

3 Analyse und Verknüpfung digitaler persönlicher Daten

„Alle Daten sind Kreditdaten, wir wissen nur noch nicht, wie wir sie richtig einsetzen“¹⁹

Douglas Merrill, ehem. Chief Information Officer (CIO) bei Google, 2012

3.1. Big Data und Verhaltensprognosen mit Statistik und Data Mining

Im Zuge der Digitalisierung haben sich in den letzten Jahrzehnten sowohl Speicher- als auch Rechenkapazitäten vervielfacht. Vor allem seit der Jahrtausendwende werden digitale Daten auf einem viel höheren Niveau gespeichert, verarbeitet und analysiert als jemals zuvor.

Unschärfer Begriff

Der Begriff **Big Data**²⁰ bezeichnet in Öffentlichkeit und Fachwelt einerseits die großen digitalen Datenmengen selbst, manchmal aber auch deren Analyse und Auswertung. Es gibt keine etablierte wissenschaftliche Definition für den Begriff, er ist unscharf und gilt als Schlagwort²¹. Nach einer Definition von *Gartner*²² bezieht sich das „Big“ im Begriff *Big Data* auf die drei Dimensionen **volume** (Umfang, Datenmenge), **velocity** (Geschwindigkeit, mit der die Daten generiert und transferiert werden) und **variety** (Bandbreite der Datentypen und -quellen). Das Beratungsunternehmen *McKinsey* verwendet eine „absichtlich subjektive“ Definition und spricht von Datenbeständen, die mit „typischen Datenbanken“ nicht mehr erfasst, gespeichert, verwaltet und analysiert werden können (vgl. Manyika 2011). Welche Datenmenge als groß eingeschätzt wird, differiere je nach Sektor, vorhandener Software und Anwendungszweck von „einigen Dutzend Terabytes bis mehreren Petabytes“.

In der Wissenschaft wird die Verarbeitung großer digitaler Datenmengen seit den 1990er-Jahren als große Herausforderung diskutiert, u.a. in der Meteorologie (Klimamodelle), der Bioinformatik (Genom-Analyse), der Physik (Simulationen) oder der Astronomie. Heute ist die Erfassung, Analyse und Auswertung großer Datenmengen in vielen Bereichen an der Tagesordnung – von staatlicher Überwachung über soziale Netzwerke und der Internet-Suche bis zur Finanzwirtschaft. Im Feld der **Business Intelligence**²³ dienen derartige Analysen der besseren Umsetzung von Unternehmenszielen.

Wahrscheinlichkeiten statt präziser Zahlen

Nach Viktor Mayer-Schönberger vom *Oxford Internet Institute* macht *Big Data* aus „präzisen Zahlen Wahrscheinlichkeiten“ und bewirkt **drei große Umwälzungen** (vgl. Mayer-Schönberger et al 2013):

- Die nicht nur auf kleine Stichproben beschränkte Analyse sehr großer Datenmengen bezogen auf ein konkretes Problem oder eine bestimmte Fragestellung.
- Die Akzeptanz einer gewissen Unschärfe anstatt von Exaktheit.
- Der „wachsende Respekt“ für Korrelationen anstatt der Suche nach Kausalitäten.

19 Hardy, Quentin (2012): Just the Facts. Yes, All of Them. New York Times, 24.03.2012. Abgerufen am 10.07.2014 von <http://www.nytimes.com/2012/03/25/business/factuals-gil-elbaz-wants-to-gather-the-data-universe.html>

20 http://en.wikipedia.org/wiki/Big_data

21 Harford, Tim (2014): Big data: are we making a big mistake? Financial Times, 28.03.2014. Abgerufen am 14.09.2014 von: <http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz3DK9lcAdI>

22 Gartner IT Glossary: „Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making“. Abgerufen am 14.09.2014 von: <http://www.gartner.com/it-glossary/big-data>

23 http://de.wikipedia.org/wiki/Business_Intelligence