

# Big Data – Zauberstab und Rohstoff des 21. Jahrhunderts

Die Informations- und Kommunikationsbranche ist ein Wirtschaftssektor, der sich schnell entwickelt und gigantische neue Wertschöpfungsmöglichkeiten eröffnet. Big Data – die maschinelle Erschliessung verborgener Schätze in grossen Datensätzen – schafft neues wirtschaftliches Potenzial. Die Entwicklung wird zunehmend als neue technologische Revolution verstanden. Die Schweiz könnte sich als Datentresor sowie als Open-Data-Pionierin Europas etablieren und zu einem führenden europäischen Standort im Bereich der Informationstechnologien werden.



Big Data eröffnet völlig neue Möglichkeiten der Prozessoptimierung, der Identifikation von Zusammenhängen und der Unterstützung von Entscheidungen. Allerdings geht das auch mit neuen Herausforderungen einher.

Foto: Keystone

Als das soziale Nachrichtenportal Whatsapp mit 450 Mio. Nutzern kürzlich von Facebook aufgekauft wurde, belief sich der Erlös auf 19 Mrd. Dollar – fast eine halbe Mrd. Dollar pro Mitarbeiter. *Big Data* verändert die Welt. Der bereits vor über 15 Jahren geprägte Begriff meint Datensätze, die so gross sind, dass sie nicht mehr mit Standardcomputerverfahren zu bewältigen sind. Immer öfter wird Big Data als «Erdöl des 21. Jahrhunderts» bezeichnet. Um davon zu profitieren, müssen wir lernen, Daten zu «fördern» und zu «raffinieren», also in nützliche Informationen und in Wissen zu verwandeln. In einem Jahr fallen weltweit nun so viel Daten an wie in der gesamten Menschheitsgeschichte zusammen, und alle 1,2 Jahre verdoppelt sich die Datenmenge.

Diese Datenmenge entstand durch vier technologische Innovationen:

- das *Internet*, welches unsere globale Kommunikation ermöglicht;
- das *World Wide Web*, ein Netzwerk von weltweit erreichbaren Webseiten, das dank der Erfindung des *Hypertext-Protokolls (http)* am Genfer Cern entstand;
- die *Entstehung sozialer Medien* wie Facebook, Google+, Whatsapp oder Twitter,

welche soziale Kommunikationsnetzwerke geschaffen haben;

- das Aufkommen des *Internets der Dinge*, welches es nun auch Gegenständen und sensorbasierten Messnetzwerken erlaubt, sich ins Internet einzuklinken. Schon bald wird es mehr maschinelle als menschliche Nutzer im Internet geben.

## Datensätze so gross wie ein Vielfaches der grössten Bibliotheken

Inzwischen erreichen Datensätze, wie sie bei Firmen wie Ebay, Walmart oder Facebook anfallen, die Grösse von Petabytes (1 Billionen Bytes). Das ist das Hundertfache des Informationsgehalts der grössten Bibliothek der Welt, der *US Library of Congress*. Die Erschliessung von Big Data eröffnet völlig neue Möglichkeiten der Prozessoptimierung, der Identifikation von Zusammenhängen und der Unterstützung von Entscheidungen. Allerdings geht sie mit neuen Herausforderungen einher, die oft durch vier Begriffe charakterisiert werden:

- Volumen – die zu bewältigenden Datenmengen sind riesig;



**Prof. Dr. Dirk Helbing**  
Professur für Soziologie,  
insbesondere Modellierung  
und Simulation,  
ETH Zürich

- Geschwindigkeit – oftmals ist eine Datenauswertung in Echtzeit erforderlich;
- Varietät – die Daten sind meist sehr verschiedenartig und unstrukturiert;
- (Un-)Zuverlässigkeit – die Daten sind vielleicht unvollständig, nicht repräsentativ und möglicherweise sogar fehlerhaft oder manipuliert.

Daher mussten völlig neue Algorithmen – also Rechenverfahren – entwickelt werden. Weil es ineffizient ist, alle relevanten Daten in einen gemeinsamen Speicher zu laden, muss ihre Verarbeitung dezentral bei den Daten stattfinden, gegebenenfalls auf Tausenden von Computern. Dies wird mit parallelen Rechenverfahren – wie *MapReduce* oder *Hadoop* – bewerkstelligt. Big-Data-Algorithmen spüren überdies interessante Zusammenhänge (Korrelationen) in den Daten auf, die oft kommerziell verwertbar sind: etwa zwischen dem Wetter und dem Kaufverhalten oder zwischen den Lebensverhältnissen und den Gesundheits- oder Kreditrisiken. Auch die Verfolgung von Kriminalität und Terrorismus stützt sich heutzutage auf die Analyse von grossen Mengen von Verhaltensdaten.

**Wie sehen konkrete Anwendungen aus?**

Big-Data-Anwendungen verbreiten sich in Windeseile. Sie ermöglichen personalisierte Angebote, Dienste und Produkte. Einer der grössten Erfolge von Big Data ist das automatische Sprachverstehen und -verarbeiten. Siri von Apple versteht unsere Worte, wenn wir nach einem spanischen Restaurant in der Nähe suchen, Google Maps wird uns hinführen. Google Translate übersetzt fremdsprachige Texte durch Vergleich mit einer riesigen Sammlung von übersetzten Texten. Der IBM-Watson-Computer versteht gesprochene Sprache und schlägt nicht nur routinierte Quizshowgegner, sondern kümmert sich bereits per Telefonhotlines um Kunden – oftmals besser als Menschen. IBM hat beschlossen, 1 Mrd. US-Dollar in die weitere Entwicklung und Vermarktung des Systems zu investieren.

Auch im Finanzsektor spielt Big Data eine wichtige Rolle. Etwa 70% aller Finanzmarkttransaktionen werden mittlerweile von automatisierten Handelsalgorithmen abgewickelt. Dabei wird täglich etwa das gesamte Geldvolumen umgesetzt, das in der Welt existiert. So viel Geld zieht auch organisierte Kriminalität an. Daher werden Finanztransaktionen mit Big-Data-Algorithmen nach Auffälligkeiten durchforstet, um verdächtige Vorgänge aufzuspüren. Mit einer ähnlichen Software namens Aladdin spekuliert die Firma Blackrock mit

15 000 Mrd. US-Dollar an Kundengeldern – mehr als dem 30-Fachen des Bruttoinlandsprodukts der Schweiz.

**Die Potenziale sind gross...**

Allein das zusätzliche ökonomische Potenzial von *Open Data* – also Datenschätzen, die für jedermann zugänglich gemacht werden – wird von McKinsey weltweit auf 3000 bis 5000 Mrd. US-Dollar pro Jahr veranschlagt.<sup>1</sup> Dieses Potenzial betrifft beinahe alle Sektoren der Gesellschaft. Beispielsweise können Energieproduktion und -verbrauch mit *Smart Metering* besser aufeinander abgestimmt und Energiespitzen vermieden werden. Ressourcen können effizienter bewirtschaftet und die Umwelt kann geschont werden. Risiken können besser erkannt und vermieden, unbeabsichtigte Folgen von Entscheidungen reduziert und Gelegenheiten, die man früher verpasst hätte, genutzt werden. Medizin lässt sich besser auf den Patienten abstimmen, und die Gesundheitsvorsorge könnte wichtiger werden als die Heilung von Krankheiten.

**... aber auch die lauernden Gefahren**

Wie alle Technologien impliziert Big Data aber auch Gefahren. Die Sicherheit der digitalen Kommunikation wurde unterminiert. Cyberkriminalität – einschliesslich Daten-, Identitäts- und finanziellen Diebstahls – nimmt schnell immer grössere Dimensionen an. Kritische Infrastrukturen – wie Energieversorgung, Finanzsystem und Kommunikation – sind ebenfalls durch Cyberangriffe bedroht. Sie könnten im Prinzip für einige Zeit ausser Funktion gesetzt werden.

Darüber hinaus decken gängige Big-Data-Algorithmen zwar Optimierungspotenziale auf; die gefundenen Zusammenhänge sind aber meist unzuverlässig und keine Kausalbeziehungen. Daher kann die naive Anwendung von Algorithmen zu falschen Schlussfolgerungen führen. Die Fehlerquote bei Einordnungen (z.B. bei der Unterscheidung von guten und schlechten Risiken) ist meist relevant. Auch kann unbemerkt ein ungeeignetes Verfahren gewählt werden. Probleme wie Fehlentscheide, Diskriminierung und Benachteiligung sind daher zu beachten. Folglich müssen wirksame Verfahren der Qualitätskontrolle aufgebaut werden. In diesem Zusammenhang dürften Universitäten eine wichtige Rolle spielen. Auch müssen wirksame Mechanismen zum Schutz der Privatsphäre und des Rechts auf informationelle Selbstbestimmung gefunden werden, etwa mittels persönlichen Daten-Postfachs (*Personal Data Purse*, siehe *Kasten 2*).

Kasten 1

**Massive Investitionen des Internet-Riesen**

Um einen Überblick über die IT-Trends zu bekommen, lohnt es sich, Google mit seinen mehr als 50 Software-Plattformen zu betrachten. Das Unternehmen investiert jährlich fast 6 Mrd. US-Dollar in Forschung und Entwicklung. Innerhalb von nur einem Jahr hat Google selbstfahrende Autos vorgestellt, stark in Robotik investiert und ein Projekt gestartet, welches dem Internet künstliche Intelligenz verleihen möchte. Ausserdem hat Google mit dem Kauf von Nest Labs 3,2 Mrd US-Dollar in das Internet der Dinge investiert.

1 Vgl. McKinsey Global Institute (2013): *Open Data: Unlocking Innovation and Performance with Liquid Information*, October. Die Analyse quantifiziert den potenziellen Wert der Verwendung von Open Data in sieben Bereichen: Bildung, Transport, Konsumprodukte, Elektrizität, Erdöl und Erdgas, Gesundheit und Finanzprodukte.

2 Es handelt sich um Personen, die gleichzeitig Konsumenten wie auch Produzenten sind. Sie könnten zum Beispiel eigene Produkte kreieren und im Internet verkaufen oder zu Hause mit einem 3D-Drucker herstellen.

## Die technologische Revolution schafft dringenden Handlungsbedarf

Die Informations- und Kommunikationstechnologien sind dabei, die meisten unserer traditionellen Institutionen zu verändern: unser Erziehungssystem (personalisiertes Lernen), die Wissenschaft (Data Science), die Mobilität (selbstfahrende Autos), den Warentransport (Drohnen), den Konsum (Amazon und Ebay), die Produktion (3-D-Drucker), das Gesundheitssystem (personalisierte Medizin), die Politik (mehr Transparenz) und die gesamte Wirtschaft (sogenannte Prosumenten<sup>2</sup>). Die Banken müssen immer mehr Terrain an algorithmischen Handel, Bitcoins, Paypal und Google Wallet abgeben. Darüber hinaus findet ein grosser Teil des Versicherungsgeschäfts mittlerweile in Finanzprodukten wie Credit Default Swaps statt. Für den wirtschaftlichen und gesellschaftlichen Transformationsprozess hin zur digitalen Gesellschaft werden wir vielleicht nur 20 Jahre oder sogar noch weniger Zeit haben. Das ist eine extrem kurze Zeitspanne, wenn man bedenkt, dass Planung und Bau einer Strasse oft 30 Jahre oder mehr erfordern.

Kasten 2

### Infrastrukturen für ein digitales Zeitalter

Wie wird die digitale Revolution unsere Wirtschaft und Gesellschaft verändern? Wie können wir dies als Chance für alle nutzen, wie Risiken reduzieren? Zur Veranschaulichung ist es hilfreich, sich an die vielen Faktoren zu erinnern, die zum Erfolg des automobilen Zeitalters führten: die Erfindung des Motors, des Autos und der Massenproduktion; den Bau von öffentlichen Strassen, Tankstellen und Parkplätzen; die Schaffung von Fahrschulen und Führerscheinen; und nicht zuletzt die Einführung von Verkehrsregeln und -schildern, Geschwindigkeitskontrollen sowie der Verkehrspolizei.

Welche technischen Infrastrukturen sowie rechtliche, wirtschaftliche und gesellschaftliche Institutionen benötigen wir, damit die digitale Gesellschaft ein grosser Erfolg wird? Insgesamt braucht das digitale Zeitalter *vertrauenswürdige, sichere, transparente, offene und partizipative IT-Systeme*. Das World Economic Forum (WEF) arbeitete dazu in einem Konsenspapier von wirtschaftlichen, politischen und wissenschaftlichen Vertretern einen *New Deal on Data* aus.<sup>3</sup> Dieses kommt zu drei Hauptergebnissen:

- Damit Big-Data-Technologien Vertrauen genießen und gesellschaftlich akzeptabel sind, muss eine bessere Balance zwischen den Interessen der Wirtschaft, des Staates und des Bürgers bzw. des Konsumenten gefunden werden.
- Betroffene müssen wieder die Kontrolle über ihre persönlichen Daten erhalten, wie es das Recht auf informationelle Selbstbestimmung verlangt.
- Sie sollten an den Gewinnen, die mit ihren persönlichen Daten erwirtschaftet wurden, fair beteiligt werden.

Wie könnten wir Informationstechnologien entwickeln, die mit unseren gesellschaftlichen Werten vereinbar sind? Zum Beispiel könnten wir das entstehende Internet der Dinge als ein Netzwerk aufbauen, das von Bürgern mitgestaltet wird. Dieses würde Echtzeit-Messungen des Zustands unserer Welt (*digitales Nervensystem*) und eine *europäische Suchmaschine* ermöglichen. Zum Schutz der Privatsphäre sollten alle über Personen gesammelten Daten in einem persönlichen Daten-Postfach (*Personal Data Purse*) abgelegt werden, über das man die Verwendungsmöglichkeiten der eigenen Daten festlegen kann. Ein *Mikro-Payment-System* würde es Datenlieferanten, Rechteinhabern und Innovatoren erlauben, eine faire Entschädigung für ihre Dienstleistungen zu erhalten. Ein *pluralistisches, benutzerkontrolliertes Reputationsystem* würde verantwortungsvolles Verhalten in der virtuellen und realen Welt fördern.

Eine *partizipative Plattform* würde jeden in die Lage versetzen, Daten, Computer-Algorithmen und Bewertungen hochzuladen und die Beiträge der anderen (kostenlos oder gegen Gebühr) weiterzuverwenden. Insgesamt würde dies ein schnell wachsendes *Informations- und Innovationsökosystem* kreieren, welches das Potenzial von Daten für Wirtschaft, Politik, Wissenschaft und Bürger erschliesst (*Innovationsbeschleuniger*). Ausserdem würden *individuell konfigurierbare Informationsfilter* und spezielle soziale Medien sowohl «Schwarmintelligenz» als auch die Erzeugung von sozialem Kapital (z. B. Vertrauen) unterstützen. Das wäre auch für die Funktionalität der Finanzmärkte wichtig. Eine *Job- und Projekt-Plattform* würde schliesslich die Voraussetzungen für den flexiblen Arbeitsmarkt 2.0 schaffen.

<sup>3</sup> Vgl. WEF: Personal Data: The Emergence of a New Asset Class.

Daraus ergibt sich ein dringender Handlungsbedarf auf technologischer, gesetzlicher und sozioökonomischer Ebene. Bereits vor Jahren starteten die USA eine Big-Data-Forschungsinitiative im Umfang von 200 Mio. US-Dollar; daneben gibt es weitere umfangreiche Forschungsprogramme. In Europa hat das *FuturICT-Projekt* ([www.futurict.eu](http://www.futurict.eu)) Konzepte für die digitale Gesellschaft entwickelt. Andere Länder setzen das FuturICT-Konzept bereits um. Beispielsweise hat Japan kürzlich am *Tokyo Institute of Technology* ein 100 Mio. US-Dollar umfassendes Zehnjahresprojekt gestartet. Daneben finden noch zahlreiche weitere Projekte statt, insbesondere im militärischen und Sicherheitsbereich, die oft ein Vielfaches des Finanzvolumens umfassen.

### Die Schweiz kann zum Innovationsmotor Europas für das digitale Zeitalter werden

Die Schweiz hat gute Voraussetzungen, um vom digitalen Zeitalter zu profitieren. Dafür ist es allerdings nicht ausreichend, bereits vorhandene Technologien auch in der Schweiz zu bauen. Es müssen neue Erfindungen entstehen, die das digitale Zeitalter prägen werden. In der Schweiz wurde einst das World Wide Web erfunden. Hier existiert mit dem Cern auch die grösste zivile Big-Data-Kompetenz der Welt. Noch sind die USA und asiatische Länder in der wirtschaftlichen Verwertung von Big Data führend. Mit dem NSA-Skandal, dem Aufkommen von drahtlos kommunizierenden Mess-Sensoren und dem Internet der Dinge werden die Karten aber neu gemischt.

Durch eine gezielte Förderung der IT-Aktivitäten an den Hochschulen könnte die Schweiz die Führungsrolle bei der Forschung und Entwicklung in Europa übernehmen. Sie sticht mit der wissenschaftlichen Leitung von drei der sechs Finalisten im europäischen Flagship-Wettbewerb akademisch hervor. Derzeit werden jedoch nur die digitale Modellierung des menschlichen Gehirns und die Robotik schwerpunktmässig gefördert. Der ETH-Bereich plant, ab 2017 verstärkt in das Gebiet *Data Science* zu investieren, ein im Entstehen begriffenes Forschungsgebiet rund um die wissenschaftliche Datenanalyse. Allerdings sind angesichts der schnellen Entwicklung im IT-Bereich, des grossen wirtschaftlichen Potenzials, aber auch der transformativen Kraft dieser Technologien eine frühere, breitere und substanzielle Forschung äusserst dringlich und im nationalen Interesse der Schweiz.

Mit ihren basisdemokratischen Werten, ihrem Rechtsrahmen und ihren IT-Unternehmen ist die Schweiz gut aufgestellt, der Innovationsmotor Europas für das digitale Zeitalter zu werden.