

Data Governance

Der Leitfaden für die Praxis

» Hier geht's
direkt
zum Buch

DIE LESEPROBE

1

Einleitung

Daten standen noch nie so hoch im Kurs wie heute. Noch nie war der Bedarf an Praxiswissen im Umgang mit Daten gefragter. Neue Trends und Technologien für Daten entwickeln sich immer schneller. Und es steigt der Bedarf, alle Möglichkeiten, die es heute bereits gibt und morgen geben wird, auszuschöpfen. Data Governance schafft in Organisationen die Voraussetzungen, um Daten wertschöpfend einzusetzen.

1.1 Daten als Schlüsselressource

In fast allen Bereichen der Wirtschaft und des privaten Lebens spielen Daten eine immer größere Rolle. Es gibt keine Organisation, welche nicht in irgendeiner Weise Daten verarbeitet. Private Nutzende teilen ihre Daten in sozialen Netzwerken, bloggen, shoppen online, bezahlen mit dem Smartphone oder investieren in digitales Geld.

Die Trends der Datenverarbeitung in der Zukunft deuten sich schon an: Diskussionen über die Abschaffung von Bargeld laufen bereits (Deutsche Bundesbank, 2018), ebenso werden die Möglichkeiten von selbstsouveränen Identitäten (SSI), also der Verwaltung der eigenen Identität im Netz, immer umfangreicher (z. B. EMIL Consortium, 2020). Die EU hat den Data Governance Act beschlossen, um die sichere, transparente und vertrauenswürdige gemeinsame Nutzung von Daten innerhalb der EU zu fördern, um Innovation und wirtschaftliche Entwicklung zu unterstützen (EU Data Governance Act, 2022). Die Verarbeitung von Daten mit Unterstützung von KI-Applikationen ist inzwischen weit verbreitet und der EU AI Act gibt die rechtlichen Vorgaben dafür (EU AI-Act, 2024). Beides, die gemeinsame Nutzung von Daten sowie die Verarbeitung von Daten durch Künstliche-Intelligenz-Lösungen sind Treiber für viele neuartige Trends und Geschäftsideen. Zum Beispiel wird mit Open-Science-Initiativen der Zugang zu Forschungsdaten aus verschiedenen europäischen Institutionen

ermöglicht und beschleunigt, was den kollaborativen Forschungsansatz fördert. Und mit Open-Data-Plattformen für KI-Trainings stehen Datenbestände zur Verfügung, die idealerweise die Entwicklung verbesserter KI-Modelle und deren algorithmische Genauigkeit durch größere und diversifizierte Datenpools ermöglichen. Mit den wachsenden Möglichkeiten, Daten zu erfassen, zu halten und zu verarbeiten, steigt unweigerlich der Bedarf nach Richtlinien, Hinweisen, Transparenz, Schutzmaßnahmen, Verantwortlichkeiten und Systemen, die den Umgang mit diesen Daten regeln.

Damit wird deutlich, dass es diese innovativen Ansätze nicht umsonst gibt. Fällt der Begriff Data Governance als Zusammenfassung all der Aktivitäten, die notwendig sind, um den vollen Wert der Daten auszuschöpfen, sind folgende Aussagen zu hören:

- Data Governance ist ein Kostentreiber,
- sie erfordert ein hohes Maß an zusätzlicher Bürokratie,
- sie benötigt neue Stellen und mehr Mitarbeitende,
- sie bremst Datenmanagement-Projekte aus,
- sie ist ohne zusätzliche Software nicht umsetzbar,
- und die Ergebnisse sind erst nach langer Zeit sichtbar.

Diese Argumente sind nur Beispiele für die skeptische Grundhaltung gegenüber Data Governance. Die Kunst ist es, einen Anreiz zu schaffen und gute Gründe zu nennen, warum es sich dennoch lohnt, in Data Governance zu investieren.

1.2 Herausforderungen und Nutzen von Data Governance

In aller Regel befinden sich die Organisationen in Bezug auf Data Governance nicht auf der grünen Wiese. Bei genauer Betrachtung gibt es in verschiedenen Bereichen Aktivitäten, die sich Data Governance durchaus zuordnen lassen. Das kann eine bereits implementierte Softwarelösung für bestimmte Datenqualitäts-Prüfungen sein. Oder es gibt bereits Kolleginnen und Kollegen, die bei Fragen zu bestimmten Daten einen guten Rat geben können. Nicht selten gibt es Customer-Relationship-Management- oder Enterprise-Ressource-Planning-Systeme, für die Datenpflege-Prozesse bereits beschrieben, geregelt und umgesetzt wurden. Je mehr man sucht und sich durchfragt, desto mehr Hinweise auf Data Governance wird man in Organisationen finden. Und die verschiedenen Aktivitäten sind oft seit längerer Zeit etabliert und auf das Management der Daten abgestimmt. Das heißt, in vielen Bereichen läuft es bereits und das Rad muss nicht neu erfunden werden. Das ist die gute Nachricht.

Die Herausforderung ist allerdings, dass bereits vorhandene Aktivitäten selten koordiniert und bereichsübergreifend stattfinden. Mitarbeitende, die sich bereits den Ruf einer „Datenexpertin“ oder eines „Datenexperten“ erarbeitet haben, finden diese

Rolle kaum in ihrer Stellenbeschreibung. Neue Anforderungen an die Daten, z. B. in Form von neuen Reports, münden in teilweise zeitaufwendiger und fehleranfälliger manueller Datenaufbereitung, die im schlimmsten Fall weder nachvollziehbar noch reproduzierbar ist. Es ist nicht bekannt, wer was in welchem Bereich für die Datenoptimierung bereits macht oder wo welche Datenpflege-Prozesse bereits implementiert sind und gut laufen. Es kommt zu Missverständnissen und doppelten bzw. mehrfachen Arbeiten, wenn die gleichen Aktivitäten an verschiedenen Stellen durchgeführt werden.

Anforderungen an die Daten sind nur aus dem eigenen Bereich bekannt. Die Anforderungen anderer Bereiche sind ungeklärt und werden somit bei der Datenerfassung und Datenverarbeitung (unwissentlich) ignoriert. Im schlimmsten Fall werden die Datenoptimierungen von Mitarbeitenden des einen Bereichs durch Mitarbeitende des anderen Bereichs überschrieben und somit zunichte gemacht. Die Koordination fehlt, und damit die Transparenz und das konsolidierte Wissen über die Möglichkeiten, was mit den Organisationsdaten heute schon gemacht wird und morgen noch getan werden soll. Die Herausforderung und Aufgabe ist also, das vermeintliche Chaos übergreifend zu strukturieren und so durch Data Governance die Effizienz des Datenmanagements zu steigern.

Tabelle 1.1 stellt die Auswirkungen von Datenmanagement ohne und mit Data Governance gegenüber.

Tabelle 1.1 Ziele von Data Governance

Ohne Data Governance	Mit Data Governance
Fehlende Finanzierung von datengetriebenen Projekten	Unterstützung und Verantwortung durch das Top-Management
Ungeklärte Verantwortlichkeiten für Daten	Data Stewardship als zentrales Organisationskonzept des Datenmanagements
Datenmanagement mit geringer Priorität	Organisationsweit abgestimmtes Management des „Vermögenswertes“ Daten
Verzögerungen bei der innovativen Nutzung von Datenprodukten	Zügige Entscheidungsfindung bei Daten-Fragen und deren Umsetzung
Ziele des Datenmanagements haben geringe Priorität in IT-Projekten	IT-Projekte mit Auswirkung auf kritische Daten finden unter Beteiligung des Data Office statt
Geschäfts- und Fachbereiche ignorieren übergreifende Auswirkungen der Datenpflege und Datennutzung	Data-Governance-Gremien stellen übergreifende Abstimmung aller Datenmanagement-Initiativen sicher
Inkonsistente Geschäftsprozesse, Erfassungsrichtlinien und Datenmodelle	Einführung und Durchsetzung von Best Practices, inklusive standardisierter Datenmodelle, Definitionen, Regeln und Geschäftsprozessen

Die Einführung von Data Governance ist nicht zu unterschätzen. Sehr schnell sind wieder die oben genannten Statements zu hören und das macht es nicht einfacher, Verbündete in der eigenen Organisation zu finden. Denn Data Governance bedeutet auch, sich mit neuen Möglichkeiten des Datenmanagements auseinanderzusetzen. Aktivitäten, die sich heute mit den Daten nicht durchführen lassen, sollen morgen möglich sein. Das bedeutet Veränderung. Veränderung braucht ein starkes Netzwerk, gute Argumente und einen langen Atem. Im Fall von Data Governance bedeutet Veränderung, dass Organisationsdaten in den Fokus rücken und deren Nutzung dem wirtschaftlichen Erfolg der Organisation unterstellt ist. Und dieser Nutzungszweck wird konsequent auf alle Bereiche des Datenmanagements über die gesamte Organisation hinweg angewendet und in vielen kleinen Maßnahmen umgesetzt. Anders gesagt, eine Datenstrategie und Maßnahmen zur Umsetzung werden definiert.

Ein einheitliches Verständnis der Strategemaßnahmen und deren Einordnung in ein Data-Governance-Rahmenwerk helfen bei der Bewältigung der genannten Herausforderungen. Voraussetzung ist, eine gemeinsame Sprache zu sprechen und das gleiche Grundverständnis von Data Governance mit seinen unterschiedlichen Handlungsfeldern zu haben. Die gemeinsame Sprache im Kontext von Data Governance bezieht sich auf Komponenten eines Data Governance Frameworks, auf die Aufgaben und Verantwortlichkeiten der definierten Rollen sowie auf die Bedeutung und den Kontext der Daten an sich, die z. B. in Business Data Dictionaries abgebildet werden. Fehlen diese gemeinsame Sprache und damit das gemeinsame Verständnis, kommt es zu Missverständnissen und Reibungsverlusten und man arbeitet den skeptischen Personen zu, die sich zu den oben genannten Aussagen verleiten lassen.

Das vorliegende Buch unterstützt dabei, bereits laufende Data-Governance-Aktivitäten in ein Framework einzuordnen. So können diese Aktivitäten besser in einen gemeinsamen Kontext gestellt und auf ein gemeinsames Ziel hin koordiniert werden. Ist diese Einordnung geschehen, kann viel einfacher identifiziert werden, an welchen Stellen von Data Governance noch Unterstützung benötigt wird. Anforderungen an interne Projekte oder externe Beratungen können präzise formuliert werden. Eine Unterhaltung auf Augenhöhe wird möglich.

Mit der Einführung von Data Governance kann die Voraussetzung für kommende Herausforderungen im Datenmanagement sichergestellt werden. Dies betrifft Aufgaben im Umfeld von Analytics und künstlicher Intelligenz. Ebenso ist absehbar, dass die Europäische Union eine Reihe von Verordnungen erlässt (z. B. European Commission, 2024), deren Umsetzung ein strukturiertes und transparentes Datenmanagement und somit auch Data Governance voraussetzt.

Die hohe Relevanz von Datenmanagement in Kombination von Data Governance wird in einer aktuellen Studie (Burbank & Knight, 2023) zum wiederholten Mal bestätigt. Demnach ist Data Governance das Topthema, welches in den Jahren 2023 bis 2025 implementiert werden wird. Weitere Topthemen sind Datenstrategie, Datenqualität, Stammdaten-Management sowie Datenarchitektur. Die gleiche Studie zeigt, dass

sich Data Governance in den meisten Organisationen nach wie vor in einem sehr frühen Implementierungsstadium befindet. Einige Organisationen planen zumindest, Data Governance einzuführen.

Damit ist jetzt der richtige Zeitpunkt, das vorliegende Buch zu lesen und mit einfachen und pragmatischen Aktivitäten Data Governance umzusetzen und das Datenmanagement zum Erfolg zu führen.

1.3 Zielgruppen und Aufbau des Buches

Das vorliegende Buch kann auf unterschiedliche Weise zum Einsatz kommen. Wer an der theoretischen Einführung zu Data Governance interessiert ist und sich aus den vorgestellten und selbst entwickelten Frameworks selbst ein Bild machen will, der kann das Buch von vorne bis hinten durchlesen.

Das Buch kann auch abschnittsweise gelesen werden. Die Themen in den Kapiteln sind so aufbereitet und beschrieben, dass sie (fast) unabhängig voneinander gelesen werden können. An manchen Stellen wird auf andere Abschnitte verwiesen und bei Bedarf kann der Wissensdurst dann an den entsprechenden Stellen gestillt werden.

Für die Praktikerinnen und Praktiker, die nach einer schnellen Möglichkeit für die Umsetzung der einzelnen Handlungsfelder suchen, ist Kapitel 7 gedacht. Hier werden Methoden, Konzepte und Tools vorgestellt, mit denen sich Aspekte von Data Governance einfach verwirklichen lassen. Und die allermeisten Methoden lassen sich mit den üblichen Office-Programmen umsetzen.

Es gibt dazu noch verschiedene Informationen, die in Kästen stehen. Hier wird zwischen Praxistipps, Hinweisen und Fallbeispielen unterschieden. Besonders interessant sind die Fallbeispiele, denn sie zeigen, wie in Projekten konkrete Data-Governance-Themen adressiert und umgesetzt wurden. Bei den Hinweisen handelt es sich um ergänzende Hinweise zu den Themen, die in Abhängigkeit des Kontexts und der Umsetzung zum Tragen kommen können oder auch nicht.

Dieses Buch enthält Hinweise für die Umsetzung von Data Governance in der Praxis für alle, die ...

- ... bei ihren täglichen Aufgaben über Ungereimtheiten bei Daten stolpern und überlegen, was die Ursache ist und wie diese behoben werden kann. Oft haben diese Personen die Rolle „Data Steward“ inne und sind (implizit) für die Qualität von Daten verantwortlich.
- ... in der Organisation bekannte Herausforderungen bei Datenmanagement-Themen fachlich einordnen möchten, um die richtigen Maßnahmen zur Verbesserung der Situation einzuleiten.

- ... oft mit großen (Stamm-)Datenbeständen aus unterschiedlichen Quellen arbeiten und diese zur Beantwortung bestimmter Fragestellungen untersuchen. Bei der Zusammenführung der Daten zur Auswertung kommt es immer wieder zu Herausforderungen, deren Auflösung sehr viel Zeit in Anspruch nimmt und manuelle Anpassungen an den Daten erfordert.
- ... in verantwortlicher Position den effizienten Einsatz von Organisationsdaten sicherstellen müssen, z. B. im Marketing oder als E-Commerce-Verantwortliche.
- ... ein neues Stammdatenmanagement (MDM)-System einführen wollen oder eingeführt haben, und nun sicherstellen müssen, dass die Stammdatenpflege nach einheitlichen Prozessen durchgeführt wird.
- ... die Rolle „Chief Data Officer“ innehaben und für das Management der Organisationsdaten verantwortlich sind. Sie geben dem Kollegium Orientierung bei der Umsetzung von Data Governance und stellen sicher, dass die Daten ein echter Produktionsfaktor sind.
- ... neu in einer verantwortlichen Rolle für Daten sind und eine Orientierungshilfe und Best-Practice-Angebote suchen.
- ... beratend tätig sind und immer wieder mit Data-Governance-Herausforderungen konfrontiert werden und praktische Hinweise für einen möglichen Lösungsweg suchen.

Aber auch alle anderen Personen, die sich nicht in den oben genannten Situationen wiederfinden, sind eingeladen, dieses Buch zu lesen. Zu Bedenken ist bei der Lektüre, dass es sich um ein Handbuch für die Praxis handelt und der Schwerpunkt auf der praktischen Anwendbarkeit und Umsetzung liegt. Wo immer es möglich ist, sind Fallbeispiele angeführt.

Das Buch beschreibt in den folgenden Kapiteln die verschiedenen Aspekte von Data Governance. Die Struktur ergibt sich wie folgt.

Kapitel 2, „Data, Governance & Co.“, erläutert die wesentlichen Grundlagen zu Daten und Informationen. Es werden die verschiedenen Datenarten beschrieben. Zudem werden die Themen Open Data und externe Daten adressiert – beides Themen, die heute bereits und in der Zukunft noch viel mehr an Relevanz gewinnen werden. Spricht man über Daten, müssen Metadaten unweigerlich genannt werden. Mit Metadaten legt man in der Regel die Grundlage für Datenqualität, da sie beschreiben, wie Daten beschaffen sein müssen. Wer Datenmanagement betreibt, kommt um Fachdatenmodellierung nicht herum. Damit ist nicht nur die Strukturierung von Attributen in Entitäten und Objekten gemeint, sondern auch die Beschreibung und Dokumentation des Datenmodells in Form von Glossaren. Im Abschnitt zur Organisationsgestaltung geht es um Fragen zu Organisation, Kompetenzen, Verantwortung und Kongruenzprinzip. All das sind Punkte, die für die Definition einer funktionierenden Data-Governance-Organisation bekannt sein sollten. Das Kapitel beschreibt zudem verschiedene Grundformen von Governance.

In **Kapitel 3, „Data Governance“**, werden verschiedene, in der Wissenschaft und Praxis bekannte Definitionen von Data Governance vorgestellt. Zusätzlich werden vier Data Governance Frameworks vorgestellt. Jedes dieser Frameworks hat einen individuellen Schwerpunkt. Damit wird deutlich, dass die Entwicklung und der Einsatz eines Data Governance Frameworks abhängig vom jeweiligen Organisations- bzw. Datenmanagement-Kontext ist. Es gibt kein Richtig und kein Falsch. Zudem wird hier die Abgrenzung von Data Governance, Datenmanagement und Data Excellence beschrieben.

Kapitel 4, „Das qualitätsorientierte Data Governance Framework“, stellt das gleichnamige Framework vor. Dieses Framework beschreibt Struktur und Rahmen für die praxisorientierte Umsetzung in den folgenden Abschnitten. Grundsätzliche Zielsetzung des Frameworks ist es, Daten in hoher Qualität für alle Nutzenden bereitzustellen. Es werden sechs Handlungsfelder auf den Ebenen Strategie, Organisation/ Prozesse und Informationssysteme beschrieben: Datenstrategie, Controlling, Organisation, Datenprozesse, Datenarchitektur und Systemarchitektur. Datenqualität ist als Querschnittsaufgabe definiert. Die Handlungsfelder oder Gestaltungsbereiche zeigen als eine Art Best Practice, woran Organisationen denken sollten, wenn sie Data Governance definieren und umsetzen wollen.

Kapitel 5, „Rollen & Gremien für Data Governance“, beschreibt eines der Kernelemente des qualitätsorientierten Data Governance Frameworks: die Struktur der Datenmanagement-Organisation und die Ausgestaltung der Rollen. Letztendlich übernehmen Inhabende von Rollen konkrete Aufgaben und Verantwortlichkeiten des Datenmanagements. Umso wichtiger ist eine gute Basis, wie die Rollenorganisation aussehen könnte. Die verschiedenen hierarchischen Ebenen werden adressiert sowie Möglichkeiten der Kommunikation zwischen diesen Ebenen. Das Rollenmodell ist flexibel gestaltet, sodass es auf Organisationen jeder Größe anpassbar ist. Wer in der eigenen Organisation nach Möglichkeiten der Definition und Besetzung von Datenmanagement-Rollen sucht, wird hier hilfreiche Informationen finden.

Im Datenmanagement nimmt die Qualität der Daten eine herausragende Rolle ein. Der Name des eingeführten qualitätsorientierten Data Governance Frameworks gibt bereits einen Hinweis darauf. **Kapitel 6, „Datenqualität“**, gibt eine ausführliche Beschreibung von Datenqualität und stellt zwei Konzepte dazu vor. Ebenso wird auf die Bedeutung von Datenqualität in der Praxis eingegangen mit Hinweisen, wie die Datenqualität zum einen gemessen und zum anderen bewertet werden kann. Auch wird die Frage adressiert, welche Kosten schlechte Datenqualität mit sich bringt. Hinweise auf die Qualität von Metadaten werden gegeben.

Das größte Kapitel des Buches, **Kapitel 7**, beschreibt diverse **„Instrumente, Techniken und Tools“** für die praktische Umsetzung von Data Governance. Für jedes der in Kapitel 4 vorgestellten Handlungsfelder werden verschiedene Möglichkeiten gezeigt, diese ganz oder teilweise in der Praxis umzusetzen. Auf die Empfehlung von zusätzlichen Softwarelösungen wird weitestgehend verzichtet. Vielmehr lassen sich die

meisten Methoden mit den üblichen Office-Programmen umsetzen. Letztendlich spielen die Idee und die Methode der Umsetzung eine größere Rolle als die Anwendung eines spezifischen Toolsets. Das bedeutet, dass dieses Kapitel für die Anwendung in der Praxis geschrieben wurde. Ideen für mögliche Templates und Fallbeispiele ergänzen die unterschiedlichen Methoden. Die vorgestellten Ansätze können sehr leicht auf die individuellen Anforderungen angepasst werden.

In **Kapitel 8, „Anwendungsszenarien von Data Governance“**, werden fünf Anwendungsszenarien aus der Praxis vorgestellt, bei denen verschiedene Methoden und Konzepte aus Kapitel 7 zum Einsatz kommen. Es geht um die Etablierung von Qualitätstoren in der Stammdatenproduktion, die DSGVO als Treiber für ein Rollen- und Berechtigungskonzept, die Einführung einer Data Quality Scorecard im Marketing, die Entwicklung einer Daten-Exzellenz-Strategie für die Stadt Freiburg i. Br. sowie die Herausforderungen beim Umgang mit KI-Lösungen und was das für Data Governance bedeutet.

Kapitel 9, „Zusammenfassung und Ausblick“, fasst das Buch kurz zusammen. Es beschreibt Themen, die bei der Weiterentwicklung des qualitätsorientierten Data Governance Frameworks berücksichtigt werden können.

2

Data, Governance und Co.

Dieses Kapitel legt die Grundlagen für die weiteren Ausführungen zu Data Governance. Es definiert den Begriff Daten und beschreibt verschiedene Arten von Daten, insbesondere Stammdaten, Metadaten und Open Data. Ein kurzer Abriss über Datenmodellierung zeigt Begriffe, die häufig im Zusammenhang mit Data Governance gebraucht werden. Da Governance auch viele Gemeinsamkeiten mit Organisationsgestaltung hat, werden zur Organisationsgestaltung ebenfalls grundlegende Konzepte erläutert. Die Einführung in die drei Grundformen von Governance – zentral, föderativ und dezentral – rundet das Kapitel ab.

2.1 Von Zeichen zu Wissen

Das der Semiotik entlehnte Ebenenmodell (siehe Bild 2.1) erklärt den Zusammenhang zwischen Zeichen, Daten, Informationen und Wissen. **Zeichen** bilden die unterste Ebene in diesem Modell, z. B. die Zeichen „.“, „0“, „2“, „8“, „M“, „r“, „z“ und „ä“. Zeichen stehen in keinem besonderen Zusammenhang. Auf der syntaktischen Ebene werden die verschiedenen Zeichen miteinander in Verbindung gebracht. Sie werden nach bestimmten Regeln in eine formale Struktur (Syntax) überführt. Durch diesen Schritt entstehen **Daten**, zum Beispiel „8. März 2020“. Entsprechend ihrer Syntax – hier TT. MMM YYYY – können Daten mit technischen Mitteln verarbeitet, also z. B. generiert, bearbeitet, weitergeleitet und gespeichert werden.

Auf der semantischen Ebene erhalten die Daten zusätzlich eine inhaltliche Bedeutung für die Empfangenden der Daten. Sie werden in einen Kontext gesetzt. Damit werden die Daten zur **Information**. Im o. g. Beispiel handelt es sich um das Datum, an dem dieser Text geschrieben wurde. Daten sind also zunächst zweckneutrale Fakten. Sie werden zur Information, wenn ihnen eine Bedeutung zugewiesen wird. Information

ist subjektiv. Was für die eine Person eine wichtige Information darstellt, schätzt eine andere Person als nebensächlich oder gar gänzlich unbedeutend ein.

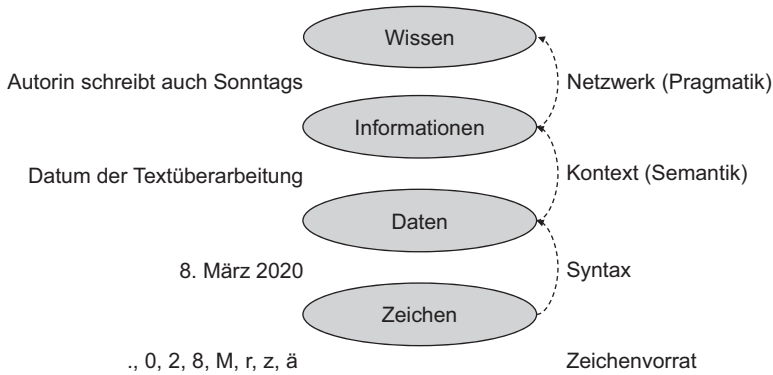


Bild 2.1 Ebenenmodell der Semiotik (in Anlehnung an Krcmar, 2015, S. 12)

Auf der obersten Ebene des Modells steht das **Wissen**. Informationen werden zu Wissen, wenn sie miteinander vernetzt bzw. interpretiert werden. Um auf das vorhergehende Beispiel zurückzukommen, könnte die Information des Datums der Überarbeitung dieses Textes mit einer weiteren verknüpft werden, nämlich, dass der 8. März 2020 ein Sonntag war. Somit könnte man daraus das Wissen ableiten, dass die Autorin auch am Sonntag schreibt – also vermutlich außerhalb ihrer regulären Arbeitszeit.

Wissen ermöglicht es Personen, Aufgaben in größere Sinnzusammenhänge, Entwicklungen, Arbeitsprozesse etc. einzuordnen (Klotz, 2011). Durch Wissen können Personen, die ihnen zur Verfügung stehenden Daten sichten, auswählen und als Informationen für ihre Aufgabenbearbeitung nutzen.

2.2 Arten von Daten

Aufgrund der prominenten Stellung des Begriffes Data (also Daten) in Data Governance lohnt sich ein näherer Blick auf die Daten. Grundsätzlich wird zwischen **strukturierten** und **unstrukturierten Daten** unterschieden. Bei ca. 80 % aller gespeicherten Daten handelt es sich um unstrukturierte Daten (Henderson et al., 2017). Beispiele für unstrukturierte Daten sind Bilder, Videos, Texte oder andere Daten, die sich in ihrer Struktur keinem Aufbau in Spalten und Zeilen wie in einer Tabelle oder Datenbank zuordnen lassen. Dennoch haben viele dieser vermeintlich unstrukturierten Daten eine gewisse Struktur. Textdokumente können eine Überschrift über dem eigentlichen Textteil haben und können in Unterkapitel eingeteilt sein. In einem

solchen Fall spricht man auch von semi-strukturierten Daten (Navarro Bullock, 2023).

Strukturierte Daten hingegen sind in Spalten und Zeilen hinterlegt, die wiederum in Datenbanken hinterlegt sind und für die Datenmodelle entwickelt wurden (vgl. Abschnitt 2.6). Vereinfacht gesagt, handelt es sich bei Daten, die in tabellarischer Form hinterlegt werden, um strukturierte Daten. In diesem Buch ist der Fokus auf strukturierte Daten gelegt.

Im unternehmerischen Kontext können vier Arten von Daten unterschieden werden (Legner & Otto, 2007; Schemm, 2009, S. 19, Bild 2.2): Stammdaten, Bewegungsdaten, Bestandsdaten und Änderungsdaten.

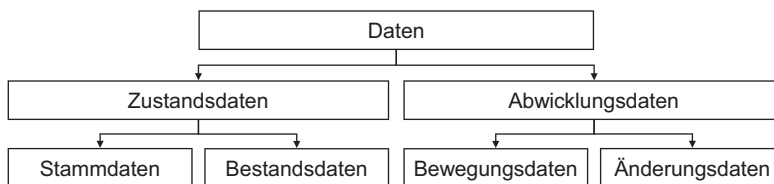


Bild 2.2 Datenarten (vgl. Schemm, 2009, S. 20)

Stammdaten repräsentieren die Kernentitäten bzw. Kernobjekte einer Organisation, z. B. Kundschaft, Materialien, Produkte, Personal und Lieferunternehmen. Stammdaten sind zustandsorientierte Daten, die sich selten ändern. Stammdaten werden von Bewegungsdaten zur Beschreibung von Geschäftsvorfällen referenziert.¹⁾



Beispiel: Verwendung von Stammdaten in Geschäftsvorfällen

Die Bestellung (= Geschäftsvorfall) einer Kundin enthält typischerweise Daten aus dem Stammdatensatz der Kundin, wie z. B. ihre Adresse und die vereinbarten Zahlungsbedingungen. Die Bestellung enthält aber auch Daten aus einem oder mehreren Produktstammdatensätzen, nämlich die Nummern, Bezeichnungen, den Preis und eventuell weitere Merkmale der bestellten Produkte.

Bestandsdaten beschreiben die betriebliche Mengen- und Wertestruktur. Typische Bestandsdaten einer Organisation sind Lagerbestand und Kontostand. Bestandsdaten sind wie Stammdaten zustandsorientiert, sie ändern sich aber häufig durch geschäftsbedingte Zu- und Abgänge.

Bewegungsdaten (auch Transaktionsdaten) bilden betriebswirtschaftliche Vorgänge ab. Typische Beispiele sind Fertigungsaufträge, Lieferscheine, Bestellungen und Rechnungen. Bewegungsdaten sind abwicklungsorientiert und verändern Bestandsdaten durch mengen- oder wertmäßige Zu- und Abgänge. Zum Beispiel erhöht ein Waren-

¹ Aufgrund der großen Bedeutung von Stammdaten für Data Governance widmet sich Abschnitt 2.3 diesem Begriff ausführlich.

eingang den Bestand an Rohstoffen und eine bezahlte Lieferantenrechnung verringert den Kontostand bei der Bank.

Änderungsdaten sind abwicklungsorientierte Daten. Sie lösen Änderungen von Stammdaten aus. Ein Beispiel ist die Hochzeit einer Mitarbeiterin, welche den Familienstand im Personalstammsatz ändert. Die Entwicklung eines besseren Wirkstoffs ändert die Zusammensetzung der Produktionsstückliste eines Düngemittels. Eine Änderung der Rechtsform einer Organisation wirkt sich auf die Bezeichnung der Organisation in dessen Lieferantenstammsatz aus.

Tabelle 2.1 fasst die verschiedenen Datenarten noch einmal zusammen und beschreibt Funktion, Auswirkung, Änderungshäufigkeit und Beispiele. Bei Gesprächen über Daten hilft die eindeutige Benennung der Datenarten, um Missverständnissen vorzubeugen und Diskussionen zielgerichtet zu führen.

Tabelle 2.1 Übersicht über typische Datenarten (in Anlehnung an Hofleithner et al., 2022, S. 29)

Merkmals	Stammdaten	Bestandsdaten	Bewegungs-/ Transaktions- daten	Änderungs- daten
Funktion	Beschreiben Objekte mit ihren Attributen	Zustands- orientiert	Beschreiben Veränderung eines Zustandes	Beschreiben Veränderungen eines Objekts
Auswirkung	Werden von vielen Prozessen referenziert	Geschäftspro- zesse bewirken Änderungen	Bilden den Geschäfts- prozess ab	Betreffen ein Stammdatum
Dynamik	Ändern sich gelegentlich	Ändern sich häufig	Wachsen sehr schnell und blei- ben konstant im Inhalt	Ändern sich selten
Beispiele	Mitarbeitende, Material	Lagerbestand, Kontostand	Buchung, Bestellung	Umzug, Hochzeit

Die Unterscheidung der Datenarten ist nicht immer eindeutig und hängt auch vom Organisationszweck ab. Beispielsweise ist ein Umzug für die meisten Organisationen ein Änderungsdatum. Für ein Postunternehmen hat die Meldung über die erfolgte oder bevorstehende Adressänderung einer Kundin den Charakter eines Bewegungsdatums. Es ist ein typischer Geschäftsvorfall, der vielfältige Aktivitäten auslöst, z. B. die Einrichtung eines Nachsendeauftrags.

Im Zusammenhang mit betrieblichen Daten können auch die Begriffe Geschäftsobjekt und Datenobjekt (bzw. Datenelement) unterschieden werden (vgl. Österle et al., 2011). **Geschäftsobjekte** sind reale oder gedachte Gegenstände, welche in Geschäftsprozessen verwendet oder bearbeitet werden. Dabei handelt es sich meistens um Stammdaten (z. B. Material, Personal, Anlage) oder Bewegungsdaten (z. B. Rechnung, Auftrag, Vertrag). **Datenobjekte** sind die informationstechnische Repräsentation dieser Geschäftsobjekte in Anwendungssystemen. Das Geschäftsobjekt Material wird demnach als Materialstammdaten in einem ERP-System abgebildet. Geschäftsobjekte beschreiben somit eine fachliche Sicht auf Datenobjekte. Sie abstrahieren von einer konkreten Repräsentation in einem Anwendungssystem.

Die Notwendigkeit dieser Unterscheidung zeigt sich speziell in der Kommunikation mit verschiedenen Zielgruppen von Data Governance. Personen der Fachabteilungen sprechen über die Verwendung von Daten (Informationen) in Geschäftsprozessen, also den Geschäftsobjekten. IT-Personal interessiert sich für die Repräsentation von Daten in Anwendungssystemen und deren Speicherung in Datenbanken und meint daher Datenobjekte.

2.3 Stammdaten

Stammdaten repräsentieren häufig die wichtigsten Geschäftsobjekte einer Organisation, z. B. Kundschaft, Artikel, Lieferunternehmen oder Personal. Sie sind für die Abwicklung von Geschäftsprozessen von zentraler Bedeutung. Stammdaten werden in Geschäftsvorfällen (Bewegungsdaten) referenziert. Sie haben eine steuernde Wirkung von IT-gestützt ablaufenden Geschäftsprozessen. Beispielsweise steuern die Maße und das Gewicht der Produkte die Beladung eines LKWs. Anhand der aufsummierten Maße und Gewichte der bestellten Produkte wird über das Versandsystem automatisch ermittelt, welche und wie viele LKWs für die Lieferung eingesetzt werden müssen.

Stammdaten durchlaufen während ihres Lebenszyklus oft verschiedene Organisationsfunktionen. Ein Beispiel sind Produktstammdaten, die in der Forschung & Entwicklung entstehen und später in der Produktion, im Marketing und in der Auftragsabwicklung genutzt werden. Die Stammdaten Name, Adresse und Geburtsdatum einer neuen Kundin erheben zum Beispiel die Mitarbeitenden des Außendienstes, das Controlling ordnet der Kundin verschiedene Kategorien für das Berichtswesen zu, der Vertrieb legt Zahlungs- und Lieferkonditionen fest, und die IT ist für die Bereitstellung und Wartung der Systeme zuständig, welche alle Stammdaten der Kundin vorhalten (z. B. ERP- und CRM-Systeme).



Beispiel: Materialstammsatz (siehe Bild 2.3)

„Ein Materialstammsatz speichert beispielsweise sämtliche Informationen zu den Artikeln, Teilen und Dienstleistungen, die eine Organisation beschafft, fertigt und lagert. Er wird von den verschiedenen funktionalen Bereichen (z. B. Einkauf, Verkauf, Logistik, Produktion oder Buchhaltung) genutzt. Diese benötigen neben den allgemeinen Grunddaten, wie z. B. einer Materialnummer zur eindeutigen Identifikation und einer Bezeichnung, in der Regel noch funktionspezifische Informationen, die in Sichten gruppiert werden: So müssen dem Einkauf insbesondere Preise und Konditionen zur Verfügung stehen, während für die Fertigungsplanung (Disposition) Sicherheitsbestände und Losgrößen relevant sind und die Buchhaltung vor allem an Kontierungsinformationen interessiert ist. Stammdaten können von der Organisationsstruktur abhängig sein bzw. in mehreren Sprachen gepflegt werden (...). Durch die Integration zum Beispiel aller materialspezifischen Informationen in einem einzigen Stammsatz entfällt dafür die redundante Datenhaltung.“ (Legner & Otto, 2007)

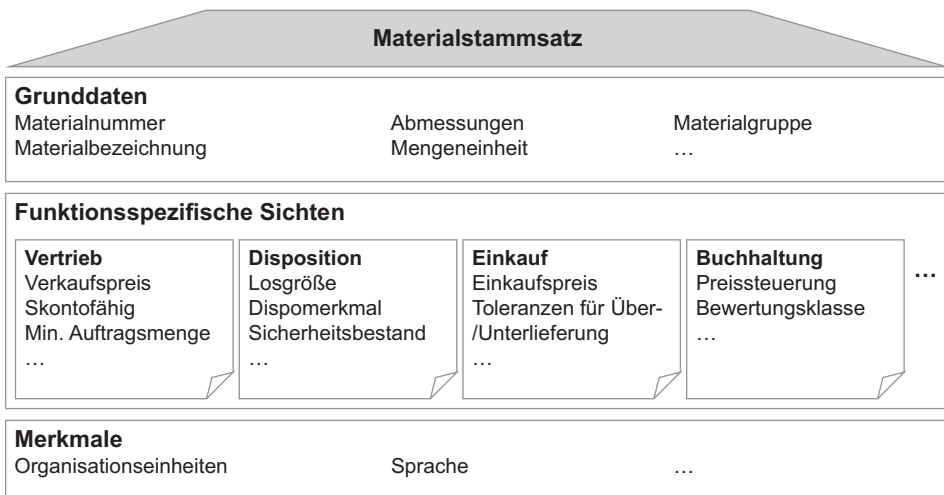


Bild 2.3 Beispiel für den Aufbau des Materialstammsatzes am Beispiel SAP R/3 (in Anlehnung an Legner & Otto, 2007)

Im Fokus von Data Governance stehen meist global gültige Stammdaten. Globale Stammdaten – oder auch **Konzerndaten** – sind Stammdaten, die für die gesamte Organisation gültig und von Bedeutung sind (Otto & Österle, 2016, S. 29 f.). Das können Daten über Produkte sein, die ein Unternehmen herstellt und vertreibt. Auch wenn die Text-Dokumentationen zu den Produkten in unterschiedliche Sprachen übersetzt und dargestellt werden, sind andere Stammdaten wie Abmessungen, Gewicht oder Bildmaterialien global gültig. Lokale Stammdaten werden hingegen nur in einem Teil

der Organisation verwendet, z. B. in einer Filiale, in einem Geschäftsbereich, in einer Abteilung oder in einem Werk.

Eine spezielle Art von Stammdaten sind **Referenzdaten**. Der Ursprung von Referenzdaten und auch die Verantwortung für ihre Qualität liegen außerhalb der eigenen Organisation (Otto & Österle, 2016, S. 30). Typische Beispiele sind Währungs- und Ländercodes, Geodaten und Klassifikationsstandards wie eCl@ss. Auch viele Adressdaten sind Referenzdaten, wie z. B. Straßennamen, Hausnummern, Ortsbezeichnungen und Postleitzahlen. Für die Vergabe (und Pflege) der deutschen Postleitzahlen ist beispielsweise die Deutsche Post AG zuständig.

Häufig werden gerade Adressdaten in Organisationen über Referenzdaten gepflegt. Das hat den Hintergrund, dass zum einen eine einheitliche Schreibweise von Straßen- und Ortsnamen sichergestellt werden soll. Zum anderen können über Referenzdaten und die Kombination verschiedener Adressbestandteile wie Straßename, Hausnummer und Ortsname die Postleitzahl eindeutig festgestellt und automatisch ergänzt werden. Aus Perspektive einer Organisation, die diese Adressen zusammen mit den Kundeninformationen wie Namen in einem CRM-System verarbeitet, ist die Gesamtheit dieser Information ein Stammdatum der Kundschaft, auch wenn ein Teil über Referenzdaten angereichert wird. Aus Perspektive der Organisation, welche die Adressdaten als Referenzdaten anbietet, sind diese Referenzdaten wiederum Stammdaten, die ebenso Änderungen erfahren können, und die ebenso durch andere Referenzen angereicht werden können.

Eine andere Sichtweise ist, dass auch organisationsinterne Daten als Referenzdaten bezeichnet werden können, wenn sie einen „referenzierenden“ Charakter haben. Sie ändern sich noch seltener als Stammdaten, sind deutlich weniger komplex und werden nicht nur in Bewegungsdaten, sondern auch in den Stammdaten selbst referenziert. Ein Beispiel sind Organisationsbezeichnungen, wie sie in ERP-Systemen verwendet werden, also z. B. Buchungskreise, Vertriebsbezirke, Kunden- oder Produktgruppen und Kostenschlüssel. Diese Referenzdaten haben meist eine ordnende Funktion oder dienen der Kategorisierung und Charakterisierung anderer Daten (Henderson et al., 2017, S. 350 ff.). In einer Hochschule könnte man Studiengänge oder Semesterzahlen (1 bis 7) als Referenzdaten bezeichnen. In dem Fall werden die Studierenden (= Stammdatum) einem Studiengang sowie dem aktuellen Semester (= Referenzdatum) zugeordnet.

Je nachdem, aus welcher der oben beschriebenen Perspektiven ein Referenzdatum vorliegt, kann die Änderung eines Referenzdatums zu einer Änderung bei sehr vielen Datensätzen führen – nämlich bei allen Datensätzen, die auf dieses Datum referenzieren. Wird z. B. ein Straßename offiziell angepasst, kann dies Auswirkungen auf viele verschiedene Kundenstammdaten haben. Ein anderes Beispiel ist die Änderung eines Farbcodes von Produkten. Wird dieser geändert, kann das Auswirkungen auf die Darstellung vieler Artikel im Online-Shop haben.

2.4 Open Data und externe Daten

Daten aus externen Quellen bekommen im Zuge der Digitalisierung eine immer größere Bedeutung. Organisationen arbeiten stark vernetzt untereinander und tauschen mit geschäftlichen Kontakten, z. B. entlang der Lieferkette, regelmäßig Daten aus (vgl. z. B. Otto et al., 2019; Schemm, 2009). Datenproduzierende Organisationen (Information Broker) stellen Daten zum Kauf bereit, z. B. Listen potenzieller Zielgruppen oder Adressdaten. Zudem gibt es immer mehr öffentliche Daten, die unter bestimmten Bedingungen frei genutzt werden dürfen. Ein weiteres Beispiel für externen Daten sind die bereits erwähnten Adress-Referenzdaten.

Entlang von Lieferketten oder Wertschöpfungsketten findet vielfältig Datenaustausch statt. Einfache Beispiele sind der Austausch von Bestellungen, Rechnungen und Zahlungen. Diese externen, von geschäftlichen Kontakten stammenden, Daten müssen in die eigenen Informationssysteme integriert werden, um sie dort weiterzuverarbeiten. Eine Bestellung löst beispielsweise einen Produktionsprozess aus oder führt dazu, dass Ware aus dem Lager entnommen, verpackt und an den Bestellenden versandt wird. Je enger die Zusammenarbeit zwischen den Parteien, umso mehr Daten werden (automatisiert) ausgetauscht (siehe Kasten).



Beispiel: Vendor Managed Inventory

Vendor Managed Inventory (VMI) dreht den klassischen Bestellprozess um und basiert auf einem erweiterten Datenaustausch zwischen Kunde und Lieferant. Der Kunde liefert seine aktuellen Bestandsdaten an den Lieferanten. Der Lieferant übernimmt daraufhin für den Kunden die Bedarfsplanung und übermittelt das Ergebnis als Auftragsbestätigung an den Kunden. In dessen Systemen wird daraufhin automatisch eine Bestellung angelegt. Durch VMI reduzieren die Kunden ihre Bestandskosten und der Lieferant profitiert von einer höheren Kundenbindung und höherer Planungssicherheit. [vgl. Senger, 2004, S. 179 ff.]

Beispiel: Blockchain-Technologie in der Agrarindustrie

Ein Beispiel für den Einsatz der Blockchain-Technologie zum Datenaustausch liefert Louis Dreyfus Co. in den USA. Der Lebensmittellieferant arbeitet mit niederländischen und französischen Banken zusammen, um Agrarrohstoffhandel auf Basis der Blockchain abzuwickeln. Sie vergleichen automatisiert Zeitstempeldaten, um Doppelarbeiten und manuelle Prüfungen zu vermeiden. Dadurch verkürzt sich die Dokumentenverarbeitung auf ein Fünftel der üblichen Zeit. Die Digitalisierung der gesamten Lebensmittelversorgungskette, vom Landwirt bis zum Verbraucher, auf Basis der Blockchain in Kombination mit Technologien wie QR-Codes, RFID, digitalen Signaturen und Mobilgeräten erlaubt eine lückenlose Überwachung des physischen Warenflusses. [Rijanto, 2021, S. 221]