und -analyse zu ihrem Alleinstellungsmerkmal ausgebaut haben bzw. ausbauen. Im Allgemeinen haben Analysespezialisten keine eigenen Daten, sondern bieten ihre Kompetenzen anderen Marktakteuren, z. B. Datensammlern, an.
- Datenaggregateure sind Akteure im Markt, die sich auf die Bereitstellung bestimmter Datenanalysen für den Massenmarkt fokussieren. Ein Beispiel hierfür sind Anbieter von Verkehrsinformationen in Echtzeit, die unterschiedliche Datenquellen (Smartphones, Kartendaten, offizielle Verkehrsinformationen) aggregieren und diese als Dienstleistung anbieten (Gustafson und Fink 2013).

- Datennutzer wiederum sind Akteure, welche die Wertschöpfungspotenziale von bestimmten Daten erkannt haben und für die Entwicklung neuer Dienstleistungen oder Produkte nutzen wollen. In der Rolle des Datennutzers haben sie also eine Idee für die Nutzung von Daten, ohne notwendigerweise die Daten bzw. die Fähigkeiten zur Nutzung der Daten zu besitzen.
- Broker sind Intermediäre, die sich darauf spezialisiert haben, zwischen den verschiedenen Akteuren zu vermitteln und temporäre Wertschöpfungsketten zu etablieren. Ein prominentes Beispiel hierfür ist kaggle.com, eine Plattform, die öffentliche Wettbewerbe zu Datenanalysen ausschreibt und so unterschiedliche Akteure zusammenbringt (Carpenter 2011).
- Regulatoren sind Akteure im Markt, die sich darauf spezialisieren, regulatorische Prämissen wie beispielsweise Datenschutz, Datensicherheit und Qualität etwa im Namen von Datennutzern zu prüfen und zu zertifizieren. Regulatoren können dabei öffentliche Institutionen auf nationaler und internationaler Ebene wie auch private Organisationen sein (Gerhardt et al. 2012).

Vor dem Hintergrund des Phänomens der „Big-Data-Hybris“ wird deutlich, dass insbesondere Regulatoren und Broker zentrale Rollen im Kontext von Big Data einnehmen. Regulatoren schaffen die notwendigen qualitativen Rahmenbedingungen, während Broker vor allem Kompetenz in der Etablierung effektiver Wertschöpfungsketten mit Big Data aufbauen.

4 Drei Thesen für die wissenschaftliche Begleitung und Gestaltung einer disruptiven Technologie

Christensen (1997) folgend, ist es typisch für disruptive Innovationen, dass sie Herausforderungen aufwerfen, die unmittelbar nicht gelöst werden können. Die bisherige Diskussion zeigt, dass Big Data damit potenziell eine disruptive Technologie darstellt. Nicht nur aus technischer, sondern auch aus rechtlicher und wirtschaftlicher Perspektive deuten sich erhebliche Verschiebungen an. Gleichzeitig sind aber erhebliche zukünftige Wertschöpfungspotenziale zu erwarten. **Tabelle 1** fasst die bisher betrachteten

Aspekte von Big Data zusammen und bewertet sie hinsichtlich postulierter Eigenschaften von disruptiven Technologien (Christensen 1997).

Einerseits hat die Wirtschaftsinformatik mit ihrer interdisziplinären Forschungstradition eine besondere Chance, aus den technologischen, rechtlichen und wirtschaftlichen Entwicklungen rund um Big Data wertvolle Erkenntnisgewinne zur Erklärung und Vorhersage von Informationssystemen in Wirtschaft und Verwaltung zu erarbeiten. Andererseits eröffnet der traditionell stark ausgeprägte Fokus der Wirtschaftsinformatik auf die Methoden der Gestaltung innovativer Informationssysteme erhebliche Potenziale hinsichtlich der Relevanz der Wirtschaftsinformatik in Wirtschaft und Verwaltung (Brinkemper 1996).

Unser Plädoyer für den methodischen und interdisziplinären Umgang mit Big Data in der Wirtschaftsinformatik besteht aus den folgenden drei Thesen.

- (1) Aus- und Weiterbildungskonzepte für den effektiven und verantwortungsvollen Umgang mit Big Data in Wissenschaft und Praxis

Das Beispiel der Big-Data-Hybris zeigt, dass entsprechende Aus- und Weiterbildungskonzepte eine wesentliche Voraussetzung für den effektiven Umgang mit Big Data in der Forschung wie auch in der Praxis sind. Unabhängig von Big Data generiert die Ubiquität moderner Informationstechnologie potenziell wertvolle Datenquellen, welche für die Erkenntnisgewinnung in der Wirtschaftsinformatik genutzt werden können. Der sichere Umgang mit modernen Methoden der Datenanalyse (z. B. Methoden des maschinellen Lernens, Data-Mining, Netzwerkanalyse) und breite Kenntnisse deren Rahmenbedingungen und Schwachstellen stellen die Voraussetzung für die Nutzung von Big Data als Chance für die Wirtschaftsinformatik dar.

- (2) Entwicklung von Modellierungswerkzeugen zur integrierten Analyse und Gestaltung von Geschäftsprozessen unter Berücksichtigung von Big Data

Ebenso zeigt das Beispiel der Big-Data-Hybris, dass durch die Verknüpfung unterschiedlicher Datenquellen allein noch keine Wertschöpfung realisiert werden kann. Die Wirtschaftsinformatik hat eine langjährige Tradition in der Entwicklung und Erprobung von Modellierungswerkzeugen sowie von Methoden der Entwicklung von Informationssystemen

Tab. 1 Bewertung von Big Data als disruptive Technologie (in Anlehnung an Christensen 1997)

Eigenschaften disruptiver Technologien	Bewertung von Big Data
Die Technologie schneidet nach traditionellen Bewertungskriterien schlechter als etablierte Technologien ab.	Traditionelle Technologien der Datenanalyse sind mittels deklarativer Sprachen von einem breiten Anwenderkreis bedienbar. Big Data hingegen führt zu probabilistischen Ergebnissen, die momentan nur von Experten erstellt und interpretiert werden können.
Die Technologie führt neue Bewertungskriterien ein.	Big Data ermöglicht die Analyse großer Datenmengen mit unterschiedlichen Datenformaten mit immer kürzeren Antwortzeiten.
Die Technologie wird von etablierten Anwendern eher zurückhaltend beobachtet.	Zwar gibt es schon eine Vielzahl von Beispielen und Pilotanwendungen für Big Data. Die Kernprozesse der Unternehmen stehen dabei noch nicht im Fokus der Anwendung von Big Data.
Die Technologie wird zunächst hauptsächlich von Start-ups und Quereinsteigern angeboten und genutzt.	Anwender von Big Data sind vor allem kleine Unternehmen, welche die Technologie als Chance für neue Geschäftsmodelle sehen. Auch unter den Anbietern von Big-Data-Technologie finden sich viele Start-ups, mittlerweile aber auch etablierte Technologieanbieter.
Die Technologie führt zu Veränderungen in den Wertschöpfungsprozessen.	Es etabliert sich eine Vielzahl von neuen Rollen im Umfeld von Big Data.

und Geschäftsprozessen. Für die Wertschöpfung mit Big Data ist es notwendig, Datenanalysen und deren Einsatz in Geschäftsprozessen integriert betrachten und gestalten zu können. Zentral für die Entwicklung entsprechender Modellierungswerkzeuge ist die Berücksichtigung von Wechselwirkungen zwischen Big Data als Initiator neuer Geschäftsprozesse sowie der Geschäftsprozesse als Anforderungsgerüst für entsprechende Datenanalysen.

(3) Entwicklung belastbarer Referenzmodelle für den verantwortungsvollen Einsatz von Big Data

Wie insbesondere die rechtliche Perspektive auf Big Data gezeigt hat, benötigt Big Data neue bzw. adaptierte Vorgaben hinsichtlich des Umgangs mit Daten hinsichtlich Datenschutz und Datensicherheit. Ebenso sind aber neue Prozesse für den Umgang mit Risiken hinsichtlich Qualität, Nutzung und Verantwortlichkeiten notwendig. Nur mit Hilfe entsprechender Prozesse wird sich Big Data als vertrauenswürdige Entscheidungsgrundlage etablieren. Hier fehlt es jedoch an belastbaren Referenzmodellen, die notwendige Mechanismen, Strukturen und Prozesse beschreiben und vorschlagen.

Insgesamt ist festzuhalten, dass diese Kernthesen unabhängig von konkreten technologischen Entwicklungen sind. Sie basieren lediglich auf der Annahme, dass Daten traditioneller Herkunft und Big Data gleichermaßen eine bedeutendere Rolle für die Erklärung und Gestaltung von Informationssystemen, Ge-

schaftsprozessen und Entscheidungsprozessen spielen werden.

Literatur

- Agrawal D, Bernstein P, Bertino E, Davidson S, Dayal U, Franklin M, Gehrke J, Haas L, Halevy A, Han J, Jagadish HV, Labrinidis A, Madden S, Papakonstantinou Y, Patel JM, Ramakrishnan R, Ross K, Shahabi C, Suci D, Vaithyanathan S, Widom J (2012) Challenges and opportunities with big data: a community white paper developed by leading researchers across the United States. Computing Research Association, Washington
- American National Standards Institute (1992) American national standard for information systems: database language SQL (ANSI X3.135-1992). American National Standards Institute, Washington
- Berglund A, Boag S, Chamberlin D, Fernández MF, Kay M, Robie J, Siméon J (2010) XML path language (XPath) 2.0. World Wide Web Consortium, Cambridge
- Boag S, Chamberlin D, Fernández MF, Florescu D, Robie J, Siméon J (2011) XQuery 1.0: an XML query language. World Wide Web Consortium, Cambridge
- Brinkkemper S (1996) Method engineering: engineering of information systems development methods and tools. *Information and Software Technology* 38(4):275–280
- Buhl HU, Röglinger M, Moser F, Heidemann J (2013) Big Data: Ein (ir-)relevanter Modebegriff für Wissenschaft und Praxis? *WIRTSCHAFTSINFORMATIK* 5(2):63–68
- Carpenter J (2011) May the best analyst win. *Science* 331(6018):698–699
- Chen H, Chiang RH, Storey VC (2012) Business intelligence and analytics: from big data to big impact. *MIS Quarterly* 36(4):1165–1188
- Christensen C (1997) The innovator's dilemma: when new technologies cause great firms to fail. Harvard Business Review, Boston
- Dean J, Ghemawat S (2004) MapReduce: simplified data processing on large clusters. In: Proc 6th symposium on operating system design and implementation, San Francisco
- Dreier T, Schulze G (2013) Urheberrechtsgesetz: Urheberrechtswahrnehmungsgesetz, Kunsturhebergesetz. Beck, München
- Eberspacher J, Wohlmuth O (Hrsg) (2013) Big Data wird neues Wissen. Münchner Kreis, München
- Gasser U (2003) Information quality and the law or how to catch a difficult horse. The Berkman Center for Internet & Society Research, Harvard Law School, Cambridge
- Gaster JL (1999) Der Rechtsschutz von Datenbanken: Kommentar zur Richtlinie 96/9/EG mit Erläuterungen zur Umsetzung in das deutsche und österreichische Recht. Heymanns, Köln
- Gerhardt B, Griffin K, Klemann R (2012) Unlocking value in the fragmented world of big data analytics: how information intermediaries will create a new data ecosystem. Cisco Internet Business Solutions Group, San Jose
- Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2009) Detecting influenza epidemics using search engine query data. *Nature* 457(7232):1012–1014
- Gustafson T, Fink D (2013) Winning within the data value chain. Innosight. <http://www.innosight.com/innovation-resources/strategy-innovation/winning-within-the-data-value-chain.cfm>. Abruf am 2014-02-28
- Klass N (2013) Neue Internettechnologien und das Urheberrecht: Die schlichte Einwilligung als Rettungsanker? *ZUM – Zeitschrift für Urheber- und Medienrecht* 57(1):1–9
- Klein D, Tran-Gia P, Hartmann M (2013) Big data. *Informatik Spektrum* 36(3):319–323
- Krcmar H (2009) *Informationsmanagement*. Springer, Heidelberg
- Lazer D, Kennedy R, King G, Vespignani A (2014) The parable of Google Flu: traps in big data analysis. *Science* 343(3):1203–1205
- Lehmann M (1992) Produkt- und Produzentenhaftung für Software. *NJW – Neue Juristische Wochenschrift* 1992(28):1721–1725
- Konstanz LG (1996) Kein Schadensersatz bei Datenverlust durch Stromausfall aufgrund Leitungsbeschädigung bei Bagger-

- arbeiten. NJW – Neue Juristische Wochenschrift 1996(40):2662
- Lycett M (2013) Datafication: making sense of (big) data in a complex world. *European Journal of Information Systems* 22(4):381–386
- Mandel M (2012) Beyond goods and services: the (unmeasured) rise of the data-driven economy. Progressive Policy Institute, Washington
- Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute, Atlanta
- Mayer-Schönberger V, Cukier K (2013) Big data: a revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, New York
- McAfee A (2013) Big data's biggest challenge? Convincing people NOT to trust their judgment. <http://blogs.hbr.org/2013/12/big-datas-biggest-challenge-convincing-people-not-to-trust-their-judgment/>. Abgerufen am 2014-02-28
- McAfee A, Brynjolfsson E (2012) Big data: the management revolution. *Harvard Business Review* 2012(October):1–9
- Meier K, Wehlau A (1998) Die zivilrechtliche Haftung für Datenlöschung, Datenverlust und Datenzerstörung. NJW – Neue Juristische Wochenschrift 1998(22):1585–1591
- Meyer A (1997) Die Haftung für fehlerhafte Aussagen in wissenschaftlichen Werken. *ZUM – Zeitschrift für Urheber- und Medienrecht* 41(1):26–33
- Office of Information and Regulatory Affairs (2002) Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. Office of Management and Budget, Executive Office of the President, Washington
- Karlsruhe OLG (1996) Haftung für Zerstörung von Computerdaten. NJW – Neue Juristische Wochenschrift 1996(3):200–201
- Prud'hommeaux E, Seaborne A (2008) SPARQL query language for RDF. World Wide Web Consortium, Cambridge
- Reese J (1994) Produkthaftung und Produzentenhaftung für Hard- und Software. *DStR – Deutsches Steuerrecht* 1994(31):1121–1126
- Schroeck M, Shockley R, Smart J, Romero-Morales D, Tufano P (2012) Analytics: the real-world use of big data. IBM Global Business Services, Somers
- Solmecke C, Wahlers J (2012) Rechtliche Situation von Social Media Monitoring-Diensten – Rechtskonforme Lösungen nach dem Datenschutz- und dem Urheberrecht. *ZD – Zeitschrift für Datenschutz* 2012(12):550–554
- Steininger K, Riedl R, Roithmayr F, Mertens P (2009) Moden und Trends in Wirtschaftsinformatik und Information Systems: Eine vergleichende Literaturanalyse. *WIRTSCHAFTSINFORMATIK* 1(6):478–495
- The Economist Intelligence Unit (2014) The data storm: retail and the big data revolution. The Economist, London
- Thomas JJ, Cook KA (2005) Illuminating the path: the research and development agenda for visual analytics. IEEE Computer Society Press, Washington
- Wandtker A-A, Bullinger W (2009) Praxiskommentar zum Urheberrecht. Beck, München
- White T (2012) Hadoop – the definitive guide: storage and analysis at internet scale. O'Reilly, Sebastopol

Zusammenfassung / Abstract

Michael Schermann, Holmer Hensen, Christoph Buchmüller, Till Bitter, Helmut Krömer, Volker Markl, Thomas Hoeren

Big Data

Eine interdisziplinäre Chance für die Wirtschaftsinformatik

Mit "Big Data" werden Technologien beschrieben, die nicht weniger als die Erfüllung eines der Kernziele der Wirtschaftsinformatik versprechen: die richtigen Informationen dem richtigen Adressaten zur richtigen Zeit in der richtigen Menge am richtigen Ort und in der erforderlichen Qualität bereitzustellen. Für die Wirtschaftsinformatik als anwendungsorientierte Wissenschaftsdisziplin entstehen durch solche technologischen Entwicklungen Chancen und Risiken. Risiken entstehen vor allem dadurch, dass möglicherweise erhebliche Ressourcen auf die Erklärung und Gestaltung von Modeerscheinungen verwendet werden. Chancen entstehen dadurch, dass die entsprechenden Ressourcen zu substanziellen Erkenntnisgewinnen führen, die dem wissenschaftlichen Fortschritt der Disziplin wie auch ihrer praktischen Relevanz dienen.

Aus Sicht der Autoren ist die Wirtschaftsinformatik ideal positioniert, um Big Data kritisch zu begleiten und Erkenntnisse für die Erklärung und Gestaltung innovativer Informationssysteme in Wirtschaft und Verwaltung zu nutzen – unabhängig davon, ob Big Data nun tatsächlich eine disruptive Technologie oder doch nur eine flüchtige Modeerscheinung ist. Die weitere Entwicklung und Adoption von Big Data wird letztendlich zeigen, ob es sich um eine Modeerscheinung oder um substanziellen Fortschritt handelt. Die aufgezeigten Thesen zeigen darüber hinaus auch, wie künftige technologische Entwicklungen für den Fortschritt der Disziplin Wirtschaftsinformatik genutzt werden können. Technologischer Fortschritt sollte für eine kumulative Ergänzung bestehender Modelle, Werkzeuge und Methoden genutzt werden. Dagegen sind wissenschaftliche Revolutionen unabhängig vom technologischen Fortschritt.

Schlüsselwörter: Big Data, Wirtschaftsinformatik, Big-Data-Hybris, Disruptive Technologie, Informationsmanagement

Big Data

An Interdisciplinary Opportunity for Information Systems Research

"Big data" describes technologies that promise to fulfill a fundamental tenet of research in information systems, which is to provide the right information to the right receiver in the right volume and quality at the right time. For information systems research as an application-oriented research discipline, opportunities and risks arise from using big data. Risks arise primarily from the considerable number of resources used for the explanation and design of fads. Opportunities arise because these resources lead to substantial knowledge gains, which support scientific progress within the discipline and are of relevance to practice as well.

From the authors' perspective, information systems research is ideally positioned to support big data critically and use the knowledge gained to explain and design innovative information systems in business and administration – regardless of whether big data is in reality a disruptive technology or a cursory fad. The continuing development and adoption of big data will ultimately provide clarity on whether big data is a fad or if it represents substantial progress in information systems research. Three theses also show how future technological developments can be used to advance the discipline of information systems. Technological progress should be used for a cumulative supplement of existing models, tools, and methods. By contrast, scientific revolutions are independent of technological progress.

Keywords: Big data, Business information systems, Systems engineering, Big data hubris, Disruptive technology, Information management