

Krotova, Alevtina; Eppelsheimer, Jan

Article

Data governance in der wissenschaftlichen Literatur: Eine Begriffsklärung anhand einer Text-Mining-basierten Literaturrecherche

IW-Trends - Vierteljahresschrift zur empirischen Wirtschaftsforschung

Provided in Cooperation with:

German Economic Institute (IW), Cologne

Suggested Citation: Krotova, Alevtina; Eppelsheimer, Jan (2019) : Data governance in der wissenschaftlichen Literatur: Eine Begriffsklärung anhand einer Text-Mining-basierten Literaturrecherche, IW-Trends - Vierteljahresschrift zur empirischen Wirtschaftsforschung, ISSN 1864-810X, Institut der deutschen Wirtschaft (IW), Köln, Vol. 46, Iss. 3, pp. 55-71, <https://doi.org/10.2373/1864-810X.19-03-04>

This Version is available at:

<https://hdl.handle.net/10419/209531>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



IW-Trends 3/2019

Data Governance in der wissenschaftlichen Literatur

Alevtina Krotova / Jan Eppelsheimer

Vorabversion aus: IW-Trends, 46. Jg. Nr. 3
Herausgegeben vom Institut der deutschen Wirtschaft Köln e. V.

Verantwortliche Redakteure:

Prof. Dr. Michael Grömling, Telefon: 0221 4981-776

Holger Schäfer, Telefon: 030 27877-124

groemling@iwkoeln.de · schaefer.holger@iwkoeln.de · www.iwkoeln.de

Die IW-Trends erscheinen viermal jährlich, Bezugspreis € 50,75/Jahr inkl. Versandkosten.

Rechte für den Nachdruck oder die elektronische Verwertung erhalten Sie über
lizenzen@iwkoeln.de.

ISSN 0941-6838 (Printversion)

ISSN 1864-810X (Onlineversion)

© 2019 Institut der deutschen Wirtschaft Köln Medien GmbH

Postfach 10 18 63, 50458 Köln

Konrad-Adenauer-Ufer 21, 50668 Köln

Telefon: 0221 4981-452

Fax: 0221 4981-445

iwmedien@iwkoeln.de

www.iwmedien.de

Data Governance in der wissenschaftlichen Literatur - Eine Begriffsklärung anhand einer Text-Mining-basierten Literaturrecherche

Alevtina Krotova / Jan Eppelsheimer, September 2019

Zusammenfassung

Data Governance bildet die Grundlage für eine effiziente Datenbewirtschaftung in Unternehmen und über Unternehmensgrenzen hinweg. Es wird in der Praxis allerdings keine eindeutige und einheitliche Definition von Data Governance und ihrer Bestandteile angewandt. Um einen umfassenden Überblick über Data Governance zu ermöglichen, die bisherigen Tendenzen in der Forschung darzustellen und den potenziellen Forschungsbedarf aufzudecken, wurde eine Text-Mining-basierte systematische Literaturrecherche durchgeführt. Data Governance hat vor allem seit dem Jahr 2012 zunehmend Raum in der wissenschaftlichen Diskussion und Literatur eingenommen. Insgesamt wurden anhand dieser Methode neun Cluster identifiziert, die bestehende Forschungsschwerpunkte darstellen. Daraus folgt, dass in der Forschung vor allem die Bedeutung von Data Governance für die digitalen Technologien thematisiert wird. Darüber hinaus sind Aspekte wie die Bewirtschaftung von Daten als Wirtschaftsgüter, das Datenmanagement, die Datenqualität, der Datenschutz, Compliance oder die Organisation von Daten im Netzwerk zu finden. In der Gesundheitsbranche spielt Data Governance relativ zu den anderen Branchen bereits eine große Rolle. Durch eine effektive Koordination der Anforderungen der unternehmerischen Praxis und der wissenschaftlichen Problemanalyse kann das Thema Data Governance besser untersucht, weiterentwickelt und umgesetzt werden.

Stichwörter: Data Economy, Data Governance, Clusteranalyse

JEL-Klassifikation: L14, L15, O30

Data Governance als Teil der Data Economy

Die Erkenntnis, dass Daten eine wichtige ökonomische Ressource sind, ist in den meisten Unternehmen angekommen. Durch eine effiziente Bewirtschaftung von Daten haben Unternehmen die Möglichkeit, ihre laufenden Prozesse zu optimieren, das bestehende Angebot zu erweitern oder komplett neue Geschäftsmodelle zu entwickeln. Die Potenziale von Daten sind längst nicht mehr nur auf interne Unternehmensprozesse begrenzt. In einer Data Economy werden die Wertschöpfungspotenziale von Daten in einem Unternehmensökosystem realisiert (Azkan et al., 2019). Eine Studie des Projekts DEMAND (2019) ergibt allerdings, dass mit 84 Prozent die meisten deutschen Unternehmen das Potenzial ihrer Daten noch nicht ausschöpfen und zu den Einsteigern in der Data Economy zählen. Bei der internen Datenbewirtschaftung besteht vor allem in den Kategorien Data Management und Data Governance ein hoher Nachholbedarf.

Der Begriff Data Governance wird derzeit oftmals abstrakt wahrgenommen. Dabei gehört Data Governance zu den festen Bestandteilen der Data Economy (Heumann/Jentzsch, 2019). Wörtlich übersetzt bedeutet Data Governance die Steuerung von oder die Herrschaft über Daten. Daraus lässt sich ableiten, dass mithilfe von Data Governance der Umgang mit Daten in Unternehmen koordiniert und überwacht wird. Die Bedeutung und die Bestandteile von Data Governance werden jedoch unterschiedlich interpretiert (Brüning et al., 2017). Laut einer Umfrage von redgate (2018) geben 52 Prozent der befragten Unternehmen an, dass es ihnen an Verständnis mangelt, was im Zuge von Data Governance von ihnen verlangt wird. Ein einheitliches Verständnis für Data Governance und ihre Bestandteile ist also erforderlich, damit alle Akteure bei dem Umgang mit Daten die gleiche Sprache sprechen.

In dem folgenden Beitrag wird zunächst eine Begriffsklärung vorgenommen. Um einen Überblick über Data Governance zu liefern, die bisherigen Tendenzen in der Forschung darzustellen und den potenziellen Forschungsbedarf aufzudecken, wird anschließend eine Text-Mining-basierte systematische Literaturrecherche durchgeführt. Anhand dieser Methode werden Cluster identifiziert, die bestehende Forschungsschwerpunkte im Kontext von Data Governance darstellen.

Eine Definition von Data Governance

Für eine erfolgreiche Implementierung von Data Governance im eigenen Unternehmen müssen die dafür zuständigen Personen eine klare Vorstellung davon haben, was der Begriff bedeutet. Obwohl Data Governance ein wichtiger Bestandteil der Datenbewirtschaftung ist, fehlt dafür sowohl in der wissenschaftlichen Literatur als auch in der betrieblichen Praxis eine einheitliche Definition. Ein funktionierendes Datenökosystem, in dem die beteiligten Unternehmen problemlos miteinander interagieren und Daten austauschen können, wird nur mit einer einheitlichen Data Governance möglich sein.

Im Rahmen des Projekts Data Economics and Management of Data Driven Business (DEMAND), das aus einem Konsortium mehrerer Institute und Unternehmen besteht und die Entwicklung eines Ansatzes zur effizienten Datenbewirtschaftung in Unternehmen zum Ziel hat (DEMAND, 2019), wurde bereits eine Definition von Data Governance entwickelt. Diese Definition berücksichtigt alle praxisrelevanten Aspekte der Datenbewirtschaftung und vereinheitlicht diese zu einem Gesamtkonzept, das sowohl für einzelne Unternehmen als auch für ganze Datenökosysteme anwendbar ist (DEMAND, 2019): Data Governance stellt demnach das Rahmenwerk dar, welches die Grundlage für den Umgang mit und die Bewirtschaftung von Daten in einem Unternehmen für alle Stakeholder bildet. Dieses Rahmenwerk besteht wiederum aus sechs Dimensionen:

- **Assets:** Definition und Identifikation von Daten und deren ökonomischen Wert, Definition von einheitlichen, unternehmensübergreifenden Standards für die Datenbewertung und Einhaltung von Datenqualitätsstandards.
- **Roles, Tasks & Responsibilities:** Festlegung von Rollen für die Data Execution und unternehmensübergreifende Rollen für das Datenökosystem, Zuweisung von Zuständigkeiten für Daten und datengetriebene Prozesse.
- **Processes:** Überwachung der internen und unternehmensübergreifenden Datenprozesse, Überwachung des Teilens und der Nutzung von Daten, Entscheidungen über das Management und die Nutzung von Daten.
- **Architecture & Tools:** Unterstützung von Data Governance durch Technologie, Definition von Standards für die technische Umsetzung und die Auswahl der genutzten Tools für die Datenbewirtschaftung.

- **Security:** Definition von internen und unternehmensübergreifenden Standards zur Datensicherheit, Festlegung der Zugriffsrechte, Vorgehensweise bei Sicherheitsverstößen.
- **Compliance:** Sicherstellung der Einhaltung von internen und externen Anforderungen/Richtlinien an das Datenmanagement und den Datenschutz.

In welchem Umfang Data Governance in der Praxis umgesetzt wird, muss jedes Unternehmen individuell entscheiden – eine universelle Implementierungsstrategie für Data Governance existiert nicht (iRights.Lab, 2018b).

Data Governance in der wissenschaftlichen Literatur

Die Kollaboration von Unternehmen und der Wissenschaft ist ein Treiber der digitalen Transformation. Aus dem Wissenschaftsbetrieb heraus entstehen Impulse und Handlungsempfehlungen für die unternehmerische Praxis. Die Unternehmen profitieren von den wissenschaftlichen Erkenntnissen, wenn die von den Wissenschaftlern behandelten Themen für die betriebliche Praxis tatsächlich relevant sind. In der vorliegenden Analyse liegt der Fokus auf einer Auswertung der wissenschaftlichen Literatur zu Data Governance. Dafür wurde die Methode der systematischen Literaturrecherche (SLR) nach Kitchenham und Charters (2007) angewendet. Darunter wird eine Form der Sekundärstudie verstanden, die eine klar definierte Methodik verwendet, um alle verfügbaren Daten im Zusammenhang mit einer bestimmten Forschungsfrage so zu identifizieren, zu analysieren und zu interpretieren, dass die Vorgehensweise unverzerrt und (bis zu einem gewissen Grad) wiederholbar ist.

Als Basis der vorliegenden Recherche wurde Scopus ausgewählt, eine der größten Abstract- und Zitationsdatenbanken für referierte Literatur mit mehr als 70 Millionen Einträgen. Dazu gehören wissenschaftliche Zeitschriften, Bücher und Tagungsberichte (Elsevier, 2019). Mit der Suchfunktion der Datenbank wurden Veröffentlichungen nach den Suchbegriffen Data Governance und Information Governance gefiltert. Information Governance wurde als Suchbegriff mitaufgenommen, da die Begriffe Daten und Information eng verwandt sind. Gemäß der Definition von Ackoff (1999) beschreiben Daten die Eigenschaften von Objekten und Ereignissen. Im Unterschied dazu bestehen Informationen aus verarbeiteten Daten, wobei die Verarbeitung darauf abzielt, den Nutzen zu erhöhen, der durch die Daten ermög-

licht wird. Durch die Filterung wurden nur Veröffentlichungen extrahiert, die im Titel, im Abstract oder als Schlagwort Data Governance und/oder Information Governance enthalten. Abgesehen davon wurden die Namen der Autoren, das Erscheinungsjahr und die von den Autoren selbst vergebenen Schlagwörter extrahiert. Auf eine Begrenzung des relevanten Zeitraums, in dem die Quellen veröffentlicht wurden, wurde bewusst verzichtet, um die zeitliche Entwicklung der Publikationen analysieren zu können.

Die Recherche ergab zum Datenstand Januar 2019 insgesamt 827 Treffer. Nach der Eliminierung von Duplikaten reduzierte sich die Anzahl der Quellen auf 814. Die anschließende Analyse wurde in zwei Schritte unterteilt. Im ersten Schritt wurde die Anzahl der Publikationen pro Jahr betrachtet. Anhand dessen konnte untersucht werden, seit wann das Thema Data Governance in der wissenschaftlichen Literatur diskutiert wird, ob das Interesse an diesem Thema mit der Zeit tendenziell gestiegen ist und ob das Interesse an Data Governance relativ zu verwandten Themen als hoch oder niedrig einzustufen ist. Im zweiten Schritt erfolgte eine inhaltliche Analyse der wissenschaftlichen Publikationen zum Thema Data Governance mithilfe einer Text-Mining-Methode. Auf Basis einer Clusteranalyse wurden Forschungsschwerpunkte und weiterer Forschungsbedarf identifiziert.

Entwicklung der wissenschaftlichen Veröffentlichungen

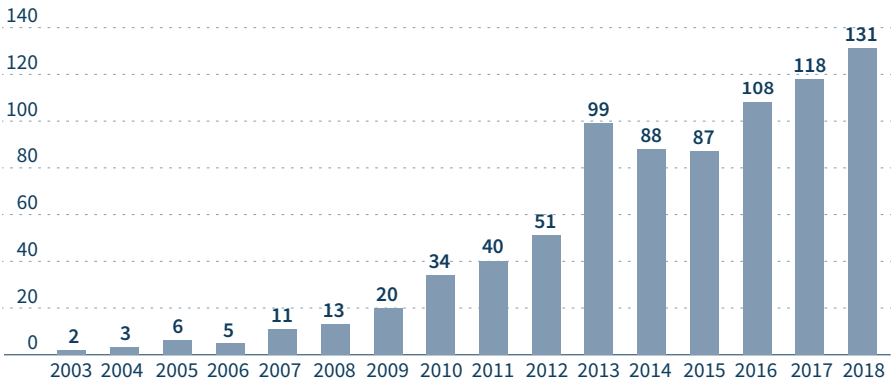
Die ersten Publikationen zu Data Governance stammen aus dem Jahr 2003. Bis zum Jahr 2018 stieg die Anzahl der Veröffentlichungen zu diesem Thema deutlich an (Abbildung 1). Im Jahr 2013 wurde der bislang größte absolute Zuwachs an vorhandener Literatur verzeichnet: Die Beiträge verdoppelten sich binnen Jahresfrist nahezu von 51 auf 99. Im Jahr 2018 erreichte die Anzahl der wissenschaftlichen Literatur zu Data Governance mit 131 Veröffentlichungen das bisherige Hoch (für eine ausführliche Beschreibung der zugrunde liegenden Literatur s. Krotova/Epelsheimer (2019)).

Um die Anzahl der Veröffentlichungen zu Data Governance in Relation zu verwandten Themen zu setzen, wurde mit der gleichen Methode nach den Publikationen zum Thema Data Management und Information Management gesucht. Die Suche ergab für den Zeitraum 2003 bis 2018 insgesamt 118.791 Veröffentlichungen zu dem Begriff Data Management, was dem 145-Fachen der Data-Governance-Ver-

Data Governance in der Literatur

Abbildung 1

Anzahl der Veröffentlichungen zu Data Governance und Information Governance



Auswertung auf Basis von 814 Veröffentlichungen in Scopus.
Quellen: Scopus; Institut der deutschen Wirtschaft

Abbildung 1: <http://dl.iwkoeln.de/index.php/s/9B4jT5foNXjzGzJ>

öffentlichungen entspricht. Die Diskrepanz zwischen den beiden Zahlen deutet darauf hin, dass das wissenschaftliche Interesse an dem Thema Datenmanagement bislang deutlich höher ist, obwohl Data Governance eine Voraussetzung für das effiziente Data Management darstellt. Der Begriff Data Management taucht bereits deutlich früher in der wissenschaftlichen Literatur auf. Die älteste identifizierte Quelle stammt von Blackburn et al. aus dem Jahr 1931. Diese frühe Studie thematisierte unter anderem die Bedeutung der Schnittstellen zwischen Abteilungen in Unternehmen, zwischen denen ein Informationsaustausch stattfindet.

Clusteranalyse der Data Governance Literatur

Ein tieferer Einblick in die relevante Literatur (s. ausführlich Krotova/Eppelsheimer, 2019) wird im zweiten Schritt der Analyse mithilfe eines Clusterungsverfahrens ermöglicht. Als Basis für diese Auswertung dienen die von den Autoren festgelegten Schlagwörter für die jeweiligen Publikationen. Insgesamt verfügten 484 Quellen über entsprechende Schlagwörter. Unter Anwendung der Software Gephi wurden die Verbindungen dieser Schlagwörter analysiert und Cluster identifiziert. Dafür wurde jedes Vorkommen von zwei Schlagwörtern in der gleichen Quelle als eine Verbindung gewertet. Die Gesamtmenge des resultierenden Netzwerks aus Schlagwörtern stellen die Verbindungen dar, die mehr als einmal aufgetreten sind.

Zur Identifikation der einzelnen Cluster wurde ein in Gephi implementierter Algorithmus von Blondel et al. (2008) benutzt. Dieser weist jedem Schlagwort ein Cluster zu und zwar so, dass die Schlagwörter innerhalb eines Clusters stärker miteinander verbunden sind als mit den Schlagwörtern außerhalb des Clusters. In diesem Fall bedeutet das, dass zwei in einem Cluster verbundene Schlagwörter mit einer höheren Wahrscheinlichkeit gemeinsam genannt werden als zwei beliebige Schlagwörter. Mit dieser Methode konnten insgesamt neun Cluster von Schlagwörtern identifiziert werden.

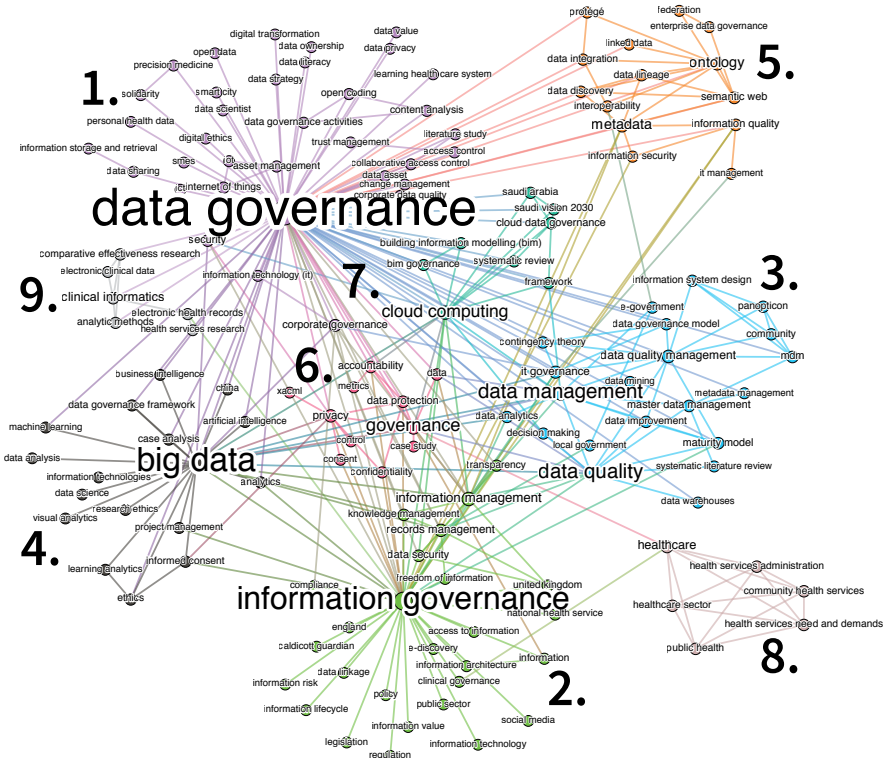
Das resultierende Netzwerk wurde in Gephi visualisiert, wobei ein gemeinsames Auftreten von Schlagwörtern durch Verbindungslinien abgebildet wurde (Abbildung 2). Die Schlagwörter, die nicht zumindest indirekt mit den Begriffen Data Governance und Information Governance verbunden waren, wurden ausgeschlossen. Die entstehenden Cluster hatten unterschiedliche Größen. Während das größte Cluster 35 Schlagwörter hatte, verfügte das kleinste lediglich über sechs Schlagwörter. Für jedes Cluster wurden darüber hinaus die Schlagwörter nach ihrem Grad, also der Anzahl an Verbindungen des Schlagworts, geordnet und das Schlagwort mit dem höchsten Grad in jedem Cluster als das Schlüsselwort bezeichnet. Im Folgenden werden die einzelnen Cluster ausführlich vorgestellt, angefangen mit dem größten Cluster und anschließend in absteigender Reihenfolge. Anhand der Schlagwörter, die zu den jeweiligen Clustern gehören, werden für diese Cluster relevante Themen und Aspekte der Data Governance identifiziert.

Cluster 1: Data Governance

Mit 35 Knoten ist das Cluster um das Schlüsselwort Data Governance das größte der neun Cluster. Die Schlagwörter in diesem Cluster sind mit einer höheren Wahrscheinlichkeit mit dem Schlagwort Data Governance verbunden als mit einem zufällig ausgewählten Schlagwort. Durch die Betrachtung der entsprechenden Begriffe können demzufolge Themenbereiche identifiziert werden, die im Zusammenhang mit Data Governance am häufigsten diskutiert werden.

Wie Abbildung 2 zeigt, sind nicht alle Schlagwörter mit dem Begriff Data Governance direkt verbunden. Dies liegt daran, dass Data Governance in manchen Publikationen im Abstract erwähnt, doch nicht als Schlagwort aufgenommen wurde. In solchen Publikationen spielt Data Governance eine eher untergeordnete Rolle.

Taxonomie für Data Governance auf Basis einer Literatursauswertung Abbildung 2
Netzwerk aus Schlagwörtern in der Literatur zu dem Thema Data Governance



Cluster zur besseren Erkennbarkeit in verschiedenen Farben; die Größe der Schlagwörter spiegelt die Häufigkeit des jeweiligen Schlagworts wider.
Auswertung auf Basis von 485 Veröffentlichungen in Scopus.
Quellen: Scopus; Institut der deutschen Wirtschaft

Die Schlagwörter, die die restlichen Cluster bilden, repräsentieren Themenbereiche, die zwar eng mit dem Begriff Data Governance verknüpft sind, allerdings nur eine indirekte Verbindung zu Data Governance aufweisen. Als einer dieser Themenbereiche kann der *Einsatz von Daten im Digitalisierungskontext* genannt werden. Daten bilden den Kern der digitalen Transformation und sind ein essenzieller Faktor bei der Entwicklung des Internets der Dinge oder von Smart Cities. Auch der Tausch von Daten ist mittlerweile für viele Unternehmen ein Bestandteil des Geschäftsmodells. Data Governance bietet in diesem Kontext das Rahmenwerk für den optimalen Dateneinsatz.

Der zweite große Themenbereich ist der *Umgang mit Daten* in Unternehmen. Die Bewirtschaftung von Daten als ein Wirtschaftsgut bildet eine der sechs Kerndimensionen von Data Governance. Dementsprechend wird die Bedeutung von Data Governance für die Verwertung des Datenbestands in Unternehmen vermehrt in der wissenschaftlichen Literatur diskutiert. Auch die Aspekte der Zuständigkeiten und Verantwortung für Daten sowie Datensicherheit finden sich in diesem Cluster wieder. In diesem Zusammenhang bietet Data Governance Mitarbeitern im Unternehmen einen Leitfaden für ihre Aufgaben und Pflichten beim Umgang mit Daten.

Cluster 2: Information Governance

Im zweitgrößten Cluster sind 27 Knoten hinsichtlich des Schlüsselworts Information Governance verbunden. Der Fokus dieses Clusters liegt auf der Verwaltung und dem *Management von Informationen*. Informationen werden als ein Produkt der Datenauswertung definiert, aus dem Wissen – mit Erfahrungswerten gebündelte Informationen – generiert wird (BVDW, 2018, 5). Basierend auf dieser Definition können Informationen als aufbereitete Daten gesehen werden, was die enge Verbindung zwischen Data Governance und Information Governance erklärt.

Im Rahmen von Information Governance werden der *Zugang zu Informationen* und die notwendige *Transparenz* der im Unternehmen verfügbaren Informationen geklärt. Erhöhte Transparenz und Regelung des Zugangs zu Informationen verbessern grundsätzlich das Vertrauen in neue Technologien und erleichtern somit den Umgang damit. Auch der Aspekt der *Compliance* ist in diesem Cluster zu finden. Compliance im Rahmen der Information Governance sorgt dafür, dass besonders schützenswerte und wissensintensive Informationen, die einen gesetzlichen Schutz genießen, von den Mitarbeitern im Unternehmen adäquat behandelt werden. Vor allem für personenbezogene Daten, wie die sensiblen und schutzwürdigen Patientendaten im Gesundheitswesen, müssen im Rahmen von Data (Information) Governance klare Regeln und Richtlinien definiert werden.

Cluster 3: Data Management und Data Quality

Die Schlüsselwörter Data Management und Data Quality verbinden 21 Schlagwörter im dritten Cluster. In der Data Economy spielt die Qualität der verfügbaren Daten eine entscheidende Rolle: Je höher die Qualität der Daten, desto höher ist ihr Nutzen für Unternehmen (Carretero et al., 2016). Data Governance sorgt durch

die Definition entsprechender Standards, Modelle und Strategien für die Erreichung einer guten *Datenqualität*.

Data Management ist ein Teil der *Exekutive der Datenbewirtschaftung* und stellt einen iterativen Prozess dar, angefangen mit der und wiederkehrend zu der Datenerhebung, endend bei der Datenveräußerung oder Löschung (ISO/IEC, 2008). In der wissenschaftlichen Literatur werden die Begriffe Data Management und Data Governance häufig als Synonyme verwendet, was vom Begriffsursprung her nicht korrekt ist: Data Governance und Data Management sind komplementär und nicht substitutiv (Al-Ruithe et al., 2018). Während Data Governance die Rahmenbedingungen für das effektive Datenmanagement definiert, steht das Data Management für die Implementierung der festgelegten Regeln und Prinzipien (Khatri/Brown, 2010). Somit wird das Datenmanagement direkt von der Data Governance beeinflusst und umgekehrt (Otto, 2011).

Cluster 4: Big Data

In diesem Cluster sind 17 Schlagwörter um das Schlüsselwort Big Data gruppiert. Unter Big Data wird nicht mehr ausschließlich ein großes Volumen an Daten verstanden, sondern auch die hohe Geschwindigkeit der Datenverarbeitung, die Vielfalt der Daten, ihre Variabilität, Verlässlichkeit, die Möglichkeiten der Visualisierung und der Datenwert (Khan, 2018).

Einen Schwerpunkt dieses Clusters bildet der Themenbereich *Business Intelligence*. Analysen von Daten werden immer häufiger zur Optimierung laufender Geschäftsprozesse eingesetzt oder bilden ein komplett neues Geschäftsmodell (Analytics-as-a-Service). Data Governance ist für die effektive Datennutzung für Analysen von entscheidender Bedeutung.

Cluster 5: Metadata und Ontology

Im fünften Cluster werden 14 Schlagwörter um die Schlüsselwörter Metadata und Ontology vernetzt. Metadaten sind sogenannte Daten über Daten und sie beschreiben ausgewählte Aspekte anderer Daten (Staab, 2002, 201). Als Ontologie wird ein formales Modell für einen Anwendungsbereich bezeichnet, das den Austausch und das Teilen von Wissen erleichtert (Staab, 2002, 200 f.). Der Fokus in diesem Cluster liegt auf den Methoden für die Organisation und dem Austausch von Daten.

Die *Organisation* von großen Mengen an unstrukturierten *Daten* in strukturiertes Wissen (z. B. in Form von Datenbanken) stellt eine der Herausforderungen dar, vor allem für Großunternehmen. Diese kann mithilfe der Data Governance bewältigt werden. Auch umgekehrt gilt, dass Data Governance in Unternehmen durch effektive Prozessmodellierung, bei der die gesamten Informationen in Ontologien organisiert sind, verbessert werden kann. Metadaten sind in diesem Zusammenhang wichtige Instrumente, um Ontologien zu vereinfachen, die eine große Anzahl an komplexen Prozessen abbilden.

Cluster 6: Governance

Im sechsten Cluster werden elf Schlagwörter um das Schlüsselwort Governance gruppiert. Im Allgemeinen ist mit dem Begriff Governance die Steuerung und die Regelung eines Unternehmens oder einer Organisation gemeint. Im Kern geht es in diesem Cluster um die *Regulierung* des Umgangs mit Daten und die Themen, die in diesem Zusammenhang verstärkt diskutiert werden: Datenschutz, Privatsphäre, Vertraulichkeit, Verantwortlichkeit und Kontrolle. Gesetzliche Regulierungsmaßnahmen für personenbezogene Daten sind in der Datenschutzgrundverordnung (DSGVO) und im Bundesdatenschutzgesetz verankert (Haar, 2017). Aber auch für nicht personenbezogene Daten spielt Regulierung eine Rolle. Denn auch diese Daten können eine hohe Wissensintensität ausweisen, wenn sie beispielsweise Firmen-Know-how oder ein Geschäftsgeheimnis beinhalten, was sie zu sensiblen und schützenswerten Daten macht.

Cluster 7: Cloud Computing

Acht Schlagwörter mit dem Schlüsselwort Cloud Computing bilden das siebte Cluster des Netzwerks. Die *Cloud* ist eine verbreitete *Technologie* für Datenspeicherung und stellt damit ein Instrument der technischen Umsetzung des Datenmanagements dar. Auch für solche technischen Komponenten müssen im Rahmen von Data Governance Regeln und Standards festgelegt werden. Cloud ist ein Beispiel für eine digitale Technologie, die eine disruptive Wirkung auf den Umgang mit Daten hat. Da die technologische Entwicklung schnell voranschreitet und Unternehmen ihre Soft- und Hardware stets an die aktuellen Entwicklungen anpassen müssen, sind auch die Anforderungen an die Data Governance kontinuierlich anzupassen. Es reicht demzufolge nicht, den Rahmen für Data Governance einmalig festzulegen. Vielmehr muss dieser gründlich geplant und nach der Implementierung

laufend evaluiert und optimiert werden. Vor allem die Sicherheitsaspekte spielen bei der Cloud-Data-Governance eine entscheidende Rolle und stellen häufig ein Hindernis für die Nutzung Cloud-basierter Lösungen in Unternehmen dar (Chihande/Van der Poll, 2017).

Cluster 8 und 9: Healthcare und Clinical Informatics

Die beiden kleinsten Cluster vernetzen je sechs Schlagwörter um die Schlüsselwörter Healthcare und Clinical Informatics. Somit handelt es sich bei den Schlagwörtern in diesen Clustern um Themen aus der *Gesundheitswirtschaft*. Abgesehen von dieser Branche hat keine andere eine so hohe Anzahl an wissenschaftlichen Publikationen, dass daraus ein separates Cluster entsteht. Der besonders hohe Stellenwert von Data Governance für die Gesundheitswirtschaft erklärt sich gemäß der Studie von Tallon et al. (2013) daraus, dass das Datenvolumen in dieser Branche stärker ansteigt als in den meisten anderen, woraus sich das Bedürfnis nach einer ausgearbeiteten Data Governance ableiten lässt. Darüber hinaus sind die durchschnittlichen Kosten eines Datenverlusts in der Gesundheitsbranche erheblich höher als in anderen Branchen (Prauß et al., 2019), was den hohen Bedarf an Datensicherheit begründet. Gesundheitsdaten sind sehr sensibel und vertraulich. Daher spielen Datenschutz, Datensicherheit und die Verantwortung für diese Daten eine ganz besondere Rolle. Vor diesem Hintergrund ist die Implementierung und Umsetzung einer effektiven Data Governance essenziell für die Gewährleistung von Integrität und Vertrauenswürdigkeit von Unternehmen aus der Gesundheitswirtschaft (Tse et al., 2018).

Forschungsausblick

Insgesamt lässt sich festhalten, dass Data Governance im wissenschaftlichen Diskurs noch ein Nischendasein führt. Obwohl die Anzahl der Veröffentlichungen zu diesem Thema in den letzten Jahren kontinuierlich gestiegen ist, ist sie in Relation zu den verwandten Themen gering. Dabei stellt Data Governance einen wichtigen Teil von Data Economy dar: Ohne eine funktionierende Data Governance ist die effiziente Datenbewirtschaftung, vor allem in komplexen Datenökosystemen, nicht möglich. Vor diesem Hintergrund ist davon auszugehen, dass das Thema in der nahen Zukunft mehr in den Fokus der Wissenschaft rückt.

Aktuell liegen die Schwerpunkte in der Forschung und der damit einhergehenden Literatur auf der Bedeutung von Data Governance für digitale Technologien im Allgemeinen. Darüber hinaus werden einige Aspekte der Data Governance detailliert beleuchtet, wie die Bewirtschaftung von Daten als Wirtschaftsgüter, Sicherung der Datenqualität, Datenschutz, Datensicherheit und Compliance. Die technische Umsetzung von Data Governance scheint in der Literatur dagegen eine nachrangige Bedeutung zu haben. Viel Aufmerksamkeit gilt dem Umgang mit Daten und besonders der Festlegung von Zuständigkeiten für Daten. Auch die gesetzliche Regulierung des Umgangs, vor allem mit sensiblen personenbezogenen Daten, bildet einen der Schwerpunkte. Ein signifikanter Anteil der Literatur beschäftigt sich mit der Organisation von großen Mengen an Daten und der Notwendigkeit einer Data Governance, um diese Mengen an Daten zu strukturieren.

Eine Forschungslücke stellt die mangelnde Abgrenzung von Data Governance und Data Management dar. Aus der synonymen Verwendung beider Begriffe, wie sie in der Literatur häufig anzutreffen ist, resultieren Irritationen und mögliches Fehlverhalten in Unternehmen in der Praxis. Aus diesem Grund sollten zukünftige Publikationen die Beziehung zwischen Data Governance und Data Management näher beleuchten und die beiden Begriffe inhaltlich besser voneinander trennen.

Die Gesundheitsbranche gilt als ein Vorreiter im Bereich Data Governance. Dies hängt mit den Eigenschaften der Daten zusammen: Patientendaten sind äußerst sensibel und schützenswert. Daher müssen für den Umgang mit solchen Daten klare Regeln und Richtlinien definiert werden. Damit wird auch das Vertrauen der Patienten gestärkt. Das Problem der sensiblen Daten besteht allerdings nicht ausschließlich im Gesundheitswesen. Auch in anderen Branchen, vor allem in Unternehmen, die über eine große Menge an personenbezogenen Kundendaten verfügen, muss der Umgang mit Daten geregelt werden. Ohne eindeutig definierte Zuständigkeiten und Regeln droht die Gefahr, dass wertvolle Daten in die falschen Hände geraten oder versehentlich gelöscht werden. Vor diesem Hintergrund müssen Unternehmen branchenübergreifend auf die Bedeutung von Data Governance aufmerksam gemacht werden.

Es stellt sich auch die Frage, ob die Themen, die in der wissenschaftlichen Literatur diskutiert werden, in der Praxis relevant sind. Wenn es sich um digitale Technolo-

gien handelt, entwickeln sich die praktische Umsetzung und die Literatur häufig parallel zueinander, anstatt miteinander verbunden zu werden und dadurch Synergieeffekte zu erzielen. Der Grund hierfür besteht darin, dass die digitale Welt sehr dynamisch ist. Digitale Technologien können in der Praxis, zum Beispiel im Rahmen von Use Cases, schnell getestet und gegebenenfalls optimiert und verworfen werden. Hier kommen wissenschaftliche Analysen häufig nicht hinterher. Vor allem die Grundlagenforschung benötigt meistens viel Zeit und Ressourcen. Damit Data Governance ihre Black Box verlässt, müssen Forscher gezielter auf den Bedarf in der Praxis eingehen. Eine bessere Koordination und Kooperation zwischen der Praxis und der Forschung können helfen, damit Unternehmen richtige Impulse und Handlungsempfehlungen erhalten, die ihnen den Umgang mit dem Thema Data Governance erleichtern.

Literatur

Ackoff, Russell Lincoln, 1999, *Ackoff's best. His classic writings on management*, New York

Al-Ruithe, Majid / Benkhelifa, Elhadj / Hameed, Khawar, 2018, A systematic literature review of data governance and cloud data governance, in: *Personal and Ubiquitous Computing*, 1. Jg., Nr. 1, S. 1–21

Azkan, Can / Spiekermann, Markus / Goecke, Henry, 2019, Uncovering research streams in Data Economy using text mining algorithms, XXX ISIPIM Innovation Conference, Celebrating Innovation - 500 Years Since Da Vinci, 16.–19.6.2019, Florenz

Blackburn, L. A. / Brussel, J. / Fors, A., 1931, *Controlling Factors in Machine-Tool Obsolescence*, Warrendale

Blondel, Vincent D. / Guillaume, Jean-Loup / Lambiotte, Renaud / Lefebvre, Etienne, 2008, Fast unfolding of communities in large networks, in: *Journal of Statistical Mechanics: Theory and Experiment*, Nr. 10, S. 1–12

Brüning, Alrik / Gluchowski, Peter / Kaiser, Andre, 2017, *Data Governance – Einordnung, Konzepte und aktuelle Herausforderungen*, Chemnitz Economic Papers, Nr. 15, Chemnitz

BVDW – Bundesverband Digitale Wirtschaft, 2018, *Data Economy – Datenwertschöpfung und Qualität von Daten*, Berlin

Carretero, Ana G. / Caballero, Ismael / Piattini, Mario, 2016, MAMD: Towards a Data Improvement Model Based on ISO 8000-6X and ISO/IEC 33000, in: Clarke, Paul / O'Connor, Rory / Rout, Terry / Dorling, Alec (Hrsg.), *Communications in Computer and Information Science*, Bd. 609, Software process improvement and capability determination, S. 241–253

Chihande, Michael / Van der Poll, John, 2017, Post cloud computing implementation benefits and challenges realised for a South African technology company, in: 2017 Conference on Information Communication Technology and Society (ICTAS), 8.–10.3.2017, Umhlanga, Durban / Piscataway, N.J., S. 1–6

DEMAND, 2019, *Data Economy. Status quo der deutschen Wirtschaft & Handlungsfelder in der Data Economy*, <https://www.demand-projekt.de/> [4.6.2019]

Elsevier, 2019, Scopus, <https://www.elsevier.com/solutions/scopus> [8.2.2019]

Haar, Tobias, 2017, Frisch gestärkt. Neues Bundesdatenschutzgesetz ergänzt EU-Grundverordnung, in: *Magazin für professionelle Informationstechnik*, 8. Jg., S. 90–93

Heumann, Stefan / Jentsch, Nicola, 2019, *Wettbewerb um Daten. Über Datenpools zu Innovationen*, Berlin

iRights.Lab, 2018b, *Data-Governance-Report. mFUND-Begleitforschung vom Think Tank iRights.Lab*, Data Governance Report, Nr. 2, Berlin

ISO / IEC, 2008, *ISO/IEC 38500: corporate governance of information technology*. ISO/IEC 38500, <https://www.iso.org/standards/51639.html> [4.9.2019]

Khan, Mudassir, 2018, Big Data Analytics Evaluation, in: *International Journal of Engineering Research in Computer Science and Engineering*, 5. Jg., Nr. 2, S. 148–152

Khatri, Vijay / Brown, Carol V., 2010, Designing data governance, in: Communications of the ACM, 53. Jg., Nr. 1, S. 148

Kitchenham, Barbara / Charters, Stuart, 2007, Guidelines for performing Systematic Literature Reviews in Software Engineering, EBSE Technical Report, Nr. 2007-01, Keele

Krotova, Alevtina / Eppelsheimer, Jan, 2019, Was bedeutet Data Governance? Eine Clusteranalyse, DEMAND White Paper, <https://www.demand-projekt.de/> [16.9.2019]

Otto, Boris, 2011, Data Governance, in: Wirtschaftsinformatik, 53. Jg., Nr. 4, S. 235–238

Prauß, Marius / Janson, Matthias / Rosenberger, Juraj, 2019, Grenzen des Speichers, <https://app.handelsblatt.com/downloads/24425600/6/grafiken-des-monats-mai-2019.pdf?ticket=ST-7349244-xuH4B-hIAWJSAaXvlgNg-ap4> [17.5.2019]

redgate, 2018, Data Governance Implementation Survey 2018, Report, Cambridge

Staab, Steffen, 2002, Wissensmanagement mit Ontologien und Metadaten, in: Informatik-Spektrum, 25. Jg., Nr. 3, S. 194–209

Tallon, Paul P. / Ramirez, Ronald V. / Short, James E., 2013, The Information Artifact in IT Governance: Toward a Theory of Information Governance., in: Journal of Management Information Systems, 30. Jg., Nr. 3, S. 141–178

Tse, Daniel et al., 2018, The Challenges of Big Data Governance in Healthcare, in: The 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom 2018), S. 1632–1636

Data Governance – A Definition of Terms Using a Text-mining-based Literature Review

Data governance forms the basis of efficient data management within companies and across company boundaries. In practice, however, there is no clear and uniform definition of data governance and its components. In order to provide a comprehensive overview of data governance, to reveal previous research trends and to uncover potential research needs, a text-mining-based systematic literature review was conducted. With this method the authors established that the place occupied by data governance in scientific debate and literature has been constantly growing, especially since 2012. They identified a total of nine clusters, each representing an existing research priority. The analysis revealed that the main topic of data governance research is its significance for digital technologies. Further aspects of data governance addressed in the literature include use of data as an economic asset, data management, data quality, data protection, compliance and organisation of data in networks. Data governance already plays a more important role in the health sector than in other industries. By effectively coordinating the requirements of business practice and scientific problem analysis, data governance can be better investigated, developed and implemented.