

Jörg Hoffmann, Angi Voss

Big Data und seine Bedeutung für das Wissensmanagement

Kann es sich ein Unternehmen in Zukunft leisten, prinzipiell verfügbare Daten nicht zu beachten und zu analysieren? Angesichts der ständigen Datenflut mag man argumentieren, dass eine Analyse nicht mehr zu leisten sei und damit Ressourcen verschwendet werden, die bei der eigentlichen Wertschöpfung der Unternehmen fehlen. Doch muss man die Frage mit einem klaren „Nein“ beantworten. So wie es sich heute Konsumartikelhersteller nicht mehr leisten können, Diskussionen und Bewertungen ihrer Produkte und Marken auf Facebook und anderen sozialen Medien nicht zu beachten, so bieten die Möglichkeiten der großskalierten Sammlung von Daten und deren Auswertung zahlreiche Chancen, sich Wettbewerbsvorteile zu verschaffen.

In einer vom Bundesministerium für Wirtschaft und Technologie geförderten Untersuchung konnte das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS das Innovationspotenzial von Big Data in der deutschen Wirtschaft ermitteln. [1] Dabei wurden neben einer internationalen Recherche zu Big-Data-Anwendungen eine Online-Befragung und Workshops mit Experten unterschiedlicher Branchen durchgeführt.

→ Ergebnisse

Aus dieser Untersuchung haben sich drei Chancen für die Nutzung von Big Data in Unternehmen herausgestellt:

- Unternehmensführung lässt sich durch Big Data effizienter gestalten,
- Big Data eröffnet neue Möglichkeiten der Massenindividualisierung,
- Produkte können durch Big Data intelligenter gestaltet werden.

Big Data ermöglicht das Ausschöpfen auch großer Datenmengen und insbe-

sondere neue und schnelle Formen der Verknüpfung und Analyse unterschiedlicher Quellen. Ein entscheidender Aspekt ist, dass mit der Analyse neue Möglichkeiten der Erschließung von Wissen gegeben sind, die jedoch auch neue Qualifikationen innerhalb der Unternehmen voraussetzen. Hier ist eine strategische Bewertung sowohl des vorhandenen Kompetenzportfolios als auch des zukünftigen Wissensbedarfs notwendig. Damit wandert Big Data fast zwangsläufig in das Blickfeld des strategischen Wissensmanagements. Die folgenden zwei Fragen markieren dieses Interesse:

- Welches verborgene Wissen erwarten wir, durch Big Data heben zu können?
- Welches Wissen benötigen wir, um Big Data nutzen zu können?

→ Ressourcen „Wissen“ und „Daten“

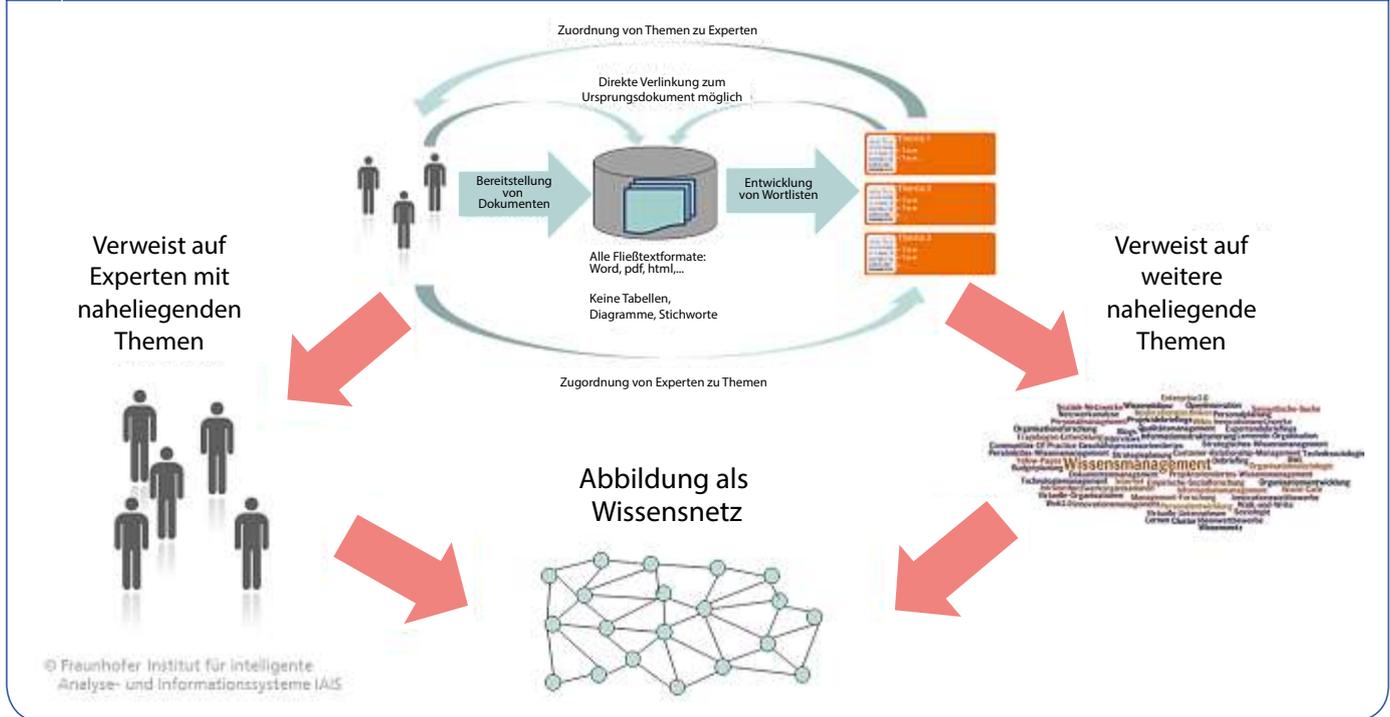
Wissensmanagement sollte durch Big Data seinen Fokus ausweiten und berei-

→ Kurz gefasst:

- *Big Data ermöglicht neue und schnelle Formen der Analyse aus unterschiedlichen internen und externen Quellen.*
- *Die Auswertung von Daten bedeutet Wissensgewinnung und unmittelbare Anwendung des Wissens.*
- *Big-Data-Analytik bietet Unternehmen Wettbewerbsvorteile durch Informationsvorsprung und Effizienzgewinn.*
- *Die Erschließung von Wissen aus Daten setzt neue Qualifikationen innerhalb der Unternehmen voraus.*

chern, indem verfügbares externes Wissen für das eigene Unternehmen noch stärker aufbereitet und mit dem internen Wissen verschmolzen wird. Neben der Aktivierung der Ressource „Wissen“ innerhalb des Unternehmens gewinnt die Ressource „Daten“ eine zunehmende Bedeutung. Während Wissensma-

→ Beispiel einer Expertensuche auf Basis unstrukturierter Daten



nagement es sich bislang vorrangig zur Aufgabe macht, das Potenzial des im Unternehmen vorhandenen Wissens zu heben, lässt sich dieses Wissen um aus Daten gewonnenen Mustern erweitern, die bislang nicht zugänglich oder aufgrund ihrer Menge und Komplexität nicht verwertbar waren. Die Integration externer Daten eröffnet neue Möglichkeiten, die bislang noch zu wenig genutzt werden konnten:

- Öffentliche Quellen und Communities öffnen ihre Schnittstellen und stellen über Open Data und Open Access beachtliche Mengen an Daten bereit,
- Sensorik und Mobilfunk produzieren mehrere Terabytes an Daten,
- das Internet bietet in Social-Media-Anwendungen und Produktportalen Einblicke in das Verhalten und die Vorlieben von Konsumenten und gewerblichen Nutzern.

Doch birgt die Auseinandersetzung mit Big Data auch ein enormes Ernüchterungspotenzial. Wird nämlich ausgeblendet, dass Wissen immer kontextabhängig und damit nicht einfrierbar und an beliebiger Stelle wiederverwertbar ist [2], so bedarf es auch bei Big-Data-Analysen eines Verständnisses für die tatsächliche Anwendung. Auch hier muss Kontext hergestellt werden, auch

hier müssen Daten mit Erfahrung verknüpft werden. Damit wird Wissen durch Big Data nicht nur durch die schiere Menge an verfügbarem Datenmaterial erzeugt, sondern insbesondere über die Kompetenz, das Material tatsächlich interpretieren und in Handlungsoptionen überführen zu können. Analog zum Wissensmanagement ist Big Data auch nicht allein eine technische Fragestellung. Wie beim Wissensmanagement handelt es sich um einen Bewusstseinsprozess, um ein neues Denken im Umgang mit Informationen und Daten.

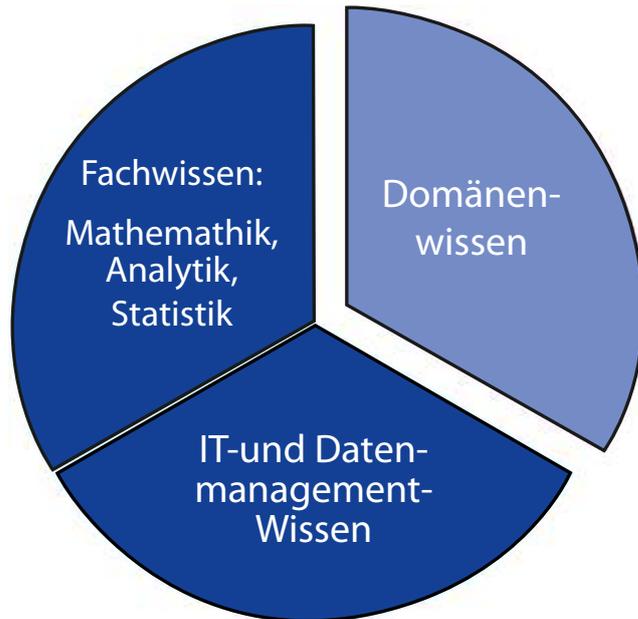
→ Begriff, Verständnis und Anwendungsbeispiele von Big Data

Big-Data-Technologien versprechen, mehr Daten schneller nutzbar zu machen. Dahinter verbergen sich drei Chancen für das Wissensmanagement:

- **Mehr Daten:** Big Data umfasst interne und externe Daten, wobei letztere vor allem aus dem Internet kommen, von Websites, Datendiensten und Social Media. Bei den internen Daten geht es darum, Silos aufzubrechen, so dass Daten aus verschiedenen Geschäfts- und Produktionssystemen zusammen analysiert und komplexe Zusammenhänge er-

kannt werden können. Dazu zählen Transaktions-, Maschinen-, Log- und Sensordaten. Bieten Altsysteme keine standardisierten Schnittstellen, so kann man die Daten mindestens als unstrukturierte Zeichenfolgen abgreifen. Sowohl externe als auch interne Daten können strukturiert und erfordern Mechanismen zur automatischen Informationsextraktion. Prominentestes Beispiel unstrukturierter Daten sind Textdaten, für die schon lange Techniken der natürlichen Sprachverarbeitung und in den letzten Jahren des automatischen Textminings entwickelt wurden. [3] Billige Speichermedien und neue Verfahren des Datenmanagements erlauben es, die extrahierten Rohdaten in großem Umfang zu speichern und zu verarbeiten (Hadoop, Map-Reduce, NoSQL-Datenbanken). Große Datenmengen beinhalten immer auch viel Rauschen, auf den ersten Blick nutzlose Daten. Andererseits bieten sie die Chance, nachträglich neue Erkenntnisse zu gewinnen und Fragen zu beantworten, die man sich zunächst (so) nicht gestellt hat: „Today’s noise can be tomorrow’s news“. Für das Wissensmanagement bedeutet dies einen Schub an Technologien für die Erschließung von Wissen aus exter-

→ Wissensfelder für Big Data



nen und unternehmensinternen Dokumenten und Daten.

- ... *schneller*: Schnelligkeit bedeutet einerseits, dass Datenströme in Echtzeit oder zumindest in geschäftsrelevanter Zeit verarbeitet werden können. Andererseits bedeutet es, Ad-hoc-Analysen auf großen Datenmengen durchführen zu können, um auch spontane Fragen zu beantworten. Man ist nicht mehr nur auf vorkonfigurierte Data Marts, Dashboards und Reports angewiesen. Für das Wissensmanagement heißt dies, das gewonnene Wissen unmittelbar nutzbar machen zu können. Wenn Datenströme oder Vorgänge automatisch überwacht werden, können ungefragt Hinweise oder Vorschläge generiert werden, ohne dass die Mitarbeiter permanent aufmerksam sein müssen, explizit digitale Inhalte durchsuchen oder Anfragen geeignet formulieren müssen. Das Wissen gelangt zur rechten Zeit zur rechten Person an den rechten Ort.
- ... *nutzbar machen*: Nutzbar machen heißt, die Daten zu analysieren. Big Data bedeutet, mehr Rohdaten speichern, und das sind vor allem Beispiele, aus denen man lernen kann: Prognosen zu erstellen, Entscheidungen zu treffen, bestimmte Arten von Informationen zu extrahieren. Lernen heißt in diesem Zusammenhang, mit Methoden der

Statistik, des Data Mining und maschinellen Lernens Modelle zu erstellen, die wegen ihres zukunftsgerichteten Charakters oft prädiktiv genannt werden.

Für das Wissensmanagement stellen sowohl extrahierte langlebige Informationen als auch die gelernten Modelle externalisiertes Wissen dar. Big-Data-Analytik heißt Wissensgewinnung aus Daten und gleichzeitig konkrete Anwendung dieses Wissens.

→ Externe und interne Daten

Modelle werden für die verschiedenen Zwecke gelernt. Hier sind einige Beispiele mit dem Schwerpunkt auf internen Daten:

- In Unternehmensprozessen Betrugsversuche oder Regelverstöße automatisch erkennen: betrügerische Kreditkartentransaktionen, Schadensmeldungen, Vortäuschen falscher Identitäten.
- Verhalten überwachen und Risiken zu erkennen: Kündigungsabsichten, E-Mails mit Eskalationsgefahr oder Maschinenteile, die auszufallen drohen.
- Verhalten überwachen, Chancen erkennen und ergreifen: Kunden spezielle Angebote unterbreiten, solange sie im Online-Shop sind oder sobald sie sich einer Filiale nähern.

Auch Mitarbeitern könnte man Vorschläge machen: „Andere, die sich dieselbe Information angeschaut haben, haben auch folgende Information nützlich gefunden: ...“.

- Informationen extrahieren: Gerade aus unstrukturierten Daten kann man langlebige Informationen gewinnen, die man den Mitarbeitern zur Verfügung stellen möchte. Klassifikation von Dokumenten und die Extraktion von Produkt-, Organisations-, Kontakt-, Orts- und Zeitangaben erzeugen zusätzliche Metadaten für die Suche bis hin zum Textverstehen für Frage-Antwort-Systeme.
- Längerfristige Prognosen machen: Produktentwicklung, Absatz- oder Personaleinsatzplanung.

Der technische Unterschied zwischen externen und internen Daten besteht vor allem in der Gewinnung der Daten. Für Speicherung und Analyse eignen sich dieselben Verfahren. Mit externen Daten steigen allerdings die Nutzungsmöglichkeiten. Längerfristige Prognosen verbessern sich, wenn man z.B. Wetterdaten einbezieht, oder Technologietrends im Netz beobachtet. Preis-, Angebots- und Nachfrageänderungen kann man schneller beobachten und entsprechend reagieren. Aus Social Media, Verbraucherportalen und Foren können konkrete Meinungen und Stimmungen zu bestimmten Produkten extrahiert werden. Setzt man Social Media auch intern ein, so kann man daraus automatisch Feedback zur internen Qualitätsverbesserung ziehen.

Ein großer Unterschied zwischen externen und internen Daten besteht allerdings im Datenschutz. Er hat in Deutschland einen ganz anderen Stellenwert als in den amerikanischen Unternehmen, die bei Big Data den Ton angeben. Die Interaktion von Mitarbeitern mit IT-Systemen kann gemäß Mitbestimmung nur mit betrieblicher Übereinkunft ausgewertet werden. Auch Daten und Dokumente mit Zugriffsbeschränkungen können nur begrenzt ausgewertet werden. Zwar gibt es Methoden des privacy preserving data mining, die sicherstellen, dass Personenbezüge nicht rekonstruierbar sind. Sie sind jedoch oft anwendungsbezogen und nicht einfach übertragbar. Hier sind juristischer Fachverstand,

weitere Forschung und Lösungen mit Vorbildcharakter gefragt.

→ Wissen für Big Data

Absehbar ist, dass die Auseinandersetzung mit Big Data für Unternehmen nicht zum Selbstläufer wird. Big Data ist ein Ansatz, sich neues Wissen zu erschließen. Doch welches Wissen steckt in dieser Menge an Daten, die in den unterschiedlichen Quellen und Formaten abgebildet werden? Die schiere Masse an Daten und die technischen Möglichkeiten der Integration und Auswertung allein bringen noch keinen Erkenntnisgewinn. Entscheidend wird es sein, die richtigen Fragen an die Daten zu stellen. Um den Wissensschatz aus den Daten zu heben, sind Kreativität und spezifisches Verständnis des Geschäfts, der Produkte, Märkte und Abläufe gefragt. Fragen dieser Art erfordern ein Vorgehen vergleichbar dem des wissenschaftlichen Erkenntnisprozesses. Denn nicht jede darstellbare Korrelation ist sinnvoll. Ohne Validitätsprüfung laufen Unternehmen Gefahr, die falschen Schlüsse zu ziehen. Der Umgang mit den Möglichkeiten von Big Data muss erst noch in den Unternehmen erlernt und die erforderlichen Kompetenzen aufgebaut werden.

Die Disziplin hinter Big Data wird Data Science genannt. Data Scientists setzen wissenschaftliche Methoden ein, um für ihr Unternehmen mehr Daten schneller nutzbar zu machen. In den USA gehören sie schon zu den meistgesuchten technisch-wissenschaftlichen IT-Fachleuten, und eine Studie von McKinsey sagt eine Lücke von über 50 Prozent für die nächsten Jahre voraus. [4] Die Ansprüche an Data Scientists sind groß: Sie müssen profunde analytische mit IT-Kompetenzen und dem Anwendungswissen aus ihrem Fachbereich vereinen, kreativ und kommunikativ sein.

Fraglich ist jedoch, ob dieser Bedarf tatsächlich gedeckt werden kann, da sich die Expertise der gesuchten Fachleute gleich auf drei Wissensfelder erstreckt. Nicht zwangsläufig können alle drei Felder von einer Person auf hohem Niveau abgedeckt werden. So mag statistisch-analytisches Fachwissen mit dem Prozesswissen des IT- und Datenmana-

gements öfters kombiniert sein. Ob das für die Analyse entscheidende jeweilige Domänenwissen bei den Data Scientists in gleichem Maße vorhanden ist, kann durchaus bezweifelt werden. Wahrscheinlicher ist vielmehr, dass neue Formen der Teamarbeit gefunden werden müssen, um den prognostizierten Bedarf der Unternehmen decken zu können. Demnach formieren sich projektbezogen gut eingespielte Teams, die über die erforderlichen Kompetenzen der drei Wissensfelder verfügen, um die Fragestellungen gemeinsam zu erarbeiten und umsetzen.

→ Fazit

Um den Transfer der Analyseergebnisse zu gewährleisten, ist auch das Management gefordert, sich mit den grundsätzlichen Wirkungsprinzipien von Data Science auseinanderzusetzen. Dies ist schließlich darin zu sehen, dass sich ganz neue Geschäftsmodelle anbieten, die bisherige Angebote und Entwicklungen in Frage stellen werden. Doch auch

das bestehende Geschäft kann von Big Data durch die klarere Analyse von Zusammenhängen und Wirkungen unterschiedlicher durch Daten beschriebener Phänomene profitieren. In jedem Fall zieht die Auseinandersetzung mit Big Data zumindest eine Überprüfung der bestehenden Unternehmens- und Wissensstrategie nach sich und wird für viele Unternehmen auch einen Wandel in der Betrachtung ihres Wissens und ihrer Kompetenzen erforderlich machen.

→ Literatur

- [1] www.iais.fraunhofer.de/bigdata-studie.html
- [2] North, K.: Wissensorientierte Unternehmensführung, Wiesbaden, 2002
- [3] Hoffmann, J.; Paaß, G.; Knapp, M.: Big [Data] Picture im Wissensmanagement. In: Bentele et al. (Hg.): Neue Horizonte für das Unternehmenswissen. Kongressband zur KnowTech 2012, S. 267–274
- [4] McKinsey: Big data: The next frontier for innovation, competition, and productivity, 2011.

→ Die Autoren



Jörg Hoffmann ist Projektleiter am IAIS und befasst sich mit den Einsatzmöglichkeiten von Text Mining im Rahmen des Wissens- und Innovationsmanagements. Nach Banklehre und sechsjähriger Berufstätigkeit studierte er Soziologie, Politik und Geographie in Bonn und Granada/Spanien. Zunächst forschte er am Fraunhofer-Institut Autonome Intelligente Systeme (AIS) und war danach über neun Jahre Leiter der Abteilung Wissensmanagement in der Fraunhofer-Zentrale. Er ist Mitglied im Industriearbeitskreis WIMIP, engagiert sich im Programmkomitee der KnowTech sowie als Mitglied der Fachjury der Initiative Exzellente Wissensorganisation (EWO).

✉ hoffmann@wissensmanagement.net



Angi Voss studierte Informatik an der Universität Bonn und promovierte 1986 an der Universität Kaiserslautern. Am Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme und seinen Vorgängern am Campus Birlinghoven leitete sie zunächst Projekte im Bereich Wissensbasierte Systeme und Knowledge Engineering. Es folgten Projekte zu ePartizipation, zur Analyse und Visualisierung von raumbezogenen Daten im Web sowie zu Open Data. Angi Voss war an der Erstellung der Big-Data-Potenzialstudie beteiligt und koordiniert das Schulungsangebot „Data Scientist“ des Instituts.

✉ voss@wissensmanagement.net