

Arbeitsgruppe 2

Vernetzte Anwendungen und Plattformen
für die digitale Gesellschaft

Projektgruppe
Smart Data

Smart Data – Potenziale und Herausforderungen



Inhalt

1	Einleitung.....	3
2	Hauptaussagen.....	3
3	Potenziale von Smart Data.....	4
	Smart Data für kleine und mittlere Unternehmen.....	4
	Beispiel: Ohne Smart Data keine Energiewende.....	4
	Beispiel: Industrie 4.0.....	5
	Beispiel: Smart Data für IT-Sicherheit und Betrugserkennung.....	5
	Beispiel: Impulse für die personalisierte Medizin.....	6
4	Technologische Grundlagen.....	7
	Datenplattformen.....	7
	Technologiekomponenten.....	9
5	Allgemeine Hürden und Hindernisse bei Smart Data.....	12
	Datenschutz und Privatsphäre.....	12
	Umsetzung von Smart-Data-Projekten.....	12
	Mangel an Experten.....	13
	Einheitliche Analyse von Unternehmensdaten.....	13
6	Branchenspezifische Chancen und Herausforderungen bei Smart Data.....	14
	Einzelhandel.....	14
	Energie.....	14
	Forschung und Wissenschaft.....	15
	Gesundheit.....	15
	Industrie 4.0.....	16
7	Handlungsempfehlungen.....	17
8	Übersicht der Beteiligten an der Projektgruppe Smart Data.....	18

1 Einleitung

In Gesellschaft und Wirtschaft gewinnen digitale Daten zunehmend an Bedeutung. Die fortschreitende Digitalisierung in allen Lebens- und Unternehmensbereichen führt zu einem rasanten Wachstum von Datenbeständen: Verbraucher äußern sich in Online-Foren oder sozialen Netzwerken über Produkte und Services, medizinische Kleinstgeräte überwachen Vitalfunktionen von Patienten, in Industrieanlagen erfassen Sensoren Parameter von Produktionsabläufen, durch das Internet der Dinge und Dienste kommt es zu einer immer stärkeren Vernetzung aller Bereiche.

Der Begriff „Big Data“ steht für eine neue Komplexität von Daten sowie deren Analyse, welche die Möglichkeiten gängiger technischer Lösungen übersteigt. Es ergeben sich etwa steigende Anforderungen bezüglich Datenvolumen, Datenraten, Datenheterogenität sowie Datenqualität. „Big Data“ ist charakterisiert durch

die drei folgenden Wesensmerkmale: a) sehr große Datenmenge, b) große Datenwachstumsraten und c) in der Regel geringe Informationsdichte bei jedoch großer Vielfalt bezüglich der Datenquellen und Datenformate. Unter „Smart Data“ wollen wir die intelligente Nutzung aller Daten verstehen, die einem Unternehmen zur Verfügung stehen. Dazu gehören auch neue Datenarten, die heute unter dem Begriff „Big Data“ subsummiert werden, sowie alle bisherigen Daten, die in Unternehmen und Behörden bereits aktiv benutzt werden. „Smart Data“ versucht dem Rechnung zu tragen und berücksichtigt neue, nicht nur technische, Aspekte sowie die wertschöpfende Datenanalyse zur Erlangung von Wissensvorteilen und zur Entscheidungsunterstützung. Dazu gehören die Datensicherheit, der Datenschutz und der Kontext, in dem die Daten erhoben wurden. Dank dieser holistischen Sicht wird der gesellschaftliche und wirtschaftliche Nutzen dieser neuen Form von Daten hervorgehoben.

2 Hauptaussagen

In diesem Strategiepapier werden Potenziale aufgezeigt, die sich aus Smart Data für die Wirtschaft in Deutschland in ausgewählten Branchen, unter anderem in Einzelhandel, Gesundheit, Energie und Industrie, ergeben. Zusätzlich werden allgemeine und branchenspezifische Herausforderungen und Hindernisse zusammengetragen, welche die Umsetzung von Smart-Data-Projekten in Deutschland erschweren. Aus Sicht der Projektgruppe bedarf es einer gemeinsamen Aktion von Wirtschaft und Politik, um die folgenden Kernpunkte weiter voranzutreiben:

- Kleine und mittlere Unternehmen stehen beim Einsatz von Smart-Data-Projekten vor besonderen Herausforderungen. Die Entwicklungen der

vergangenen Jahre in Bezug auf Cloud Computing sowie bei Datenmarktplätzen haben hier die Einstiegshürden verkleinert. Dennoch ist eine Förderung weiterhin unabdingbar.

- Wir wollen die große Vielfalt von Daten für unsere Gesellschaft und für das Individuum deutlich stärker nutzbar machen. Dafür müssen intelligente Lösungen zur Anwendung kommen, die ein hohes Maß an Flexibilität zulassen und das Prinzip der digitalen Selbstbestimmung stützen, wie etwa Pseudonymisierungs- oder Anonymisierungslösungen.
- Es braucht eine verstärkte Förderung in der universitären sowie berufsbegleitenden Ausbildung um dem Mangel an Experten für den Umgang mit Smart Data entgegenzuwirken.

3 Potenziale von Smart Data

Die Verwendung von Smart Data ermöglicht die Erkennung von Zusammenhängen und Mustern, damit bessere Entscheidungen oder Vorhersagen für die Zukunft getroffen werden können. Entscheidend ist hierbei die Datenqualität. Nur qualitativ hochwertige Daten können die Basis für korrekte Entscheidungen bilden. Im Folgenden werden die Potenziale von Smart-Data-Anwendungen an ausgewählten Anwendungsfällen vorgestellt.

Smart Data für kleine und mittlere Unternehmen

Es gibt zwei bedeutende Entwicklungen, die es kleinen und mittleren Unternehmen ermöglichen, die Potenziale von Smart Data zur Entscheidungsunterstützung genauso erfolgreich zu nutzen, wie es Großunternehmen bereits tun. Zum einen werden die technischen und finanziellen Anforderungen für die Einführung von Smart-Data-Technologien in den Unternehmen dank voranschreitender Etablierung von Cloud-Computing-Lösungen (Infrastructure, Platform sowie Software as a Service) immer geringer. Zum anderen gibt es mit den aufkommenden Informationsmarktplätzen zunehmend bessere Möglichkeiten zum gegenseitigen Austausch von Daten, was zur Folge hat, dass auch kleinen Unternehmen immer mehr Daten zur Verfügung stehen. Durch politische Rahmenbedingungen können sich hier Wettbewerbsvorteile für den deutschen Mittelstand einstellen. Diese Rahmenbedingungen müssen vor allem die rechtlichen (Datenschutz, Datensicherheit und Datenaustausch) sowie personellen Anforderungen berücksichtigen.¹

Beispiel: Ohne Smart Data keine Energiewende

Mit einer Optimierung des privaten Verbrauchs von Elektroenergie unter Nutzung von Smart Metering in deutschen Haushalten könnte ein bedeutender Beitrag zur Energiewende geleistet werden. Im vorigen Jahr stellte das Bundesministerium für Wirtschaft und Energie eine Studie zum flächendeckenden Einsatz von Smart Metering in Deutschland vor. Ein Szenario geht davon aus, dass bis 2022 mehr als zehn Millionen digitale Zähler installiert und somit rund zwei Drittel aller Haushalte und Unternehmen von der Innovation erfasst werden.

Smart Meter bilden die Grundlage für ein intelligentes Stromnetz der Zukunft, auch Smart Grid genannt. Das intelligente Stromnetz verbindet Erzeuger und Verbraucher sowie Energiespeicher und ermöglicht den Austausch von Zustandsinformationen. So wird es möglich, Angebot und Nachfrage besser aufeinander abzustimmen und zeitnah auf Schwankungen zu reagieren. Dies kann nur funktionieren, wenn es gelingt, die Daten von Millionen Verbrauchern in kurzer Zeit zu verarbeiten. Benötigt wird auch eine Rückkopplung an die Verbraucher, z. B. in Form von Preisinformationen, um die Motivation zur Verminderung des Verbrauchs in Zeiten von vergleichsweise geringem Angebot zu stärken. Das intelligente Stromnetz der Zukunft besteht aus einer großen Zahl von technischen Systemen, die durch Daten-Dienste miteinander verbunden sind. Dazu gehören die Erfassung, Speicherung, Verarbeitung und Visualisierung aller technischen und betriebswirtschaftlichen Daten und komplexe Optimierungsrechnungen. Smart Data und Cloud Computing gehören zu den Technologien, die im intelligenten Stromnetz eingesetzt werden. Energieunternehmen können Smart-Grid-Lösungen

¹ Ein Ergebnis der vom BMWi in Auftrag gegebenen Innovationspotenzialanalyse in Bezug auf Technologien zum Verwalten und Analysieren von großen Datenmengen (Big Data Management); http://www.dima.tu-berlin.de/menue/research/big_data_management_report/.

in der Cloud mit Ketten von modular aufgebauten IT-Services umsetzen. Smart-Data-Analysen können eingesetzt werden, um die Energiepreise besser auf bestimmte Kundengruppen auszurichten.

Beispiel: Industrie 4.0

Seit einiger Zeit macht der Verband Deutscher Maschinen- und Anlagenbau (VDMA) auf neue Wettbewerber aufmerksam: Chinesische Maschinenbauer sind nicht mehr nur im Segment der preiswerten Massenware tätig. Sie dringen schrittweise in die Segmente der Qualitätsspitzenprodukte ein – seit Jahrzehnten eine Festung der deutschen Hersteller. Aus diesem Grund müssen deutsche Maschinenbauer mit innovativen Produkten immer mindestens einen Schritt voraus sein.

Diese Chance bietet sich mit Industrie 4.0. Mit Intelligenz ausgestattete Produktionsanlagen werden über das Internet miteinander vernetzt, verschiedene Wertschöpfungsstufen horizontal und vertikal zu hochflexiblen Fertigungs- und Logistiksystemen integriert. Smart Data, Cloud Computing, RFID, Internet der Dinge und Dienste – alle diese Technologien gehören zu den Treibern. Das Konzept Industrie 4.0 wird im kommenden Jahrzehnt schrittweise umgesetzt. Prototypische Realisierungen einzelner Komponenten lassen sich in der Wirtschaft bereits besichtigen. Schrittweise entstehen Smart Factories, in denen intelligente Produktionsanlagen, Produkte und Ressourcen sowie Menschen interagieren und mit anderen intelligenten Systemen – Smart Grids, Smart Logistics – in Verbindung stehen. In allen Prozessschritten einer so hochintegrierten Produktion entstehen massenhaft Daten, die durch Smart-Data-Technologien erfasst, gespeichert und verarbeitet sowie für Optimierungen und Simulationen genutzt werden. Aus all diesen Gründen scheint es gerechtfertigt, von grundlegenden Veränderungen in der Art zu sprechen, wie Produkte und Dienstleistungen entwickelt, produziert und auch konsumiert werden.

Beispiel: Smart Data für IT-Sicherheit und Betrugserkennung

Smart-Data-Technologien lassen sich bei der Aufspürung von hochentwickelten Bedrohungen, Insider-Angriffen oder Konto-Übernahmen nutzen. Hier ist die Echtzeitfähigkeit von besonderer Bedeutung, da auch Kriminelle mit steigender Geschwindigkeit Verbrechen begehen. Mit Smart Data lassen sich Muster von Übergriffen herausarbeiten – so könnten Organisationen sich schnell ein Bild über die Sicherheitslage in ihrem Unternehmen machen und damit zeitnah auf Bedrohungen reagieren.

Ein konkreter Anwendungsfall stellt beispielsweise die Verwaltung von Server-Logdaten dar, wobei große Volumen an computergenerierten Daten anfallen. Es umfasst das Sammeln, die zentrale Aggregation, die Langzeitspeicherung, die Analyse der Logdaten in Echtzeit sowie die Suche in den Logdaten und daraus generierten Berichten. Nicht nur die großen Logvolumina, sondern auch die Zahl der unterschiedlichen Logformate stellen hierbei erhebliche Herausforderungen dar. Die Möglichkeit einer sinnvollen manuellen Auswertung und Analyse wird dagegen immer geringer. Dies gilt besonders dann, wenn Logdaten miteinander korreliert werden müssen, um Beurteilungen einer Situation oder Auswertungen von Fehlerzuständen vorzunehmen. Verdächtige Veränderungen oder Vorgänge möglichst rasch und am besten in Echtzeit zu erkennen ist der Anspruch von IT-Sicherheitsexperten. Solche Anwendungen stehen zurzeit noch am Anfang, werden jedoch in ihrer Verbreitung und Bedeutung stark zunehmen, da die Auswertung von Logdaten mittels Smart Data einen sehr guten Ansatzpunkt dafür bietet.

Beispiel: Impulse für die personalisierte Medizin

Besonders mit Blick auf die alternde Gesellschaft in Deutschland werden Smart-Data-Anwendungen in der Medizin für viele Menschen immer wichtiger. Konkret helfen Smart-Data-Analysen bei der Überwachung von Vitalparametern, personalisierter Medizin sowie der Erforschung von Erbkrankheiten.

Die personalisierte Medizin ermöglicht eine an den einzelnen Patienten angepasste Behandlung und Prognose von Krankheiten. Erhebliche Fortschritte im Bereich der Molekularbiologie sowie der Biotechnologie haben die personalisierte Medizin als eine praktikable Alternative ermöglicht. Besonders die Kombination der Biotechnologie mit IT-gestützten Analysen und Prozessen birgt vielversprechende Potenziale. Darüber hinaus eröffnet die personalisierte Medizin für Krankenkassen die Möglichkeit einer verbesserten und individuelleren Versorgung von Patienten.

Beispielsweise ermöglicht die personalisierte Medizin bei Krebspatienten besser angepasste und wirkungsvollere Therapien. Heutzutage spricht nur ein Teil der Patienten tatsächlich auf die Chemotherapie-Behandlung an. Da sich jeder Tumor von den anderen unterscheidet, beispielsweise von seinem genetischen Material her, müssen die behandelnden Ärzte die individuellen Aspekte der Erkrankung bei der Therapie berücksichtigen, damit sie Wirkung zeigt. Hierbei

helfen IT-gestützte DNA-Analysen dabei potenziell wirkungslose Behandlungsmethoden mit jedoch immensen Nebenwirkungen für den Patienten zu vermeiden.

Ein weiteres Beispiel stellt die Behandlung der Erbkrankheit Mukoviszidose dar. Diese Krankheit behindert die Lungenfunktion, rund die Hälfte der Patienten mit dieser Diagnose erreicht nicht das vierzigste Lebensjahr. Die Fortschritte bei der Gensequenzierung und bei den Smart-Data-Methoden erlauben es, eine auf den Patienten zugeschnittene Therapie zu entwickeln und ihnen so ein deutlich längeres und besseres Leben zu ermöglichen.

Smart Data kann auch für Infarkt-Betroffene nützlich sein. So erlaubt es die automatische Analyse von EKG-Daten von Infarkt-Betroffenen, frühzeitig Abweichungen von einem „normalen“ Verlauf zu erkennen, die das Risiko vervielfachen, innerhalb eines Jahres einen erneuten Infarkt zu erleiden.

Als letztes Beispiel seien Analysen der DNA von an Parkinson-Erkrankten genannt. So bauen Forscher mit prominenter Unterstützung des selbst betroffenen Schwergewichtsweltmeisters Muhammad Ali und das Unternehmen 23andMe die weltweit größte Datenbank für die Parkinson-Forschung auf. Erste bahnbrechende Erkenntnisse sind schon vorhanden: Zwei Gene wurden identifiziert, die mit der Krankheit verknüpft sind, und ein Gen wurde gefunden, das Schutz gegen Parkinson bietet.

4 Technologische Grundlagen

Dieses Kapitel umreißt den Stand der Technik. Es werden die wichtigsten Arten von Plattformen für das Speichern und Analysieren von Smart Data beschrieben sowie deren typische Komponenten erläutert.²

Datenplattformen

In den vergangenen Jahren wurden eine Reihe von Datenplattformen entwickelt, die den Betrieb und das Management der neuen Datenklassen auf eine möglichst effiziente und kostengünstige Art und Weise ermöglichen. Dazu gehören Technologien wie Hadoop und noSQL-Datenbanken, die sich einen festen Platz in Smart-Data-Architekturen gesichert haben. Diese Technologien sind exzellente Ergänzungen zu den bisherigen Datenhaltungssystemen, die den Großteil der heute in Unternehmen zur Analyse anstehenden Daten verwalten. Diese Data-Warehouse-Systeme verfügen in der Regel über exzellente analytische Verfahren und Algorithmen und eine breite Basis. Nahezu alle diese Systeme verfügen über entsprechende In-Memory-Technologien, die für eine schnelle Auswertung unabdingbar sind. In Abbildung 1 sind die entsprechenden Technologiekomponenten dargestellt. Entscheidend für den Erfolg von Smart Data wird u. a. sein, ob es gelingt, diese unterschiedlichen Datenhaltungs-Technologien auf einfache Art und Weise interagieren zu lassen. Daten jedweder Art und Daten aus unterschiedlichen Quellen, wie beispielsweise Hadoop, noSQL, relationale Systeme und Dateisysteme, müssen gemeinsam nutzbar sein.

Massiv parallele Systeme: Eine hervorzuhebende Plattform ist Hadoop der Apache Software Foundation. Hadoop bildet zunehmend die Grundlage von Systemen und Teilkomponenten für das Speichern und Verarbeiten von Smart Data, die in offen verfügbaren sowie kommerziellen Lösungen Anwendung finden. Konzeptionell beruht Hadoop auf dem verteilten Dateisystem HDFS sowie dem MapReduce-

Programmiermodell und -Ausführungssystem. HDFS ist in der Lage, sehr große Mengen von Daten in beliebigen Formaten verteilt und fehlertolerant auf tausenden von Rechnern zu speichern. Dabei werden individuelle Dateien redundant im Rechnerverbund gehalten. Das MapReduce-Programmiermodell ermöglicht eine einfache Spezifikation von parallelen Analyseprogrammen, bei denen der Nutzer die normalerweise sehr komplexen Aspekte der parallelen Programmierung nicht zu berücksichtigen braucht. Diese Spezifikation folgt einem strikten Schema, das sich jedoch für vielfältige Anwendungsfälle als gut geeignet erwiesen hat. Es können nahezu beliebige Operationen auf beliebig strukturierten Daten ausgeführt werden und das parallel und kostengünstig auf bis zu tausenden von Rechnern. Da das MapReduce-Ausführungssystem für Batch-basierte Analysevorgänge mit sehr großen Datenmengen optimiert wurde, ist eine interaktive Datenanalyse mit kurzen Antwortzeiten im Sekundenbereich nicht möglich. Üblich sind Ausführungszeiten im Bereich von Minuten bis Stunden.

Eine andere Innovation ist der High-Performance Computing Cluster (HPCC), eine Open-Source-Supercomputing-Plattform für datenintensive Anwendungen, die Unternehmen dabei unterstützt, Smart-Data-Probleme zu lösen. Die HPCC-Plattform zeichnet sich insbesondere durch ihre konsistente, datenzentrierte Programmiersprache (ECL) sowie eine Architektur aus, die eine effiziente Datenverarbeitung ermöglicht. Sie bietet damit Geschwindigkeit, Skalierbarkeit und eine schnelle Analyse großer Datenmengen.

noSQL-Datenbanksysteme: noSQL-Datenbanksysteme („noSQL“ steht für „Not only SQL“) dienen, ähnlich wie Hadoop, der Speicherung großer Mengen von Daten mit geringer Dichte und besitzen im Gegensatz zu relationalen Systemen keine Schemabindung. In den letzten Jahren hat sich eine Reihe von unterschiedlichen noSQL-Datenbanken ausgebildet, wobei die Key-Value-

² Für weiterführende Details wird auf den BITKOM-Leitfaden „Big-Data-Technologien – Wissen für Entscheider“ verwiesen.

Smart Data – Potenziale und Herausforderungen

Store-Technologie am weitesten verbreitet ist. Daneben sind weitere Kategorien zu nennen wie z. B. Document-, Column- oder Graph-Store-Datenbanken. Ähnlich wie bei Hadoop können noSQL-Datenbanken auf einem Netz von Standard-Rechnersystemen ausgeführt werden, was den Betrieb großer Systeme sehr kosteneffizient macht.

Hauptspeicher-Datenbanksysteme (In-Memory-Datenbanksysteme): Für Anwendungsfälle, in denen eine interaktive Analyse von Daten notwendig ist, die im relationalen Datenschema abgebildet werden können, haben sich relationale, spaltenbasierte Hauptspeicher-Datenbanksysteme etabliert. Diese Datenbanksysteme halten ihren Datenbestand primär im zuletzt deutlich günstiger gewordenen Hauptspeicher der Rechner vor, was gegenüber Festplatten die Latenzen für Datenzugriffe verringert, Datendurchsätze steigert und somit eine sehr effiziente Analyse ermöglicht.

Analytische Datenbanken: Diese Systeme bilden im Wesentlichen die heutige Basis für Data-Warehouse- und Business-Intelligence-Anwendungen. Dabei werden aus unterschiedlichsten Quellen Daten unterschiedlichster Art in einem Data Warehouse (oder Data Mart) konsolidiert und mit unterschiedlichsten Algorithmen (Data Mining, Analytics etc.) ausgewertet. Zunehmend verfügen diese Systeme über optimierte spaltenorientierte Hauptspeicherdatenbanken, die für Near-Realtime-Analytics unabdingbar sind. Des Weiteren hat sich gezeigt, dass eine Einbindung von Hadoop- und noSQL-Datenquellen in eine bestehende oder neu zu schaffende analytische Datenbankumgebung eine sehr effiziente und äußerst tragfähige Anwendungsplattform schaffen kann.

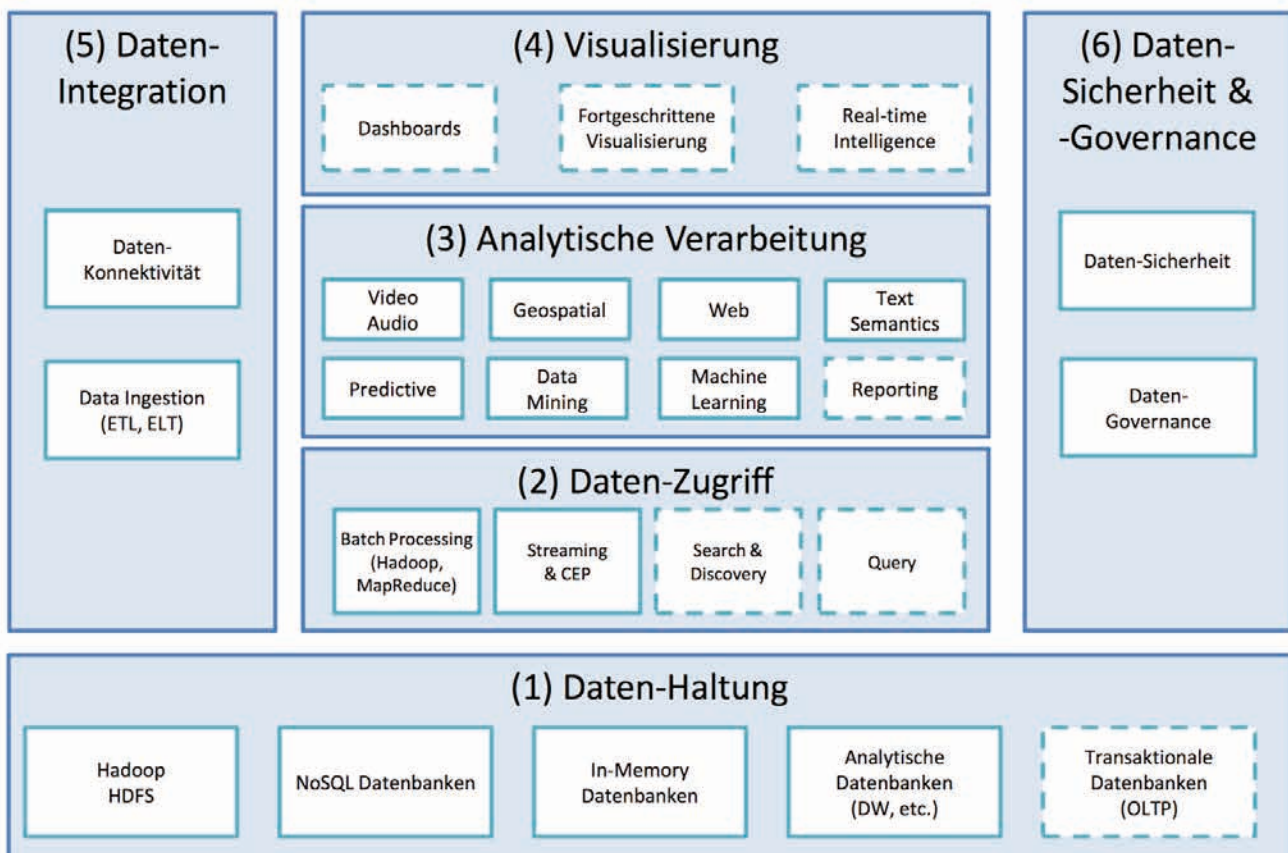


Abbildung 1: Taxonomie von Smart-Data-Plattformen

Technologiekomponenten

Der Zweck jeder Smart-Data-Plattform ist es, Daten in entscheidungsrelevante Informationen umzuwandeln. Voraussetzung dafür ist, dass es sich um qualitativ hochwertige Daten handelt. Die Vielfalt an Einsatzszenarien erfordert vielfältige Werkzeuge auf jeder Ebene.

Abbildung 1 stellt eine sehr vereinfachte Taxonomie der Technologien dar, die für eine Komplettlösung benötigt werden. Die Darstellung ist als modularer Baukasten zu verstehen. Die Schichten 1 bis 4 repräsentieren den Weg von den Rohdaten zu geschäftsrelevanten Erkenntnissen. Dieser Weg wird flankiert von den Schichten 5 (Datenintegration) und 6 (Datensicherheit sowie -Governance), die auf ein harmonisches Zusammenspiel von Smart-Data-Plattformen mit existierenden Technologien, Standards, Prozessen und Compliance-Vorgaben zielen.

Schicht 1: Datenhaltung

Viele Unternehmen sind von der Vielfalt, Geschwindigkeit und Menge an Daten überfordert. Ein echter Mehrwert für das Geschäft kann aus diesen Daten nur gewonnen werden, wenn es gelingt: a) die Daten kostengünstig zu speichern, b) die Daten auf vielfältige und flexible Weise zu „befragen“, um wertvolle und umsetzbare Erkenntnisse zu gewinnen und c) diese Erkenntnisse zeitnah zur Verfügung zu stellen. Die kostengünstigen, exponentiell wachsenden Kapazitäten von Online-Storage ermöglichen die Speicherung immer größerer Datenmengen als Basis für analytische Fragestellungen.

Schicht 2: Datenzugriff

Was typischerweise nicht wächst, ist die zur Verfügung stehende Zeit für die Beantwortung aufkommender Fragen. Die technische Herausforderung besteht also darin, eine massiv zunehmende Menge an Eingabedaten zu verarbeiten, ohne dabei den zur Verfügung stehenden Zeitrahmen zu überschreiten. Ein solches System hat die Verarbeitung und den Datenfluss einer großen Datenmenge zu organisieren. Hier seien zwei Verfahren beispielhaft genannt:

Batch Processing: Bei der Batchverarbeitung oder auch Stapelverarbeitung werden Geschäftsvorfälle gesammelt und – in häufig nächtlichen, vom Online-Betrieb freien – Batch-Läufen verarbeitet. So werden durch Batch-Skripte und Werkzeuge für ETL (Extract, Transform, Load) immer wiederkehrend die neu angefallenen Daten aus den operativen Systemen abgezogen und für entsprechende analytische Zielsysteme aufbereitet. Die neu berechnete Datenbasis steht den Analysenwerkzeugen mit zeitlicher Verzögerung zur Verfügung. Zur Reduzierung dieser Verzögerung beim Laden von Daten müssen die Batch-Läufe weiter parallelisiert und optimiert werden.

In einem Smart-Data-Szenario stößt man schnell an die Grenzen der traditionellen Batch-Verarbeitung. Ein Hadoop-basierter Sammel Speicher hingegen vereinigt die Zwischenspeicherung der Originaldaten aus den Quellsystemen im HDFS mit deren Transformation mit Hilfe der schnellen parallelen Batch-Verarbeitung.

Streaming und Complex Event Processing: Der traditionelle Ansatz von Data Warehousing analysiert ruhende Daten, in denen der Anwender mit verschiedenen Techniken nach den gewünschten Informationen sucht. Die Überwachung und Steuerung dynamischer Prozesse bedarf eines anderen Ansatzes. Hierbei werden zeitlich geordnete Ereignisse überwacht, verdichtet, gefiltert und korreliert. Muster und Regel sind hier konstant, die Daten dagegen in schneller Bewegung, daher auch der Begriff Streaming.

Schicht 3: Analytische Verarbeitung

Text- und Semantische Analyse, Video- und Audio-Analyse: Neben Datenströmen, die z. B. aus der Vernetzung von Geräten stammen oder von Sensoren generiert werden, stellt die Auswertung von Daten, die in textueller Form vorliegen, ein großes wirtschaftliches Potenzial für Unternehmen dar. So lassen sich beispielsweise in Social-Media-Daten Produktrends erkennen oder Dokumente mit Hilfe von Textanalyse und semantischen Technologien durch Zusatzinformation anreichern, so dass die angereicherten Daten neue Möglichkeiten für Anwendungen und Analysen bieten.

Smart Data – Potenziale und Herausforderungen

Neben den Datenströmen, die unmittelbar auf Text-Dokumenten basieren, wachsen insbesondere Audio- und Video-Inhalte explosionsartig an. Nach Extraktion der Text-Informationen lassen sich diese Schätze unter Nutzung linguistischer und semantischer Verfahren heben.

Orts- und raumbezogene Datenanalyse: Viele Daten weisen einen Ortsbezug auf. Die schnell steigenden Nutzerzahlen von Smartphones und Digitalkameras bieten die Möglichkeit, viele Informationen zusammen mit GPS-Koordinaten zu erheben. So sind z.B. im Handel Analysen zur Bewertung neuer Standorte unter Einbeziehung der räumlichen Dimension seit vielen Jahren Standard. In die Absatzprognosen für einzelne Artikel gehen Wettermodelle ebenso ein wie die Wohnverteilung bestimmter Einkommensklassen. Versicherungen ermitteln anhand von Geolokationen zusammen mit Überflutungsmodellen die Schadensrisiken von Immobilien. Mit Smart-Data-Methoden werden nun auch andere Daten, die nur implizit über einen Ortsbezug verfügen, mit Geoinformationen verknüpft.

Web-Analyse: Führende Online-Händler experimentieren täglich auf ihren Seiten. Sie wollen herausfinden, ob die Änderungen zu gewünschten Kundenreaktionen führen. Die Vielzahl an Erkenntnissen hinsichtlich einzelner Benutzergruppen wird dazu genutzt, um in Echtzeit den Inhalt einer Webseite anhand der Zuordnung eines Benutzers zu einer gewissen Benutzergruppe dynamisch anzupassen. Graphbasierte Methoden eignen sich zur Modellierung der Web-Daten, um mehr als die offensichtlichen Abhängigkeiten z.B. von Produkten im Webshop zu erkennen und so Benutzer über die Platzierung von Angeboten gezielter zu steuern. Des Weiteren wird die Web-Analyse immer mehr mit anderen Analysetechniken wie z.B. Predictive Analytics verknüpft, um nicht nur das Verhalten eines Kunden zu analysieren und damit Webseiten zu optimieren, sondern auch Voraussagen über seine nächsten Schritte zu treffen und ihm speziell zugeschnittene Angebote zu präsentieren.

Predictive Analytics: Mithilfe von Predictive Analytics können aus einem Datensatz Trends und Verhaltensmuster abgeleitet und vorhergesagt werden. Hierbei kommen je nach Anwendungsszenario verschiedene statistische Algorithmen und Modellierungs-Techniken zum Einsatz, die darauf abzielen, Muster in aktuellen oder historischen Daten zu erkennen und daraus Ableitungen für das zukünftige Verhalten treffen zu können.

Data Mining und Machine Learning: Data Mining ist ein bildlicher Oberbegriff für eine Vielzahl von verschiedenen Methoden, Verfahren und Techniken, die die Intention zusammenfasst – geradezu im Sinne eines „Daten-Bergbaus“ – Schätze, also verwertbares Wissen, aus den Daten des Unternehmens zu fördern. Insbesondere bezeichnet Data Mining das intelligente, größtenteils automatisierte Aufspüren und die Extraktion von wirtschaftlich nutzbaren Mustern und Zusammenhängen in großen Datenbeständen. Der Begriff Machine Learning beschreibt eine Vielzahl von Anwendungen und Methoden, in denen Computerprogramme selbständig neues Wissen erwerben.

Schicht 4: Visualisierung

Das Hauptziel der Visualisierung ist die Multiplikation des analytischen Potenzials von Mensch und Computer durch eine effektive Kombination interaktiver Visualisierungstechniken mit rechnergestützter Datenanalyse. Bei der visuellen Analyse kommen unterschiedliche Techniken zum Einsatz. Sie lassen sich nach folgenden Kriterien unterscheiden: a) Visualisierung zur Informationsbereitstellung versus visuelle Analyse zur Wissensaufbereitung, b) Visualisierungsnutzung durch den Entscheidungsträger versus durch den Analysten und Domänen-Experten sowie c) Informationskonsum oder Analyse ortsgebunden am Arbeitsplatz versus mobil.

Schicht 5: Datenintegration

Die herkömmlichen Technologien zur Integration von Anwendungen und Daten sind für Smart-Data-Lösungen weiterhin von Bedeutung. Diese Technologien sind heute im unternehmensweiten produktiven Einsatz in verschiedenen Business-Intelligence-, SOA-, Anwendungs- und B2B-Integrationsszenarien. Eine Wiederverwendung in Smart-Data-Architekturen ist notwendig, da die meisten werthaltigen Unternehmensdaten heute in strukturierter Form in existierenden Anwendungen (z. B. ERP, CRM) und Datenbanken vorliegen. Etablierte Integrationstechnologien sind in der Lage, zahlreiche Datenquellen zu integrieren und sie einer Smart-Data-Umgebung zur Verfügung zu stellen. Durch den Einsatz von Integrations-Middleware können diese unterschiedlichen Datenquellen über standardisierte Schnittstellen zugänglich gemacht werden. Hier haben sich Enterprise-Service-Bus-Technologien (ESB) etabliert.

Die andere wesentliche Technologie zur Anbindung verschiedener Datenquellen sind Datenintegrations- oder ETL-Plattformen. Im Unterschied zu einem ESB liegt bei ETL-Plattformen der Fokus auf dem Transfer und der Transformation von großen und komplexen Datenmengen, die häufig im Batch-Verfahren durchgeführt werden.

Im Rahmen von Smart Data kommen zu den existierenden Integrationslösungen neue Konnektivitäts-Anforderungen hinzu. Die beschriebenen Faktoren – hohe Datenvolumina und -Raten, unvollständige Schemata – führen insgesamt zu einer Neudefinition der Datenintegration: weg vom Vorgehen des „Extract-Transform-Load“, hin zu einem „Extract-Load-Transform“ (ELT). In einem Smart-Data-Umfeld ist der Aufwand für Extraktion (Extract) der Daten häufig vernachlässigbar. Sensorik-Daten werden bereits in großen Mengen geliefert, Social-Media-Dienste haben

einfach abzufragende APIs, Webserver protokollieren die Benutzerzugriffe sehr umfangreich und die Daten liegen oftmals in Form von Textdateien bereits vor. Folgt man dem Paradigma „Speichere jetzt – Verarbeite später“, dann reduziert sich das Laden der Daten (Load) auf einen einfachen Transportvorgang. Die Transformationen schließlich werden aufwendiger. Die Verarbeitung wenig strukturierter Daten lässt sich schlechter optimieren und verbraucht deutlich mehr Ressourcen. Selbst wenn die Daten nicht in eine strukturierte Form gebracht werden müssen, so sind doch auch für statistische und andere Datenanalysen oft umfangreiche Vorverarbeitungen erforderlich. In einer massiv parallelen Umgebung, die viel Rechenleistung zur Verfügung stellt, können diese Transformationen auch bei großen Datenmengen noch hinreichend schnell durchgeführt werden.

Schicht 6: Datensicherheit und -Governance

Datensicherheit: Bei der Umsetzung von Smart-Data-Projekten sind die Themen Datenschutz und IT-Sicherheit stark in den Vordergrund gerückt. Im Entwurf der Datenschutz-Grundverordnung sind zudem die Anforderungen an „Data protection by design and by default“ ausdrücklich verankert.

Daten-Governance: Unter Daten-Governance versteht man eine Kombination von Prozessen, Technologien und Wissen, mit der sich nachhaltig wertvolle und qualitativ hochwertige Informationen gewinnen lassen.

5 Allgemeine Hürden und Hindernisse bei Smart Data

Die Umsetzung von Smart-Data-Projekten birgt große Potenziale für Unternehmen in Deutschland, jedoch treffen viele Projekte auf dieselben Hindernisse. Des Weiteren fehlt es an Experten für die Datenanalyse.

Datenschutz und Privatsphäre

Bei Smart-Data-Anwendungen, bei denen personenbezogene Daten verwendet werden, muss wie auch bei anderen Datenverarbeitungsprozessen das deutsche Datenschutzrecht berücksichtigt werden. Deutsches Datenschutzrecht ist auch dann anwendbar, wenn ein nicht in der EU ansässiges Unternehmen im Inland Daten erhebt, verarbeitet oder nutzt. Um sowohl für Unternehmen als auch für Verbraucher zu größerer Rechtssicherheit beizutragen, können intensivere Aufklärung und, sofern im Einzelfall gewünscht, auch Unterstützung bei der Rechtsanwendung sinnvoll sein. Hierzu gehört auch, dass illegale Datennutzungen und Datenmissbrauch effektiv verfolgt und geahndet werden.

Dies gilt auch für die teilweise vorgebrachten Forderungen nach einer Nutzung der Daten nur nach Einwilligung des Betroffenen oder einer generellen Anonymisierung der Daten. Diese Vorgaben, die zudem selbst vom deutschen Datenschutzrecht nicht vorgesehen sind, würden aber den Einsatz von Big-Data-Anwendungen in wesentlichen Bereichen verunmöglichen. Hier gilt es, sachgerechte Lösungen zu finden, die einen hohen Schutz der Privatsphäre des Einzelnen ermöglichen, aber gleichzeitig die Nutzung der erheblichen gesellschaftlichen und wirtschaftlichen Potenziale von Smart Data nicht ausschließen. Dies sollte auch bei der derzeitigen Überarbeitung des EU-Datenschutzrechts berücksichtigt werden. Wichtig ist auch hier, dass Ansätze gefunden werden, die hinreichend flexibel sind, um Innovationen und Wachstum in der EU weiterhin zu ermöglichen. Dafür

müssen intelligente Lösungen zur Anwendung kommen, die ein hohes Maß an Flexibilität zulassen und das Prinzip der digitalen Selbstbestimmung stützen, wie etwa Pseudonymisierungs- und Anonymisierungslösungen. Auch freiwillige Lösungen der Industrie können in diesem Zusammenhang eine sinnvolle Lösung darstellen.

Neben einer notwendigen Expertise im Bereich der technischen Verarbeitung sind klare Regeln und Leitlinien beim Umgang mit allen Daten, insbesondere bei personenbezogenen Daten, unverzichtbar. Dazu gehört nicht nur der technische Datenschutz, sondern auch die Implementierung von Datenschutzregeln in der eigenen Organisation, z.B. in Form von innerbetrieblichen Richtlinien und Schulungen der Mitarbeiter zum Umgang mit Daten.

Umsetzung von Smart-Data-Projekten

„Datenspeicherung, -management, -sicherheit und -bereitstellung fordern von Unternehmen einen Aufwand, den einige von ihnen scheuen“, berichtet IDC in einer seiner aktuellen Umfragen.³ Dass Investitionen bei der Einführung von Smart-Data-Analysen anfallen ist klar, jedoch ist für deren Durchführung nicht automatisch der Austausch vorhandener Systeme erforderlich. Häufig ist es nur eine gezielte Ergänzung. Die Kosten hängen dabei von den individuellen Anforderungen und Gegebenheiten des jeweiligen Anwenderunternehmens ab.

Die Menge an Daten, die Unternehmen im Blick haben müssen, wächst unaufhaltsam, etwa in Form von Maschinendaten oder Informationen aus sozialen Netzwerken. Diese Daten können bei der Produktentwicklung hilfreich bzw. unverzichtbar sein und damit die Produkteinführungszeit für neue Produkte verkürzen. Doch für die Analyse der riesigen Datenmengen fehlen häufig die technischen Möglichkeiten und Fähigkeiten. Dieses Problem kann durch die Verbindung von Smart Data mit

³ IDC-Studie: Big Data – Business Value in deutschen Unternehmen auf dem Prüfstand, Frankfurt am Main, 13. Dezember 2013.

Elementen des Cloud Computings gelöst werden. Bei Cloud-Lösungen stehen IT-Ressourcen schnell, skalierbar und nahezu in unbegrenztem Umfang zur Verfügung.

Mangel an Experten

Ein großes Hemmnis bei der Umsetzung von Smart-Data-Projekte ist der Mangel an Experten, die in der Lage sind, entsprechende Analysen mit großen und heterogenen Datenmengen durchzuführen.⁴ Häufig werden stattdessen Personen aus den eigenen IT-Reihen des Unternehmens eingesetzt, die noch keine oder wenig Erfahrung mit Smart-Data-Projekten haben. Es ist dringend zu empfehlen, dass Unternehmen erst mit kleinen und mittleren Smart-Data-Lösungen starten. So können Mitarbeiter reale Erfahrung sammeln und diese in späteren größeren Projekten erweitern.

Die „Datenwissenschaft“ ist ein relativ neues und ein sehr komplexes Thema, das Wissen aus unterschiedlichen Bereichen voraussetzt. Hier treffen sich Informatik, Mathematik, Statistik, Datenschutz, Sicherheit und im besten Fall spezifisches Branchenwissen. Häufig werden Mitarbeiter für Smart-Data-Projekte eingestellt, die Experten in ihrem speziellen Bereich sind, aber kein nötiges Branchenwissen aufweisen. Dies führt dazu, dass Smart-Data-Anwendungen im schlimmsten Falle nicht den gewünschten inhaltlichen oder monetären Erfolg bringen.

Einheitliche Analyse von Unternehmensdaten

Neben fehlendem Expertenwissen sind die fehlende unternehmensweite Datensammlung sowie mangelhafte Datenintegration große Hürden. Jedoch ist genau dies der Schlüssel für ein umfassendes Auswerten der Unternehmens- und Kundendaten im Rahmen der gesetzlichen Regelungen. Meist führen einzelne Abteilungen ihre eigenen Datenbanken, auf die jedoch kein anderer Bereich Zugriff hat. Für die Durchführung einer unternehmensweiten Analyse der Daten ist die Infrastruktur vieler Unternehmen jedoch noch nicht ausgereift. Nicht nur der Datenfluss zwischen den Abteilungen ist eingeschränkt, auch unterschiedliche Prioritäten der Smart-Data-Anwendungen verzögern das Implementieren.

Eine zusätzliche Herausforderung ist die große Vielfalt der Daten und Datenformate. Dies ist eine Problematik, die nicht erst durch Smart-Data-Analysen akut wurde, sondern bereits seit der Einführung von Data-Warehouses eine entscheidende Herausforderung darstellt.

⁴ Teradata-Umfrage: Big Data Analytics ist Chefsache, Oktober 2013.

6 Branchenspezifische Chancen und Herausforderungen bei Smart Data

Neben allgemeinen Hürden und Hindernissen bei Smart-Data-Projekten gibt es zahlreiche spezifische Herausforderungen in den einzelnen Branchen. Im Folgenden werden spezifische Hürden und Hindernisse für die Branchen Energie, Forschung, Gesundheit, Einzelhandel und Industrie 4.0 zusammengetragen.

Einzelhandel

In Zeiten einer wachsenden Digitalisierung von Handelspartnern und Konsumenten steht der Einzelhandel vor neuen Herausforderungen. Es geht nicht nur darum, die klassischen Stärken, wie z. B. die Verfügbarkeit der Waren, zu verbessern, sondern darum, dem Kunden am Point of Sale über existierende und neue Vertriebskanäle, wie zum Beispiel über mobile Geräte, hinweg einen Mehrwert zu bieten.

Die Bereitstellung und Verknüpfung von geeigneten Kanälen soll die Zukunfts- und Wettbewerbsfähigkeit verbessern, die Kundenorientierung erhöhen und eine stärkere Ausrichtung an eine sich schnell ändernde digitalisierte Konsumentenwelt gewährleisten. So kann Smart Data u. a. die Entwicklung einer umfassenderen Sicht auf den Kunden und eines besseren Marktverständnisses sowie die Durchführung von Echtzeitanalysen über alle Absatzkanäle oder eine dynamische Sortiments- oder Preisplanung unterstützen.

Der Einsatz von Smart Data im Einzelhandel muss einhergehen mit der Integration von heterogen gewachsenen und dezentral verteilten ERP-Geschäftsanwendungen, welche die heutigen Kernprozesse und -funktionalitäten abbilden. Zur Ausnutzung des Potenzials von Smart Data muss ein Verständnis gewonnen werden, welche Konsumentendaten relevant und unter Berücksichtigung der datenschutzrechtlichen Rahmenbedingungen nutzbar sind sowie einen Mehrwert für das operative Geschäft bringen können. Smart Data kann sowohl die herkömmlichen Prozesse, wie zum Beispiel die Sortimentsplanung, optimieren als

auch neue innovative Geschäftsmodelle ermöglichen, beispielsweise personalisierte Angebote und Werbung.

Energie

Die Energiewirtschaft in Deutschland ist nach § 7 EnWG entbündelt, d. h., das Netz (Transport- und Verteilnetz), der Vertrieb, die Energieerzeugung und der Energiehandel sind rechtlich und operationell voneinander entflochten. Ein freier Datenaustausch insbesondere zwischen dem Netzbetreiber und den anderen Gesellschaften der Energieversorgung ist ausschließlich im Rahmen der rechtlichen Notwendigkeiten erlaubt. Beispielsweise werden Zählerdaten bei Privat- und Gewerbekunden vom Netzbetreiber erhoben und mittels marktüblicher Prozesse an den jeweils zuständigen Energievertrieb zwecks Erstellung einer Energieabrechnung weitergeleitet. Dem Netzbetreiber ist es nicht gestattet, die von ihm erhobenen Daten ohne Zustimmung des Kunden für andere Zwecke zu nutzen.

Im Gegensatz zum Übertragungsnetz ist das Verteilnetz bisher mit wenig Sensorik ausgestattet. Der Grund hierfür liegt in der räumlich begrenzten Ausdehnung möglicher Netzstörungen und der damit verbundenen Notwendigkeit, Netzzustände zeitnah zu erfassen. Aufgrund der zunehmenden dezentralen Einspeisung durch EEG-Anlagen und der damit teilweise verbundenen Notwendigkeit der Steuerung des Netzes (Smart Grid) wird es in Zukunft aber notwendig sein, Netzzustandsdaten zu erheben und diese dann weiter zu verarbeiten.

Die Steuerung der Betriebselemente sowohl im Übertragungs- als auch im Verteilnetz erfolgt über ein separates Kommunikationsnetz (PDV-Netz), das vom öffentlichen Kommunikationsnetz abgetrennt ist. Der Grund hierfür liegt im vorbeugenden Schutz der Energienetze vor möglichen Manipulationsangriffen aus öffentlichen Kommunikationsnetzen heraus. Die Daten aus dem PDV-Netz werden bisher für öffentliche Analysen nicht zur Verfügung gestellt. Die sichere Energieversorgung steht bisher über allen Bedürfnissen, mit analytischen Methoden weitergehende Erkenntnisse

zu sammeln, dabei aber mögliche Sicherheitsrisiken in Kauf zu nehmen. Die Nutzung netzrelevanter Daten mittels Data-Analytics-Anwendungen für das Transport- und Verteilnetz muss gegen die Sicherheitsrisiken abgewogen werden. Zukünftige steuerungsrelevante Einheiten zur Einspeisebegrenzung von EEG-Strom im Verteilnetz werden aber die Frage nach der Trennung und den Schnittstellen der beiden Kommunikationsnetze neu stellen.

Forschung und Wissenschaft

Die Menge der in Wissenschaft und Forschung anfallenden Daten hat in den vergangenen Jahren erheblich zugenommen. Die Auswertung und Analyse der riesigen Datenmengen könnte dabei helfen, Antworten auf grundlegende Fragen zu finden und neue wissenschaftliche Erkenntnisse zu gewinnen. Entscheidungen, die bislang auf der Basis von Schätzungen oder theoretischen Modellen getroffen werden mussten, können heute auf der Basis von konkreten Daten gefällt werden. Smart-Data-Anwendungen bieten daher ein erhebliches Potenzial für Wissenschaft und Forschung. Einen ersten Eindruck, welches Potenzial in der Analyse großer Datenmengen steckt, vermitteln etwa Projekte in der Astrophysik, wie der Sloan Digital Sky Survey⁵, aber auch Projekte in der Medizin, wie das Human Brain Project⁶. Auch anderen Disziplinen eröffnen Smart-Data-Anwendungen die Möglichkeit, die Art der wissenschaftlichen Erkenntnisgewinnung von Grund auf zu revolutionieren. Eine der Haupthürden bei der Zusammenarbeit zwischen Wirtschaft und Wissenschaft bei der Nutzung der Potenziale von Smart Data ist die Unsicherheit bei der Weitergabe von Daten. Hier sind klare, aber standardisierte technische Prozesse und juristische Regulierungen notwendig, die sowohl eine vertrauensvolle und rechtlich abgesicherte Grundlage als auch eine Basis zur zeitnahen Aufnahme der Zusammenarbeit zwischen Industrie und Wissenschaft ermöglichen. Dass dazu hohe technische Sicherheit gehört, versteht sich von selbst.

⁵ Ein Projekt zur Vermessung und Kartierung des Himmels: <http://sdss.org>

⁶ Ein Projekt zur computerbasierten Modellierung und Simulation des menschlichen Gehirns: <https://www.humanbrainproject.eu>

⁷ Smart Data Innovation Lab: <http://www.sdil.de>.

Ein Beispiel für eine positive Zusammenarbeit zwischen Wirtschaft und Wissenschaft ist das Smart Data Innovation Lab. Das Smart Data Innovation Lab (SDIL) wurde gegründet⁷, um die Nutzung von Smart Data zu ermöglichen, deren Potenzial für die Industrie zu erschließen und exzellente Forschung zu unterstützen. Die im SDIL angelegte enge Zusammenarbeit von Industrie und Wissenschaft schafft hier verbesserte Voraussetzungen für Spitzenforschung und Innovation im Bereich Data Engineering und Data Analytics für Smart Data. Um diese Zusammenarbeit zu beschleunigen, hat sich das SDIL zum Ziel gesetzt die Hürden bei der Zusammenarbeit zwischen Wirtschaft und Wissenschaft zu identifizieren und anzugehen. Im SDIL können geschlossene Netzwerke von Wirtschaft und Wissenschaft schnell eingerichtet werden, sodass die Kooperation in einem abgesicherten Rechtsrahmen vonstatten gehen kann. Die Bereitstellung von technischer Ausrüstung und professionellem Personal ist ebenso entscheidend wie klare und einfache Prozesse bei der Durchführung von Smart-Data-Analysen.

Gesundheit

Steigende Kosten und der zunehmende Wunsch nach einer effizienteren und individualisierten Versorgung stellen die modernen Gesundheitssysteme vor große Herausforderungen. Smart-Data-Anwendungen bieten in diesem Zusammenhang das Potenzial, die Qualität der medizinischen Versorgung zu erhöhen und zu einer Verringerung der Kosten beizutragen. Egal ob in der Diagnostik, im Behandlungsprozess oder der Prävention – im Versorgungsprozess rund um den Versicherten werden heutzutage eine Menge Daten generiert, die es sinnvoll zu nutzen gilt. Die Analyse und Integration dieser Daten mit weiteren Daten kann dazu beitragen, die Versorgung in höherem Maße patientenorientiert und personalisiert zu gestalten. Smart-Data-Anwendungen spielen dabei eine wichtige Rolle. Auch die sogenannte evidenzbasierte Medizin profitiert erheblich von einer umfassenden Analyse unterschiedlicher Daten.

Bei der Entwicklung und Erprobung neuer Medikamente können Smart-Data-Anwendungen zu effektiveren Wirkungsanalysen und klinischen Studien von Medikamenten beitragen. Auch die tägliche Arbeit der Beteiligten am Versorgungsprozess des Versicherten kann durch Smart-Data-Anwendungen positiv beeinflusst werden: Strukturierte Datenanalysen und effiziente digitale Programme bieten die Chance, zeitintensive Verwaltungstätigkeiten effektiver zu gestalten und durch datenbankgestützte Anwendungen personalisierte Therapien und Behandlungen besser zu unterstützen.

Eine große Herausforderung für die Nutzung von Smart-Data-Anwendungen im Gesundheitswesen ist die Heterogenität und Fragmentierung der medizinischen Systeme und Anwendungen. Dringend gebraucht werden in Deutschland verlässliche Standards, um Daten semantisch interoperabel nutzen zu können. Verpflichtende Standards fördern Innovation und geben der Industrie Investitions- und Planungssicherheit. Dabei darf es keine deutsche Sonderlösung geben – internationale anerkannte Standards müssen lediglich verbindlich und verpflichtend festgelegt werden. Hinzu kommt, dass eine Integration der Daten oftmals aus anderen Gründen scheitert, wie etwa wegen des Standortes oder der Verbindung, über welche die Daten weiter verarbeitet werden, oder auch wegen rechtlicher und regulatorischer Beschränkungen der Verarbeitung der Daten.

Die im Gesundheitswesen anfallenden Daten enthalten sogenannte Sozialdaten, dies sind stets personenbezogene und datenschutzrechtlich sensible Informationen, so dass ihre Verarbeitung höheren Anforderungen unterliegt. Pseudonymisierung und Anonymisierung kommen in diesem Zusammenhang ein hoher Stellenwert zu und sie bilden die Grundlage für neue Ansätze einer Public-Health Versorgungsforschung. Sowohl die Telematikinfrastruktur als auch alle geplanten und zukünftigen Anwendungen der Gesundheitskarte müssen und werden die höchsten Sicherheitsstandards erfüllen. Welche Anwendungen aber jeder Patient individuell nutzen möchte, hat der Gesetzgeber jedem Einzelnen selbst zugestanden. Der Patient ist Herr seiner Daten.

Klarheit geschaffen werden muss allerdings bei der Weiterentwicklung bestehender Gesetze. So enthalten das Bundesdatenschutzgesetz sowie andere Datenschutzvorschriften keine ausreichenden Rechtsgrund-

lagen für die Erhebung, Verbreitung und Nutzung von Gesundheitsdaten. Auch das ärztliche Berufsrecht muss im Sinne einer modernen, vernetzten Behandlung („Fernbehandlungsverbot“) geändert werden.

Nicht zuletzt wirft die Integration und Analyse von Gesundheitsdaten auch ethische und gesellschaftspolitische Fragen auf. In Zukunft wird es daher entscheidend darauf ankommen, Lösungen zu finden, die eine Balance zwischen den mit einer Verarbeitung von Gesundheitsdaten einhergehenden Bedenken und der durch die Technologie ermöglichten verbesserten Versorgung herstellen.

Industrie 4.0

Durch die zunehmende Digitalisierung von Maschinen, Infrastrukturen, Dienstleistungen und Managementprozessen bis hin zu HR-Prozessen entstehen neue Möglichkeiten einer intelligenten und vernetzten Produktions- und Prozessautomatisierung. Das Konzept Industrie 4.0 soll durch die gezielte Unterstützung von Informations- und Kommunikationstechnologien ein neues Innovations- und Wachstumspotenzial ermöglichen. Smart Data spielt in diesem Zusammenhang eine wichtige Rolle, da ein nutzenorientiertes Zusammenspiel aller Industrie-4.0-Komponenten nur durch eine leistungsstarke, sichere und flexible Daten-Infrastruktur voll ausgeschöpft werden kann. Auf Basis dieser Infrastruktur können neue wettbewerbsfähige Produkte und Dienstleistungen schneller entwickelt werden.

Der Einsatz von Smart Data in Industrie-4.0-Unternehmen trifft auf eine sehr heterogen geprägte Landschaft von Software-Systemen und Maschinen. Diese ist geprägt von unterschiedlichen Technologieparadigmen und einer Vielzahl unterschiedlicher Kommunikationsprotokolle, auf deren Basis Produktionsanlagen heute Daten bereitstellen oder austauschen. Smart Data muss mit dieser Heterogenität sowie mit immensen Datenmengen und Kommunikationsgeschwindigkeiten umgehen können, ohne die Industrieunternehmen mit der IT-Komplexität zu überfordern. Speziell in kleinen und mittelständischen Unternehmen ist die Verfügbarkeit von IT-Spezialisten und von notwendigen IT-Systemen nur bedingt gegeben.

7 Handlungsempfehlungen

Um deutsche Unternehmen dabei zu unterstützen, die mit Smart Data verbundenen Herausforderungen zu bewältigen und die aufgezeigten Potenziale zu nutzen, bedarf es Maßnahmen seitens der öffentlichen Institutionen und der privaten Wirtschaft.

- **Ausbildung und Qualifizierung von Fachkräften:** Die Aus- und Weiterbildung von Smart-Data-Experten mit Schlüsselkompetenzen in den Bereichen Systemprogrammierung, qualitative und quantitative Analyse von Daten aus verschiedenen Domänen sowie Recht und Wirtschaft sollte gefördert werden.
 - **Weiterbildung des Führungspersonals:** Interdisziplinäres, unternehmerisches Denken und Handeln wird eine wesentliche Rolle bei dem Vorhaben spielen, Smart Data in Unternehmen gewinnbringend zu nutzen. Es sollten Initiativen in Gang gesetzt werden, die auf die Weiterbildung des Führungspersonals zielen, damit die Chancen und Potenziale von Smart Data besser identifiziert, kommuniziert und umgesetzt werden. Der volle Mehrwert von Smart Data wird in Unternehmen entstehen, die bereit sind ihre Organisation, Methodik und Prozesse neu auszurichten.
 - **Qualifikation im sicheren Umgang mit Daten:** Neben der Ausbildung von Experten ist die Qualifikation im sicheren Umgang mit Daten sowohl von Bürgern im Privatem als auch von Mitarbeitern in Unternehmen ein essenzieller Punkt für die Schaffung eines vertrauenswürdigen Umfelds für Smart Data in Deutschland. Hierbei geht es insbesondere um die Schaffung des Bewusstseins für die Sensibilität von Daten und für die Sicherheitsfaktoren bei Abläufen und Prozessen.
 - **Schaffung und Berücksichtigung rechtlicher und ethischer Rahmenbedingungen:** Angesichts der Vielzahl möglicher Smart-Data-Anwendungen und der Komplexität der damit einhergehenden Datenverarbeitungsprozesse sind für Unternehmen und Verbraucher rechtssichere Rahmenbedingungen unerlässlich. Hierzu gehört auch Aufklärung über den gegenwärtigen Rechtsrahmen und gegebenenfalls Unterstützung bei der Rechtsanwendung. Wichtig
- ist jedoch auch, einen rechtssicheren Raum für den effektiven Einsatz von Smart-Data-Anwendungen zu finden. Die derzeitige Überarbeitung des EU-Datenschutzrechts bietet die Möglichkeit sachgerechter und angemessener Lösungen. Das Ziel sollte ein einheitliches europäisches Datenschutzrecht sein sowie eine weitere internationale Harmonisierung rechtlicher Grundlagen. Wir wollen die große Vielfalt von Daten für unsere Gesellschaft und für das Individuum deutlich stärker nutzbar machen. Dafür müssen intelligente Lösungen zur Anwendung kommen, die ein hohes Maß an Flexibilität zulassen und das Prinzip der digitalen Selbstbestimmung stützen, wie etwa Pseudonymisierungs- oder Anonymisierungslösungen.
- **Standardisierung:** Es sollten Standards entwickelt und weiterentwickelt werden, die den Zugang zu Daten unter Einhaltung der bestehenden Datenschutz- sowie Datensicherheitsbestimmungen ermöglichen. Auch werden Standards für Datenaustauschformate sowie die Qualität von Daten benötigt. Dazu muss eine Standardisierung bei den datenverarbeitenden Smart-Data-Diensten, gerade im Kontext von Cloud-Computing, erfolgen. Zu guter Letzt sollten Standardisierungsprozesse unterstützt werden, die der Innovationsgeschwindigkeit bei Smart-Data-Technologien besser entsprechen.
 - **Ausbau von Cyber-Abwehr:** Um Angriffe auf kritische Systeme und sensible Daten zu verhindern und abzuwehren, ist es unerlässlich, Mechanismen zur Cyberabwehr in Deutschland weiter auszubauen. Insbesondere im Zuge von Industrie 4.0 wird das Thema zunehmend brisant.
 - **Infrastruktur zur Unterstützung von Smart-Data-Services:** Der Aufbau einer leistungsfähigen und sicheren IKT- und Cloud-Infrastruktur zur Unterstützung von Smart-Data-Services ist essenziell. Es sollte gewährleistet werden, dass auch Unternehmen mit kleinen IT-Budgets Zugang zu skalierbaren Datenspeicher-, Datentransport- und Datenanalyse-Diensten erhalten, der mit dem von global tätigen Unternehmen vergleichbar ist.

Übersicht der Beteiligten an der Projektgruppe Smart Data

Leitung Projektgruppe Smart Data der AG2 im Nationalen IT-Gipfel

Prof. Dr. Christoph Meinel
Hasso-Plattner-Institut

Daniel Schneiss
SAP

Mitglieder der Projektgruppe Smart Data

Thomas Bodner
SAP

Guido Falkenberg
Software AG

Prof. Dr. Michael Laskowski
RWE

Dr. Karina Lott
Reed Elsevier

Sven Löffler
T-Systems

Dr. Jürgen Müller
SAP

David Schwalb
Hasso-Plattner-Institut

Günther Stürner
Oracle

Stefan Vaillant
Cumulocity

Dr. Mathias Weber
BITKOM

Markus Widmer
Intel

