

## **Interview mit Interview mit Professor Volker Markl von der Technischen Universität Berlin, künftiger Leiter des Big-Data-Kompetenzzentrums in Berlin**

*Frage: Herr Prof. Markl, unter dem Begriff „Big Data“ können sich die meisten Menschen nicht viel vorstellen. Könnten Sie den Begriff einmal erläutern?*

**Herr Prof. Markl:** In Wirtschaft, Wissenschaft und Gesellschaft stehen durch die fortschreitende Digitalisierung immer größere Datenmengen zur Verfügung, die zur Verbesserung oder auch zu Effizienzsteigerung von ökonomischen, wissenschaftlichen, sozialen Prozessen verwendet werden können. Dabei kann in der Wissenschaft z.B. aus Daten eine Hypothese abgeleitet und verifiziert werden. Das kehrt den üblichen Wissenschaftsprozess um. Denn üblicherweise denkt sich ein Wissenschaftler zunächst Hypothesen und Theorien aus und erhebt dann Daten durch Experimente. Dadurch werden die Theorien dann bestätigt oder entkräftet. Big Data kann nun Theorien vorschlagen, für die es einen gewissen Grad der Bestätigung bereits gibt. So etwas Ähnliches kann man sich dann auch in der Wirtschaft vorstellen, wo durch Massenindividualisierung aufgrund der Analyse von Daten maßgeschneiderte Produkte entstehen können. Dabei wäre ein extremes Beispiel die Auswahl von Sitzplatznachbarn in Flugzeugen anhand von Interessen, um dann interessante Gespräche mit den Nachbarn während des Fluges führen zu können. Ein anderes Beispiel wären Vertriebs- und Produktionsprozesse wo man durch datengetriebene Vorhersagemodelle besser planbare Prozesse erreichen kann und auch besser auf Störungen reagieren kann. Und das geht soweit, dass man im Heimbereich sogar mit Big Data den Energieverbrauch optimieren kann und so wirklich persönlich zur Energiewende beitragen kann.

*Frage: Warum ist der Umgang mit großen Datenmengen ein Thema für die Forschung? Geht es vor allem um methodische Fragen oder die konkrete Analyse der Daten?*

**Herr Prof. Markl:** Also das Thema Big Data ist eigentlich sehr umfassend. Die Forschung befasst sich dabei mit wirtschaftlichen, rechtlichen, gesellschaftlichen und technischen Fragen. Und in all diesen Bereichen gibt es große Herausforderungen. So ist im vorherigen Beispiel der Auswahl der Sitzplatznachbarn natürlich der Datenschutz ein großes Thema. Dabei muss eine Abwägung zwischen dem Nutzen von Big Data und dem Schutz der Privatsphäre getroffen werden. Denn selbst bei persönlichen Daten kann ich die Analyse von personenbezogenen Daten durch Dritte in einigen Bereichen wollen, in anderen Bereichen aber nicht. Wenn ich mir dazu als Beispiel vorstelle würde ich mir im Allgemeinen nicht wünschen, dass Sie meine Krankenakte kennen. Falls ich jetzt jedoch hier während diese Interviews ohnmächtig umfalle, wäre ich sehr froh, wenn Sie meine Krankengeschichte und Medikamente kennen und entsprechend reagieren könnten. Zu mindestens der Zugriff auf Daten kontextabhängig und das Konzept des Verbots mit Erlaubnisvorbehalt des Datenschutzes greift hier nicht wirklich. Denn die Daten sind eigentlich neutral und nicht wirklich der Schutzgegenstand. Die Beschränkung der Erhebung von Daten ist im Zeitalter von Big Data nicht mehr die richtige Vorgehensweise, sonst könnte Datenschutz – in diesem

Beispiel – vielleicht sogar Leben kosten und das können wir uns als Gesellschaft natürlich nicht leisten. Was wir stattdessen vielleicht machen könnten, ist um die informationellen Selbstbestimmung gewährleisten zu können, ist Analyse zu schützen. Das heißt sicherzustellen, dass nicht alles, was analysiert werden kann auch analysiert werden darf. Und ich sollte festlegen können, wer meine Daten unter welchen Umständen in welcher Form analysieren darf. Dies erfordert ein drastisches Umdenken im Datenschutz. Und im rechtlichen Bereich gibt es neben den Datenschutzfragen auch noch weitere dringende Fragen, zum Beispiel: Wer haftet, wenn aufgrund einer Analyse von Big Data aus Datenquellen von vielen Anbietern eine falsche Entscheidung getroffen wurde? Oder haftet dann der Entscheider, haftet der Entwickler der Analysealgorithmen oder haftet der Lieferanten der Daten? Aber daneben, neben diesen rechtlichen Fragen, gibt es auch noch große technologische Herausforderungen, die wir beispielsweise in diesen Big Data Kompetenzzentren in Berlin und Dresden/Leipzig angehen werden.

**Frage:** *Was haben Betriebe davon, was nützt das dem normalen Bürger?*

**Herr Prof. Markl:** Die Bürger werden durch Big Data bessere, maßgeschneiderte Produkte, einen effizienteren Umgang mit Ressourcen wie Elektrizität und Wasser und besseres Verkehrsmanagement und insgesamt leichteren Zugang zu Informationen haben. Betriebe können ihre Produktionsprozesse optimieren, passgenauere Produkte für Kunden entwickeln und neue Geschäftsmodelle und Firmen werden in der Informationsökonomie entstehen. Dabei sind viele Geschäftsmodelle vorstellbar, in denen Firmen Daten und Analysen in Echtzeit als Produkt anbieten, beispielsweise auf Informationsmarktplätzen. So ein Informationsmarktplatz ist eine Plattform, auf der Anbieter und Nachfrager von Daten und Analysen zusammengebracht werden. So könnte ich zum Beispiel an der Frage interessiert sein, welche Fusionen und Akquisitionen im Bereich von nachhaltigen Technologien sich im letzten halben Jahr ereignet haben. Und auf einem Informationsmarktplatz, der Datenquellen und technischen Verfahren zur Extraktion und Integration bereitstellt, könnte ich so eine Antwort erhalten. Das ist insbesondere für den Mittelstand interessant und wichtig, da mittelständische Unternehmen häufig nicht die finanziellen Ressourcen und auch nicht die Kompetenz haben, um derartige Daten zu einem zu speichern, vorzuhalten und die Analysen zu programmieren. Und hier werden hier völlig neue, oftmals internetgetriebene Dienstleistungsmodelle und Geschäftsmodelle entstehen.

**Frage:** *Im Rahmen der Hightech-Strategie will das Bundesforschungsministerium zwei Kompetenzzentren in Berlin und Dresden/Leipzig einrichten. Warum ist dies notwendig und welche Partner beteiligen sich daran?*

**Herr Prof. Markl:** Derzeit haben wir als führende Wirtschaftsnation nicht eine ausreichende Zahl an Menschen mit den Kompetenzen, um mit Big Data souverän umgehen zu können. Somit müssen wir zum einen durch Ausbildung Kompetenzen aufbauen. Und das passiert derzeit, viele Universitäten und Weiterbildungseinrichtungen nun die Ausbildung zu einem sogenannten „Data Scientist“ anbieten. Das Problem dabei ist aber, dass so ein Data Scientist eigentlich eine eierlegende Wollmichsau ist. Ein Data Scientist muss nämlich Kenntnisse im Umgang mit großen Datenmengen haben.

Daneben benötigt so ein Data Scientist aber auch noch mathematisch-statistische Kenntnisse im Bereich der Datenanalyse, aber selbst das reicht nicht noch nicht. Sondern ein Data Scientist braucht auch noch Kenntnisse in einem bestimmten Anwendungsproblem lösen, denn die Analysen auf Big Data finden ja nicht in einem luftleeren Raum statt, sondern sollen ein bestimmtes Anwendungsproblem lösen. Und Menschen, die tiefe System- und Technologiekenntnisse mit Datenanalysekenntnissen und Anwendungswissen vereinen, sind sehr rar.

Und das Bundesforschungsministerium hat diese Mängel und Herausforderungen erkannt und nun zwei Kompetenzzentren in Dresden/Leipzig und Berlin eingerichtet, die Big Data-Technologien erforschen, um Big Data in Zukunft beherrschbar zu machen. Das Berliner Kompetenzzentrum entwickelte eine Technologie, die es Data Scientists nun ermöglichen soll ohne Systemprogrammierung sehr einfach Datenanalysen ausführen zu können. Das ermöglicht es jetzt dem Data Scientist sich auf die Entwicklung des Datenanalyseverfahrens und auf das genaue Anwendungsproblem zu konzentrieren ohne tiefes Technik-Knowhow Systemprogrammierung zu haben. Kurz gesagt: Insgesamt wird sich durch unsere Technologieentwicklung die Menge an Menschen, die Analysen auf Big Data durchführen können, drastisch erhöhen.