BIG DATA – TEIL I

Big Data ist ein Thema, das derzeit in den Medien kontrovers diskutiert wird. Denn die Auswertung und Verknüpfung großer Datenmengen birgt zwar Chancen, aber auch ernst zu nehmende Risiken für die informationelle Selbstbestimmung jedes Einzelnen.

| von DR. DIRK JÄGER und NEDISLAV NEDYALKOV

Der Aufstieg von Unternehmen wie z. B. Google zeigt, welches Potenzial Big Data, also die Auswertung großer Datenmengen für den Erfolg einer Organisation hat. Gerade auch die öffentliche Verwaltung verfügt über eine Vielzahl von Daten aus unterschiedlichen Quellen. Bisher werden diese Quellen kaum genutzt. Das liegt auch daran, dass das Bundesdatenschutzgesetz die Nutzung von Daten standardmäßig nur für die Zwecke vorsieht, für die die Daten ursprünglich erhoben wurden.[1] Dieser Artikel zeigt, wie Big Data auf Basis dieser Daten einen Mehrwert für die Verwaltung und damit auch für den Bürger schaffen kann, ohne die informationelle Selbstbestimmung des Einzelnen zu kompromittieren.

Von Big Data spricht man dann, wenn die Datenmenge (Volume), die Verschiedenheit der Daten (Variety) oder die notwendige Geschwindigkeit (Velocity) der Verarbeitung zu groß für klassische IT-Architekturen wird. Aus den Anfangsbuchstaben der englischen Begriffe leitet sich das Schlagwort der drei „Vs von

1 Ausnahmen davon sind jedoch unter Umständen zulässig.

Big Data ab. Dabei verbergen sich hinter Big Data sowohl eine Menge neuer Technologien, wie z. B. nichtrelationale Datenbanken, als auch eine neue Herangehensweise an die Auswertung von Daten. Big-Data-Technologien können die Kapazitäten der klassischen Datenverarbeitung erweitern, z. B. durch Kopplung relationaler und nichtrelationaler Datenbanken. Das volle Potenzial von Big Data erschließt sich aber erst durch seine neuartigen Analyseverfahren.

Online-Händler nutzen solche Verfahren, um beispielsweise das Kaufverhalten von Kunden vorherzusagen. Dazu werden nicht nur Daten bisheriger Einkäufe, sondern auch andere Daten des Kunden herangezogen. Ein Beispiel: Wenn ein Kunde Produkt A innerhalb der letzten drei Wochen gekauft hat, in einer Stadt mit mehr als 100.000 Einwohnern lebt und bevorzugt mit Kreditkarte bezahlt, dann kauft er mit 80-prozentiger Wahrscheinlichkeit auch Produkt B.

Gespeichert werden können große Datenmengen schon länger, Analysen jedoch waren in der Vergangenheit aufwendig und meist nur für Stichproben möglich. Das heißt, man musste zunächst festlegen, welche Zusammenhänge man in den Daten überhaupt vermutet, und dann eine repräsentative Stichprobe auswählen. Mit Big Data ist es nun möglich, statt nur einer Stichprobe alle Datensätze zu analysieren. Damit können sehr viele Analysemodelle auf dem Gesamtbestand der Daten durchgerechnet und miteinander verglichen werden.

Es ist charakteristisch für Big Data, dass die Kausalitäten, die hinter den Zusammenhängen stehen, zweitrangig sind. Warum ein Kunde nach Produkt A mit hoher Wahrscheinlichkeit auch Produkt B kaufen wird, ist für einen Online-Händler nicht wichtig. Für den Händler ist nur wichtig, dass diese Vorhersage zutrifft. Im Blickpunkt kommerzieller Anwendungen steht dabei heute die Vorhersage des Verhaltens von Menschen als Konsumenten. Allgemein geht es bei Big Data aber darum, das Verhalten von sozialen oder technischen Systemen zu prognostizieren. Gemessen wird der Erfolg von Big Data daran, ob die Qualität der gefällten Entscheidungen sich verbessert oder nicht. Solche Vorhersagen und das Verständnis von Zusammenhängen sind auch für die öffentliche Verwaltung von großem Interesse und können helfen, Aufgaben effizienter zu erfüllen.

RAHMENBEDINGUNGEN IM ÖFFENTLICHEN SEKTOR

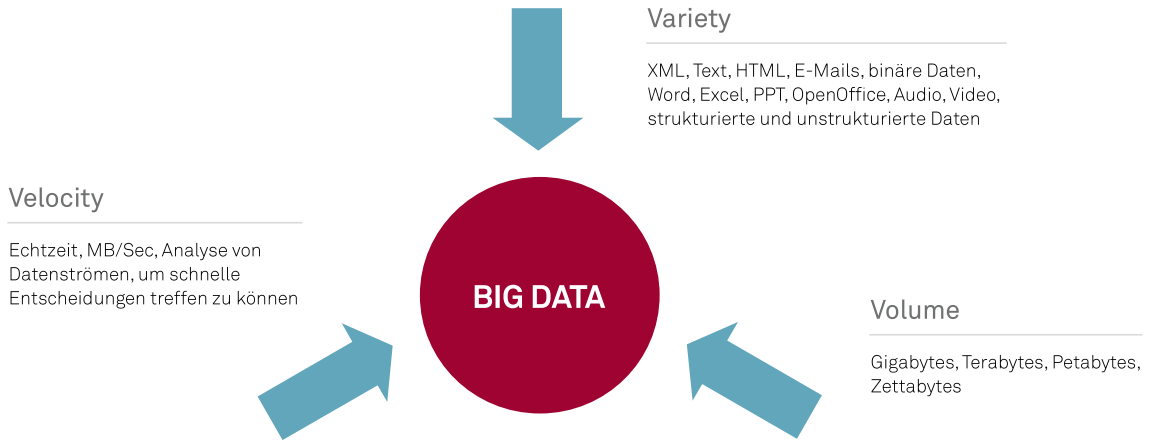
Bund, Länder und Kommunen in Deutschland verfügen über große Datenbestände, die das gesamte Spektrum öffentlicher Aufgaben abbilden. Sie umfassen beispielsweise Daten von Bau-, Gesundheits-, Sicherheitsbehörden, kommunalen Energieversorgern und Verkehrsbetrieben. Rein praktisch lässt sich aufgrund organisatorischer und gesetzlicher Abgrenzungen nur ein Teil dieser Daten unmittelbar verknüpfen und auswerten. Aber auch innerhalb einer Organisation gibt es in der Regel viele Datenquellen, die kombiniert ausgewertet werden könnten. Besonders die Kommunen verfügen über umfangreiche Detailinformationen aus unterschiedlichen Aufgabenfeldern – zumeist ohne sich dessen bewusst zu sein. Und gerade auf kommunaler Ebene haben Big-Data-Projekte eine hohe Sichtbarkeit, da sie sich unmittelbar auf das Leben der Bürger auswirken.

Die Daten, über die der öffentliche Bereich verfügt, sind in der Regel hochwertig, denn ihre Korrektheit ist für die meisten Verwaltungsvorgänge essenziell. Behörden und betroffene Bürger haben ein hohes Interesse daran, Fehler schnell zu berichtigen. Da öffentliche Institutionen beständiger sind als die meisten Unternehmen, liegen Daten für längere Zeiträume vor. Dies ermöglicht auch das Ableiten von langfristigen Trends. Insofern sind die Voraussetzungen für Big Data im öffentlichen Bereich sehr gut.

In vielen Fällen speichern öffentliche Stellen personenbezogene Daten, für die besondere Bestimmungen des Bundesdatenschutzgesetzes (BDSG) gelten.[2] Kennzeichnend für Big Data ist, möglichst viele Daten zu erheben und zu speichern und erst später nach möglichen Anwendungen zu suchen. Der größte Mehrwert im Sinne von neuen und überraschenden Erkenntnissen lässt sich generieren, wenn man Daten aus verschiedenen Kontexten verknüpft. Die rechtlichen Rahmenbedingungen müssen immer für den Einzelfall juristisch geprüft und beurteilt werden. Von einem generellen Verbot der Verknüpfung auch von personenbezogenen Daten aufgrund des BDSG muss zunächst nicht ausgegangen werden.

Organisationen des öffentlichen Sektors haben genau definierte Aufgaben und ein vorgegebenes Budget. Insofern kann der öffentliche Bereich nicht in dem Maße in neue „Geschäftsmodelle“ auf Basis von Big Data investieren wie die Privatwirtschaft. Eine Big-Data-Lösung muss daher eine unmittelbare Verbesserung bei der Erfüllung der gegenwärtigen Aufgaben einer öffentlichen Stelle mit sich bringen und sich in die vorhandenen IT-Systeme integrieren.

2 Das Bundesdatenschutzgesetz (BDSG) beschränkt die Nutzung personenbezogener Daten durch öffentliche Stellen grundsätzlich auf den Zweck, für den die Daten ursprünglich erhoben wurden (§ 14 Abs. 1 BDSG). Es erlaubt jedoch Ausnahmen, z. B. zur Gefahrenabwehr, bei der Verfolgung von Straftaten und Ordnungswidrigkeiten oder allgemein zur Wahrung „erheblicher Belange des Gemeinwohls“ (§ 14 Abs. 2 BDSG).



Big-Data-Einflussfaktoren

POTENZIALE UND RISIKEN

Die Entscheidung über den Einsatz von Methoden und Technologien aus dem Big-Data-Werkzeugkasten erfordert Mut zum Umdenken. Denn eine Umstellung zum Beispiel von relationalen auf verteilte Datenbanken bringt eine andere Art der Nutzung und der Auswertung von Daten mit sich. Damit können einerseits große Potenziale entfaltet werden, andererseits müssen die ernst zu nehmenden Risiken bewertet und minimiert werden.

Potenziale und mögliche Einsatzgebiete

Der Einsatz von Big-Data-Technologien verspricht eine höhere Effizienz der Prozesse, Individualisierung der Produktpalette sowie das Angebot intelligenter, maßgeschneiderter Produkte. Der öffentliche Bereich bietet viele Anwendungsgebiete, in denen Big-Data-Szenarien denkbar sind:

- **Einsatz begrenzter Ressourcen:** Dabei handelt es sich um die Analyse, Planung, Steuerung und Optimierung des Umgangs mit wertvollen und begrenzten Ressourcen aller Art. Angestrebt werden Einsparungen durch Verringern oder Erhöhen des Einsatzes von Ressourcen zum richtigen Zeitpunkt. Beispiele dafür sind der Einsatz von Sachbearbeitern zur Prüfung von Anträgen oder die Optimierung von Polizei- und Feuerwehreinsätzen durch Vorhersage- und Frühwarnsysteme.

- **Gesundheitsvorsorge:** Durch die Analyse gesundheitsrelevanter Daten kann die Ausbreitung von Krankheiten in der Bevölkerung beobachtet und prognostiziert werden. Ein weiteres Beispiel ist die Durchführung von Notfallanalysen und -vorhersagen. Außerdem können durch die Auswertung persönlicher Gesundheitsdaten eines Patienten dessen Vorsorge- und Therapiemaßnahmen gezielt gesteuert und verbessert werden.

- **Verkehrssteuerung und Smart Mobility:** Der Verkehrsfluss kann durch die Auswertung dynamischer Umwelt- und Verkehrsdaten vorhergesagt sowie gesteuert und optimiert werden.

- **Energieversorgung:** Durch intelligente Stromzähler und verteilte Sensoren eröffnen sich in Verbindung mit Big Data neue Möglichkeiten, den Verbrauch und die Netzauslastung zu steuern. In einem intelligenten Netz (Smart Grid) können dezentrale Verbraucher, wie z. B. Warmwasserspeicher oder Elektroautos an Ladestationen, gezielt zugeschaltet werden, wenn es ein Überangebot an Strom gibt.

- **Inspektionen:** Die öffentliche Verwaltung kann in vielfältiger Weise beim Erkennen von Missbrauch und Betrug unterstützt werden, beispielsweise der Zoll bei der Wareneinfuhr oder Finanzämter bei der Analyse von Steuerfällen.

Risiken und Nebenwirkungen

Beim Einsatz von Big Data müssen allerdings viele Aspekte datenschutzrechtlicher und ethischer Natur in Betracht gezogen werden. Ein Verlust der Privatsphäre durch Überwachung und Kontrolle des Einzelnen darf weder das Ziel sein noch passieren. Das Gefühl der Sicherheit und das Vertrauen der Bürger in die öffentliche Verwaltung muss bewahrt und sogar noch gestärkt werden. Aspekte, die Risiken bergen, sind beispielsweise:

- **Speicherung, Verarbeitung, Verknüpfung personenbezogener Daten:** Das Verhalten einzelner Personen kann aus einer Vielzahl von Daten vorausberechnet werden. Menschen folgen Gewohnheiten; wo sich eine Person in den letzten Wochen regelmäßig aufgehalten hat, wird sie sich in den nächsten Wochen mit hoher Wahrscheinlichkeit wieder aufhalten.

- **Benachteiligung aufgrund von Vorhersagen:** Big Data trifft Vorhersagen durch Korrelationen. Eine Analyse aller Lebensdaten eines Jugendlichen könnte ergeben, dass er mit hoher Wahrscheinlichkeit straffällig werden wird. Wie geht eine Gesellschaft damit um?

- **Re-Anonymisierbarkeit:** Obwohl Big Data als statistisches Instrument a priori nicht an einer Person als Individuum und der Verknüpfung seiner personenbezogenen Daten selbst interessiert ist, wird es durch die große Datenbasis zunehmend schwieriger, diese Daten zu anonymisieren. Nur den Namen aus den Daten zu entfernen und alle Fakten zu belassen, reicht nicht aus. Werden diese anonymen Daten mit öffentlichen Informationen verglichen, können sie einer Person wieder eindeutig zugeordnet werden.

- **Verlust von Kreativität und Entscheidungsfreiheit:** Wenn wir uns daran gewöhnen, dass Big-Data-Vorhersagen überwiegend zutreffen und unser Leben zuverlässig vorhersagen, dann bewegen wir uns möglicherweise nur noch auf den vorhergesagten Pfaden.

- **Datenhoheit wird zum Machtfaktor:** Je mehr Daten zur Analyse herangezogen werden können, desto treffsicherer werden die Vorhersagen. Wer die meisten Daten hat, trifft die besten (ökonomischen) Entscheidungen. Daraus resultieren Geld und Einfluss.

Und schließlich ist auch der Faktor „Know-how" ein Risiko bei der Projektdurchführung. Big-Data-Projekte erfordern eine hohe Zahl an Spezialisten, die neue Technologien aus den Bereichen verteilte Datenbanken, Netzwerke, Mathematik, Algorithmik und Statistik beherrschen. Bis diese in ausreichendem Maße zur Verfügung stehen, können Big-Data-Projekte nur mit Verzögerung besetzt und durchgeführt werden.

FAZIT: BIG DATA BIETET EIN GROSSES POTENZIAL FÜR DEN ÖFFENTLICHEN BEREICH

Big Data steht für eine ganze Reihe von Technologien und Verfahren, die in den letzten Jahren neue Möglichkeiten im Bereich der Datenverarbeitung eröffnet haben. Diese Entwicklung wurde maßgeblich von der Privatwirtschaft vorangetrieben. Die dabei entstandenen Lösungsansätze können jedoch auch auf Problemstellungen des öffentlichen Bereichs übertragen werden. Im ersten Teil des Artikels haben wir die wesentlichen Potenziale und Risiken von Big Data aufgezeigt und die spezifischen Rahmenbedingungen beleuchtet, denen Big-Data-Anwendungen im öffentlichen Sektor unterliegen. Im **zweiten Teil**[3] stellen wir konkrete Beispiele für Big-Data-Anwendungen im öffentlichen Sektor vor und beschreiben den Weg zu einem ersten Big-Data-Projekt. ●

ANSPRECHPARTNER – DR. DIRK JÄGER
Lead IT Consultant
- +49 2233 9721-6212
- dirk.jaeger@msg-systems.com

3 Erscheint in .public 02-2015