



# Versuch einer „Multi-Layer“-Betrachtung von Daten im Lichte von Big Data

Erwin Hoffmann · Wolfgang Rams  
Martin Rupp · Nikolai Schwandt

## Zum aktuellen Verständnis von Big Data

Sehr spät, aber in Konsequenz, hat selbst Kanzlerin Merkel den Vergleich aufgeführt, dass Big Data der Treibstoff der digitalen Ökonomie sei [21]. Während diese Aussage vor Wirtschaftskapitänen durchaus Anklang findet, wird sie jedoch nicht immer als uneingeschränkt positiv betrachtet, sondern weckt auch Befürchtungen [25]. Abhängig vom Standort eines Betrachters (oder auch Betroffenen/Nutzers) drängen sich unterschiedliche Attribute (Dimensionen) von Big Data in den Vordergrund. Allerdings besitzt die ausschließlich technische Betrachtungsweise speziell im Hinblick auf die Datenanalyse (Analytics) in der aktuellen Diskussion ein überragendes Interesse, wie dies in [10] nachvollzogen werden kann.

Daten können nicht nur als ein primäres Ergebnis rechnergestützter Operationen sein (input/output<sup>1</sup>), sondern die Berechnung kann selbst wieder sensitive *Metadaten* liefern: *Prozessdaten*, die quasi als „Beifang“ zu den zu verarbeitenden Daten technisch bedingt anfallen. Den Prozessdaten kommt somit die Rolle von *side-channel data* inne, die sowohl die algebraische Berechnung als „computational process“ als auch den Dateninput und dessen (algebraische) Komplexität reflektiert. Sind *Metadaten* einer Operation bekannt, können mittels dieser im Rahmen eines *Reverse-Engineering* Rückschlüsse über den Verarbeitungsprozess und zudem über die verarbeiteten Daten bzw. die Datenquelle gezogen werden: *Profiling*.

<sup>1</sup> Im Bereich der *Business Information Systems* (BIS) werden diese *Operationelle Daten* betrachtet.

Big Data – unabhängig davon, ob diese aus einer oder mehreren unterschiedlichen Quellen stammen – erlaubt nun durch Identifikation ähnlicher Attribute in der Datenmenge ein *Profiling*, um spezifische interessante Aussagen zu erwirken. Zwar kann die Datenmenge, die der Big Data Analyse zugrunde liegen, *anonymisiert* oder auch nur *pseudonymisiert* werden, die strukturellen Eigenschaften der Daten sind hiervon jedoch nicht betroffen. Diese Maßnahmen haben sich bekanntermaßen als wenig wirksam herausgestellt [1].

Hieraus ergibt sich, dass eine A-priori-Weitergabe von Daten weder im Interesse von Privatpersonen noch für juristische Personen wie etwas Unternehmen, Kapitalgesellschaften oder anderen Rechtsformen ist. Alle diese Beteiligten agieren allerdings in der „Internetökonomie“ und stellen selbst einen Teil des Internets dar.

Die Weitergabe von Daten kann allerdings auch nützlich und erforderlich sein, um zu einer Prozessverbesserung zu gelangen, wie dies im Rahmen des PDCA-Zyklus angestrebt wird [11]. Daher muss in der Praxis eine Balance zwischen „offenen“ und „gedeckelten“ Daten angestrebt werden.

<https://doi.org/10.1007/s00287-019-01193-y>

© Die Autoren 2019.

Erwin Hoffmann  
Now at Frankfurt University of Applied Sciences,  
Frankfurt am Main  
E-Mail: ehoffmann@fb2.fra-uas.de

Erwin Hoffmann · Wolfgang Rams · Martin Rupp  
Nikolai Schwandt  
Proবাদis School of International Management  
& Technology AG,  
Frankfurt am Main  
E-Mail: {Wolfgang.Rams, Martin.Rupp}  
@proবাদis-hochschule.de, niko.schwandt@googlemail.com

## Zusammenfassung

Während Big Data häufig lediglich im Kontext „technischer“ Attribute gesehen wird, soll hier der Versuch unternommen werden, ein konsistentes Multi-Layer-Modell für Daten zu entwickeln, das einem systemischen Ansatz folgt. Dieses Modell wird mit seinen Dimensionen vorgestellt und die Relevanz der einzelnen Dimensionen im Hinblick auf ihre Anwendbarkeit und die aktuelle Gesetzeslage diskutiert.

Dies verlangt aber ein explizites Datenmodell, wie z. B. in [23] gefordert, das sowohl adäquat als auch konkret anwendbar ist. In dieser Abhandlung versuchen wir, dieses Modell in einem Bottom-up Ansatz zu entwickeln und seine potenzielle Nützlichkeit im Kontext der deutschen Datenschutzgesetze zu evaluieren.

## Schichtenmodell von Big Data

Bei der Diskussion um Big Data werden in der Literatur häufig vereinfacht die Merkmale

- Variety und
- Velocity

Layer	Stakeholder		
Application Layer	Originator	Principal	Community
Commercial Layer			
Legal Layer			
Technical Layer			
Domain Layer			

Abb. 2 Big Data Reference Model [24]; zur Bedeutung der Schichten siehe Tab. 1

betrachtet (Abb. 1) und diesen eine hervorgehobene Bedeutung zugesprochen. Beschränkt man sich lediglich hierauf, werden nur die technischen Aspekte des Datenzugriffs in Form des logischen Zugangs (*Methode*) und der Verarbeitungsgeschwindigkeit betrachtet.

Ergänzend sind auch weitere Attribute sowie die „Eigentümerrechte“ an den Daten (*Ownership*) von Belang, die für ein differenzierteres und vollständigeres Bild notwendig sind und die wir entsprechend Abb. 2 in Schichten und Ebenen organisiert haben:

Unter der Annahme, dass diese Datenattribute unabhängig voneinander sind, kann die Projektion auf jede Dimension als gerichteter *View* oder als (materialisierter) *Layer* betrachtet werden.

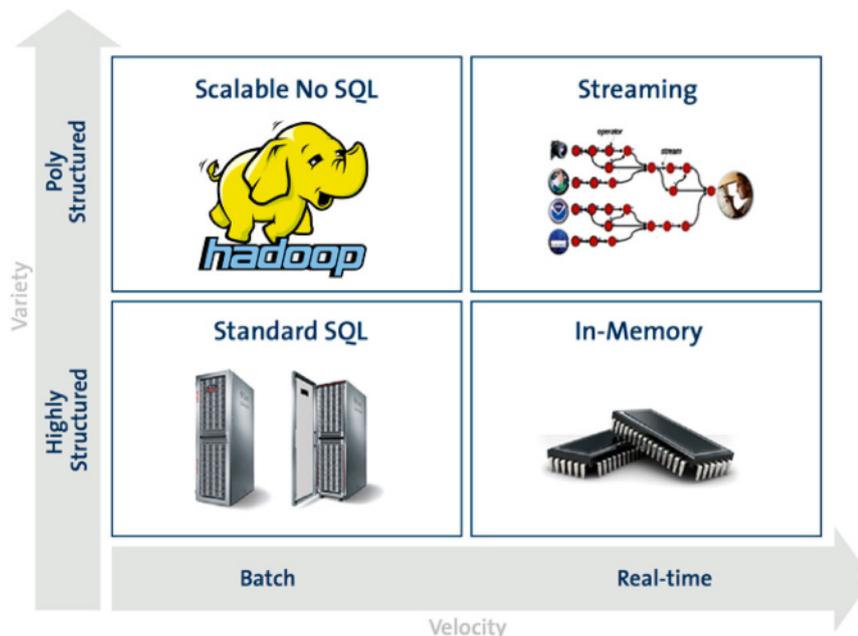


Abb. 1 Dimensionen des technischen Big-Data-Zugriffs [18]



## Schichten und ihre Bedeutung sowie Rolle im Datenverarbeitungsprozess

Schicht	Bedeutung	Rolle
Application Layer	Verwertungsarten	Worin besteht der intendierte Zweck der Daten?
Commercial Layer	Verwertungsnutzen	Was ist der potenzielle kommerzielle Nutzwert der Daten?
Legal Layer	Verwertungsrechte	Worin bestehen die rechtlichen Einschränkungen für die Nutzung der Daten?
Technical Layer	Verwendbarkeit	Wie können die Daten (weiter-)verarbeitet werden?
Domain Layer	Herkunft	Aus welcher Quelle stammen die Daten?

Ergänzend wird ein Nutzungsmodell benötigt, das die Verwendungs-/Verwertungsinteressen berücksichtigt, die wir als Ebene (*Plane*) bezüglich folgender Rollen (*Stakeholder*) betrachten wollen:

- dem *Erzeuger*,
- dem *Prinzipal* (Nutzer) sowie
- der *Gemeinschaft*.

Im Allgemeinen kann die Herkunft bzw. die *Ownership* von Daten zweifelsfrei ermittelt werden. Dem gegenüber ist die Nutzung der Daten – unter Berücksichtigung unterschiedlicher Verwendungs-/Verwertungsinteressen – derzeit in Diskussion.

### Herkunft: Domain Layer

Beim aktuellen – gerade auch öffentlich diskutierten – Umgang von Big Data fehlt elementar das Verständnis für – oder die Sicht auf – die Herkunft der Daten.

Dieses Attribut kann sehr wohl auch als „Kategorie“ aufgefasst werden, sind mit diesem doch nicht nur die Eigentümerrechte, sondern auch die Qualität der Daten verknüpft. Hiermit wird auch das „Vertrauensverhältnis“ zu den Daten beschrieben bzw. welche Operationen notwendig sind diese aufzubereiten, bevor sie einer Analyse zugewiesen werden können.

Eine grobe hierarchische Einordnung in drei Datenkategorien liefert Abb. 3 zusammen mit deren verknüpften *Typen*:

- Bei Stammdaten („reference data“) steht das beschriebene *Subject* im Vordergrund: Sie beschreiben eine Abstraktion der *Wirklichkeit*.
- Bei Mediadaten („media data“) ist deren *ideeller Inhalt* von Bedeutung und die (menschliche)

Rezeption ist mit dem technischen *Format* verknüpft.

- Prozessdaten („process data“) entstammen einer sie hervorbringenden Verarbeitungsquelle und bedingen typischerweise IT-Systeme, um sie weiter zu verarbeiten.

Aus einer Studie des Deutschen Telekom „Research Lab for Digital Business“ [22] lässt sich entnehmen, aus welcher Quelle bzw. Kategorie die Daten stammen, die im Rahmen einer Big-Data-Analyse genutzt wurden. Beachtlich ist der umfangreiche Einsatz von Daten aus unterschiedlichen Quellen (= *Typen*). Allerdings beinhaltet die Erhebung lediglich Firmen aus den Bereichen

- Produkthersteller,
- Energieversorger sowie
- Telekommunikationsdienstleister

und ist somit nur hierfür aussagekräftig. Abb. 4 lässt aber deutlich erkennen, dass deutsche Firmen vor allem (interne) Prozessdaten für ihre Big-Data-Analyse einsetzen.

### Verwendbarkeit: Technical Layer

Auf der technischen Ebene bzw. bei der Verarbeitung von Daten haben wir es üblicherweise zu tun mit den Schritten

- Beschaffung,
- Filtern sowie
- Verarbeitung,

speziell wenn diese aus unterschiedlichen Quellen stammen. Im Datenbankjargon spricht man hier von *ETL*(Extract/Transform/Load)-Operationen. Allerdings verlangt die Datenaufbereitung in der Regel zusätzlich die

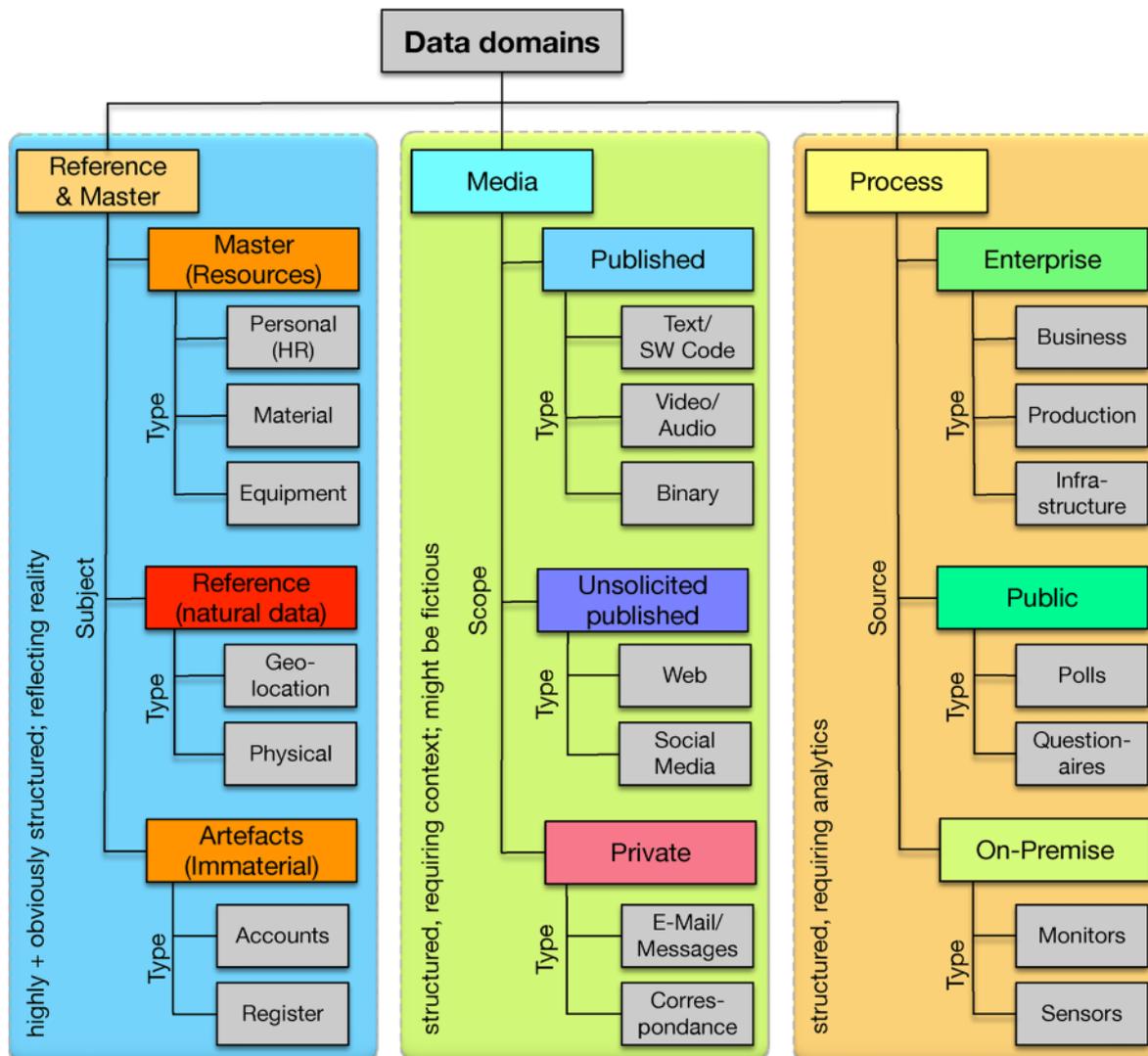


Abb. 3 Datendomänmodell für Big Data unter Einschluss einer ungefähren Einordnung der Datentypen; „Text“ ist das geschriebene Wort, wie Bücher und Fachartikel, „Audio“ und „Video“ beinhalten Mediaformatdaten mit spezifischer Metainformation, „Binary“ bezeichnen ausführbare Dateien oder andere binäre Datenformate, wie Dumps.

- Kanonisierung bzgl. der Syntax und die
- Harmonisierung im Hinblick auf ihre Semantik

was eine Herausforderung speziell für *Business Information Systems* (BIS) darstellt. Ausgehend von den Daten sollen zunächst Informationen erzeugt und hieraus schließlich Handlungsoptionen für Entscheidungen abgeleitet werden. Dies häufig unter den Verarbeitungsbedingungen großer sowie inkonsistenter Datenmengen. Wie bereits in Abb. 1 dargestellt, erfordert dieses Vorgehen unterschiedliche Technologien und Ansätze.

Stammdaten liefern im Gegensatz zu den anderen Datentypen durch ihre semantischen Informationen bereits eine Struktur mit. Dies entspricht einem inhärenten *Klassifikationsschema* für die spätere Analyse, also landläufig die Beistellung eines *Kontextes*. Andere Algorithmen der Datenanalyse, die hierauf verzichten und kontextfrei agieren – wie Deep Learning<sup>2</sup> – arbeiten ausschließlich mit dem einlaufenden Datenstrom „on-the fly“, ggf. aber unter Einbeziehung initialer, strukturierter Referenzdaten.

<sup>2</sup> Auch als „Machine Learning“ bezeichnet.

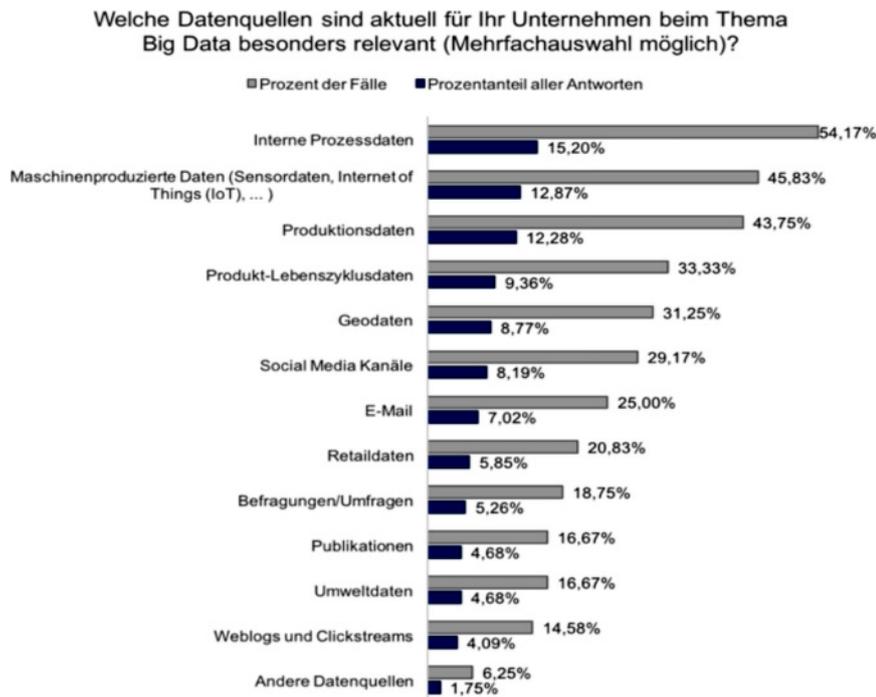


Abb. 4 Daten-Domains für die Big Data Analyse [22]

## Verwertungsrechte: Legal Layer

Richten wir unser Augenmerk auf die Herkunft (Abb. 3) der Daten als spezifische *Kategorie* in Bezug auf

- Stammdaten,
- Mediadaten sowie
- Prozessdaten,

so können klare domänbezogene Nutzungsrechte festgemacht werden.

In den meisten Ländern der Erde, speziell in denen der Europäischen Union, genießen personenbezogene *Stammdaten* eine hohe Schwärzwürdigkeit im Hinblick auf deren Verwendung und ihre Weitergabe (Proliferation). Diese Daten werden somit als „privat“ betrachtet und unter den umfangreichen Schutzschirm der „personenbezogenen Daten“ [13] gestellt.

*Mediadaten* sind im Allgemeinen Gegenstand von *Intellectual-Property*(IP)-Rechten. Selbst das Referenzieren dieser Daten in anderen Veröffentlichungen kann restriktiv gehandhabt werden, was im Zuge des Verbots des „deep linking“ [19] bekannt geworden ist: Die Verwendung von Kartenmaterial, einschließlich von Auszügen, Musikmaterial und das Verwenden (klei-

nerer) Zitate anderer Webseiten wird umfangreich eingeschränkt.

Im Gegenzug hierzu bedeutet die rechtliche Freigabe nicht unbedingt eine „kostenlose“ Nutzung, sondern ist an eine *Lizenzierung* gebunden, mit der die Rechte des Eigentümers gewahrt werden sollen. Dies stellt ein generelles Problem im Internet dar, wo Inhalte sowohl *ubiquitous* als auch kostenfrei angeboten werden. Eine mögliche Lösung stellt das Creative-Commons-Modell [26] dar, das von der *Wikimedia Foundation* ins Leben gerufen wurde und ein rechtliches Rahmenmodell für einen gemeinfreien Content formuliert.

Die abschließend zu betrachtenden Kategorien stellen die *Prozessdaten* dar. Diese sind *per se* technische Daten und deren weitergehende (rechtmäßige) Nutzung bleibt unklar. Die Zweckbindung von Daten, die noch in älteren Versionen des BSDGs in § 28 Abs. 1 Nr. 1 BDSG zu finden ist (vgl. [15, Abschnitt VIII/4]) wurde in seiner aktuellen Version – nach Adaption an die DSGVO – entfernt. Dies gilt auch für den Erlaubnisvorhalt, also eine explizite Einwilligung für die Weiterverarbeitung der Daten (vgl. [16, Abschnitt IV/4]).

In der Praxis werden aber Prozessdaten umfanglich im Rahmen der *Business Analytics* wie

bei Google<sup>3</sup> eingesetzt. Dies erfolgt häufig mit dem Verweis, dass diese nur *anonym* Verwendung finden. Andere Anwender, wie Amazon, nutzen diese aber gezielt für personalisierte Werbezwecke im Internet.

### Verwertungsnutzen: Commercial Layer

Das „mobile“ Internet hat dazu geführt, das „IT“ (in Form eines Smartphones) nun eine *Commodity* darstellt und eine ganze Generation von Anwendern geprägt hat. Die umfangreiche Nutzung von Big Data wird voraussichtlich dazu führen, dass nunmehr *Daten als Commodity* betrachtet werden, die für unser Leben unverzichtbar werden. Hiervon werden auch umfangreiche Impulse nicht nur für Produkte, sondern auch für die Wirtschaft in Gänze erwartet:

- *Mediadaten*: Video- und Audiomaterial, entweder „streamed“ oder als klassischer Download angeboten, stellen zusammen mit E-Books und Applikationen einen relevanten Markt für Umsätze im Internet dar, so wie dies *Apple*<sup>4</sup> oder *YouTube* derzeit vorexerzieren. Diese Darstellungsform ergänzt bzw. substituiert deren Legacy-Formen, die mit einem physikalischen Datenträger verbunden sind. Gelöst werden muss speziell bei der datenträgerfreien Nutzung bzw. Distribution das Problem der „Intellectual Property“ bzw. der Verwertungsrechte. Ob Ansätze, wie dies die GEMA verfolgt, Media auf bestimmte Länder zu beschränken, von Erfolg gekrönt sein werden, darf aber getrost bezweifelt werden. IT-Technik führt in der Regel dazu, den Konsumptionsbereich der Media zu erweitern; eine technische Reglementierung kann zwar angestrebt werden, ist aber nur bei geschlossenen Systemen, wie z. B. *Skype* oder *Apple*, erfolgreich zu realisieren.
- *Stammdaten*: Der Handel mit Adressdaten, E-Mail-Konten und Kreditkartennummern umfasst ein durchaus beachtliches Volumen im Internet, geht es doch immer darum, seine Kunden zu adressieren und zu kennen. Aufgrund der unterschiedlichen Gesetzgebungen [3, 4, 17] und des Verständnisses von Stammdaten ist ein gemeinsames reglementiertes Nutzungskonzept noch in weite Ferne gerückt.

- *Prozessdaten*: Mit der Einführung des *World Wide Web* besitzen Prozessdaten eine zentrale Bedeutung, verbinden sie doch den (potenziellen) Kunden mit seinen Transaktionen, sprich Kaufgewohnheiten bzw. -absichten. Hieraus hat sich ein eigenes Geschäftsmodell entwickelt [2], das derzeit noch weitgehend unreguliert abläuft.

Die Nutzung von Daten, die im Zuge einer Geschäftstätigkeit IT-technisch anfallen, verlangt jedoch nach einem rechtlichen Rahmen. Gerade bei Prozessdaten ist dieses noch nicht definiert und die grundlegenden Charakteristika und Anforderungen müssen erst bestimmt werden. Es ist allerdings offenkundig, dass die Verwendung von Prozessdaten in jedem Fall eine Auswirkung auf diejenigen haben wird, der am Ursprung der Datenkette steht und somit auch zum Missbrauch verwendet werden kann.

### Verwertungsarten: Application Layer

Die Zweckbindung von Daten hängt naturgemäß in erster Linie von ihrer Kategorie (Domain) ab.

Werden *Mediadaten* im Hinblick auf ihren *Inhalt* betrachtet, lässt sich differenzieren zwischen

- Lern- bzw. wissensvermittelnde Media; im Besonderen natürlich wissenschaftliche Artikel,
- Inhalte, die sich auf Kunst, Freizeit (auch Religion) beziehen, wie dies bei Musik und Filmen ausgeprägt ist, sowie
- maschineninterpretierbare Inhalte, und hierbei im Besonderen Software,

die ihren jeweils eigenen Verwertungsbedingungen folgen, konsumiert, gesammelt und zur Wissensverbreitung oder als Ausgangsmaterial für logische oder mathematische Operationen genutzt werden (oder eine Kombination dieser).

*Stammdaten* im Gegenzug besitzen einen ausgeprägten informativen Charakter. Obwohl sowohl „Harry Potter“ als auch die „Gelben Seiten“ als Printausgaben vorliegen und somit als Media betrachtet werden können, ist die Antizipation von Daten aus diesen Domänen jedoch grundsätzlich verschieden. Für jede funktionierende Gesellschaft sind Stammdaten unverzichtbar, repräsentieren diese doch das Gerüst, um sachlich qualifizierte Entscheidungen überhaupt treffen zu können. Die Verwischung von Stammdaten und fiktiven Daten

<sup>3</sup> See: <https://analytics.google.com/>.

<sup>4</sup> Das Streaming-Format verdrängt zunehmend den klassischen Datei-Download.

hat in der Menschheitsgeschichte schon viel Unheil gebracht, was sich z. B. durch die Hexenverfolgung im Mittelalter belegen lässt. Aber auch der aktuelle Trend in der Politik, Anti-Wissenschaften hoffähig zu machen (z. B. in Bezug auf Klimaänderungen), zeigt ein durchaus problematisches Szenario auf.

Prozessdaten schaffen in diesem Zusammenhang eine neue Datenkategorie, da diese Daten sowohl Subjekt als auch Objekt von IT-Operationen sind und sich zudem in Gänze nur mit IT erschließen lassen. Zwar gibt es bereits umfangreiche Nutzungsszenarien für Prozessdaten, die im Rahmen einer Big Data Analyse zugänglich gemacht werden; im Hinblick auf den fehlenden „Input“ stehen wir aber erst am Anfang, die Bedeutung und den Nutzen dieser Daten vollumfänglich zu erfassen.

Ein interessanter Aspekt in Bezug auf die Verwertbarkeit von Daten ergibt sich unter Berücksichtigung ihrer Relevanz bzw. ihrer *Nützlichkeit* für zukünftige Handlungsmöglichkeiten. Kann die Relevanz von Daten unter diesem Aspekt grob in die Bereiche

- verarbeitungsrelevant, also aussagekräftig und positiv korrelierend,
- eher irrelevant, also entweder redundant oder vernachlässigbar sowie
- eher falsch, d. h. entweder auszufiltern oder negativ korrelierend

eingeteilt werden, spielt speziell die zeitliche Nähe der Daten und ihre Lebensdauer im Hinblick auf ihren Nutzungszeitpunkt eine bedeutende Rolle: *Volatilität*. Stammdaten und Mediadata sind in der Regel lange gültig, während Prozessdaten nur für eine beschränkte Zeitdauer von Bedeutung sind; also solange wie mithilfe dieser auf den (laufenden Meta-)Prozess Einfluss genommen werden kann.

## Die Stakeholder der Daten

Nachdem wir die Dimensionen unseres Modells ausgeleuchtet haben, verbleibt in Ergänzung die Fragen nach den Dateneigentümer- und -verwertungsrechten zu klären (Abb. 5):

Auch hier verfolgen wir einen geschichteten Ansatz der

- den *Erzeuger* der Daten,
- den *Prinzipalnutzer* der Daten und

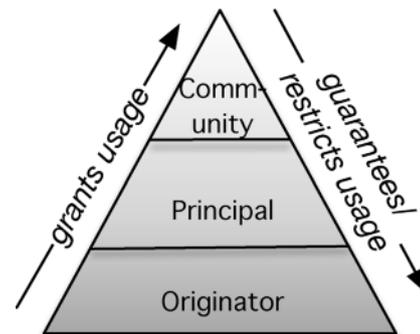


Abb. 5 Stakeholder für Big Data und ihr Zusammenspiel

- die *Gesellschaft* in Bezug auf die systemische Nutzung der Daten

adressiert.

Alle diese drei Parteien betrachten wir als *Stakeholder*, die wiederum ihren eigenen Interessen und Nutzungsmöglichkeiten bzw. Compliance-Regeln folgen. Unter Differenzierung der Producer-/Consumer-Verhältnisse beleuchtet Abb. 5 diese Rollen in Bezug auf

- das Vergeben („granting roles“) von Nutzungsrechten an den Daten sowie
- der Sicherstellung („guaranteeing legal usage“) der rechtmäßigen Nutzung der Daten.

Beide Aktionen verlangen ein rechtlich abgesichertes Rahmenwerk. Hierfür kann das Lizenz- und Patentrecht Pate stehen: Der (verbriefte) Besitzer eines Markenrechts garantiert im Rahmen von Lizenzvereinbarung die Nutzung und zugleich kann eine nicht rechtmäßige Nutzung sanktioniert werden. Historisch betrachtet findet dies aber nur für Sachverhalte mit abschätzbarem (und angemeldetem) Nutzen statt (was z. B. durch eine ISBN-Nummer verbrieft wird). Was ist aber mit Daten, für die dieser Nutzwert nicht erkannt oder kein Nutzungsrecht proklamiert wird?

## Der Originator

Im Rahmen unseres Herkunftsmodells

- beschreiben *Stammdaten* den Erzeuger (Subjekt) oder einen Sachverhalt,
- werden *Mediadata* von Erzeugern generiert (Objekte von Subjekten) oder

- beinhalten *Prozessdaten* Informationen spezielle über die Quelle und das Ziel von Transaktionen, wobei die Quelle als auslösender Erzeuger aufgefasst werden kann.

Somit sind die Daten entweder ein Abbild des Erzeugers oder aber sind werden durch primäre Aktionen des Erzeugers generiert. Sie unterliegen damit dem „Schutz personenbezogener Daten“, sofern sie Eigenschaften von Menschen beschreiben.

Sind menschliche Erzeuger im Spiel<sup>5</sup>, sind sowohl

- Stammdaten als auch
- Mediadaten

zunächst a-priori als persönlich und damit als „none-disclosure“ Daten zu betrachten, für die explizite Verwertungsrechte eingeräumt werden müssen.

Die deutsche Gesetzgebung folgt dem umfänglich:

- (Mediadaten) – *Grundgesetz* [Artikel 10 (1)] [8]: „Das Briefgeheimnis sowie das Post- und Fernmeldegeheimnis sind unverletzlich.“
- (Stammdaten) – *Bundesdatenschutzgesetz* [§ 22] [9]: „Die Erhebung, Verarbeitung und Nutzung personenbezogener Daten sind nur zulässig, soweit dieses Gesetz oder eine andere Rechtsvorschrift dies erlaubt oder anordnet oder der Betroffene eingewilligt hat.“
- (Unspezifiziert) – *Gesetz gegen den unlauteren Wettbewerb* [§ 17] [6]: „Verrat von Geschäfts- und Betriebsgeheimnissen“.

*Private Mediadaten* wie Korrespondenzen (Briefe, E-Mails) unterliegen entsprechend dem Gesetzgeber grundsätzlich der Vertraulichkeit, speziell bei der Übertragung über öffentliche Kanäle (traditionell der Post).

Für *Öffentliche Mediadaten* bzw. deren Inhalte gilt die „Meinungsfreiheit“ wie dies beispielsweise im Grundgesetz [7] garantiert wird. Zugleich werden aber auch Auflagen nicht nur im Hinblick auf deren Inhalt, sondern auch im Hinblick auf die Rechte Dritter formuliert.

Bezüglich der *Prozessdaten*, existieren derzeit keine entsprechenden Regelungen in deutschen Gesetzen (vgl. [23]); allerdings ist die Reichweite von [6] dergestalt, dass Prozessdaten hierunter fallen können, sofern sie „Geheimnisse“ darstellen.

### Der Prinzipal

Wir betrachten als *Prinzipal* denjenigen Nutzer, für den oder mit dessen Hilfe die Daten (explizit) erstellt bzw. (implizit) erzeugt wurden.

Im besonderen für *Prozessdaten* – wie in Abb. 3 gezeigt – lässt sich die Herkunft der Daten attribuieren, d. h. ob diese auch einem betrieblichen Kontext, aus öffentlichen Informationsquellen oder aus Kundenbeziehungen stammen. Daten aus Geschäftsbeziehungen (speziell mit Kunden) besitzen eine hervorgehobene Rolle, da diese die Wechselbeziehung zwischen zwei Subjekten beschreiben, die entweder als natürliche Person oder als Organisation (juristische Person) in Erscheinung treten.

Während [6] pauschal die Weitergabe von Firmengeheimnissen unter Strafe stellt, ist nicht klar, inwieweit dies auf Prozessdaten zutrifft. Das Verhältnis zwischen *Originator* und *Prinzipal* – konkret, zwischen Kunde und Firma – wird in der Praxis durch ein uni-laterales Abkommen im Rahmen der „Allgemeine Geschäftsbindungen“ (AGB) geregelt. Uni-lateral deshalb, weil die AGBs in der Regel nicht verhandelbar sind. Prozessdaten – wie hier vorgestellt – als spezifische Datenkategorie, wenn diese aus Geschäftstransaktionen mit Privatkunden stammen, sind Gegenstand intensiver öffentlicher Diskussion, speziell dann, wenn deren Nutzung Rückschlüsse auf die Identität des Geschäftspartners ermöglicht, ggf. diesen kompromittieren kann und dessen Recht auf Vertraulichkeit (Privacy) durch den Prinzipal einschränkt.

Diese Disparität kann dadurch ausgedrückt werden, dass der dominierende Geschäftspartner (Prinzipal) einfach verlangen kann: „Wenn Du nicht unseren AGBs zustimmst, können wir leider kein Geschäft mit Dir abschließen und Du musst Dir einen andern Geschäftspartner suchen.“ AGBs drehen somit das in Abb. 5 auf der linken Flanke dargestellt Verhältnis um. Der Unmut war groß, als die Europäische Kommission eine Cookie-Richtlinie [12] in Kraft gesetzt hat, die im Grunde ein *Opt-In* für Cookies fordert und die Parität wieder teilweise herstellte.

<sup>5</sup> Vgl. die Posse um das „Affenselfie“ <http://www.sueddeutsche.de/panorama/rechtsstreit-um-affen-selfie-streit-ueber-affen-selfie-beendet-1.3662934>.

Allenthalben ist die Notwendigkeit der Verfügbarkeit von Prozessdaten, die bei geschäftlichen Transaktionen abfallen, in einer IT-getriebenen Geschäftswelt unverzichtbar. Diese Daten sind den beteiligten Parteien offenzulegen, bzw. gehören beiden Partnern; zumindest im Prinzip. Werden diese Daten an Dritte weitergereicht, sollte dies nicht ohne gesetzliche Regularien erfolgen, die unabhängig von den AGBs gültig sind. Geschäftsmodelle, die dies im Grundsatz missachten und die Transaktionsdaten als Beutegut betrachten und diese veräußern, haben auf lange Sicht keinen Bestand.

## Die Community

Die Gesellschaft – und hier speziell gesellschaftliche Organisationen wie Staaten und deren Regierungen – bestimmen den (gesetzlichen) Kontext, unter dem Daten weitergegeben und genutzt werden dürfen. In Deutschland liegen z. B. folgende normative Regelungen vor:

- Im Hinblick auf *Stammdaten* greift das BDSG [9] und bestimmt, unter welchen Umständen *Stammdaten* erhoben, gespeichert und genutzt werden dürfen.
- Die organisatorische Umsetzung von Nutzungs- und Verbreitungsrechten von *Mediadaten* werden in Deutschland nicht staatlichen Gesellschaften (öffentlichen Rechts) als Treuhänder zugesprochen, wobei insbesondere die *VG Wort*<sup>6</sup> für literarische und die *GEMA*<sup>7</sup> für musikalische Inhalte zu benennen sind. Als Treuhänder-Organisation fällt diesen die Rolle zu, die Rechte des Urhebers (*Originators*) gegenüber dem nutzenden Prinzipal zu regeln; speziell im Hinblick auf die anfallenden Nutzungsentgelte (*Royalties*).
- Das Augenmerk in Bezug auf *Prozess- bzw. Transaktionsdaten*, die häufig auch als Metadaten bezeichnet werden, lag bislang ausschließlich auf den Daten, die bei Telko- und Internet-Service-Provider anfallen, wie dies beispielsweise im deutschen „Telekommunikationsgesetz“ [5] geregelt ist. Das aktuelle Dokument der Europäischen Kommission „Proposal for a Regulation on Privacy and Electronic Communications“ [14] adressiert Originator und Prinzipal und weist dem Originator Rechte zu, während der Prinzipal zum sorgfältigen

<sup>6</sup> Siehe <http://www.vgwort.de/startseite.html>.

<sup>7</sup> Siehe <https://www.gema.de>.

Umgang mit Stamm- und Prozessdaten verpflichtet wird.

## Fazit und Vergleich mit anderen Modellen

Wir haben ein Modell entwickelt, das Daten attribuiert und speziell auch für Big Data angewandt werden kann sowie im Besonderen die Interessen der Stakeholder beleuchtet (Abb. 5). Das Modell fußt zudem auf einem Schichtenmodell der Daten (Abb. 2) und trägt im Rahmen eines Domain-Modells (Abb. 3) auch der historischen Entwicklung von „Daten“ in ihrem allgemeinsten Sinn Rechnung. Andere Modelle, wie z. B. in [20] dargestellt, bevorzugen statt des Terminus „Schichten“ sog. „Datendimensionen“; allerdings mit durchaus vergleichbarer Attribuierung.

Zudem wird eine ähnliche – aber nicht so explizit formulierte Sichtweise – auch in der aktuellen europäischen Rechtsprechung formuliert, so wie sie im EU-DSGVO [14] zu finden ist. Für alle Stakeholder im Bereich der Datennutzung ist relevant, die eigene Rolle im Erstellungs-/Verarbeitungs-/Verbreitungskontext gerade auch von Big Data zu kennen und die spezifischen Merkmale der Daten zu beachten. Im Rahmen dieser Ausarbeitung hoffen wir, hierzu einen Beitrag geleistet zu haben.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Literatur

1. Barbaro M, Zeller T Jr (2006) A Face Is Exposed for AOL Searcher No. 4417749. <http://www.nytimes.com/2006/08/09/technology/09aol.html?r=1>, last access: 20.8.2017
2. Barger J. Hungriges Datenmonster. <https://www.heise.de/ct/ausgabe/2016-24-Was-Facebook-mit-Ihren-Daten-macht-3458219.html>, letzter Zugriff: 20.8.2019
3. Bundesbeauftragte für den Datenschutz und die Informationsfreiheit. [https://www.bfdi.bund.de/DE/Europa\\_International/International/Artikel/SafeHarbor.html](https://www.bfdi.bund.de/DE/Europa_International/International/Artikel/SafeHarbor.html), letzter Zugriff: 5.12.2017
4. Bundesbeauftragte für den Datenschutz und die Informationsfreiheit. EU-US Privacy Shield und Datenübermittlungen in die USA. [https://www.bfdi.bund.de/DE/Europa\\_International/International/Artikel/EU-US\\_PrivacyShield\\_Daten\\_übermittlungenUSA.html](https://www.bfdi.bund.de/DE/Europa_International/International/Artikel/EU-US_PrivacyShield_Daten_übermittlungenUSA.html), letzter Zugriff: 20.8.2019
5. Bundesregierung. Telekommunikationsgesetz. [https://www.gesetze-im-internet.de/bundesrecht/tkg\\_2004/gesamt.pdf](https://www.gesetze-im-internet.de/bundesrecht/tkg_2004/gesamt.pdf), letzter Zugriff: 20.8.2019
6. Bundesregierung (2017) Gesetz gegen den unlauteren Wettbewerb. <https://dejure.org/gesetze/UWG/17.html>, letzter Zugriff: 20.8.2019
7. Bundesregierung (2017) Grundgesetz der Bundesrepublik Deutschland. [https://www.gesetze-im-internet.de/gg/art\\_5.html](https://www.gesetze-im-internet.de/gg/art_5.html), letzter Zugriff: 20.8.2019

8. Bundesregierung (2017) Grundgesetz der Bundesrepublik Deutschland. [https://www.gesetze-im-internet.de/gg/art\\_10.html](https://www.gesetze-im-internet.de/gg/art_10.html), letzter Zugriff: 20.8.2019
9. Bundesregierung (2018) Bundesdatenschutzgesetz. [https://www.gesetze-im-internet.de/bdsg\\_2018/](https://www.gesetze-im-internet.de/bdsg_2018/), letzter Zugriff: 20.8.2019
10. Chen M, Mao S, Liu Y (2014) Big Data: A Survey. <https://link.springer.com/article/10.1007/s11036-013-0489-0>, last access: 20.8.2019
11. Demings EJ (2000) Out of the Crisis. First MIT Press edition, ISBN-13: 978-0-262-54114-2
12. European Comission. Cookies. [http://ec.europa.eu/ipg/basics/legal/cookies/index\\_en.htm](http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm), last access: 20.8.2019
13. European Commission (2014) Protection of Personal Data. <http://eur-lex.europa.eu/legal-content/EN/LSU/?uri=celex:31995L0046>, last access: 20.8.2019
14. European Commission (2017) Proposal for a Regulation on Privacy and Electronic Communications. <https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-privacy-and-electronic-communications>, last access: 20.8.2019
15. Hoeren T (2008) Grundzüge des Internet Rechts, 2. Auflage. C.H. Beck
16. Hoeren T (Hrsg) (2014) Big Data und Recht, vol 83. C.H. Beck
17. International Trade Administration. Privacy shield overview. <https://www.privacyshield.gov/Program-Overview>, last access: 20.8.2019
18. Kisker H, Sridharan S (2013) Big Data: Gold Rush Or Illusion? <http://docplayer.net/3190497-Complimentary-webinar-big-data-gold-rush-or-illusion-bigger-better-stronger-faster.html>, last access: 20.8.2019
19. Madder C (2015) A Brief History of Deep Linking. <https://techcrunch.com/2015/06/12/a-brief-history-of-deep-linking/>, last access: 20.8.2019
20. Markl V (2015) Gesprengte Ketten. Informatik Spektrum. <http://www.bbd.cerlin.de/newsevents/blog-artikel/>, letzter Zugriff: 20.8.2019
21. Morvan L. Data: The fuel of the Digital Economy and SME Growth. [https://www.accenture.com/t20160901T103414\\_\\_w\\_\\_us-en/\\_acnmedia/PDF-29/Accenture-Data-The-Fuel-Of-The-Digital-Economy-And-SME-Growth.pdf](https://www.accenture.com/t20160901T103414__w__us-en/_acnmedia/PDF-29/Accenture-Data-The-Fuel-Of-The-Digital-Economy-And-SME-Growth.pdf), last access: 20.8.2019
22. Research Lab for Digital Business (2015) Big Data Report (Part 1). [http://www.t-systems-mms.com/fileadmin/mms\\_upload/04\\_Unternehmen/Downloads/Big\\_Data\\_Report\\_Teil\\_1.pdf](http://www.t-systems-mms.com/fileadmin/mms_upload/04_Unternehmen/Downloads/Big_Data_Report_Teil_1.pdf), last access: 20.8.2019
23. Schermann M, Hensen H, Markl V, Buchmüller C, Bitter T, Hoeren T (2014) Big Data – eine Interdisziplinäre Chance für die Wirtschaftsinformatik. *Wirtschaftsinformatik* 56(5):281–287
24. Schwandt N (2016) Big Data vs. Datenschutz - Widerspruch oder Bedingung für den Erfolg?, Provadis School of International Management & Technology AG
25. Smith M, Szongott C, Henne B, von Voigt G (2012) Big Data Privacy Issues in Public Social Media. In: 2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST). IEEE, IEEE
26. Wikimedia. Share Your Work. <https://creativecommons.org/licenses>, last access: 20.8.2019