

MASTER THESIS  
Simon Hamma

# Data Governance in kleinen und mittelständischen Unternehmen: Entwicklung und Evaluierung eines adaptiven Frameworks

---

FAKULTÄT TECHNIK UND INFORMATIK  
Department Informatik

Faculty of Engineering and Computer Science  
Department Computer Science

Simon Hamma

# Data Governance in kleinen und mittelständischen Unternehmen: Entwicklung und Evaluierung eines adaptiven Frameworks

Masterarbeit eingereicht im Rahmen der Masterprüfung  
im Studiengang *Master of Science Informatik*  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Olaf Zukunft  
Zweitgutachterin: Prof. Dr. Ulrike Steffens

Eingereicht am: 30. August 2024

**Simon Hamma**

## **Thema der Arbeit**

Data Governance in kleinen und mittelständischen Unternehmen: Entwicklung und Evaluierung eines adaptiven Frameworks

## **Stichworte**

Data Governance, KMU, Framework, Datenqualität, Compliance, Datenmanagement, Datenflüsse

## **Kurzzusammenfassung**

Um den größten Mehrwert aus Daten zu ziehen und deren Sicherheit zu gewährleisten, ist die Betrachtung von Daten als Vermögenswert auch für kleine und mittelständische Unternehmen (KMU) ein unverzichtbarer Aspekt geworden. Die in der Forschung und Industrie diskutierten und verwendeten Rahmenwerke richten sich jedoch besonders auf große Unternehmen und sind für kleinere Unternehmen aufgrund ihrer Komplexität und ihrem Umfang schwer umzusetzen. Diese Arbeit beschäftigt sich aus diesem Grund mit der Entwicklung und Evaluierung eines Data Governance Frameworks für KMUs. Ziel des sogenannten *CoreGovernance* Frameworks ist es, eine sichere Verwaltung, Datensicherheit und die Einhaltung gesetzlicher Vorschriften KMUs mit ihren besonderen Voraussetzungen und Herausforderungen zu ermöglichen. Dies wird durch die Fokussierung auf die wesentlichen Aspekte von Data Governance und deren Anpassung an die Besonderheiten von KMUs erreicht. Um eine weitere Forschungslücke zu schließen, wird zur Evaluierung eine praxisorientierte Testumgebung mit Metriken erstellt. Dies dient ebenso dem Vergleich zukünftiger Arbeiten mit dem CoreGovernance Framework. Nach der Erstellung und Definition des theoretischen Frameworks wird dieses in einem in der Testumgebung erstellten fiktiven Unternehmen angewendet, um die Machbarkeit untersuchen zu können. Die Ergebnisse der Analysen ergeben, dass das Framework kleineren Unternehmen in realitätsnahen Szenarien bei der sicheren Verwaltung der Daten deutlich unterstützen kann und dabei die wesentlichen Elemente von Data Governance beinhaltet.

---

**Simon Hamma**

**Title of Thesis**

Data governance in small and medium-sized enterprises: Development and evaluation of an adaptive framework

**Keywords**

Data Governance, SME, Framework, Data Quality, Compliance, Data Management, Data Flows

**Abstract**

In order to get the most value out of data and ensure its security, viewing data as an asset has also become an essential aspect for small and medium-sized enterprises (SMEs). However, the frameworks discussed and used in research and industry are particularly focussed on large companies and are difficult to implement for smaller companies due to their complexity and scope. For this reason, this thesis deals with the development and evaluation of a data governance framework for SMEs. The goal of the so called *CoreGovernance* framework is to enable secure management, data security and compliance with legal regulations for SMEs with their specific requirements and challenges. This is achieved by focussing on the essential aspects of data governance and adapting them to the specific characteristics of SMEs. In order to close another research gap, a practice-orientated test environment with metrics will be created for evaluation purposes. This also serves to compare future work with the CoreGovernance framework. Once the theoretical framework has been created and defined, it is applied in a fictional company created in the test environment in order to analyse its feasibility. The results of the analyses show that the framework can significantly support smaller companies in the secure management of data in realistic scenarios and includes the essential elements of data governance.

# Inhaltsverzeichnis

Abbildungsverzeichnis	viii
Tabellenverzeichnis	x
Abkürzungen	xi
<b>1 Einleitung</b>	<b>1</b>
<b>2 Stand der Forschung</b>	<b>3</b>
2.1 Grundlagen . . . . .	3
2.1.1 Data Governance . . . . .	3
2.1.2 Vorteile und Auswirkungen . . . . .	8
2.1.3 Kleine- und Mittelständische Unternehmen . . . . .	10
2.2 Komponenten von Data Governance . . . . .	11
2.2.1 Standards und Richtlinien . . . . .	11
2.2.2 Prozesse . . . . .	12
2.2.3 Verantwortlichkeiten und Rollen . . . . .	13
2.3 Überblick über bestehende Frameworks . . . . .	16
2.3.1 DAMA-DMBoK . . . . .	16
2.3.2 Data Governance Institute . . . . .	21
2.3.3 PwC . . . . .	24
2.4 Data Governance in KMUs . . . . .	28
2.4.1 Rahmenbedingungen . . . . .	28
2.4.2 Einordnung der bestehenden Literatur . . . . .	29
2.4.3 Forschungsfragen . . . . .	31
<b>3 Design einer Testumgebung</b>	<b>33</b>
3.1 Aufbau und Ziele der Testumgebung . . . . .	33
3.2 Fiktives KMU . . . . .	34
3.2.1 Unternehmensstruktur . . . . .	35

3.2.2	Datenflüsse . . . . .	36
3.2.3	Data Lake . . . . .	38
3.2.4	Technologien . . . . .	40
3.2.5	Herausforderungen . . . . .	44
3.3	Metriken . . . . .	49
3.3.1	Domänenmatrix . . . . .	49
3.3.2	Evaluierungsszenarien . . . . .	51
<b>4</b>	<b>Konzeption eines Frameworks</b>	<b>55</b>
4.1	Analyse bestehender Frameworks . . . . .	55
4.1.1	DAMA . . . . .	56
4.1.2	DGI . . . . .	57
4.1.3	PwC . . . . .	59
4.2	Bewertung der Frameworks . . . . .	61
4.3	CoreGovernance Framework . . . . .	63
4.3.1	Übersicht . . . . .	63
4.3.2	Kernkomponenten . . . . .	65
4.3.3	Überwachung und Reifegradmodell . . . . .	76
<b>5</b>	<b>Case Study: Zusammenführen von Theorie und Praxis</b>	<b>82</b>
5.1	Vorgehensweise . . . . .	82
5.2	Ziele und Strategie . . . . .	84
5.3	Rollen und Verantwortlichkeiten . . . . .	86
5.4	Reifegradmodell . . . . .	90
5.5	Operatives Modell . . . . .	93
5.6	Technologische Umsetzung . . . . .	95
<b>6</b>	<b>Ergebnisse &amp; Analyse</b>	<b>103</b>
6.1	Qualitative Untersuchung . . . . .	103
6.1.1	Domänenmatrix . . . . .	103
6.1.2	Vergleich mit den vorgestellten Frameworks . . . . .	107
6.2	Quantitative Untersuchung . . . . .	108
6.2.1	Bewertung der Entscheidungsfindung . . . . .	109
6.2.2	Evaluierungsszenarien . . . . .	109
6.2.3	Fortschritt im Reifegradmodell . . . . .	113
6.3	Diskussion der Ergebnisse . . . . .	113

<b>7 Fazit</b>	<b>116</b>
7.1 Zusammenfassung . . . . .	116
7.2 Ausblick . . . . .	118
<b>Literaturverzeichnis</b>	<b>119</b>
<b>A Anhang</b>	<b>127</b>
A.1 Abbildungen . . . . .	127
A.1.1 Data Governance Modell . . . . .	127
A.1.2 Prozessabläufe . . . . .	129
A.1.3 Implementierung in DataHub . . . . .	132
A.1.4 Soda Core Ergebnis . . . . .	135
A.1.5 Apache Airflow Workflow . . . . .	135
A.2 Listings . . . . .	136
A.2.1 Rechtezuweisung in MinIO . . . . .	136
<b>Selbstständigkeitserklärung</b>	<b>137</b>

# Abbildungsverzeichnis

2.1	Anzahl der Veröffentlichungen mit direktem Bezug zu Data Governance pro Jahr [1] . . . . .	4
2.2	DAMA DMBok Data Management Framework [34] . . . . .	16
2.3	Framework des Data Governance Institute [22] . . . . .	22
2.4	Framework des Beratungsunternehmens PwC [65] . . . . .	25
2.5	Schwerpunkte wissenschaftlicher und praxisorientierter Publikationen [4] .	30
3.1	Datenflüsse des Unternehmens . . . . .	37
3.2	Vergleich Data Warehouse und Data Lake [51] . . . . .	39
3.3	Data Lake Architektur der AquaFlow GmbH . . . . .	41
3.4	Einflüsse auf Entscheidungsfindung in der Marketing-Abteilung . . . . .	46
3.5	Einflüsse auf Entscheidungsfindung in der Entwicklungsabteilung . . . . .	46
3.6	Skizze der Einflüsse auf die Entscheidungsfindung . . . . .	48
3.7	Die fünf Entscheidungsdomänen nach Khatri und Brown [41] . . . . .	50
3.8	Domänenmatrix, angelehnt an [4] . . . . .	52
4.1	Vergleich der Kernelemente, relevanten Aspekte für KMUs und Anpassungsbedarf der Data Governance Frameworks von DAMA, DGI und PwC	62
4.2	Übersicht des CoreGovernance Frameworks . . . . .	64
4.3	Übersicht über anfallende Aufgaben im operativem Modell des Core Governance Frameworks . . . . .	71
4.4	Visualisierung der Reifegrade in den verschiedenen Kernbereichen [34] . .	80
4.5	Visuelles Modell für die Erfassung des Reifegrads . . . . .	80
5.1	Übersicht der Data Governance Rollen in der AquaFlow GmbH . . . . .	87
5.2	Aktueller Stand des Reifegrads der AquaFlow GmbH . . . . .	92
5.3	Prozessablauf für Probleme mit der Datenqualität . . . . .	94
6.1	Domänenmatrix des CoreGovernance Frameworks . . . . .	104
6.2	Einflüsse auf die Entscheidungsfindung mit Data Governance . . . . .	110

6.3	Aktuelles Reifegradmodell der AquaFlow GmbH . . . . .	114
A.1	Übersicht über Data Governance Modell [34] . . . . .	128
A.2	Prozessablauf für Integration neuer Datensätze . . . . .	130
A.3	Prozessablauf für die Einführung neuer Datenkontrollen . . . . .	131
A.4	Ausschnitt der Startseite von Datahub mit Suchleiste . . . . .	132
A.5	Übersicht über vorhandene Datenquellen in Listenform . . . . .	133
A.6	Begriffsgruppen des erstellten Business Glossars . . . . .	133
A.7	Übersicht der rechtlichen Richtlinien und Normen . . . . .	134
A.8	Detaillierte Informationen eines Datensatzes . . . . .	134
A.9	Ergebnisse der ersten Qualitätsmessung . . . . .	135
A.10	Graphische Visualisierung eines Datenqualitäts-Workflows . . . . .	135

# Tabellenverzeichnis

2.1	Schwellenwerte für die Einstufung von KMUs [19] . . . . .	10
3.1	Data Lake Komponenten und die verwendeten Technologien . . . . .	41
5.1	Zuweisung der Datenquellen an die Abteilungen . . . . .	89
5.2	Assessment der Data Governance Charakteristiken . . . . .	90

# Abkürzungen

**AWS** Amazon Web Services.

**BI** Business Intelligence.

**CDO** Chief Data Officer.

**CIO** Chief Information Officer.

**CMMI** Capability Maturity Model Integration.

**CRM-System** Customer-Relationship-Management-System.

**DAMA** Data Management Association.

**DG** Data Governance.

**DGC** Data Governance Council.

**DGF** Data Governance Framework.

**DGI** Data Governance Institute.

**DGO** Data Governance Office.

**DMBoK** Data Management Body of Knowledge.

**DSGVO** Datenschutz-Grundverordnung.

**ELT** Extract, Load, Transform.

**ERP-System** Enterprise-Resource-Planning-System.

**ETL** Extract, Transform, Load.

**EU** Europäische Union.

**FF** Forschungsfrage.

**GDPR** General Data Protection Regulation.

**IAM** Identity and Access Management.

**IoT** Internet of Things.

**ISO** International Organization for Standardization.

**JSON** JavaScript Object Notation.

**KI** Künstliche Intelligenz.

**KMU** kleine und mittelständische Unternehmen.

**PwC** PricewaterhouseCoopers.

**RDBM-System** relationales Datenbankmanagement-System.

**S3** Simple-Storage-Service.

**SQL** Structured Query Language.

# 1 Einleitung

In den vergangenen Jahren ist die Datenmenge, die von Unternehmen genutzt wird, stark angestiegen und nimmt eine immer größer werdende Rolle in den Unternehmensprozessen ein [62]. Datenbasierte Entscheidungen sind ein unverzichtbarer Bestandteil der heutigen Geschäftswelt [47]. Dadurch entstehen immer neue Anforderungen an die Daten, um diese bearbeiten, speichern und analysieren zu können [2]. Die Bewältigung der auftretenden Herausforderungen und Probleme im Zusammenhang mit Daten erfolgt häufig dezentral und situationsbezogen in den einzelnen Abteilungen [40]. Basierend auf diesen Erkenntnissen ist die Notwendigkeit identifiziert worden, Daten als einen wertvollen Vermögenswert innerhalb des Unternehmens anzusehen [41]. Daraus ergibt sich, dass Daten im Unternehmen ebenso sorgfältig verwaltet und geschützt werden müssen wie andere Vermögenswerte [56].

Dies führt zu dem Bedarf nach klaren Strukturen und Vorgaben, die die Daten systematisch verwalten und schützen, was zur Entwicklung des Forschungsgebiets der Data Governance geführt hat. Dadurch erhoffen sich Forscher und Anwender eine Verbesserung der Qualität, Nutzung und Sicherheit der Daten innerhalb des Unternehmens. Zudem gibt es immer stärkere Richtlinien und Gesetze für Unternehmen im Bezug auf personenbezogene Daten. Aufgrund der Bedeutung und dem Potenzial von Daten für geschäftliche Entscheidungen hat das Forschungsgebiet stark an Popularität gewonnen [41][72].

In Reaktion auf die Notwendigkeit und die Anforderungen wurden eine Vielzahl an sogenannten Rahmenwerken (engl. Frameworks) entwickelt, die durch Vorgaben für Prozesse und Standards eine strukturierte und systematische Verwaltung der Daten in Unternehmen ermöglichen. Diese Frameworks wurden dabei häufig für und von großen Unternehmen konzipiert, da in diesen die Datenprobleme zuerst auftraten und sie durch ihre größere Präsenz und Einfluss eine Rolle bei der Entwicklung solcher Standards spielen können.

Durch die technologischen Fortschritte und die voranschreitende Digitalisierung haben KMUs ebenfalls Zugang zu leistungsfähiger Infrastruktur und großen Datenmengen. Diese Unternehmen können dabei ebenfalls von Informationen in den Daten profitieren und schnelle, datengetriebene Entscheidungen treffen, die ihre Wettbewerbsfähigkeit verbessert [36]. KMUs weisen jedoch im Gegensatz zu großen Unternehmen andere Strukturen und Ressourcen auf, was die Einführung der bestehenden Frameworks erschwert [11].

Die fehlende Aufmerksamkeit, die KMUs im Bezug auf Data Governance erhalten, ist insbesondere angesichts deren Beitrag zur Wirtschaft und der stetig wachsenden Rolle von Daten in der gesamten Geschäftswelt bedauerlich. Diese Arbeit beschäftigt sich aus diesem Grund intensiv mit der Rolle von Data Governance für KMUs und stellt das speziell auf deren Anforderungen zugeschnittene CoreGovernance Framework vor.

Die Arbeit strukturiert sich dabei wie folgt. In Kapitel 2 werden grundlegende Konzepte von Data Governance erörtert und bestehende Frameworks aus unterschiedlichen Bereichen vorgestellt. Ebenso wird auf die Anforderungen und Herausforderungen von KMUs näher eingegangen und anhand davon die Forschungsfragen formuliert. In Kapitel 3 wird eine Testumgebung für die Evaluierung von Frameworks, die in KMUs eingesetzt werden sollen, entwickelt. Dabei wird ein fiktives Unternehmen mit den dazugehörigen Herausforderungen und Problemen erstellt, die von dem zu entwickelnden und künftigen Frameworks adressiert werden müssen. Kapitel 4 befasst sich mit der Entwicklung des theoretischen Teils des CoreGovernance Frameworks, das auf die Bedürfnisse von KMUs zugeschnitten ist. Dafür werden Stärken und Schwächen bestehender Frameworks analysiert und Elemente identifiziert, die für KMUs besonders relevant sind. Im Anschluss daran wird das CoreGovernance Framework und seine Bestandteile vorgestellt. In Kapitel 5 wird das theoretische Framework auf die zuvor entwickelte Testumgebung angewandt, um die praktische Anwendbarkeit zu demonstrieren. Die Evaluation der Theorie und der praktischen Umsetzung erfolgt in Kapitel 6. Die Ergebnisse der Arbeit werden in Kapitel 7 zusammengefasst und durch Ansätze für die weiterführende Forschung ergänzt.

## 2 Stand der Forschung

In diesem Kapitel wird der aktuelle Stand der Forschung im Bereich Data Governance untersucht. Ziel ist es, ein fundiertes Verständnis der bestehenden Theorien, Modelle und Frameworks zu entwickeln. Dafür werden zunächst verschiedene Grundlagen, die für das Verständnis der vorliegenden Arbeit relevant sind, erklärt. In Abschnitt 2.2 sind die wesentlichen Bestandteile, die in allen Data Governance Frameworks auftreten, erläutert. Um das Verständnis für den Aufbau einer Data Governance Strategie zu erlangen und aktuelle Best-Practices aufzuzeigen, werden in Abschnitt 2.3 Frameworks unterschiedlicher Herkunft analysiert. Im Anschluss daran wird in Abschnitt 2.4 die wachsende Bedeutung von Data Governance für KMU untersucht sowie bestehende Literatur evaluiert. Anhanddessen wird der Bedarf nach mehr Forschung in diesem Bereich dargelegt und es werden die Forschungsfragen definiert.

### 2.1 Grundlagen

Im folgenden Abschnitt werden die theoretischen Grundlagen geschaffen, die für das Verständnis und die Entwicklung eines Data Governance Frameworks benötigt werden. Dabei werden zunächst unterschiedliche Definitionen und Ansätze von Data Governance untersucht. Im Anschluss daran werden die erhofften Vorteile und Auswirkungen durch die Implementierung der Data Governance Aktivitäten näher betrachtet. Abschließend erfolgt eine nähere Definition des Begriffs KMU, um deren Herausforderungen und Probleme verstehen zu können.

#### 2.1.1 Data Governance

Für Data Governance gibt es viele verschiedene Definitionen, die sich auch im Laufe der Zeit stetig weiterentwickeln, da es sich gerade im Kontext Digitalisierung und Künstliche Intelligenz (KI) um ein immer relevanter werdendes Thema handelt. Daher werden

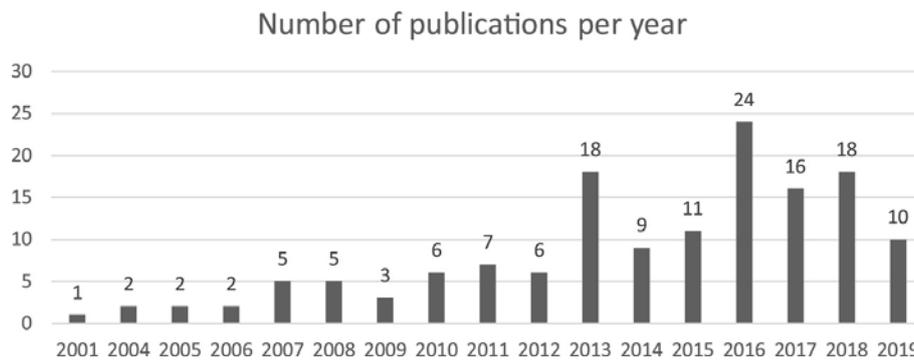


Abbildung 2.1: Anzahl der Veröffentlichungen mit direktem Bezug zu Data Governance pro Jahr [1]

wie im gesamten Rahmen der Arbeit Quellen aus verschiedenen Bereichen für Definitionen und Begriffserklärungen genommen. Bei den Quellen handelt es sich um Werke aus der Grundlagenliteratur wie von der Data Management Association (DAMA), aktuelle Forschungsliteratur sowie praxisorientierte Veröffentlichungen von Beratungsfirmen oder Vendors. Ein Start mit der Grundlagenliteratur Data Management Body of Knowledge (DMBoK) von DAMA ist angemessen, da sich nahezu alle anderen Quellen darauf beziehen. Die erste Ausgabe von DMBoK wurde bereits 2009 veröffentlicht, als weniger Augenmerk auf das Thema gelegt wurde, was Untersuchungen von Abraham et al. in [1] bestätigen. In der Literaturübersicht haben die Autoren unter anderem die Anzahl der Veröffentlichungen pro Jahr analysiert, die in direktem Bezug zu Data Governance stehen. Das Ergebnis davon ist in Abbildung 2.1 zu sehen. Durch den Einfluss von weiteren Veröffentlichungen und zugewonnene Erfahrungen erschien 2017 eine überarbeitete Version von DMBoK. In diesem wird Data Governance wie folgt definiert [34]:

*Data Governance (DG) is defined as the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets.*

Dabei wird zunächst der wesentliche Punkt für die heutige Sichtweise auf Daten deutlich. Während Daten früher noch als Nebenprodukt der verschiedenen Businessabläufe gesehen wurden, so sollten sie heutzutage als Vermögenswert oder *Asset* behandelt werden [2]. Das Konzept dafür, Informationen als Asset zu sehen, wurde bereits 1994 vom Hawley Komitee entwickelt [44]. Oppenheim et al. dehnen die Definition von Vermögenswerten auf Daten als Ressource aus [56]. Auch andere Quellen unterscheiden in späteren Veröffentlichungen nicht mehr zwischen Daten und Informationen [33][41]. Da Daten bzw. Informationen nun Vermögenswerte sind, sollten sie von der Managementebene aus eben-

so wie andere Vermögenswerte, wie Anlagen oder Technologien, behandelt werden [56]. Um den Wert von den Assets zu erhalten und zu steigern, ist eine sorgfältige Erfassung, Bewertung und fortlaufende Pflege notwendig. Dieses Konzept muss nun auf die Daten eines Unternehmens ausgeweitet werden. Das bedeutet, dass konkrete Mechanismen zur Datenerfassung und -analyse während des gesamten Lebenszyklus zur Sicherstellung von Qualität und Relevanz implementiert werden müssen. Und genau dort kommt Data Governance ins Spiel. Data Governance bietet dem Unternehmen und seinen Verantwortlichen ein strukturiertes System, das Strategien, Rollen und Prozesse definiert, um den Umgang mit allen Daten zu organisieren und zu steuern. Durch diese klaren Vorgaben und Richtlinien von Data Governance können Unternehmen sicherstellen, dass ihre Daten den gesetzlichen Anforderungen sowie externen und internen Standards entsprechen [40]. Die daraus resultierende erhöhte Qualität der Daten fördert das Vertrauen in diese und die daraus abgeleiteten Entscheidungen [61][14].

In der Definition von [34] wird klar, dass DAMA bei Data Governance rein die Ausführung der Autorität und Kontrolle über Verwalten der Daten sieht und nicht die direkte Umsetzung, also Datenmanagement, selber. Der Schwerpunkt liegt also in der Festlegung von Rahmenbedingungen, Richtlinien und Prozessen, die eine effektive Verwaltung und Verarbeitung der Daten gewährleisten soll. Die konkrete Umsetzung auf Informations- und Technologieebene wird dem Bereich des Datenmanagements überlassen. Für eine klarere Unterscheidung und ein besseres Verständnis sollten die beiden Konzepte von Data Governance und Data Management kurz erläutert werden. Auch in der Forschung und in der Praxis werden die Begriffe häufig uneinheitlich verwendet [2]. DAMA definiert Data Management als die Entwicklung, Umsetzung und die Überwachung von Plänen, Regeln, Programmen und Prozessen zur Kontrolle, zum Schutz, zur Bereitstellung und zur Steigerung des Werts von Daten und Informationsbeständen [34]. Insgesamt fokussiert sich Data Management mehr auf die konkreten Datenelemente und wie sie gespeichert, strukturiert und bewegt werden [2]. Data Governance übernimmt dabei die Planung und Kontrolle der Datenverwaltung auf höherer, abstrakterer Ebene. Laut Brous et al. sollte Data Governance sicherstellen, dass alle operativen Funktionen des Data Managements zur Erreichung der Unternehmensvision, -mission und -strategie beitragen [12]. Data Governance sorgt also dafür, dass die angestrebten Ziele des Unternehmens durch eine klare Definition von Richtlinien und Strukturen für das Datenmanagement unterstützt und erreicht werden können. Generell lässt sich sagen, dass Data Governance das Rahmenwerk mit einer Strategieausrichtung, Regularien und Zielen liefert, während Data Management die Entwicklung und Umsetzung dieser Vorgaben auf operativer Ebene darstellt

[1][14][40][27]. Khatri und Brown liefern die simplere Erklärung, dass Data Governance sagt, welche Entscheidungen von wem getroffen werden müssen, und Data Management diese Entscheidungen umsetzt [41]. Insgesamt lässt sich sagen, dass die Definitionen von DAMA ebenso wie die Erläuterungen im Anschluss generisch und allgemein gültig sind. Die abstrakten Konzepte können es jedoch Anwendern schwer machen, die allgemeingültigen Prinzipien ohne tieferes Verständnis und praktische Erfahrung umzusetzen. Aus diesem Grund werden im Folgenden weitere Definitionen und Erläuterungen von Data Governance betrachtet.

Eine ausführlichere Definition liefern Abraham et al. in [1]:

*Data governance specifies a cross-functional framework for managing data as a strategic enterprise asset. In doing so, data governance specifies decision rights and accountabilities for an organization's decision-making about its data. Furthermore, data governance formalizes data policies, standards, and procedures and monitors compliance.*

Die Autoren teilen die Definition in sechs Abschnitte auf, die alle relevanten Aspekte abdecken. Data Governance soll als ganzheitlicher, unternehmensweiter Ansatz umgesetzt werden. Dadurch wird eine Zusammenarbeit über die Grenzen der Abteilungen und datenspezifischen Domänen ermöglicht. Als Rahmenwerk soll es dabei eine Struktur und Regularien für den Umgang mit Daten schaffen. Ebenso betont es die Bedeutung der Daten als Vermögenswert innerhalb eines Unternehmens, der entsprechend behandelt und verwaltet werden sollte. Um festzulegen, welche datenbezogenen Entscheidungen von wem in dem Unternehmen getroffen werden dürfen, muss Data Governance verschiedene Entscheidungsrechte und Verantwortlichkeiten spezifizieren. Das Herausarbeiten der Artefakte Datenrichtlinien-, -standards, und -prozessen ist ein wesentlicher Bestandteil von Data Governance, um einen gewünschten, mit der Unternehmensstrategie im Einklang stehenden Umgang mit Daten zu fördern. Zum Schluss steht die Aufgabe der Überwachung der Einhaltung der definierten Regularien und Standards. Dies beinhaltet implementierte Kontrollmechanismen. Insgesamt liefern Abraham et al. in [1] eine präzisere und konkretere Definition als DAMA. Die Aufgaben der einzelnen Bestandteile und Artefakte werden genannt und bieten Anwendern einen ersten Leitfaden für die Umsetzung einer eigenen Data Governance Strategie in ihrer Organisation.

Mit dem Ziel, konkrete Lösungen für Unternehmen in der Industrie zu entwickeln und zu fördern, befasst sich die Organisation Data Governance Institute (DGI) intensiv mit

dem Thema Data Governance. Eigenen Aussagen nach handelt es sich bei DGI um die älteste und bekannteste Quelle für fundierte Best-Practices und Anleitungen zur Data Governance [22]. Da sie den Anspruch hat, als Quelle für Anwender von Data Governance zu dienen, wird sie hier als repräsentative Quelle für einen Vertreter aus der Industrie genommen. DGI definiert Data Governance wie folgt:

*Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods.*

Der Fokus von Data Governance liegt dabei auf den Entscheidungsrechten und Verantwortlichkeiten innerhalb eines Unternehmens. Diese legen ein System fest, das auf vereinbarten Modellen basiert und angibt, welche Handlungen mit welchen Informationen unter welchen Umständen von wem durchgeführt werden können. DGI sieht den Kern des Rahmenwerkes in Managern auf mittlerer Ebene einer Organisation und möchte diesen klare und ausformulierte Leitlinien zur Handhabung von Informationen und zur Entscheidungsfindung bieten. Diese sollen auch insbesondere für die täglich anfallenden Aufgaben bezüglich Data Governance gelten. Die Leitlinien sind näher in Unterabschnitt 2.3.2 beschrieben. Laut DGI hilft dieser Ansatz bei der Akzeptanz von Data Governance Aktivitäten in ergebnisorientierten Unternehmen, da ein Fokus auf autoritäre Aspekte wie bei DAMA bei den Umsetzenden schlechter aufgenommen wird als ein gemeinschaftliches Erarbeiten und Umsetzen von Richtlinien und Standards.

Insgesamt lässt sich sagen, dass Data Governance ein dynamisches Feld ist, dass durch die immer weiter voranschreitende Digitalisierung und Einsatz von KI immer bedeutender wird. Dies wird durch die Anzahl der Veröffentlichungen mit konkretem Bezug zu Data Governance, visualisiert in Abbildung 2.1, gestützt. Da die Arbeit eine ganzheitliche Untersuchung des Themas anstrebt, wurden Quellen aus verschiedenen Bereichen untersucht. Der Konsens aller Definitionen dabei ist, dass Daten als Asset angesehen werden müssen und Data Governance ein strukturiertes Rahmenwerk für den Umgang damit liefert. Das Ziel dabei ist, mehr Informationen aus den Daten zu erlangen und den Wert von diesen für das Unternehmen zu erhöhen. Der Fokus und die beteiligten Komponenten unterscheiden sich dabei zum Teil zwischen den Definitionen. DAMA legt den Fokus von Data Governance auf Kontrolle und Überwachung aller Datenmanagement verwandten Aufgaben, wobei das Vorantreiben und Definieren hauptsächlich in höheren Ebenen innerhalb einer Organisation stattfindet. Abraham et al. und DGI erweitern

den Ansatz und schlagen eine ganzheitliche, unternehmensweite Lösung vor, die neben der Verwaltung auch die Nutzung und die Vorteile von Data Governance miteinbezieht. Dabei liegt der Fokus insbesondere auf der Entscheidungsfindung und den dafür benötigten Verantwortlichkeiten. Data Governance muss dabei von jedem mitgetragen und ausgeführt werden, wobei jedoch von einer geeigneten Komponente klare Richtlinien und Standards definiert werden müssen.

### 2.1.2 Vorteile und Auswirkungen

Um zu verstehen, warum die Einführung von Data Governance so entscheidend für Unternehmen ist, wird im Folgenden kurz auf die erhofften Auswirkungen eingegangen. Diese lassen sich generell in zwei Kategorien aufteilen - das Monetarisieren der Daten sowie die Risikominimierung durch das Einhalten von Regularien und Standards [70]. Durch die ständig wachsende Menge an Daten erhoffen sich Unternehmen einen Wettbewerbsvorteil insbesondere durch KI und datenbasierte Tools zur Entscheidungsfindung [22][23][2]. Um dies zu erreichen, müssen die Daten von hoher Qualität sein und es muss klar sein, woher die Daten stammen und wie sie verarbeitet wurden. Diese Kenntnisse schaffen Vertrauen in die Daten und verhindern verzögerte oder ungenaue Entscheidungen [14]. Das Konzept, dass schlechte Daten zu schlechten Ergebnissen führen, ist auch als *garbage in, garbage out* bekannt [43]. Werden bereits qualitativ schlechte Daten für eine Entscheidungsfindung herangezogen, so kann auch die daraus resultierende Entscheidung nur von schlechter Qualität sein. Dass eine unternehmensweite Erhöhung der Datenqualität eine der Hauptvorteile von einem Data Governance Framework ist, zeigt die Umfrage des LeBow Business College. Diese besagt in [45], dass 66% der Daten- und Analyseverantwortlichen eine verbesserte Datenqualität als führenden Vorteil erlebten. Neben der sichergestellten Qualität der Daten, müssen diese außerdem gut dokumentiert sein und für die richtigen Nutzer zur Verfügung stehen. Dies steigert die Effizienz der Analysten, die dadurch nicht kostbare Zeit für das Suchen sowie Verstehen der vorliegenden Daten verschwenden. Eine Studie von McKinsey dazu ergab, dass im Durchschnitt 30 Prozent der gesamten Unternehmenszeit mit nicht-wertschöpfender Arbeit, ausgelöst durch schlechte Verfügbarkeit und geringe Datenqualität, verbraucht wird [63]. Für Unternehmen kann es zu Beginn jedoch schwierig sein, konkrete und insbesondere messbare Vorteile zu bestimmen [63][55], da sich beispielsweise Datenqualität auf alle Domänen und Projekte des Unternehmens auswirkt. Daraus folgt, dass die Akzeptanz für die Einführung erschwert

sein könnte [60][34]. Aus diesem Grund ist eines der Hauptbestandteile eines erfolgreichen Data Governance Frameworks die Definition von Metriken und Reifemodellen.

Zusätzlich zu den erwähnten Vorteilen, die zur Wertsteigerung der Daten beitragen, spielt die Risikominimierung eine wesentliche Rolle. Diese wird durch die Einhaltung datenschutzrechtlicher Bestimmungen und die Prävention von Datenlecks erreicht. Dadurch, dass immer mehr Aspekte des Lebens sich in der digitalen Welt abspielen und die Daten von Kunden in Zeiten von Big Data, Internet of Things (IoT) und KI immer relevanter für Unternehmen werden, entwickelte die Europäische Union (EU) die Datenschutz-Grundverordnung (DSGVO) oder auch *General Data Protection Regulation (GDPR)*. Diese regelt den Umgang von Unternehmen mit personenbezogenen Daten, was sowohl den Kunden schützt als auch den Unternehmen klare Richtlinien und Vorgaben bereitstellt, um ihre rechtlichen Risiken zu minimieren. Hohe finanzielle Sanktionen sollen Unternehmen zur Einhaltung der Regularien verpflichten. Dabei können sich die Strafen auf bis zu 20 Millionen Euro beziehungsweise 4% des weltweiten Umsatzes eines Unternehmens, je nachdem welches höher ist, belaufen [28]. Einer Studie nach, durchgeführt von der Anwaltskanzlei DLA Piper, haben EU-Datenschutzbehörden im Zeitraum eines Jahres von Januar 2021 bis Januar 2022 für Verstöße gegen die DSGVO Strafen in Höhe von 1,2 Milliarden US-Dollar verhängt [18]. Das Ziel der Einhaltung der Richtlinien und die Vermeidung der hohen finanziellen Folgen ist in vielen Unternehmen laut DAMA der Hauptgrund für die Einführung von Data Governance [34]. Data Governance soll dabei eine konkrete Vorgabe für das Etablieren und Umsetzen von für das Unternehmen relevanten Standards, Regularien und Prozessen liefern [60]. Darüber hinaus sollte ein geeignetes Framework Mechanismen zur Kontrolle und Überwachung der Einhaltung der Vorgaben bereitstellen. Diese Automatisierung der datenschutzbezogenen Aufgaben senkt den Aufwand und Kosten für das Unternehmen und legt klare Verantwortlichkeiten fest [60]. Die Nicht-Einhaltung der Vorgaben und Regularien erhöht sowohl das Risiko eines Datenlecks als auch die potenziellen resultierenden Folgen und Schäden. Denn Datenlecks können nicht nur weitere Strafen mit sich führen, sondern auch weitreichende Folgen für die Zukunft wie der Verlust von Kunden sowie eine Rufschädigung mit sich bringen [34][50]. Diese Folgen sind schwer kalkulierbar und vermutlich noch größer als die auferlegten direkten Strafen und sollten daher unter allen Umständen vermieden werden [14]. Zusammenfassend lässt sich sagen, dass die Aufgabe von Data Governance im Bezug auf Risikominimierung die Schaffung eines robusten Rahmens für die Erkundung, Einhaltung und Überwachung gesetzlicher Vorgaben und Gesetzen ist. Mit klaren Prozessen soll der transparente Umgang mit Daten sichergestellt und nachvollziehbar werden sowie

finanzielle und gesellschaftliche Konsequenzen minimiert werden. Durch einen proaktiven Umgang mit Datenschutz können Unternehmen sich an ändernde Regularien schnell anpassen [38].

### 2.1.3 Kleine- und Mittelständische Unternehmen

Um die spezifischen Anforderungen und Bedingungen von KMUs im Bereich Data Governance zu verstehen, ist es zunächst wichtig, den Begriff KMU zu definieren und ihre Bedeutung in der Geschäftswelt zu verdeutlichen. Angesichts der bedeutenden Rolle von KMUs auf dem europäischen Wirtschaftsmarkt hat die Europäische Kommission eine Definition für diese Unternehmen veröffentlicht. Der EU nach handelt es sich bei einem Unternehmen um ein KMU, wenn dieses über weniger als 250 Mitarbeiter verfügt und einen Jahresumsatz kleiner 50 Millionen Euro oder eine Jahresbilanzsumme kleiner 43 Millionen Euro besitzt [19]. Dabei ist die Mitarbeiterzahl ein zwingendes Kriterium für eine Einstufung. Ob ein Unternehmen den Grenzwert der Jahresbilanzsumme oder den des Jahresumsatzes herbei nimmt, spielt keine Rolle. Lediglich eines der beiden Grenzwerte muss eingehalten werden. KMUs werden je nach Größe in drei Untergruppen eingeteilt. Diese sind in Tabelle 2.1 zu sehen.

Kategorie	Mitarbeiteranzahl	Jahresumsatz	Jahresbilanzsumme
Mittelgroß	< 250	$\leq 50$ Mio. EUR	$\leq 43$ Mio. EUR
Klein	< 50	$\leq 10$ Mio. EUR	$\leq 10$ Mio. EUR
Kleinst	< 10	$\leq 2$ Mio. EUR	$\leq 2$ Mio. EUR

Tabelle 2.1: Schwellenwerte für die Einstufung von KMUs [19]

Zusätzlich berücksichtigt die Europäische Kommission für die Einteilung die wirtschaftlichen Verflechtungen eines Unternehmens. Dies stellt sicher, dass nur wirklich unabhängige und eigenständige KMUs von den spezifischen Vorteilen profitieren. Erfüllt ein Unternehmen zwar die oben genannten Kriterien der Mitarbeiterzahl sowie eines der beiden Schwellenwerte, besitzt jedoch durch Zugehörigkeit zu einem oder mehreren größeren Unternehmen Zugriff auf zusätzliche Ressourcen, wird diesem kein KMU-Status verliehen. Auf die näheren Details, die für eine Unterscheidung anhand der Verflechtungen relevant sind, wird nicht weiter eingegangen, da sie für die Arbeit nicht von Bedeutung sind. Es ist lediglich anzumerken, dass vorausgesetzt wird, dass die KMUs sich keine Expertise direkt von verwandten Unternehmen einholen können. In [19] beschreibt die

Europäische Kommission KMUs als das Rückgrat des europäischen Marktes. Über 99% aller Unternehmen innerhalb der EU lassen sich diesen zuordnen. Über 100 Millionen Menschen arbeiten in KMUs, was insgesamt zwei Drittel aller Beschäftigten ausmacht. Dabei tragen sie zu mehr als der Hälfte des Bruttoinlandsprodukts der EU bei. Diese Zahlen verdeutlichen, wie essentiell KMUs für Wachstum, Innovation und Beschäftigung innerhalb der Europäischen Union sind und wieso es relevant ist, dass der Fokus der Forschung bei der Erstellung von Strategien, Vorgaben und Richtlinien sich vermehrt auf diese Gruppe verschiebt. Dies ist insbesondere von Bedeutung, da KMUs vor einer Vielzahl an Herausforderungen stehen, die eine Adaption zu neuen Trends und Best-Practices erschwert. Auf diese und ihre Rolle im Bezug auf die Einführung einer Data Governance Strategie wird in Unterabschnitt 2.4.1 eingegangen.

## 2.2 Komponenten von Data Governance

Ähnlich wie bei den Definitionen von Data Governance aus Unterabschnitt 2.1.1 unterscheiden sich bestehende Frameworks in der konkreten Formulierung. Im Kern jedoch bestehen sie in der Regel aus den gleichen Komponenten, die das Grundgerüst für die verschiedenen Ausprägungen der Frameworks bilden. Für ein besseres Verständnis werden deswegen im folgenden Abschnitt diese Bestandteile definiert.

### 2.2.1 Standards und Richtlinien

Das allgemeine Ziel von Standards ist die Vereinheitlichung von Prozessen und Methoden in allen Bereichen eines Unternehmens. Die Prozesse und Methoden sollen dabei einem zuvor erforschten und optimierten Rahmen folgen, der Konsistenz und Qualität sicherstellt. Bezogen auf den Umgang mit Daten bedeutet dies, dass nicht jeder Nutzer der Daten alle Best-Practices der Datenverarbeitung und -speicherung kennen muss, sondern dass sich ein geeignetes Gremium, bestehend aus Experten aller Bereiche, damit befasst und sicherstellt, dass der Umgang mit Daten im Einklang mit branchenspezifischen Vorgaben und der Unternehmensstrategie steht [34][41][60]. In einem datengetriebenen Unternehmen gibt es dazu verschiedene Einsatzmöglichkeiten von Standards. Beispiele sind die Standardisierung von Datenformaten und Datenelementen. So kann beispielsweise klar geregelt werden, in welchem Format Telefonnummern oder Adressen gespeichert werden müssen. Diese Standardisierung ermöglicht eine Überprüfung der Datenqualität,

da nun ein Soll-Zustand gegeben ist, an dem sich orientiert werden kann. Dabei kann die Strenge der Einhaltung in verschiedenen Bereichen eines Unternehmens unterschiedlich sein. So kann eine falsche Adresse im Controlling mehr Schaden anrichten als in der Werbeabteilung. Wie hoch die Anforderungen an die Datenqualität in den Unternehmensbereichen sind, sollte ebenfalls durch Standards definiert werden. Weitere Metriken zur Feststellung der Qualität der Daten sind Genauigkeit, Vollständigkeit, Konsistenz, Relevanz sowie Zugänglichkeit [3][9]. Eine weitere wichtige Anwendung ist die Standardisierung der Terminologie innerhalb eines Unternehmens. Entscheidungen in verschiedenen Abteilungen basieren häufig auf den gleichen Daten. Diese werden jedoch häufig unterschiedlich genannt. In den Einträgen einer Abteilung ist der Jahresumsatz als *Umsatz* aufgelistet, in dem Bericht eines Analysten steht jedoch *Einnahmen*. Dies Unklarheiten können zu Kommunikationsproblemen und ineffizienten Entscheidungsfindungen führen [60]. Für eine stärkere Kooperationen und Zusammenarbeit zwischen den Abteilungen ist daher die Einführung eines allgemeingültigen Geschäftsvokabulares notwendig [50][4]. In dieser *Glossar* genannten Sammlung sind alle relevanten Begriffe aufgelistet und näher definiert [50][9]. Es handelt sich also dabei um eine Standardisierung der Bezeichnungen für Begriffe im Kontext des Unternehmens. Dieses Glossar sollte allen zugänglich sein und befindet sich in der Regel in einem Datenkatalog (siehe Kapitel 3). Neben Standards, die Vorgaben häufig auf technischer und operationaler Ebene festlegen, legen Richtlinien oder sogenannte *Policies* allgemeine Ziele und Prinzipien fest. Diese sollen das *was - also was man tun sollte* und *was man nicht tun sollte* im Umgang mit Daten - definieren während Standards und Prozesse das *wie sollte man es tun* beschreiben [34]. Ein Beispiel hierfür könnte das Ziel des Unternehmens sein, Daten zu schützen und sensitive Informationen wirklich nur bei konkretem Bedarf einzusetzen. Eine Richtlinie hierfür könnte lauten: *Alle sensitiven, personenbezogenen Daten müssen geschützt und vertraulich behandelt werden*. Daraus lassen sich Standards ableiten wie die Beschränkung von Zugriffen auf sensitive Daten für wenige Personen oder das Anonymisieren dieser Daten bei Analysen und Aufgaben, für die diese keine Rolle spielen. So sollten Daten für Datenanalysen im Marketingbereich, bei denen es um die Beobachtung von allgemeinen Verhaltensweisen geht, keine dafür irrelevanten personenbezogenen Informationen beinhalten.

### 2.2.2 Prozesse

Die International Organization for Standardization (ISO) definiert einen Prozess als eine Reihe von miteinander verbundenen oder interagierenden Aktivitäten, die Eingaben in

Ausgaben umwandeln [35]. Im Kontext von Data Governance sind das standardisierte, dokumentierte und wiederholbare Methoden, die das Ziel haben, die definierten Richtlinien und Standards umzusetzen [39][2]. Auf die Rolle der Standardisierung wurde in Unterabschnitt 2.2.1 eingegangen. Eine klare Dokumentation ist wichtig, um Entscheidungen zu einem späteren Zeitpunkt nachvollziehen zu können. Wiederholbarkeit bedeutet, dass durch klare Vorgaben, Richtlinien und Verantwortlichkeiten jeder Prozess bei Wiederholung unter gleichen Bedingungen das gleiche Ergebnis bzw. die gleiche Ausgabe liefert. Die Prozesse sind dabei vielfältig und existieren in allen Bereichen des Lebenszyklus des Data Governance Programmes. Hierbei ist nochmal hervorzuheben, dass Data Governance kein abschließbares Projekt ist, sondern es sich vielmehr um einen kontinuierlichen Vorgang handelt [15][34]. Dieser bedarf einer ständigen Überwachung und Anpassung, um die gesteckten Ziele zu erreichen [2]. Beispiele für geregelte Prozesse sind die Erstellung von Datenrichtlinien sowie die Überwachung ihrer Einhaltung. Außerdem sollten bereits zu Beginn Abläufe festgelegt werden, wie der Fortschritt und die implementierten Verbesserungen des aktuellen Programmes gemessen werden können [2]. Dafür müssen Reifegradmodelle und Metriken definiert werden [40][57]. Insgesamt lässt sich sagen, dass Prozesse in Data Governance Frameworks sicherstellen, dass jeder Schritt anhand von klaren Vorgaben durchgeführt wird und dieser durch geeignete Dokumentation nachvollziehbar und wiederholbar ist.

### **2.2.3 Verantwortlichkeiten und Rollen**

Eine, wenn nicht sogar die Hauptaufgabe von Data Governance ist die Verteilung von klaren Verantwortlichkeiten auf verschiedene Gremien, Rollen und Personen. In den verschiedenen Definitionen aus Unterabschnitt 2.1.1 sind die Begriffe Verantwortlichkeiten und Entscheidungsrechte wesentliche Bestandteile davon. Unklarheiten in diesem Gebiet können weitreichende Folgen mit sich bringen. Diese können beispielsweise entstehen, da Datenflüsse häufig nicht der Logik des Unternehmens folgen. Dies kann zu Daten-Silos, Duplikaten sowie zu fehlender Kontrolle über den gesamten Lebenszyklus von Daten führen [37]. Bei auftretenden Problemen oder benötigten Entscheidungen fühlt sich keine der Parteien verantwortlich, diese zu beheben bzw. zu treffen. Auch ist die Überwachung der Datenqualität dadurch erschwert [37]. Die Einführung von Verantwortlichkeiten gibt klare Strukturen vor, wer welche Entscheidungen treffen soll. Data Governance Themen finden täglich Anwendung im operativen Geschäft und können nicht von wenigen Personen erstellt und überwacht werden [60]. Daher kann man den Aufbau der Verantwortlichkeiten

auch als ein *Teile-und-Herrsche-System* sehen [50]. In der Regel hat ein Data Governance Programm unterschiedliche Schichten, um Probleme auf unterschiedlichen Ebenen lösen zu können [34]. Der konkrete Aufbau und der Fokus des Entscheidungssystems unterscheidet sich von Unternehmen zu Unternehmen aufgrund unterschiedlicher Strategien, Kulturen, auftretenden Problemen sowie Reifegrades bisheriger Programme [34], jedoch haben die meisten die gleichen Rollen, die im Folgenden näher erläutert werden.

**Data Governance Steering Committee** Dieses Gremium ist in größeren Unternehmen die höchste Autorität für Data Governance innerhalb einer Organisation. Dabei ist es für die Beaufsichtigung und Finanzierung der Aktivitäten verantwortlich und kein Teil des Entwicklungsprozesses von konkreten Richtlinien, Standards oder Prozessen [34]. Das Komitee erhält Vorschläge für Data Governance Aktivitäten vom Data Governance Council (DGC) und dem Chief Data Officer (CDO). Basierend auf diesen Empfehlungen stellen sie Ressourcen zur Verfügung, damit die Initiativen umgesetzt werden können. Es besteht aus einer funktionsübergreifenden Gruppe von Führungskräften aus der C-Suite [34].

**Data Governance Council** Übersetzt mit Data Governance Rat ist der DGC ein hierarchieübergreifendes und funktionsübergreifendes Gremium, dass für die strategische Ausrichtung des gesamten Data Governance Programms verantwortlich ist [59][9]. Alle Richtlinien, Metriken und Standards werden hier verabschiedet und priorisiert [34]. Außerdem überwacht der Rat den Fortschritt des Programmes und stellt sicher, dass alle Vorgaben und Aktivitäten korrekt umgesetzt werden [1]. Zusätzlich behandelt er alle Fälle, die nicht auf operativer Ebene geregelt werden konnten und nach oben eskaliert wurden [34][9]. Üblicherweise besteht er aus leitenden Datenverantwortlichen - *Senior Data Stakeholdern*, dem CDO, Data Ownern von wichtigen Datendomänen sowie, falls existent, dem Chief Information Officer (CIO) [9].

**Chief Data Owner** Der CDO ist eine Führungskraft aus der C-Suite, der für die gesamte Umsetzung des Data Governance Programmes verantwortlich ist [9]. Die Person versteht die Strategie des Unternehmens und konzentriert sich darauf, diese mit Daten zu unterstützen und fördern. Außerdem besitzt der CDO die finanziellen Möglichkeiten, Data Governance Initiativen und Aktivitäten zu sponsorn und voranzutreiben [1].

**Data Governance Office** Diese Arbeitsgruppe übernimmt die Verwaltungsaufgaben auf operativer Ebene und beschäftigt sich mit der Einhaltung von Standards und Richtlinien im täglichen Geschäft [34]. Es dient dabei als Stütze und Anlaufstelle bei Fragen für die täglichen Aktivitäten von Data Stewards und Data Owner und koordiniert die Kommunikationskanäle zwischen verantwortlichen Personen und Rollen [22]. Außerdem informiert das Data Governance Office (DGO) den DGC über die Fortschritte des Programmes [50]. Es besteht aus den Data Owner, Data Stewards sowie Data Custodians [34].

**Data Owner** Das Konzept, dass Personen Besitzer von Daten in einer Domäne werden, ist ein zentraler Bestandteil von Data Governance. Dadurch trägt eine Person die ultimative Verantwortung über die Daten sowie ihren Inhalt und ihre Qualität [9]. Sie müssen sicherstellen, dass Nutzer dieser Daten vorgeschriebene Richtlinien und Standards einhalten. Außerdem entscheiden sie, wer Zugriff auf die Daten erhält und wie sie genutzt werden dürfen [50]. In den meisten Fällen handelt es sich bei Data Owner um Führungskräfte aus der C-Suite oder Manager der Fachabteilung, die die anfallenden Aufgaben im täglichen Geschäft an andere Personen, den sogenannten Data Stewards, delegieren. Insgesamt sind durch die klare Zuordnung der Daten zu einem Besitzer die Verantwortlichkeiten bei auftretenden Problemen und Fragen eindeutig geregelt.

**Data Steward** Diese Person verwaltet die Daten im Auftrag eines oder mehrerer Data Owner [48]. Sie sorgen mit ihrer Arbeit im täglichen, operativen Geschäft für die Umsetzung der Ziele und Strategien, die vom Data Owner spezifiziert wurden [9]. Das beinhaltet die Anwendung von Richtlinien und Standards auf die Daten, um die ausgegebene Datenqualität zu erreichen, für das sie auch rechenschaftspflichtig sind [34]. Sie bringen sowohl großes Domänenfachwissen als auch technisches Fachwissen mit, da sie die High-Level Vorgaben und Ansprüche in technische Spezifikationen überführen müssen [16][34]. Sie haben einen festgelegten Zeitanteil, den sie für die datenbezogenen Aufgaben aufwenden [34]. Je nach Umfang können dies 20% bis 30% ihrer Arbeitszeit sein [9]. Häufig werden Data Stewards in Business Data Stewards und Technical Data Stewards oder auch Data Custodian unterschieden [1]. Business Data Stewards sind dabei Domänenexperten, die sich insbesondere mit dem zugrunde liegenden Kontext der Daten und ihren Einsatzgebieten auskennen. Sie arbeiten mit den technisch ausgerichteten Data Custodians zusammen, die in der Regel aus dem Umfeld der IT stammen und für die technische Machbarkeit verantwortlich sind [34].

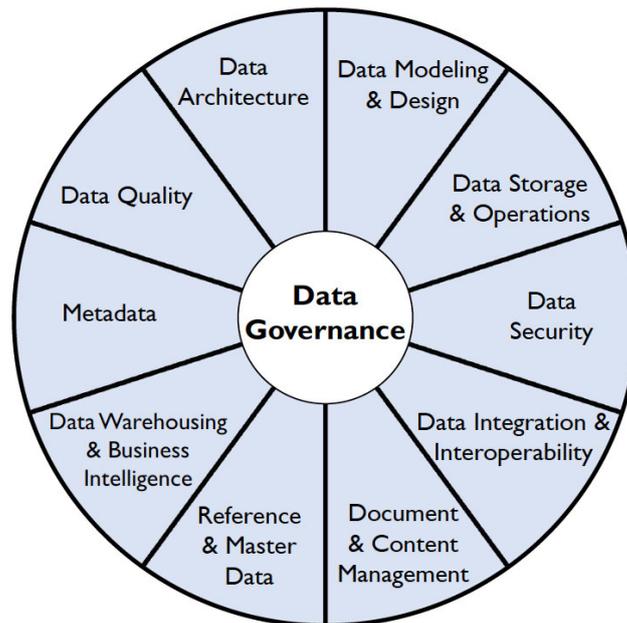


Abbildung 2.2: DAMA DMBoK Data Management Framework [34]

## 2.3 Überblick über bestehende Frameworks

Um einen Überblick darüber zu gewinnen, wie Data Governance Frameworks aufgebaut sein können, werden im folgenden Abschnitt etablierte Frameworks aus verschiedenen Bereichen vorgestellt. Zu den untersuchten Modellen gehören das DAMA DMBoK Framework, das DGI Data Governance Framework sowie das praxisorientierte Modell von PricewaterhouseCoopers (PwC).

### 2.3.1 DAMA-DMBoK

Das DAMA DMBoK Data Management Framework ist ein umfassender Leitfaden, der einen ganzheitlichen Ansatz für den Umgang mit Daten bietet. Ziel von DAMA ist es, eine standardisierte Sprache und klare Richtlinien für das Datenmanagement zu schaffen. Dies ermöglicht einen professionellen und einheitlichen Austausch zwischen den Anwendern innerhalb eines Unternehmens sowie in der Forschung und Literatur. Das Framework gliedert die Aspekte des Datenmanagements in elf Schlüsselbereiche, wobei Data Governance als zentrales Bindeglied dient. Dies ist in dem *DAMA Rad der Data Governance* in Abbildung 2.2 zu sehen.

Die gezeigten Bereiche stehen in engem Zusammenhang zueinander, wobei Data Governance die übergreifende Strategie definiert. Im Folgenden erfolgt eine kurze Erläuterung der elf Bereiche, bevor der Fokus stärker auf den Schlüsselbereich Data Governance gelegt wird.

**Data Governance** Wie in der Definition von DAMA bereits beschrieben, ist Data Governance verantwortlich für die Einführung und Pflege von Richtlinien, Standards und Prozessen für die Verwaltung von Daten innerhalb einer Organisation. Dies beinhaltet die Verteilung von Rollen und Prozessen, das Sicherstellen einer guten Datenqualität sowie die Förderung von klaren Verantwortlichkeiten für alle Datenbestände.

**Datenarchitektur** Diese beschreibt die logische und physische Struktur der Datenbestände des Unternehmens. Die Datenarchitektur ist dabei ein wesentlicher Bestandteil für die Gesamtstrategie und Kontrolle der Daten. Wichtig dabei ist, dass der Architekt eng mit den Business Data Stewards zusammenarbeitet, damit die Datenarchitektur den Anforderungen und der Ausrichtung des Unternehmens gerecht wird. Es besteht dabei aus Artefakten und Dokumenten auf verschiedenen Abstraktionsebenen, um Modelle und Konzepte für unterschiedliche Zielgruppen verständlich zu machen. Diese Dokumente und Artefakte beinhalten Spezifizierungen über den aktuellen Zustand, definieren Anforderungen an Daten sowie Datenflüsse und geben Richtlinien für die Datenintegration vor. Die Datenarchitektur stellt sicher, dass Datenstrategien und Governance-Aktivitäten effektiv umgesetzt werden können, indem sie klare Vorgaben für die Strukturierung von Datenflüssen, Datennutzung und Datenspeicherung macht.

**Datenmodellierung und -design** Dieser Bereich umfasst die Erstellung und Verwaltung von Datenmodellen, die die Struktur, die Beziehungen und Beschränkungen von Daten innerhalb des Unternehmens darstellen. Dadurch sollen Unternehmen verstehen, wie unterschiedliche Daten zusammenpassen und effektiv genutzt werden können. Die erstellten Datenmodelle sollen dabei helfen, Anforderungen an Daten zu verstehen und zu dokumentieren. Zusätzlich unterstützt es bei dem Design von effektiven Datenspeichern und der Kommunikation zwischen geschäftlichen und technischen Teams. Insgesamt gewährleistet die Modellierung der Daten die Integrität und Qualität der Daten, erleichtert die Datenintegration und fördert die Wiederverwendbarkeit der Daten.

**Datenspeicherung und -betrieb** Datenspeicherung und -betrieb beschreibt das Design, die Umsetzung und die Pflege der Datenspeicher. Es wird das Ziel verfolgt, den Mehrwert der Daten über die gesamte Lebensspanne, also von Erstellung bis hin zur Beseitigung, zu maximieren. Konkrete Aufgaben umfassen dabei die Wahl geeigneter Speichertechnologien, die Umsetzung vorgegebener Datenstrukturen und die Verwaltung von Datenzugriffen. Insgesamt soll dieser Bereich sicherstellen, dass Daten sicher und effizient gespeichert und zugänglich gemacht werden.

**Datensicherheit** Der Bereich Datensicherheit fokussiert sich auf einen sicheren Datenzugriff und verfolgt das Ziel, Informationen und Daten im Einklang mit Datenschutzvorschriften und Geschäftsanforderungen zu schützen. Dazu können Umsetzungen von Zugriffskontrollen, Authentifizierungsmechanismen, Verschlüsselungen und Datenmaskierung zum Schutz sensibler Daten sowie regelmäßige Sicherheitsprüfungen und Audits gehören. Dadurch soll die Vertraulichkeit und Integrität der Daten sichergestellt und das Risiko und potenzielle Konsequenzen eines Datenlecks verringert werden.

**Datenintegration und Interoperabilität** Dieser Bereich umfasst das Design von Prozessen, die bei der Bewegung und Zusammenführung von Daten innerhalb und zwischen Datenspeichern, Anwendungen und Organisationen, ausgeführt werden. Dies beinhaltet in der Regel Vorgaben und Techniken für die Integration und Transformation von Daten sowie eine Standardisierung von Datenformaten. Dadurch soll gewährleistet werden, dass Daten aus unterschiedlichen Datenquellen kombiniert oder zusammengeführt werden können und dadurch effizient innerhalb der Organisation genutzt werden.

**Dokumenten- und Inhaltverwaltung** Hiermit ist die Kontrolle der Erfassung, Speicherung, Zugriffs und Nutzung von Daten und Informationen gemeint, die nicht in relationalen Datenbanken gespeichert sind. Beispiele hierfür sind unstrukturierte Daten wie Dokumente, Multimedia-Inhalte oder Log-Dateien. Ziel ist es, dass diese Daten ebenfalls konsistent und leicht zugänglich sind. Konkrete Aufgaben umfassen die Implementierung von Managementsystemen, die Definition von Metadatenstandards, das Erschaffen von Versionierungskontrollen sowie Strategien zur langfristigen Aufbewahrung und Archivierung der Daten. Insgesamt stellt eine effektive Dokumenten- und Inhaltsverwaltung die Einhaltung gesetzlicher Anforderung sicher und erhöht die Zusammenarbeit innerhalb des Unternehmens.

**Referenz- und Stammdaten** Dieser Bereich beschäftigt sich mit der Verwaltung und Pflege der wichtigsten Daten, die abteilungsübergreifend gemeinsam genutzt werden. Zu diesen Stammdaten gehören beispielsweise Kundendaten, einheitliche Bezeichnungen und Codes sowie Teilelisten. Durch die ständige Weiterentwicklung der Daten und Systeme können sich Inkonsistenzen der Datenstrukturen und Datenwerte zwischen den Systemen entwickeln. Das Aufbauen von Referenz- und Stammdaten ist aus diesem Grund unvermeidbar. Da diese Daten von vielen Parteien genutzt werden, sind eine klare Verwaltung und Zuordnung von Verantwortlichkeiten erforderlich.

**Data Warehousing und Business Intelligence** Business Intelligence (BI) beschreibt das Konzept, dass Unternehmen durch die Analyse der eigenen Daten Einsichten über ihre Produkte, Dienstleistungen und Kunden erlangen und dadurch bessere Entscheidungen treffen können. Ein Data Warehouse ist ein Datenspeicher, der aus Daten aus verschiedenen Quellen besteht, die in ein gemeinsames Datenmodell überführt wurden. Es unterstützt BI, da es als zentraler Datenspeicher dient, welcher effiziente Abfragen und Untersuchungen ermöglicht.

**Metadaten** Metadaten enthalten dabei Informationen über die Erstellung, das Format, den Namen, Merkmale, die Herkunft und vieles mehr. Diese sind von essentieller Bedeutung, um relevante Inhalte zu finden, wichtige Informationen darüber zu erhalten und diese mit anderen Nutzern zu teilen. Metadaten werden auch als *Daten über Daten* bezeichnet [66].

**Datenqualität** Die anderen Bereiche zielen alle darauf ab, dass die Daten leicht zugänglich sind, sicher gespeichert und einfach verteilt werden können sowie den Geschäfts- und regulatorischen Ansprüchen entsprechen. Damit aus den Daten jedoch richtige Entscheidungen abgeleitet werden können, müssen sie von hoher Qualität sein. Dafür müssen Prozesse und Maßnahmen entwickelt werden, die Dinge wie Genauigkeit, Vollständigkeit, Konsistenz und zeitliche Relevanz messen und sicherstellen. Dazu können das Definieren von Datenqualitätsstandards, regelmäßige Überprüfungen, Datenbereinigung sowie Werkzeuge für die Überwachung und Darstellung der Datenqualität gehören. Governance ist hier von entscheidender Bedeutung, da klare Verantwortlichkeiten für die Pflege und Überwachung notwendig sind.

Alle Bereiche werden von DAMA in einem SIPOC-Diagramm (suppliers, inputs, process, outputs, customers) dargestellt. DAMA gliedert das SIPOC-Diagramm über Data Governance, gezeigt in Abbildung A.1, in drei Spalten, wobei oben die Objekte und Aktivitäten stehen, darunter involvierte Personen, Rollen und Gremien und ganz unten technische Unterstützer und Förderer. Zu allererst werden die Eingaben und Einflüsse für Data Governance aufgeführt. Dazu gehören Geschäftsstrategien und -ziele, Vorschriften innerhalb eines Unternehmens, gesetzliche Regularien und die interne Geschäftskultur. Involvierte Personen und Rollen, die in diesem Fall als Zulieferer bezeichnet werden, umfassen unter anderem Personen aus der exekutiven Ebene, Data Stewards und Data Owner, Gesetzgeber sowie Mitarbeiter. Hierbei unterstützen können klare Kommunikation und Botschaften sowie gute Kontakte, die den Austausch und die Zusammenarbeit zwischen den Beteiligten erleichtern.

Der Kern von Data Governance besteht in den Aktivitäten, die für eine erfolgreiche Umsetzung notwendig sind. Zunächst muss ein dazu generelles Vorgehen für Data Governance innerhalb des Unternehmens entworfen werden. Dies umfasst die Strategieentwicklung, die Messung der Bereitschaft für das Umsetzen, die Sicherstellung der Übereinstimmung der Data Governance Strategie mit den Geschäftszielen sowie der Aufbau organisatorischer Kontaktpunkte innerhalb des Unternehmens für eine effektive Kommunikation und Zusammenarbeit im Bereich Data Governance. Nach diesen Planungsaktivitäten folgt deren Implementierung. Dazu gehören die Definition und Förderung von Datenstandards und -verfahren, die Erstellung eines Business Glossars, die Koordination von Gremien und die Unterstützung aller Prozesse bei der Einordnung von Daten als Vermögenswert. Beteiligte Rollen dabei sind verschiedene Gremien, der CDO, führende Data Stewards, Compliance Teams, Software- und Datenarchitekten sowie Experten der Datendomänen. Technische Unterstützer dabei können Websites, Business Glossar und Arbeitsfluss-Werkzeuge sowie Data Governance Bewertungskarten sein.

Die Aktivitäten liefern verschiedene Ergebnisse und Ausgaben. Diese umfassen die generelle Data Governance Strategie und Roadmap, Datenprinzipien sowie Richtlinien und Prozesse, das operative Framework, Business Glossare und definierte Metriken und Eskalationspläne. Bei den involvierten Personen bzw. Konsumenten dieser Ausgaben handelt es sich unter anderem um unterschiedliche Data Governance Rollen, Projekt-Manager, Compliance-Teams, Mitarbeiter, das Management sowie Partnerorganisationen. Technische Unterstützungen können verschieden Metriken wie die Einhaltung regulatorischer und interner Datenrichtlinien, messbarer Wert, Effektivität und die Nachhaltigkeit sein.

Damit Unternehmen ihre Entwicklung und Effizienz im Umgang mit Daten stets im Blick haben und verbessern können, richtet DAMA den Fokus auf ein Reifegradmodell. Dieses Modell unterstützt die Unternehmen dabei, den Grad ihrer Prozesse und Aktivitäten im Bereich Data Governance und Datenmanagement zu bewerten und kontinuierlich weiterzuentwickeln. Es besteht aus sechs Stufen, die den Fortschritt von Prozessmerkmalen beschreiben. Mit Erreichen jeder neuen Stufe werden die Prozesse konsistenter, vorhersehbarer und zuverlässiger.

Das DAMA DMBOK Framework verfolgt einen ganzheitlichen Ansatz für die Verwaltung von Daten. Data Governance spielt die zentrale Rolle dabei und hat Wechselwirkungen mit den anderen zehn Bereichen. Es fungiert als Bindeglied zwischen diesen und es ist nur schwer möglich, Data Governance isoliert zu betrachten. Insgesamt ist das Framework umfangreich und strukturiert und liefert eine einheitliche Sprache sowie klare Richtlinien für die Verwaltung und Nutzung von Daten innerhalb eines Unternehmens.

### 2.3.2 Data Governance Institute

Das DGI Data Governance Framework ist ein logisches Modell für die Klassifizierung, Organisation und Kommunikation komplexer Aktivitäten, die für die Entscheidungsfindung auf der Grundlage von Unternehmensdaten erforderlich sind. Es kann sowohl gänzlich von Unternehmen übernommen werden, als auch an ihre Bedürfnisse angepasst werden. Ziel des Frameworks ist es ebenfalls, den Nutzen von Daten zu maximieren und Risiken zu minimieren. Das DGI ist der Ansicht, dass in der heutigen, dynamischen und digitalen Welt jedes Unternehmen und jede Organisation in gewissem Umfang ein Data Governance Framework benötigt, um den immer strengeren Vorschriften und Gesetzen gerecht zu werden. Das gesamte Programm ist wertbasiert, d.h., der Fokus liegt auf den Vorteilen und Ausgaben, die dem Unternehmen einen Nutzen bringen. Alle anderen definierten Aktivitäten sind darauf ausgerichtet, diese Werte zu erreichen. Eine Übersicht über das Framework ist in Abbildung 2.3 zu sehen.

Wichtig für das Programm ist, dass die Ausgaben in messbaren Ergebnisgrößen definiert werden, so dass die Wirkung des Modells kontinuierlich überprüft werden kann. Diese quantifizierbaren Ausgaben werden dabei in *kleines g* und *Großes G* klassifiziert. Die *kleinen g's*, also kleineren Governance Aktivitäten, umfassen Produkte und Ausgaben, die sich aus konkreten datenbasierten Aktivitäten ergeben. Dazu gehören Datenprodukte

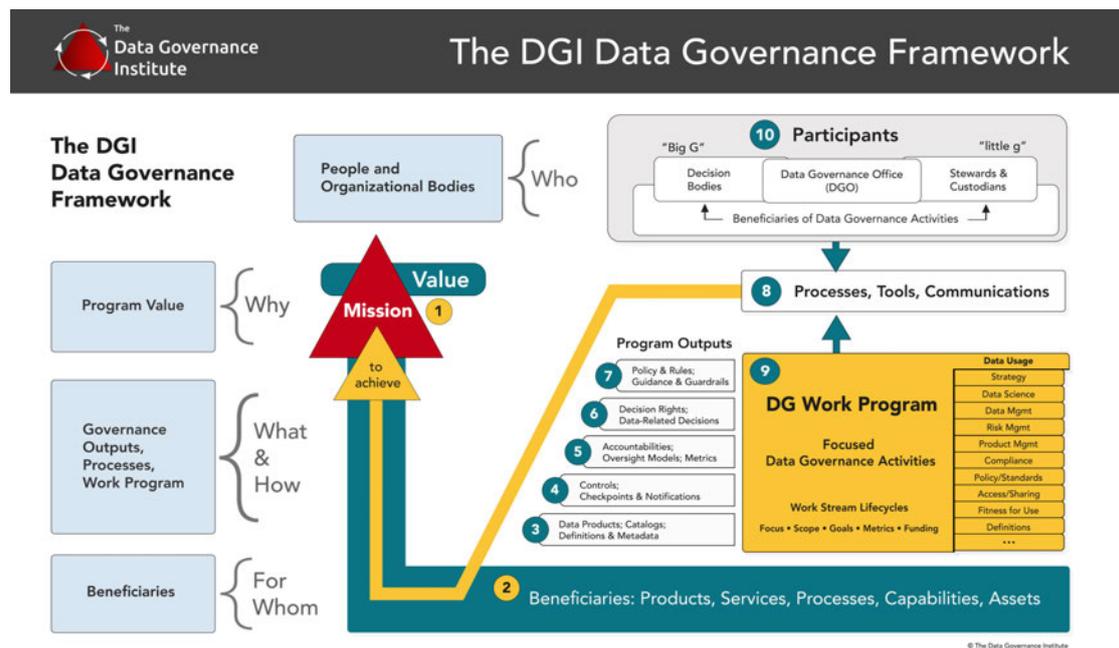


Abbildung 2.3: Framework des Data Governance Institute [22]

wie Datenkataloge, Datendefinitionen, Metadaten, Messungen und Metriken wie die Datenqualität. Die großen Governance Aktivitäten betreffen Unternehmensstrukturen und zielen darauf ab, die Effizienz und Effektivität im Umgang mit Daten zu erhöhen. Ergebnisse daraus sind Entscheidungsrechte über Daten, Verantwortlichkeiten für die Qualität, Zugriff und Nutzen von Daten sowie Richtlinien und Regeln.

Insgesamt setzt sich das Framework aus zehn Komponenten zusammen, die je nach Fokus und den individuellen Rahmenbedingungen der Organisation unterschiedlich ausgeprägt sind. Im Folgenden werden die zehn Komponenten kurz erläutert. Die Reihenfolge der Bestandteile orientiert sich an der Sichtweise des DGI auf das Framework. Das DGI fragt sich, *warum* ein Programm existiert, *was* es macht, *wer* daran beteiligt ist und *wie* die Prozesse ablaufen. Das Ziel von Data Governance ist es offensichtlich, einen Mehrwert für das Unternehmen zu schaffen, wie z.B. die Verbesserung von Produkten und Dienstleistungen, die Erhöhung der Datenqualität und die Reduzierung von Risiken und Kosten. Diese vagen Aussagen reichen jedoch nicht aus, um als Grundlage für die Einführung von Data Governance zu dienen. Vielmehr müssen konkrete Werte formuliert werden, die auch die unternehmensspezifischen Gegebenheiten und Probleme berücksichtigen, die verbessert oder beseitigt werden sollen. Das Unternehmen bzw. die Programmverantwort-

lichen formulieren eine *Mission und konkrete Werte* und geben allen den Auftrag, diese Werte zu erreichen. Die zweite Komponente konzentriert sich auf die *Begünstigten des Data Governance-Programms* und komplementiert das *warum* eines Data Governance Frameworks. Dabei kann es sich sowohl um interne Stakeholder handeln, die von einer höheren Datenqualität oder daraus resultierenden Produkten wie einem Datenkatalog profitieren, als auch um externe Stakeholder wie Kunden oder Aufsichtsbehörden. Der Nutzen bezieht sich jedoch nicht nur auf die Daten und die Datenqualität, da Personen auf Managementebene von klaren Richtlinien in Entscheidungsprozessen profitieren oder auch Entwickler durch Standards und Prozessspezifikationen einen Leitfaden erhalten, den sie sonst selbst entwickeln müssten. Häufig sind *Datenprodukte* ein Hilfsmittel, um die Ziele der Data Governance Aktivitäten zu erreichen. Dazu gehören z.B. Datenkataloge, Self-Service Portale sowie Dashboards, in denen qualitativ hochwertige Daten aus unterschiedlichen Quellen bereitgestellt oder visualisiert werden. Mit Hilfe von Metadaten werden die Qualität, die enthaltenen Informationen sowie die Herkunft der Daten beschrieben. Ein Glossar sorgt für eine einheitliche Sprache. Insgesamt soll so eine Demokratisierung der Daten erreicht und den Nutzern Zeit bei der Suche nach relevanten Daten erspart werden. *Kontrollmechanismen* sind ebenfalls ein wichtiger Bestandteil erfolgreicher Data Governance Programme. Der Besitz von Daten birgt auch Risiken, denn Datenlecks sowie die Missachtung von Vorschriften und Gesetzen können weitreichende Folgen haben. Mechanismen müssen über den gesamten Lebenszyklus der Daten vorhanden sein. Diese können präventiv oder investigativ sowie automatisiert oder manuell sein. Beispiele sind Zugriffskontrollen, Schulungen sowie die Verschlüsselung von Daten. Die Bedeutung von *Verantwortlichkeiten* und *Entscheidungsrechten* sowie die verschiedenen Rollen in einem Data Governance Programm wurden bereits in Unterabschnitt 2.2.3 erläutert. Die Ansichten von DGI zu diesen Themen unterscheiden sich nicht zu den dort formulierten Aussagen. Ein weiteres Ergebnis eines Data Governance Programms sind *Richtlinien und Regeln*. Diese beantworten alle datenbezogenen Themen sowohl auf höherer Ebene wie generelle Strategie oder Vorgehen als auch auf detaillierter und technischer Ebene wie Datenstandards oder Namenskonventionen. Dies ist in Unterabschnitt 2.2.1 thematisiert worden. Die Komponenten Datenprodukte, Kontrollmechanismen, Verantwortlichkeiten, Entscheidungsrechte sowie Richtlinien und Regeln sind die Ergebnisse des Data Governance Programms. Diese dienen als Grundlage und Leitfaden für konkrete Handlungen im operativen Geschäft. *Prozesse, Werkzeuge und Kommunikation* beschreiben, wie Data Governance ablaufen soll. Wie in Unterabschnitt 2.2.2 beschrieben, sollen Prozesse verständlich, dokumentiert und wiederholbar sein. Dadurch soll das Einhalten von Datenschutz-, Datenqualitäts- und Sicherheitsanforderungen gewährleistet sein.

Zwar sind Data Governance Programme häufig technologieneutral, die Ergebnisse jedoch hängen oft von spezialisierten Werkzeugen ab. Beispiele hierfür sind Projektmanagement-Software, BI-Tools sowie Datenkataloge mit Zugriffskontrollen und Glossaren. Durch die große Anzahl an Personen, die in Data Governance Aktivitäten involviert sind, spielt Kommunikation eine Schlüsselrolle für den Erfolg eines Programms. Diese sollte sowohl schriftlich, visuell als auch mündlich auf hohem Niveau erfolgen und regelmäßig überprüft und geschult werden. Um den Überblick über die verschiedenen Beteiligten und die erwarteten Werte und Ergebnisse in realen Projekten nicht zu verlieren, bieten viele Data-Governance-Programme einen Leitfaden, ein *Datenarbeitsprogramm*, für die Organisation der Arbeit. Dabei handelt es sich in der Regel um einen Arbeitsablauf mit eigenem Umfang, Schwerpunkt, Messgrößen und Verantwortlichkeiten. In *Data Governance Beteiligte* werden die involvierten Gremien und Rollen wie in Unterabschnitt 2.2.3 beschrieben.

Das DGI Data Governance Framework bietet einen ganzheitlichen Ansatz für den Umgang mit Daten als Vermögenswert. Durch den wertbasierten Ansatz liegt der Fokus stets auf dem Nutzen und den Vorteilen für das Unternehmen, was sicherstellt, dass Ressourcen effizient eingesetzt werden und strategische Ziele nicht aus den Augen verloren werden. Mit seinen zehn Komponenten ist es sehr umfangreich und detailliert.

### 2.3.3 PwC

PwC ist eine weltweit tätige Gesellschaft von Wirtschaftsprüfern und Beratern [65]. Mit umfassenden Dienstleistungen in den Bereichen Wirtschaftsprüfung, Steuerberatung und Unternehmensberatung unterstützen sie Unternehmen dabei, ihre Geschäftsprozesse zu optimieren und ihre strategischen Ziele zu erreichen. Im Bereich der Data Governance hat PwC ein praxisorientiertes Framework entwickelt, das eine vielschichtige Herangehensweise an das Thema bietet. Es zeigt, wie Data Governance in bestehende Unternehmensprozesse eingebettet werden kann. Dabei wird deutlich, dass Data Governance nicht isoliert als einzelnes Projekt, das irgendwann abgeschlossen ist, betrachtet werden kann, sondern als täglicher Bestandteil der Geschäftsprozesse angesehen werden muss. Die Visualisierung des Frameworks ist in Abbildung 2.4 zu sehen.

In der obersten Schicht der Abbildung 2.4 und zum Start der Data Governance Einführung steht im Framework von PwC ebenfalls die Strategieausrichtung. Diese umfasst konkrete Geschäftsziele, die durch Data Governance erreicht werden sollen. Zudem werden dazugehörige Meilensteine festgelegt und Metriken definiert, um den Fortschritt zu

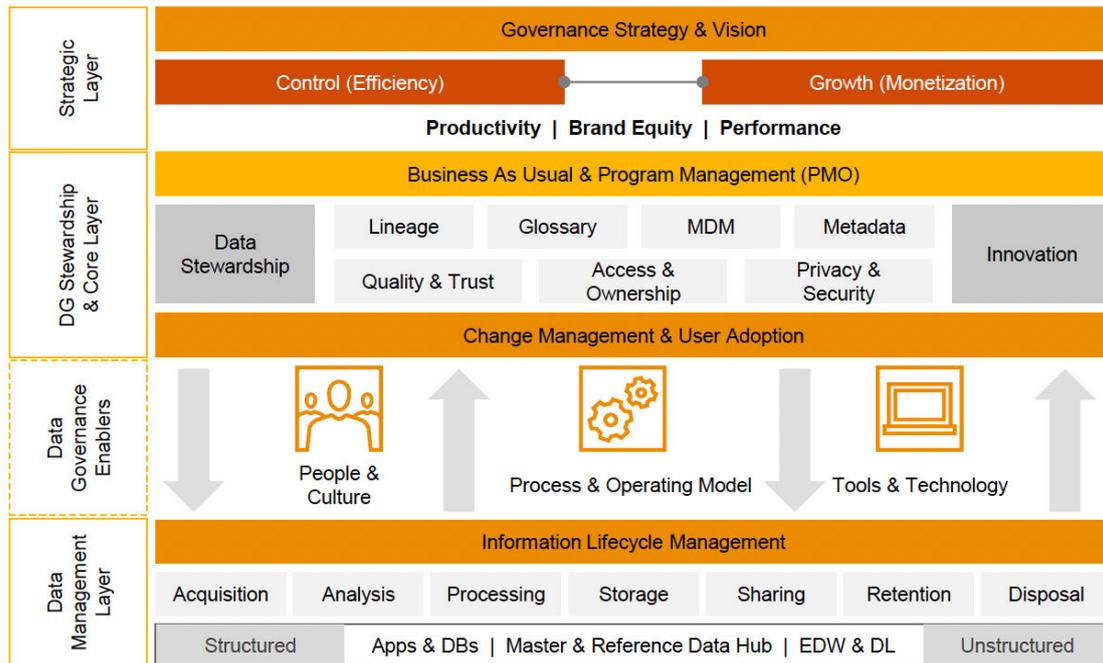


Abbildung 2.4: Framework des Beratungsunternehmens PwC [65]

überwachen. In der obersten Schicht sind ebenfalls die beiden großen Treiber von Data Governance zu sehen: Kontrolle und Wachstum. Kontrolle soll sicherstellen, dass alle Prozesse und Daten den gesetzlichen Vorschriften und internen Richtlinien entsprechen, wodurch Risiken wie Datenlecks, Datenverlust und Missbrauch von Daten minimiert werden. Durch die Kontrolle der Daten soll zudem die Effizienz der Datenverarbeitung und -nutzung erhöht und Kosten reduziert werden. Der Aspekt Wachstum zielt auf die Wertsteigerung der Daten ab. Mit einer höheren Qualität und einem höheren Vertrauen in die Daten, können neue Geschäftsfelder erschlossen und der finanzielle Wert der Daten gesteigert werden.

In der nächsten Schicht des Frameworks befinden sich die Stewardship und die Kernpunkte von Data Governance. Hier sind die wesentlichen Funktionen und Maßnahmen angesiedelt, die für eine effektive und erfolgreiche Umsetzung von Data Governance notwendig sind. Die beiden Haupttreiber sind die Data Stewardship und Innovation. Data Stewardship umfasst die verantwortliche Verwaltung und Überwachung von Daten. Innovation beschreibt die kontinuierliche Weiterentwicklung und Verbesserung der Datenverwaltung und -nutzung durch neue Technologien, Methoden und Prozesse. Dadurch soll sichergestellt werden, dass die Daten korrekt, sicher und jederzeit zugänglich sind.

Mithilfe von Data Lineage wird Klarheit über die Datenherkunft und -verwendung geschaffen. Der Weg der Daten von der Entstehung bis hin zur Archivierung soll transparent sein, was die Integrität erhöht und die Auswirkungen von Änderungen nachvollziehbar macht. Mithilfe eines Business Glossars werden Begriffe und Definitionen innerhalb des Unternehmens vereinheitlicht. Dies fördert das gemeinsame Verständnis und eine klare Kommunikation zwischen den verschiedenen Teams und Rollen. Master Data Management oder auch Stammdatenmanagement soll ebenso wie bei DAMA für konsistente, aktuelle und genaue Stammdaten sorgen, um abteilungsübergreifend auf den gleichen Daten arbeiten zu können. Auch bei PwC spielen Metadaten eine Rolle, um das Auffinden, die Verwaltung und die Nutzung der Daten zu vereinfachen. Regelmäßige und automatisierte Qualitätskontrollen fördern eine hohe Datenqualität und damit das Vertrauen in die Daten. Mit Zugriffskontrollen kann selektiert werden, welche Nutzer und Systeme auf bestimmte Daten zugreifen können. Damit wird das Risiko eines Datenmissbrauchs und eines Datenverlustes minimiert und sensible Informationen geschützt. Wie bei allen anderen Data Governance Framework sind die Zuweisung von klaren Verantwortlichkeiten entscheidend für den Erfolg. Dadurch gibt es feste Ansprechpartner, die für die Qualität und Verfügbarkeit der Daten verantwortlich sind und dafür Rechenschaft ablegen. Konkrete Maßnahmen und Kontrollen bezogen auf Datenschutz und Sicherheit stellen sicher, dass personenbezogene und sensible Daten geschützt und gesetzliche Datenschutzbestimmungen wie die DSGVO in allen Prozessen und Systemen eingehalten werden.

Ebenso von großer Bedeutung sind die Bereitschaft des Unternehmens zu Veränderungen und die Benutzerakzeptanz. Die Umsetzung der verschiedenen Maßnahmen erfordert von den Mitarbeitern und Anwendern zusätzlichen Aufwand im Arbeitsalltag und eine Einarbeitung in neue Prozesse und Systeme. Zusätzlich sind die positiven Effekte von Data Governance häufig nicht sofort sichtbar, sondern zeigen sich erst langfristig. Daher müssen Unternehmen eine Kultur der Veränderungsbereitschaft fördern und ihre Mitarbeiter durch Schulungen und klare Visionen in den Prozess einbeziehen.

Am unteren Ende befindet sich die Datenmanagementschicht des PwC Frameworks. Diese Schicht umfasst einerseits alle wesentlichen Bereiche, die in typischen Datenökosystemen auftreten wie Datenarchitekturen, Datenmodelle, Datenintegrationen, Referenzdatenmanagement sowie BI. Andererseits legt das PwC Data Governance Framework einen großen Wert auf einen Ansatz zur Verwaltung des Datenflusses während des gesamten Lebenszyklus. Dieser Ansatz ist richtlinienbasiert und erstreckt sich von der Erstellung und Erfassung bis zu dem Zeitpunkt, an dem die Daten so veraltet sind, dass sie entweder

archiviert oder gelöscht werden. Häufig konzentrieren sich die Prozesse der Datenverwaltung nur auf die aktive Nutzung und Speicherung der Daten. Richtlinien, Standards und Vorgaben für andere Phasen des Lebenszyklus wie die Archivierung sind folglich häufig nicht vorhanden. Dies führt dazu, dass alte und ungenutzte Daten im Datenökosystem verbleiben und zu ineffizient genutzten Speicherressourcen, erhöhten Sicherheitsrisiken und verminderter Datenqualität führen können.

Das PwC Framework hingegen betont die Wichtigkeit der ganzheitlichen Betrachtung aller Phasen des Datenlebenszyklus, um eine durchgängige Einhaltung aller Regularien und Vorschriften sicherzustellen. Zudem wird dadurch die Kosteneffizienz und die Datenintegrität erhöht. Die genannten Maßnahmen und Ansätze beziehen sich dabei nicht nur auf einzelne Datenquellen, sondern umfassen alle Arten von diesen. Dies schließt sowohl strukturierte Daten, wie sie in Datenbanken oder Data Warehouses vorliegen, als auch unstrukturierte Daten aus unterschiedlichen Anwendungen und Data Lakes ein.

Alle zuvor genannten Komponenten drehen sich um drei zentrale Säulen: Personen und Unternehmenskultur, das operative Modell sowie Werkzeuge und Technologien. Diese Säulen sind die Enabler von Data Governance, indem sie alle geplanten Strategien und Richtlinien im Arbeitsalltag umsetzen müssen und so Data Governance zum Leben erwecken. Personen nehmen die definierten Rollen wahr und sitzen in Gremien, die über Data Governance Aktivitäten entscheiden. Außerdem bedienen sie die Systeme und Programme und halten sich an festgelegte Standards und Richtlinien. Zudem bestimmt die Unternehmenskultur das Verhalten und die Einstellung der Mitarbeiter gegenüber neuen Prozessen und Innovationen. Daher muss im Unternehmen eine Kultur der Offenheit und Innovation herrschen, die von den Beteiligten und Entscheidern vorgelebt werden muss. Das operative Modell sorgt für einen reibungslosen Ablauf und stellt sicher, dass der Fortschritt der Data Governance Bemühungen regelmäßig evaluiert wird. Zudem fördert es den Austausch und die Kommunikation zwischen den Beteiligten, indem es regelmäßige Treffen und Besprechungen organisiert. Die dritte Säule Werkzeuge und Technologien unterstützt die Data Governance Bemühungen durch die Bereitstellung der benötigten technischen Infrastruktur und Anwendungen. Dazu gehören beispielsweise Datenkataloge, Metadatenmanagement-Werkzeuge und Systeme zur Messung der Datenqualität.

Die drei Säulen stehen in Wechselwirkung mit den anderen Schichten des Data Governance Frameworks und stehen daher zwischen den Schichten. Die Schichten sind dabei nicht isoliert zu betrachten, sondern interagieren kontinuierlich miteinander über die Data Governance Enabler. Dieser ständige Austausch ist entscheidend für stetige Anpassungen

und Verbesserungen. Insgesamt sind die drei Säulen von elementarer Bedeutung für den Erfolg des Data Governance Programmes.

Insgesamt verfolgt das PwC Data Governance Framework einen umfassenden und praxisnahen Ansatz für einen effektiven Umgang der vorhandenen Datenressourcen. Dabei berücksichtigt es die übergeordnete strategische Ausrichtung des Unternehmens, fördert effiziente Prozesse und erkennt Daten als Vermögenswert an, wodurch der gesamte Datenlebenszyklus einbezogen wird. Die drei Enabler, Personen und Unternehmenskultur, das operative Modell sowie Werkzeuge und Technologien werden als essenziell für eine erfolgreiche Umsetzung und nachhaltige Entwicklung von Data Governance angesehen.

### **2.4 Data Governance in KMUs**

Im folgenden Abschnitt wird der Einsatz von Data Governance in KMUs untersucht. Dafür werden zunächst die Rahmenbedingungen und Herausforderungen, mit denen diese kleine Unternehmen im Vergleich zu großen konfrontiert sind, analysiert. Im Anschluss daran werden die gewonnenen Erkenntnisse aus den vorherigen Unterkapiteln in Kontext zu KMUs gestellt und bewertet. Außerdem werden bisherige Bemühungen in der Literatur zur Integration von Data Governance Praktiken in kleineren Unternehmen untersucht. Anhand von den gewonnenen Erkenntnissen wird die Forschungslücke identifiziert sowie die für diese Arbeit relevanten Forschungsfragen formuliert.

#### **2.4.1 Rahmenbedingungen**

KMUs haben aufgrund ihrer Struktur und Größe andere Voraussetzungen und stehen vor einzigartigen Herausforderungen. So sind ihre finanziellen Ressourcen begrenzt, was zu eingeschränkten Investitionen in Technologie und Fachpersonal führt [36][11]. Dies hemmt die Adaption von neuen Best-Practices, die in größeren Unternehmen bereits etabliert sind. Diese Entwicklung wird durch Marktversagen verstärkt. Marktversagen beschreibt Situationen, in denen der Markt nicht effizient funktioniert und dadurch den Geschäftsbetrieb von KMUs erschwert [19]. Zudem sind sie durch begrenzte Infrastruktur und Ressourcen oft nicht in der Lage, den Anforderungen komplexer Vorschriften gerecht zu werden. Des Weiteren fehlt es KMUs häufig an den notwendigen Führungs- und technischen Kenntnissen. Hinzu kommt, dass es der starre Arbeitsmarkt ihnen erschwert, qualifiziertes Fachpersonal zu finden [19]. Diese Situation führt dazu, dass Mitarbeiter

oft als interdisziplinäre Experten agieren müssen, da spezialisiertes Fachpersonal weder erschwinglich noch wirtschaftlich sinnvoll ist [36]. Darüber hinaus sind viele der aus der Forschung und Wirtschaft stammenden Tools und Methoden für KMUs oft zu komplex oder umfangreich, um für sie relevant zu sein. Während vereinfachte Tools zwar leichter zu bedienen sind, mangelt es ihnen häufig an Effektivität [36]. Eine Umfrage aus dem Jahr 2020 zeigt, dass nur 17 % der KMUs digitale Technologien umgesetzt haben, während dieser Anteil bei großen Unternehmen bei 54 % liegt [20]. Dazu fehlt es KMUs oft an der notwendigen Infrastruktur, um große Analysensysteme oder andere komplexe Technologien zu betreiben. Der Weg in die Cloud bietet zwar eine mögliche Lösung, erfordert jedoch neue Expertise, die KMUs zunächst entwickeln müssen [36].

Diese Herausforderungen wirken sich auch auf die Umsetzung von Data Governance aus. Denn ein effektives und erfolgreiches Data Governance Framework erfordert technologische Ressourcen, spezialisiertes Fachwissen und eine strategische Führung, die datenbezogene Entscheidungen bewerten und überwachen kann. KMUs haben oft Schwierigkeiten, diese Voraussetzungen zu erfüllen und sind dadurch nicht in der Lage, umfassende und komplexe Data Governance Programme umzusetzen.

### 2.4.2 Einordnung der bestehenden Literatur

Die existierende Literatur zum Thema Data Governance zeigt eine Vielzahl von Ansätzen und Vorschlägen für den Umgang mit Daten. Die Quellen stammen aus unterschiedlichen Bereichen, darunter Forschung, Organisationen und Unternehmen, und bieten verschiedene Ansätze und Herangehensweisen. In ihrer Literaturrecherche [4] haben Alhassan et al. den Fokus von Veröffentlichungen sowohl aus Forschung als auch aus praxisorientierten Publikationen untersucht. Die Ergebnisse davon sind in Abbildung 2.5 dargestellt. Dabei fällt auf, dass insbesondere in den wissenschaftlichen Veröffentlichungen der Fokus hauptsächlich auf der Definition von Data Governance Bestandteilen liegt und konkrete Vorschläge, wie diese implementiert und überwacht werden können, rar sind. Praxisorientierte Beiträge hingegen bieten eine ausgewogenere Mischung, wobei der Fokus noch weiter zu Gunsten der Implementierung und Überwachung verschoben werden könnte. Dies ist begründet, dass das Definieren bereits ausreichend durch wissenschaftliche Veröffentlichungen abgedeckt wird.

Wichtig für KMUs ist, dass Frameworks an ihre spezifischen Bedürfnisse angepasst und vor allem herunterskaliert werden können [11]. KMUs haben weniger involvierte Perso-

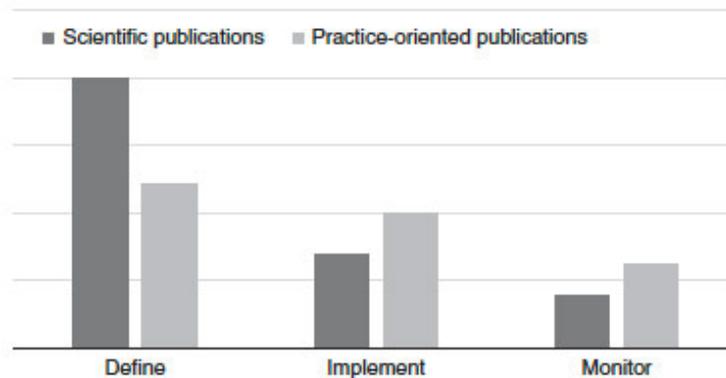


Abbildung 2.5: Schwerpunkte wissenschaftlicher und praxisorientierter Publikationen [4]

nen, Technologien und Prozesse, weshalb der Fokus auf konkreten Vorteilen liegen sollte. Begg und Caira untersuchten in [11], ob die verbreitete These, dass Data Governance Lösungen, die für große Unternehmen konzipiert wurden, auch auf KMUs übertragen werden können, zutrifft. Obwohl in der Praxis häufig die Skalierbarkeit solcher Lösungen angegeben wird, fehlen in der Literatur bislang konkrete Beispiele. Begg und Caira wählten das ihrer Meinung nach einfachste Framework aus, die fünf Datendomänen von Khatri und Brown [41], um die Anwendbarkeit eines Frameworks für KMUs zu testen. Sie stellten jedoch fest, dass dies zum Zeitpunkt der Veröffentlichung mit den durchgeführten Unternehmen nicht umsetzbar war. Dies unterstreicht die Notwendigkeit, spezifische und praktikable Lösungen für KMUs zu entwickeln, die den besonderen Herausforderungen und begrenzten Ressourcen dieser Unternehmen gerecht werden.

Ein Beispiel für die potenzielle Herausforderungen, denen KMUs bei der Implementierung von Data Governance Frameworks gegenüberstehen, wird im Folgenden durch die hypothetische Umsetzung von DAMA gezeigt. Der große Umfang, den DAMA DMBok abdeckt, kann für Neulinge von Data Governance, wie es Anwender in KMUs häufig sind, überfordernd sein. DAMA liefert allgemeine Richtlinien, die für alle Arten von Organisationen gelten, ohne konkret zu werden, wie diese auf operativer Ebene angewendet oder umgesetzt werden sollen. Darüber hinaus betrachtet DAMA Datenqualität und Metadaten nicht als integralen Bestandteil von Data Governance, obwohl gerade diese Aspekte für KMUs relevant sind, um messbare Ergebnisse und Vorteile zu erkennen. DAMA fokussiert sich stattdessen auf die organisatorische und strukturelle Ebene, die in KMUs weniger ausgeprägt ist und daher weniger Aufmerksamkeit benötigt. Diese allgemeinen und umfangreichen Ansätze können KMUs überfordern und ihre Fähigkeit einschrän-

ken, effektive Data Governance Strategien zu implementieren. Das DGI gibt an, dass für kleinere Organisationen oder solche mit einfacheren Datenprozessen möglicherweise ein informelles System der Governance ausreichen kann. Ihr Framework ist speziell für große Organisationen und Unternehmen konzipiert. In der heutigen Zeit der Digitalisierung, Big Data und KI benötigen jedoch auch KMUs formelle Richtlinien und Werkzeuge, um den Anforderungen der Regularien, Gesetze und den eigenen Ansprüchen an Datenqualität und Datenschutz gerecht zu werden.

### 2.4.3 Forschungsfragen

Die Definitionen von Data Governance in Unterabschnitt 2.1.1 legen nahe, dass neben einem geeigneten Data Governance Framework auch ein passender Datenmanagement Ansatz fehlt, der einen Leitfaden für konkrete Umsetzungen bereitstellt. Datenmanagement beschäftigt sich mit der praktischen Umsetzung der im Rahmen der Data Governance festgelegten Richtlinien, Standards und Strategien. Die Forderung nach einem, für KMU geeigneten Data Governance Ansatz, der nicht nur High-Level-Aktivitäten definiert, sondern auch Leitfäden für die Implementierung und Überwachung des Programms liefert, ist in der Forschungsfrage (FF) 3 ausformuliert. Der Bedarf nach einem passenden Datenmanagement-Ansatz spiegelt sich in FF4 wider. Um ein neues Framework zu testen, muss jedoch zunächst ein geeigneter Rahmen geschaffen werden, anhand dessen Frameworks evaluiert werden können. Die Testumgebung muss realistische Bedingungen von KMUs abbilden und wird in FF1 behandelt. Im Anschluss daran wird untersucht, was von bereits eingesetzten Frameworks sich auch für KMUs eignet und für ein eigenes Framework übertragbar ist. Die Ergebnisse davon beantworten die FF2. Es sei darauf hingewiesen, dass es auch innerhalb der Definition von KMUs aus Unterabschnitt 2.1.3 Unterschiede für Voraussetzungen und Bedingungen gibt. In einem Unternehmen mit 200 Mitarbeitern, welches weiterhin in die Definition fällt, herrschen andere Bedingungen als in einem Unternehmen mit einigen wenigen Mitarbeitern. Der Fokus dieser Arbeit liegt dabei nicht auf Kleinstunternehmen, sondern vielmehr auf kleinen und mittelständischen Unternehmen, wie in Tabelle 2.1 definiert. Daraus ergeben sich folgende Forschungsfragen:

**Forschungsfrage 1** Wie sieht eine passende Testumgebung aus, um bestehende und künftige Frameworks mit Bezug auf KMU-Größe untersuchen und validieren zu können?

**Forschungsfrage 2** Welche Aspekte bestehender Frameworks sind für KMUs relevant und welche Elemente können übernommen werden?

**Forschungsfrage 3** Wie sieht ein Data Governance Framework aus, das speziell auf die Bedürfnisse von KMUs zugeschnitten ist?

**Forschungsfrage 4** Wie kann eine praxisgerechte Umsetzung eines Data Management Programms gestaltet werden, um die Konzepte des Data Governance Frameworks im Unternehmensalltag anzuwenden?

## 3 Design einer Testumgebung

Um die Funktionalität des zu entwickelnden Data Governance Frameworks zu testen und zu evaluieren, ist die Erstellung eines Testumfeldes erforderlich, das auf die Besonderheiten von Frameworks in KMUs ausgerichtet ist. Da der Fokus der Arbeit auf einer einsetzbaren und praktikablen Lösung liegt, soll das Szenario einen möglichst realistischen Unternehmenskontext darstellen. Dazu wird ein fiktives Unternehmen erstellt, in dem auftretende Herausforderungen und Probleme eines KMUs abgebildet werden. Dieses Testumfeld kann in Zukunft auch für die Bewertung neuer Frameworks eingesetzt werden, um die Anwendbarkeit und den Mehrwert dieser für KMUs zu überprüfen und mit dem in dieser Arbeit entwickelten oder anderen Frameworks zu vergleichen.

### 3.1 Aufbau und Ziele der Testumgebung

Im Unterschied zu anderen Arbeiten und Publikationen wird in dieser Arbeit nicht allein auf die theoretischen Konzepte und die theoretische Anwendung eines Data Governance Frameworks eingegangen, sondern auch die praktische Umsetzbarkeit der Theorie untersucht. In Bezug auf KMUs weisen Beggs und Cairns in [11] darauf hin, dass eine Adaption der vorgestellten Frameworks zwar häufig eine Anwendung in kleineren Unternehmen ermöglichen soll, jedoch die praktische Umsetzung in diesem Kontext nicht gezeigt wird. Daher wird im Rahmen des Testumfeldes die Situation in einem Unternehmen dargestellt, das sich der Rolle von Daten und den damit verbundenen Aspekten bewusst ist oder durch Evaluierung der Prozesse bewusst wird. Dies führt zu dem Ziel in dem Unternehmen, ein formelles Data Governance Programm einzuführen. Das Data Governance Programm soll Unterstützung in Aspekten wie Datenqualität, Datenschutz, Zugriffskontrollen, Compliance, Metadaten und Entscheidungsrechten geben. Die formulierten Herausforderungen des Unternehmens und die definierten Metriken ermöglichen eine Untersuchung und Evaluierung des Frameworks. Reale Anwender und Unternehmen können die beschriebenen Vorgänge als Vorlage nehmen und die Schritte und Abläufe an

ihre spezifischen Bedingungen anpassen. Dazu wird zunächst ein fiktives Unternehmen definiert, wobei die Anzahl der Mitarbeiter pro Abteilung geschätzt und das geschätzte Maß an Expertise im Kontext von Data Governance Aktivitäten angegeben wird. Im Anschluss daran werden Datenflüsse und -abläufe detailliert beschrieben. Im nächsten Schritt werden auftretende Probleme und Herausforderungen identifiziert. Diese umfassen Schwierigkeiten in der Datenverarbeitung, mangelnde Datenqualität, fehlende Datenschutzmaßnahmen sowie unklare Verantwortlichkeiten. Mithilfe dieser Problembeschreibungen können die Auswirkungen der Einführung von Data Governance evaluiert werden. Im Anschluss werden Evaluierungsszenarien definiert, die sowohl alltägliche Situationen als auch Sonderfälle abbilden sollen.

## 3.2 Fiktives KMU

Bei dem fiktiven Unternehmen wird ein kleines Unternehmen gewählt, das sich gerade an der Grenze zum mittelständischen Unternehmen befindet. Dadurch soll es eine gewisse Allgemeingültigkeit besitzen, die es ermöglicht, relevante Frameworks sowohl für kleine (weniger als 50 Mitarbeiter) als auch für mittelgroße Unternehmen (50 bis 250 Mitarbeiter) zu untersuchen. Das Unternehmen existiert schon länger, ist familiengeführt und das Kerngeschäft besteht hauptsächlich aus dem Produzieren von Geräten. Die Frameworks sollen nicht nur in modernen, dynamischen und technologieaffinen Umgebungen einsetzbar sein. Sie müssen auch in traditionelleren, handwerklicheren Unternehmen, bei denen der Ursprung nicht aus dem digitalen Kontext stammt, funktionieren. In diesen gibt es in der Regel feste Strukturen und Prozesse, die sich im Laufe der Unternehmensgeschichte bewährt haben. Frameworks müssen daher anpassungsfähig und praktikabel sein, um den Anforderungen dieser Unternehmen gerecht zu werden. Zunächst wird dafür in Unterabschnitt 3.2.1 ein Unternehmen entworfen. Hierfür werden kurz die Geschäftsfelder erläutert, die Gründe für die Einführung eines formellen Data Governance dargelegt sowie die Struktur und Aufgaben der einzelnen Abteilungen vorgestellt. Im Anschluss werden in Unterabschnitt 3.2.2 die internen Datenflüsse untersucht. In Unterabschnitt 3.2.3 und Unterabschnitt 3.2.4 werden die in dem Unternehmen eingesetzten Technologien in der Datenverwaltung näher betrachtet. In Unterabschnitt 3.2.5 werden auftretende Probleme und Herausforderungen im Kontext der Datenverwaltung identifiziert und beschrieben.

#### 3.2.1 Unternehmensstruktur

Die AquaFlow Solutions GmbH ist ein traditionsreiches Unternehmen, das seit über 40 Jahren Bewässerungssysteme für den Garten im kleinen und großen Stil herstellt. Das Unternehmen hat 50 Angestellte und hat sich von der Produktion mechanischer und elektrischer Bewässerungssysteme zu modernen, IoT-unterstützten Lösungen weiterentwickelt. Diese ermöglichen eine erhöhte Effizienz und Effektivität der Systeme sowie die Erschließung neuer Geschäftsfelder. Durch eine Vielzahl an Sensoren, die Daten verschiedener Art kontinuierlich an eine zentrale Plattform senden, ist die Überwachung und Steuerung der Bewässerungsanlagen von überall in Echtzeit möglich. Durch die großen Datenmengen, die mit dem Einsatz der neuen Produkte anfallen, ist die Einführung eines formellen Data Governance Programms unabdingbar geworden. Im Folgenden werden die einzelnen Abteilungen vorgestellt sowie ihr Einsatz von Daten jeglicher Art untersucht.

**Geschäftsführung** Diese besteht aus insgesamt drei Personen, die für die generelle strategische Ausrichtung des Unternehmens verantwortlich sind. Dabei stehen die Planung, Koordination und Überwachung aller Geschäftsaktivitäten im Vordergrund. Sie greifen dafür auf Informationen und Daten aus verschiedenen Bereichen zurück. Analysen, erstellt von den jeweiligen Ressorts sowie Berichte und Dashboards geben ihnen einen Überblick über den aktuellen Stand und unterstützen sie bei Entscheidungen.

**Marketing** Das Marketingteam besteht aus sechs Mitarbeitern und ist für die Entwicklung und Umsetzung von Marketingstrategien und -kampagnen verantwortlich. Dafür nutzt es Daten und Informationen aus verschiedenen Quellen, um geeignete Aktionen und Maßnahmen zu entwickeln. Beispiele für Datenquellen sind Kundendaten, Markttrends sowie die Ergebnisse der Kampagnen.

**Vertrieb** Die Vertriebsabteilung umfasst zehn Mitarbeiter, die sich um den Verkauf aller Bewässerungssysteme, die Kundenakquise und die Pflege von Kundenbeziehungen kümmern. Dabei nutzen sie Kundendaten aus dem Customer-Relationship-Management-System (CRM-System), Produktinformationen sowie Analysen aus den Nutzungsdaten der IoT-Systeme.

**Produktion** Die Produktionsabteilung besteht aus 15 Mitarbeitern und ist für die Herstellung der Bewässerungssysteme verantwortlich. Dabei werden Sensordaten der Produktionsanlagen verwendet, um einen reibungslosen Ablauf und die Qualität der Produkte zu gewährleisten. Außerdem sollen so benötigte Wartungsarbeiten antizipiert sowie die Produktionsprozesse stetig verbessert werden.

**IT-Abteilung** Sie besteht aus sieben Mitarbeitern und verwaltet die technische Infrastruktur des Unternehmens. Dazu gehören die Wartung von Servern, Netzwerken und Datenbanken sowie die Entwicklung von digitalen Services und Produkten. Daten umfassen Nutzungsinformationen aller lokalen Systeme sowie der IoT-Geräte und Logs der Server, Netzwerke und Datenbanken. Sie spielen die zentrale Rolle auf dem Weg der Digitalisierung, da sie den anderen Abteilungen Analysetools und andere Datenmanagementlösungen bereitstellen.

**Entwicklung** Die Forschungs- und Entwicklungsabteilung umfasst acht Mitarbeiter, die für Innovation und die Verbesserung der Produkte verantwortlich sind. Sie arbeiten dafür mittlerweile eng mit der IT-Abteilung zusammen, um aus den Analysen der Sensordaten stetig neue Erkenntnisse gewinnen und diese in die Produktentwicklung einfließen zu lassen.

**Personalabteilung** Die Personalabteilung besteht aus zwei Mitarbeitern, die für die Rekrutierung, Schulungen und Betreuung der Mitarbeiter verantwortlich sind. Dabei verwaltet sie alle Mitarbeiterdaten und stellt sicher, dass alle personenbezogenen Daten den bestehenden Datenschutzrichtlinien entsprechend behandelt und gespeichert werden.

#### 3.2.2 Datenflüsse

Die Datenflüsse und Interaktionen innerhalb und zwischen den Abteilungen liefern Rückschlüsse auf potenzielle Herausforderungen im Umgang mit Daten. Eine Analyse zeigt auf, welche Anforderungen an die technische Infrastruktur gestellt werden und welche Prozesse für eine reibungslose Interaktion zwischen den Abteilungen notwendig sind. Zudem zeigt sich dadurch, dass jede Abteilung durch ein formelles Data Governance profitieren kann. In Abbildung 3.1 werden die Interaktionen zwischen den Abteilungen sowie die Nutzung verschiedener Datenquellen durch die Abteilungen visualisiert.

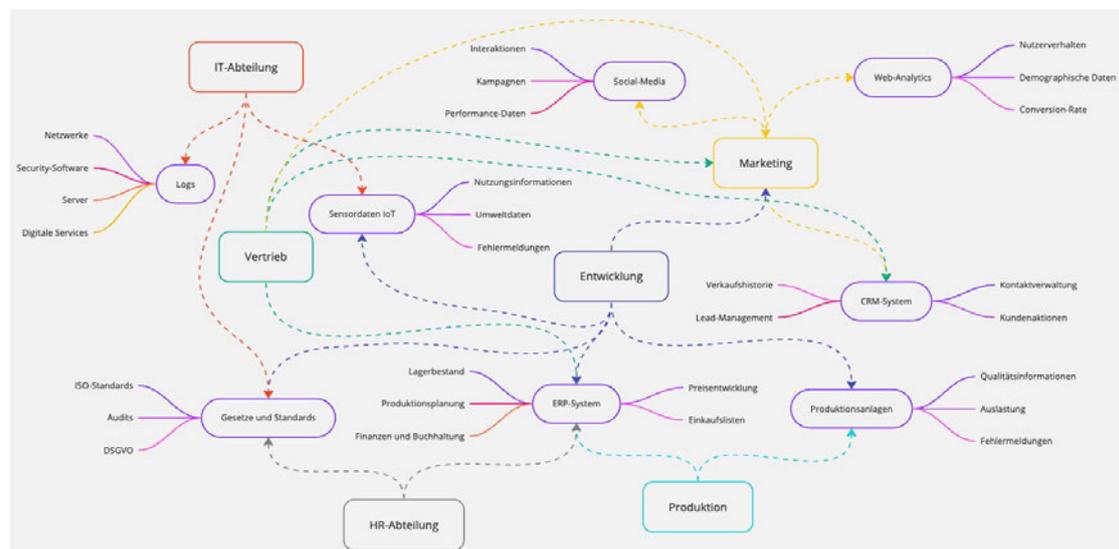


Abbildung 3.1: Datenflüsse des Unternehmens

Die einzelnen Abteilungen aus Unterabschnitt 3.2.1 sind dabei in farbigen Kästchen dargestellt, während die unterschiedlichen Datenquellen mit Ovalen visualisiert werden. Für eine bessere Nachvollziehbarkeit sind die Abteilungen sowie ihre Interaktionen mit Datenquellen und anderen Abteilungen in jeweils unterschiedlichen Farben visualisiert. Im Rahmen dieser Arbeit ist es nicht möglich, alle Datenquellen eines realen Unternehmens abzubilden. Es wird daher ein Fokus auf die wichtigsten und repräsentativsten Datenquellen gelegt. Dazu zählen Daten und Informationen aus Social-Media und Web-Analytics Programmen, Sensordaten der Bewässerungssysteme im realen Betrieb, Logs verschiedener interner Systeme, Meldungen der Produktionsanlagen, Gesetze und Verordnungen sowie Informationen aus dem CRM-System und Enterprise-Resource-Planning-System (ERP-System). Weitere Details zu den verschiedenen Datenquellen sind in Abbildung 3.1 dargestellt. Im Folgenden wird kurz auf die Vorgänge der einzelnen Abteilungen eingegangen.

Das Ziel des Marketingteams ist die Erhöhung der Bekanntheit des Unternehmen sowie die Gewinnung von Kunden. Für die Erstellung der Kampagnen und Maßnahmen nutzen sie Daten aus Social-Media und Web-Analytics-Programmen. Dabei handelt es sich um Informationen über das Nutzungsverhalten und die Demographie von Anwendern. Diese Daten werden mit Informationen aus dem CRM-System kombiniert, um gezielte und personalisierte Marketingstrategien zu entwickeln. In ihren Vorgängen und Prozessen nutzt die Vertriebsabteilung Informationen aus dem CRM-System, um Informationen der

Kunden sowie ihrer Verkaufshistorie zu erhalten und zu pflegen. Das ERP-System liefert Informationen über aktuelle Lagerbestände und Lieferzeiten. Außerdem werden darin die Aufträge abgewickelt und Preise kalkuliert. Der Vertrieb steht dabei in engem Kontakt mit der Marketing-Abteilung, um sich über Kundenfeedback und Markttrends auszutauschen. Mithilfe von Sensordaten der Produktionsanlagen überwacht und optimiert die Produktionsabteilung die Prozesse und Arbeitsschritte. Die Produktionsplanung und Lagerbestände werden durch das ERP-System erfasst. Die IT-Abteilung überwacht die technische Infrastruktur des Unternehmens mithilfe von Logs aus Servern, Netzwerken und digitalen Produkten. Durch die Analyse dieser können Schwachstellen identifiziert und behoben werden. Außerdem müssen die Aktivitäten stets im Einklang mit den geltenden Gesetzen und Verordnungen stehen. Für die Entwicklung neuer und Optimierung aktueller Produkte werden Sensordaten der IoT-Bewässerungssystemen analysiert. Bei Markttrends und Kundenwünsche werden Berichte der Marketing-Abteilung herangezogen. Außerdem müssen die entwickelten Produkte stets Gesetze und geltende Normen einhalten. Die Personalabteilung interagiert mit dem ERP-System, um die Mitarbeiterdaten zu verwalten und Schulungen zu planen. Feedback von Abteilungen und Mitarbeitern werden bei der Personalplanung berücksichtigt. Die Abteilung muss dabei im Zusammenspiel mit der IT-Abteilung sicherstellen, dass die sensitiven Mitarbeiterinformationen geschützt sind.

#### 3.2.3 Data Lake

Ein Data Lake ist ein zentrales Datenlager, in dem sowohl strukturierte als auch semi-strukturierte und unstrukturierte Daten gespeichert werden können. Die Analyse von Daten aus verschiedenen Datenquellen ermöglicht es Unternehmen, Einblicke in ihre eigenen Operationen sowie in die ihrer Kunden zu gewinnen. Auf Basis dieser Analysen können Trends abgeleitet und künftige Entscheidungen getroffen werden. Dafür wurden zunächst sogenannte *Data Warehouses* erstellt. In diesen werden die Daten gespeichert, um künftig bessere BI Entscheidungen treffen zu können [69][54]. Die heterogenen Daten müssen zunächst aufbereitet werden, um sie für die weitere Verarbeitung nutzbar zu machen. Dazu wird das sogenannte *Extract, Transform, Load (ETL)*-Pattern angewendet. Zunächst werden die Daten aus der Datenquelle extrahiert, anschließend bearbeitet, gereinigt und in das gewünschte Schema transformiert. Schließlich werden sie in das Warehouse geladen. Das daraus resultierende *schema-on-write*, also das Anwenden eines Schemas beim Schreiben bereits, bringt jedoch einige Einschränkungen mit sich [42]. Zum

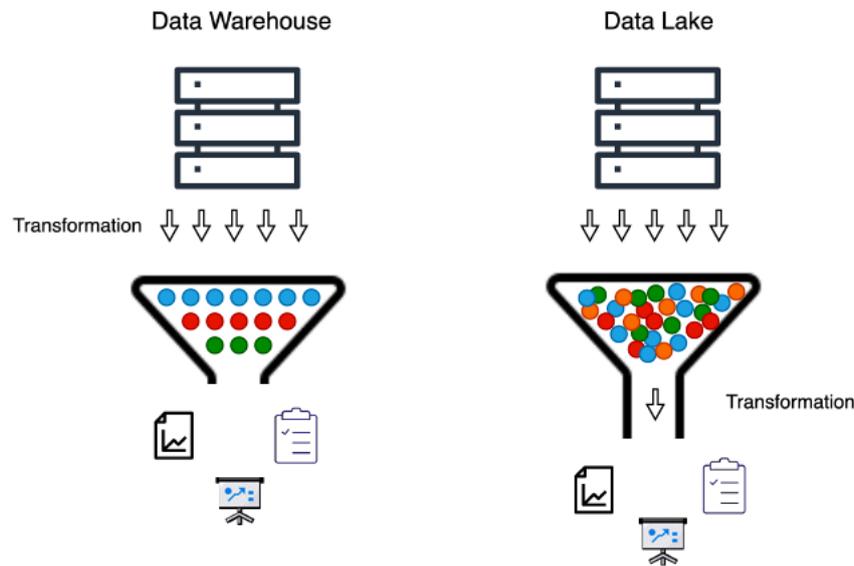


Abbildung 3.2: Vergleich Data Warehouse und Data Lake [51]

einen werden lediglich die strukturierten Daten gespeichert und die restlichen verworfen [54]. Somit sind diese für künftige, fortgeschrittenere Analysen verloren [30]. Durch die festgelegte Struktur ist es außerdem zeitaufwändig und teuer, Änderungen am Design vorzunehmen [42]. Um diese Probleme zu beheben und insbesondere die immer größer werdende Menge an Daten speichern und nutzen zu können, wurden sogenannte Data Lakes eingeführt. Im Unterschied zum Data Warehouse werden beim Data Lake die Daten aus den unterschiedlichen Quellen in ihrem ursprünglichen Zustand gespeichert. Damit ist dieser in der Lage, sowohl strukturierte, semi-strukturierte als auch völlig unstrukturierte Daten speichern zu können [8]. Folglich können Daten schnell und kosten-effektiv abgelegt werden, ohne vorher eine aufwändige Transformation durchlaufen zu müssen. In Abbildung 3.2 sind die unterschiedlichen Konzepte von einem Data Warehouse und einem Data Lake dargestellt. Die Daten werden im Data Warehouse vor dem Laden transformiert und liegen in diesem strukturiert vor. Im Data Lake werden die Daten in ihrer ursprünglichen Form geladen und erst bei Bedarf transformiert. Dies wird das *Extract, Load, Transform (ELT)*-Pattern genannt. Auf den Daten von Data Lakes können dynamisch Analysen durchgeführt werden [24], wobei die Daten erst bei der Anfrage transformiert werden. Dieses Vorgehen wird *schema-on-read* genannt. Dadurch können Informationen kombiniert aus strukturierten sowie unstrukturierten Daten erlangt werden [42]. Ein Data Lake ist jedoch nicht ausschließlich ein Datenspeicher. Ein effektives

Metadaten-Management ist elementar für das Wiederfinden, die Nachverfolgung der Herkunft sowie das Qualitätsmanagement der Daten [31]. Aufgrund des fehlenden Schemas beim Laden der Daten müssen so viele Informationen wie möglich aus den Datenquellen gewonnen werden, um die Entstehung eines *Datensumpfes (Data Swamp)* zu vermeiden. Dies bezeichnet Data Lakes, die unstrukturiert und unübersichtlich sind und in denen es für Nutzer nur mit großem Aufwand möglich ist, relevante Inhalte zu finden [30][31].

Die steigende Menge an Daten, die durch die Digitalisierung der Bewässerungssysteme generiert wird, sowie die wachsende Relevanz von Analysen für Entscheidungen haben die AquaFlow GmbH dazu motiviert, den Schritt zu einem Data Lake zu gehen. Diverse Datenquellen wie die IoT-Bewässerungssysteme, Web-Analysen und Logs von Produktionsanlagen und digitalen Produkten senden kontinuierlich Daten. Die direkte Transformation und Verarbeitung, die bei traditionellen Systeme wie einem Data Warehouse und relationalen Datenbanken notwendig ist, ist durch die große Menge an Daten zu aufwendig und kostspielig. Stattdessen werden die Daten in ein Data Lake geladen. Auf Abruf können auf den Daten Transformationen und Analysen ausgeführt werden. Die Architektur des Systems ist in Abbildung 3.3 dargestellt. Die Daten der verschiedenen Datenquellen werden in dem Data Lake zusammengeführt. Mithilfe von Transformationen der unstrukturierten Rohdaten zu strukturierten Datenmodellen sowie der Aggregation von Informationen aus verschiedenen Datensätzen können Dashboards zur Übersicht und Geschäftsberichte erstellt werden. Diese ermöglichen es, fundierte Entscheidungen zu treffen, die Effizienz der Systeme und Prozesse zu erhöhen sowie die Produkte zu verbessern.

#### 3.2.4 Technologien

Im Folgenden wird der technische Aufbau der AquaFlow GmbH dargestellt, der exemplarisch für die Infrastruktur und Systeme in Unternehmen solcher Art stehen soll. Aufgrund der Komplexität der Umstände und Gegebenheiten eines realen Unternehmens kann an dieser Stelle nicht jedes Detail umgesetzt und beschrieben werden. Der Fokus liegt auf der technischen Umsetzung der in Unterabschnitt 3.2.2 beschriebenen Datenflüsse in dem in Unterabschnitt 3.2.3 vorgestellten Data Lake. In Tabelle 3.1 sind die verwendeten Programme für die jeweiligen Bestandteile des Data Lakes aufgelistet.

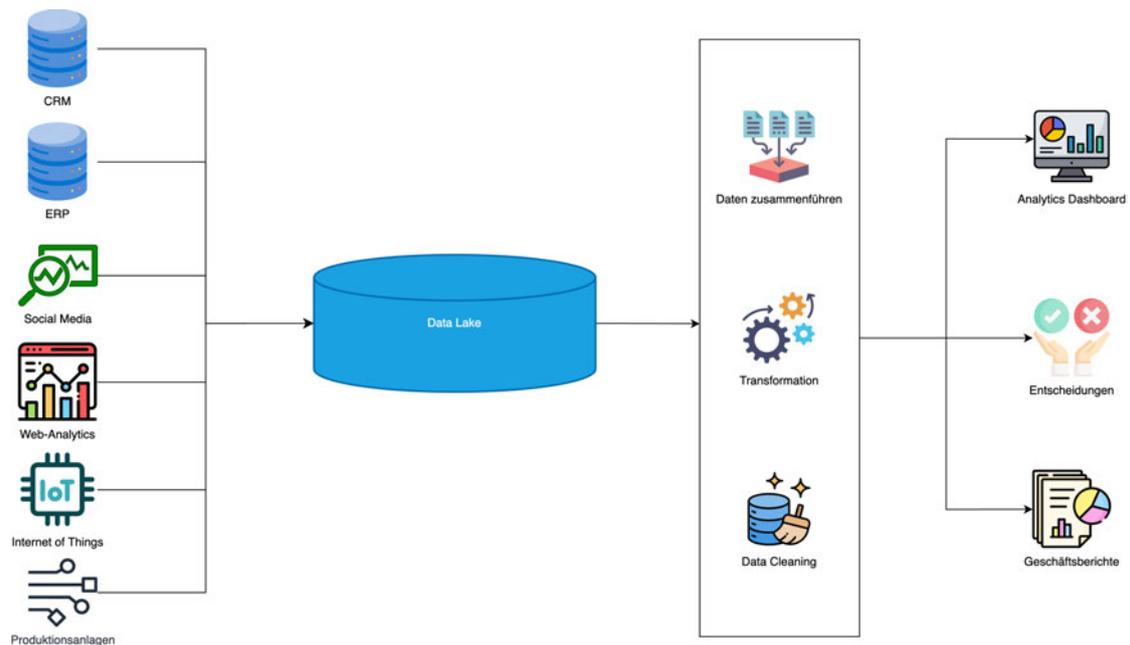


Abbildung 3.3: Data Lake Architektur der AquaFlow GmbH

Data Lake Komponente	Verwendete Technologie
Deployment	Docker & Docker-Compose
ERP-System	Mocked durch MySQL
CRM-System	Mocked durch Postgres
Social-Media Daten	Faker
Web-Analytics Daten	Faker
IoT-Sensoren	Faker
Produktionsanlagen	Faker
Data Lake	MinIO
Datenverarbeitung	Spark Cluster

Tabelle 3.1: Data Lake Komponenten und die verwendeten Technologien

**Docker & Docker-Compose** Docker ist ein System für die Entwicklung, Bereitstellung und Ausführung von Anwendungen und basiert auf dem Konzept der Containerisierung [25]. Dieses verfolgt den Ansatz, Anwendungen und die dazugehörigen Abhängigkeiten in isolierte Umgebungen, sogenannte Container, zu bündeln. Das hat den Vorteil, dass die Anwendungen unabhängig von der zugrundeliegenden Infrastruktur werden und auf nahezu allen Systemen ausgeführt werden können. Ohne Containerisierung müssen für

jedes Betriebssystem die spezifischen Versionen der Bibliotheken installiert werden. Mit Docker reicht die Erstellung eines einzelnen Softwarepakets aus, das dann mithilfe der Docker-Engine überall ausführbar ist. Die Container sind leichtgewichtig und können im Vergleich zu echten virtuellen Maschinen schnell gestartet und gestoppt werden, da kein eigenes Betriebssystem gestartet werden muss. Dadurch sind sie effizient in der Nutzung der vorliegenden Ressourcen und ermöglichen eine gute Skalierbarkeit. Je nach Bedarf können sie repliziert und parallel nebeneinander ausgeführt werden [5]. Für die Steuerung und Verwaltung von Anwendungen mit mehreren Containern ist die Nutzung eines Orchestrierungstools wie Docker-Compose sinnvoll. Über Konfigurationsdateien lassen sich Startbedingungen, Abhängigkeiten zwischen Containern sowie Speicher- und Netzwerkeinstellungen festlegen [26]. Aufgrund der Skalierbarkeit und Portabilität werden alle Anwendungen der AquaFlow GmbH in Docker-Containern ausgeführt und mithilfe von Docker-Compose gesteuert.

**MinIO** Das Speichermedium eines Data Lakes muss eine Reihe von Anforderungen erfüllen [73]. Aufgrund der großen und stetig wachsenden Datenmengen muss es skalierbar sein, ohne dabei an Leistung zu verlieren. Zudem muss der Speicher aufgrund der unvorhersehbaren Struktur der Daten in der Lage sein, strukturierte, semi-strukturierte und unstrukturierte Daten effizient zu speichern. Sowohl für das Laden der Daten in den Data Lake als auch aus diesem für die Datenverarbeitung und die Datenanalyse wird die Unterstützung oder Bereitstellung gängiger Formate zur Interaktion benötigt. MinIO ist eine performante, verteilte Objektspeicherlösung und ist vollständig kompatibel mit der Amazon Simple-Storage-Service (S3)-API [52]. Als Objektspeicherlösung werden Dateien nicht mehr in hierarchischen Strukturen wie Ordnern organisiert, sondern formen zusammen mit verschiedenen Metadaten ein Objekt [68]. Dieses erhält eine eigene Adresse, so dass es nicht in einer Dateistruktur gespeichert werden muss. Dadurch entfallen die hierarchischen Pfade beim Zugriff auf die Daten für bessere Performance und der Speicher ist leicht skalierbar [68]. Abgestufte Zugriffskontrollen über ein Identity and Access Management (IAM)-System ermöglichen eine präzise Zuweisung von Lese- und Schreibberechtigungen. Dies stellt sicher, dass nur autorisierte Nutzer Zugang zu den verschiedenen Daten haben und mindert das Risiko eines Datenlecks. Für die AquaFlow GmbH ist MinIO die optimale Lösung für ihren wachsenden Speicherbedarf. Aufgrund der einfachen Einrichtung und Handhabung ist eine schnelle Entwicklung möglich. Außerdem ermöglicht die Unterstützung der gängigen S3-API einen potenziellen Übergang zu einer Cloud oder Hybrid-Cloud Lösung.

**MySQL** MySQL ist ein weitverbreitetes relationales Datenbankmanagement-System (RDBM-System) und basiert auf der Structured Query Language (SQL). Die Tatsache, dass MySQL Open-Source ist, einfach zu bedienen ist und eine schnelle Performance bietet, macht es zu einer weitverbreiteten und beliebten Wahl für viele Anwendungen [58]. Im Rahmen des Testumfeldes der AquaFlow GmbH repräsentiert MySQL ein vollwertiges ERP-System. Die Funktionalitäten dabei sind ausreichend, um die Kernaspekte eines solchen Systems zu simulieren und ermöglichen den Fokus auf die Kernaspekte von Data Governance Funktionen, ohne dabei ein spezialisiertes ERP-System implementieren zu müssen.

**PostgreSQL** Ähnlich wie MySQL ist PostgreSQL ebenfalls ein Open-Source RDBM-System, das auf SQL basiert und dessen Funktionalitäten erweitert. Es unterstützt sowohl relationale als auch nicht-relationale Abfragen, wodurch auch verschachtelte Objekte wie JavaScript Object Notation (JSON)-Objekte effizient gespeichert und abgerufen werden können [64]. Ebenso können eigene Datentypen definiert werden, wodurch Anwender eine höhere Flexibilität für eigene Implementationen haben. Insgesamt ist PostgreSQL weitverbreitet und verfügt über Integratoren in den gängigsten Programmiersprachen. Ähnlich wie bei MySQL und dem ERP-System wird PostgreSQL anstelle eines vollwertigen, meist kommerziellen CRM-Systems eingesetzt, um eine flexible und leichtgewichtige Lösung für die Repräsentation der grundlegenden Aspekte eines CRM-Systems zu erhalten.

**Apache Spark** Apache Spark ist ein Open-Source Datenverarbeitungs- und Analyseprogramm, das für die Handhabung von großen Datenmengen konzipiert wurde. Die dafür benötigte Performanz erreicht es durch die Kombination eines Kernmoduls, das die Verteilung der Rechenlast auf mehrere Knoten ermöglicht und eines Programmiermodells für die Verarbeitung von Daten im Arbeitsspeicher [67]. Dadurch ist es in der Lage, über 100 Terabyte an Daten in wenigen Minuten zu sortieren [6]. Spark besitzt dafür eine Architektur mit drei verschiedenen Komponenten. Der Master-Knoten (engl. master node) ist für die Verwaltung des Cluster verantwortlich, koordiniert Arbeitsabläufe und kontrolliert die Aufgabenverteilung der anderen Knoten. Die Arbeiterknoten (engl. worker nodes) führen die eigentliche Datenverarbeitungsarbeit aus. Diese bekommen sie vom Master-Knoten zugewiesen. Der Client-Knoten (engl. client node) übergibt Aufgaben an den Master-Knoten und initiiert damit die Spark-Anwendung [7]. Durch diese verteilte

Architektur ist Spark in der Lage, effizient und skalierbar große Datenmengen zu verarbeiten und zu analysieren. Dabei unterstützt es mehrere Programmiersprachen wie Scala, Java, Python und R und besitzt integrierte Bibliotheken für SQL-Anfragen, maschinelles Lernen, Graphverarbeitung sowie die Verarbeitung von Streaming-Daten. Die stetig steigende Anzahl an IoT-Geräten sowie die beständige Erweiterung von Datenquellen als Grundlage für Entscheidungsfindungen hat zu einem erhöhten Bedarf nach einem skalierenden Datenverarbeitungsprogramm geführt. Aus diesem Grund wird ein Spark-Cluster erstellt, das aufgrund der Verwendung von Docker-Containern mit wenig Aufwand und Ressourcen skaliert werden kann, indem neue Arbeiterknoten hinzugefügt werden.

**Faker-API** Faker ist eine Bibliothek, die es ermöglicht, fiktive aber realistische Daten zu erzeugen. Dabei können eine Vielzahl von Datentypen wie Namen, Adressen, Telefonnummern, E-Mail-Adressen, Kreditkartennummern und Produktnamen erstellt werden. Diese können für Test- und Entwicklungszwecke eingesetzt werden, bei denen realistische Daten benötigt werden, echte Daten aber jedoch noch nicht vorhanden sind oder aus Datenschutzgründen nicht verwendet werden dürfen. Anwender haben die Möglichkeit, sogenannte interne Datenanbieter (engl. provider) zu modifizieren oder eigene zu erstellen. Auf diese Weise können Daten nach den jeweiligen spezifischen Anforderungen generiert und somit ein realistisches Testumfeld erzeugt werden. In der AquaFlow GmbH werden im Laufe des Testszenarios verschiedene Provider der Faker-API verwendet, um realistische Daten für das CRM-System, das ERP-System, Web-Analytics, Social-Media Plattformen und IoT-Sensordaten zu generieren.

#### 3.2.5 Herausforderungen

Um die Wirksamkeit von Data Governance Frameworks in einem passenden Kontext validieren zu können, müssen diese mit realistischen Herausforderungen und Problemen getestet werden. In vielen wissenschaftlichen Publikationen werden vorgestellte oder untersuchte Data Governance Frameworks rein qualitativ bewertet. Dabei erfolgt eine Diskussion über die theoretischen Modelle und anhand verschiedener Annahmen wird eine qualitative Bewertung vorgenommen. Dadurch wird ein grundlegendes Verständnis geschaffen. Dieses ist jedoch zu abstrakt, um die praktische Anwendbarkeit im realen Kontext bewerten zu können. Die Testumgebung in dieser Arbeit strebt neben der qualitativen auch eine quantitative Bewertung an. Dafür werden zunächst relevante Pain-Points und auftretende Probleme der AquaFlow GmbH in ihrem Umgang mit Daten sowie

Entscheidungsfindungsprozessen identifiziert. Anhand dieser Herausforderungen werden in Abschnitt 3.3 Metriken abgeleitet, die später für die Evaluierung der verschiedenen Frameworks verwendet werden. In Unterabschnitt 3.3.2 werden die Probleme, Herausforderungen und Metriken in verschiedene Evaluierungsszenarien überführt. An diesen können sich die Frameworks erproben und es ergibt sich ein erster Eindruck, wie die Frameworks Data Governance in der Realität umsetzen.

Für das Identifizieren der Probleme und Herausforderungen werden die Entscheidungsfindungsprozesse verschiedener Abteilungen untersucht und auf Gemeinsamkeiten und wiederkehrende Probleme analysiert. Dabei sollen universelle Herausforderungen und grundlegende Muster erkannt werden, die in allen Bereichen des Unternehmens auftreten und den Einsatz eines unternehmensweiten Data Governance Ansatzes erfordern.

Für Entscheidungen und die Erstellung von geeigneten Maßnahmen spielen in der Marketingabteilung verschiedene Faktoren eine Rolle. In Abbildung 3.4 werden diese Einflüsse graphisch dargestellt. Daten aus dem CRM-System liefern Informationen über das aktuelle Kundenverhalten und deren Bedürfnisse. Diese werden durch Erkenntnisse aus Kundengesprächen und Marktanforderungen, die aus Berichten der Vertriebsabteilung stammen, ergänzt. Social-Media Plattformen liefern aktuelle Trends und Meinungen. Nutzungs- und Performancedaten von den verschiedenen Online-Marktplätzen sowie dem eigenen Onlineshops geben mithilfe von Web-Analytics Aufschluss über die Wirksamkeit des Onlineauftritts sowie den Marketingstrategien. Die Kombination der verschiedenen Informationsquellen erhöht die Effektivität und hilft, bessere Prognosen und Maßnahmen zu entwickeln. Gleichzeitig gibt es verschiedene äußere Einflüsse wie generelle Markttrends, Strategien von Konkurrenten sowie Datenschutzbestimmungen.

In der Entwicklung und Forschung sind Informationen aus nahezu allen anderen Bereichen entscheidend für Entscheidungen und die Planung neuer Produkte. In Abbildung 3.5 sind die Einflüsse abgebildet. Anonymisierte Nutzerdaten und Fehlermeldungen der IoT-Geräte liefern Informationen über die Funktionalität und Effektivität sowie Indizien über mögliche Schwachstellen der Bewässerungssysteme. Zusätzlich stehen die Mitarbeiter der Abteilung in stetigem Austausch mit dem Marketing und Vertrieb, um aus Verkaufsstatistiken und Markttrends Erkenntnisse über aktuelle Marktanforderungen sowie mögliche Verbesserungen zu gewinnen. Neben diesen Einflüssen spielen auch externe Faktoren wie gesetzliche Anforderungen, Normen und Standards eine Rolle. Beispielsweise müssen umweltschonende Materialien eingesetzt und die DSGVO eingehalten werden.

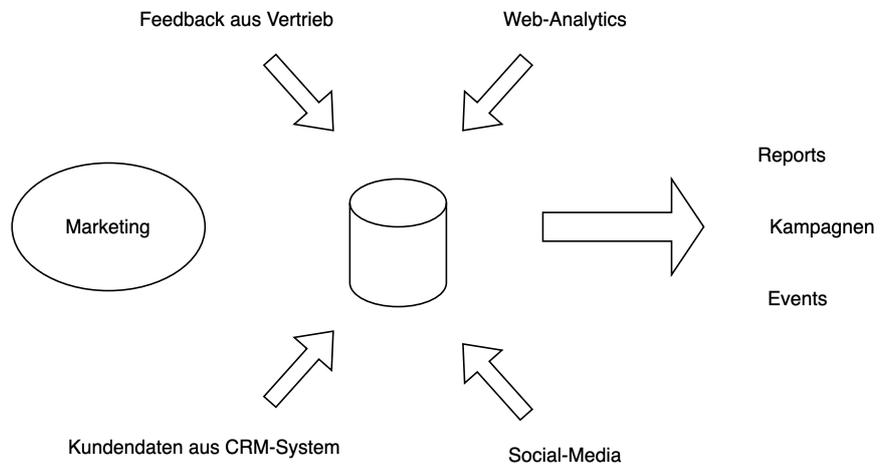


Abbildung 3.4: Einflüsse auf Entscheidungsfindung in der Marketing-Abteilung

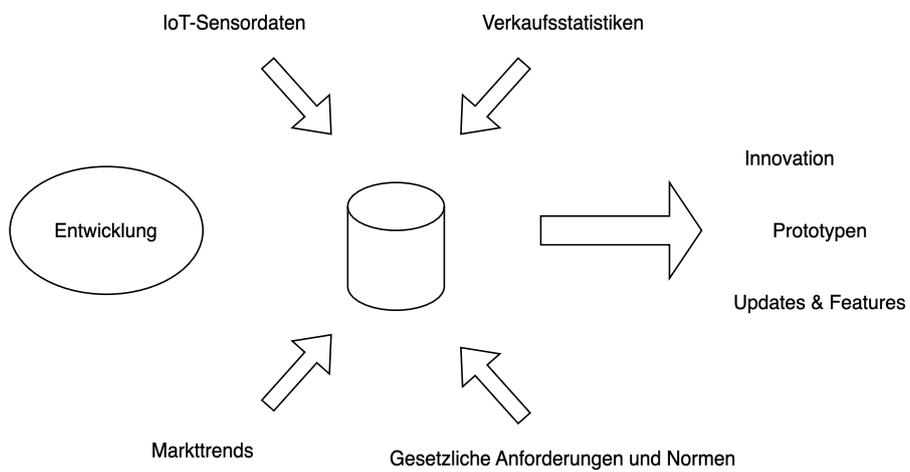


Abbildung 3.5: Einflüsse auf Entscheidungsfindung in der Entwicklungsabteilung

Näher betrachtet ergeben sich bei der Untersuchung der Einflüsse auf die Entscheidungsprozesse gemeinsame Muster und Einflussfaktoren, die in allen Abteilungen auftreten. Nahezu alle Abteilungen sind auf Daten angewiesen, die nicht direkt in ihrem Verantwortungsbereich liegen und nicht von ihnen erstellt werden. So ist die Marketingabteilung von Kundendaten aus dem CRM-System abhängig, die in der Regel im Vertrieb angesiedelt sind. Die Entwicklungsabteilung verwendet beispielsweise die Sensordaten der Produktionsanlagen sowie Kundenservice und -supportdaten aus der Vertriebsabteilung. Daneben verwenden beide Abteilungen Datenquellen, die von ihnen erstellt werden bzw. in ihrem Verantwortungsbereich liegen. Hier sind sie für die Qualität der Daten verantwortlich, müssen jedoch auch Anforderungen von anderen Abteilungen, die die Datenquellen verwenden, berücksichtigen. In der Marketingabteilung wären Daten von Social-Media Plattformen sowie Informationen aus Web-Analyticswerkzeugen Beispiele. Da bei der Einführung der IoT-Funktionalitäten keine spezielle Abteilung für die Analyse der großen Datenmengen eingerichtet wurde, liegt die Verantwortung bisher sowohl bei der Entwicklungs- als auch der IT-Abteilung. Die Entwicklungsabteilung konzentriert sich auf den Nutzen der Daten, die IT-Abteilung ist für die technische Infrastruktur und die Verwaltung der Daten verantwortlich. Ein weiterer Einflussfaktor sind Kooperationen zwischen den Abteilungen, die durch den Bedarf an Informationen, die nicht direkt in Datensätzen liegen, von anderen Abteilungen entstehen. Dabei handelt es sich in der Regel um Meetings und den Austausch von Berichten und Analysen. Dazu kommen technische Limitierungen und zur Verfügung stehende Kapazitäten. Beispiele hier sind begrenzt zur Verfügung stehende Informationen oder Daten der Analysewerkzeuge der Marketingabteilung sowie die limitierte Bandbreite und Systemkapazitäten in der Entwicklungsabteilung. Etablierte Strukturen und Prozesse spielen ebenfalls eine wichtige Rolle in der Entscheidungsfindung. In traditionelleren Strukturen können Datenanalysen möglicherweise nicht ausreichend in den Entscheidungsprozessen verankert sein. Dies kann verschiedene Gründe haben, beispielsweise ein geringes Vertrauen in die Daten aufgrund fehlender Datenqualität oder eine zu geringe Datenaffinität. Ebenso spielt die Organisationskultur und -strategie eine entscheidende Rolle. Innovative Umgebungen fördern den Einsatz neuer Methoden und Prozesse. Zu guter Letzt beeinflussen externe Faktoren, die außerhalb des Unternehmens liegen, die verschiedenen Abteilungen. Gesetzliche Vorschriften betreffen dabei alle Abteilungen. Datenschutz und die Verwendung von Standards im Umgang mit sensiblen Daten bestimmen gewisse Prozesse und Vorgehen. Markttrends und Erfolgsfaktoren von Konkurrenten beeinflussen die Marketingabteilung. Vorgaben und Normen für den Einsatz von Materialien müssen von der Entwicklung berücksichtigt werden.

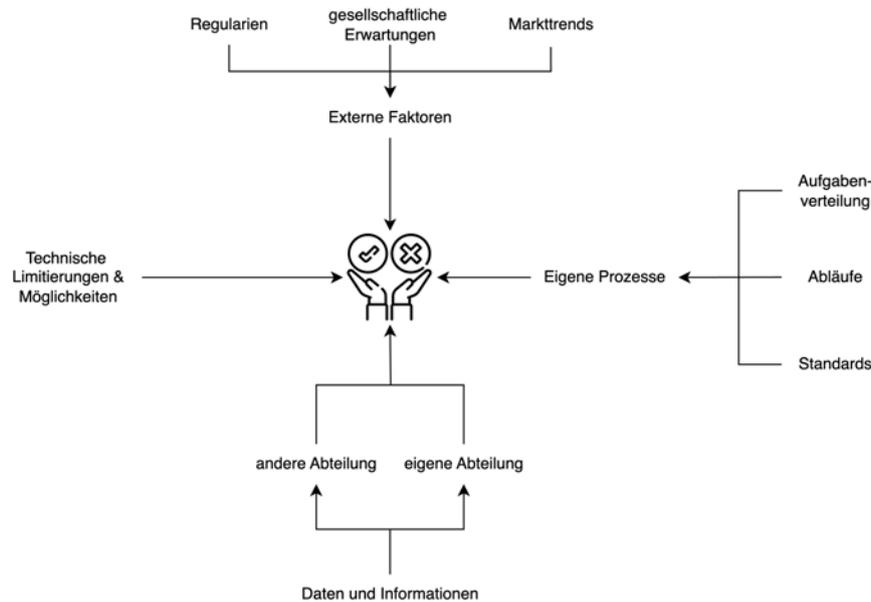


Abbildung 3.6: Skizze der Einflüsse auf die Entscheidungsfindung

Diese Faktoren lassen sich, wie in Abbildung 3.6 dargestellt, in vier übergeordnete Kategorien zusammenfassen. Diese zeigen, dass die Entscheidungsfindung unabhängig von der Abteilung von verschiedenen internen und externen Faktoren abhängt. Jede dieser Einflüsse erhöht die Komplexität des Entscheidungsprozesses und führt zu Unsicherheiten. Eine wesentliche davon ist die Frage nach der Qualität und Herkunft der Daten. Für ein hohes Vertrauen in die Daten müssen Fragen über die Genauigkeit, Aktualität und Vollständigkeit beantwortet sein. Stammen die Datensätze aus anderen Bereichen, treten Unklarheiten auf, ob die Daten den eigenen Qualitätsanforderungen entsprechen. Untersuchungen oder der Austausch mit den Verantwortlichen verzögern den Entscheidungsprozess und mindern die Effizienz der Abläufe. Ebenso führen inkorrekte Daten zu suboptimalen Entscheidungen und ineffizienten Prozessen. Im Gegensatz dazu erhöht die Verwaltung einer eigenen Datenquelle oder eines eigenen Datensatzes die Verantwortung gegenüber den anderen Abteilungen. Die Abteilung muss genannte Kriterien wie Genauigkeit und Aktualität gewährleisten, um nicht die Prozesse der anderen Bereiche zu beeinträchtigen. Suboptimale und verzögerte Entscheidungen entstehen ebenfalls durch Unklarheiten bei Verantwortlichkeiten und Entscheidungsträgern. Sofern keine Zuordnung von den Datensätzen und Datenquellen zu Personen oder Bereichen erfolgt, fühlt sich niemand verantwortlich und niemand pflegt die Daten. Auch innerhalb der Bereiche und Abteilungen müssen klare Verantwortlichkeiten hinsichtlich der Einhaltung von

Standards, des Umgangs mit sensitiven Daten, der Zugriffsrechte und weiterer wichtiger Aspekte definiert sein, da ansonsten jeder nach eigenem Ermessen möglicherweise relevante Entscheidungen trifft. Auf die beiden genannten Herausforderungen und Unklarheiten aufbauend entstehen Unsicherheiten bezüglich der Einhaltung gesetzlicher Vorschriften und branchenspezifischen Standards. Aufgrund der Vielzahl der Vorgaben und Normen ist womöglich nicht jedem Nutzer und Ersteller zu jederzeit klar, welche von diesen er berücksichtigen muss und ob er sie bei bestehenden Prozessen und Schritten eingehalten hat. Dadurch drohen dem Unternehmen rechtliche und finanzielle Konsequenzen.

## 3.3 Metriken

Um das zu entwickelnde Framework sowie weitere Frameworks bewerten zu können, ist es erforderlich, allgemeingültige Metriken zu definieren. In bisherigen Arbeiten wurden die Data Governance Frameworks meist qualitativ bewertet, ohne den Einsatz unter realen Bedingungen zu berücksichtigen. Aufgrund dessen sollen in dieser Arbeit Frameworks mit einer Kombination von qualitativen und quantitativen Methoden bewertet werden, um möglichen Anwendern auch eine fundierte Grundlage für die Entscheidung des geeigneten Data Governance Frameworks zu bieten. Ebenso wird dadurch eine Vorstellung vermittelt, wie diese in der Praxis umgesetzt werden. Dafür wird in Unterabschnitt 3.3.1 eine Domänenmatrix vorgestellt, die die Abdeckung der wichtigsten Aspekte von Data Governance darstellt. Im Anschluss daran werden aus den in Unterabschnitt 3.2.5 identifizierten Problemen und Herausforderungen in Unterabschnitt 3.3.2 Evaluierungsszenarien abgeleitet, die den Einsatz des Frameworks zur Lösung realer Probleme zeigen sollen.

### 3.3.1 Domänenmatrix

Khatri und Brown haben mit ihrer Arbeit in [41] eine Grundlage für viele nachfolgende Studien und Frameworks im Bereich der Data Governance geschaffen. In ihrer Publikation identifizieren sie fünf zentrale Domänen im Kontext von datenbasierten Entscheidungen, die für ein effektives Data Governance Framework entscheidend sind. Diese Domänen sollen Wissenschaftlern einen klaren Fokus auf die wesentlichen Themen und Praktikern eine Struktur zur Schaffung eines geeigneten und effektiven Frameworks bieten. Die fünf Domänen von Khatri und Brown, die in Abbildung 3.7 visualisiert sind, sind allgemeingültig

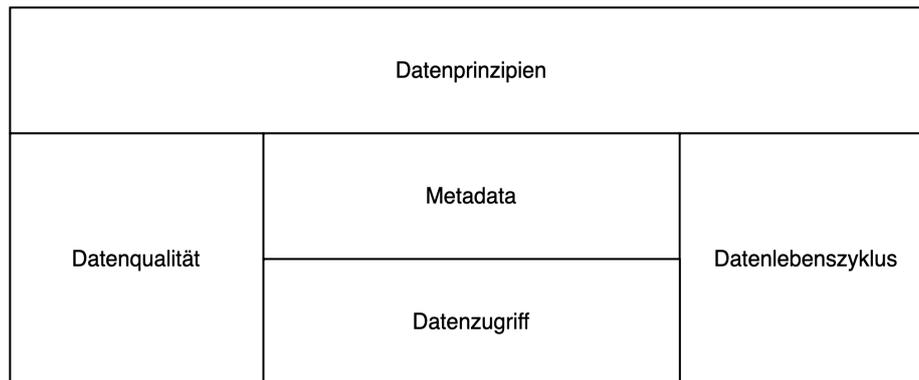


Abbildung 3.7: Die fünf Entscheidungsdomänen nach Khatri und Brown [41]

und dienen daher als Grundlage für die in dieser Arbeit verwendeten Domänenmatrix. Die fünf Domänen Datenprinzipien, Datenqualität, Metadata, Datenzugriff und Datenlebenszyklus sind dabei miteinander verknüpft. Datenprinzipien legen allgemeingültige Regeln und Richtlinien im Umgang mit Daten als strategisches Vermögensgut fest und dienen als Leitfaden für die anderen Domänen. Datenqualität ist für die Zuverlässigkeit und Genauigkeit der Daten verantwortlich, um fundierte Entscheidungen treffen zu können. Diese sind gleichzeitig die Grundlage für die Domänen Metadaten und Datenzugriff. Nur qualitativ hochwertige Daten liefern genaue und zuverlässige Metadaten. Ebenso müssen Qualitätsstandards eingehalten werden, um Zugriffsrechte verteilen zu können. Unternehmen und Organisationen müssen für einen sicheren Umgang mit Daten den gesamten Lebenszyklus dieser berücksichtigen.

Alhassan et al. haben mit ihrem Literaturvergleich in [4] unter anderem untersucht, wie viele Publikationen sich auf die unterschiedlichen Phasen von Data Governance, Definition, Implementierung und Überwachung, in Unternehmen fokussieren (siehe Abbildung 2.1). Die Autoren projizieren diese Phasen auf die fünf Entscheidungsdomänen von Khatri und Brown und verwenden diese Einteilung zur Analyse der bestehenden Literatur. Alle identifizierten Konzepte von Data Governance werden in sogenannte acht Bereiche der Data Governance eingeteilt. Diese umfassen die eigentlichen Maßnahmen und Strukturen, die für die Umsetzung der Data Governance Strategie notwendig sind. Um die Analyse der Frameworks übersichtlicher zu gestalten, werden die acht Bereiche zu drei übergeordneten Kategorien zusammengefasst. Governance Struktur setzt sich zusammen aus den Bereichen Datenrollen und Verantwortlichkeiten, Datenrichtlinien und Datenleitlinien, Datenoperationen besteht aus den Bereichen Datenprozessen, Datenstandards und Datenanforderungen und Datenstrategie und Technologien beschreibt die

beiden einzelnen Komponenten. Diese Reduktion ermöglicht eine fokussiertere Analyse, ohne dabei die entscheidenden Elemente von Data Governance zu vernachlässigen.

Aufgrund dessen, dass diese Arbeit den Anspruch hat, praktikable Lösungen zu präsentieren, wird diese Matrix zur Bewertung der Frameworks verwendet. Darin wird abgebildet, ob ein Framework die relevanten Datendomänen von Khatri und Brown in den verschiedenen Phasen von Data Governance abdeckt. Damit soll die Praktikabilität der Frameworks untersucht werden, und sicher stellen, dass nicht ausschließlich theoretische Konzepte entwickelt werden, sondern auch klare Anweisungen und Richtlinien zur Umsetzung und Verbesserung der Prozesse vorhanden sind. In Abbildung 3.8 ist die Domänenmatrix, die die Abdeckung der fünf zentralen Entscheidungsdomänen mit den acht Bereichen von Data Governance in allen Phasen visualisiert.

#### 3.3.2 Evaluierungsszenarien

In den nachfolgenden Evaluierungsszenarien kann die Praktikabilität der Frameworks an realen Herausforderungen untersucht werden. Die Szenarien basieren auf den in Unterabschnitt 3.2.5 identifizierten Problemen und Herausforderungen und sollen die Herangehensweise der Data Governance Strategien demonstrieren. Dadurch können potenzielle Anwender nicht nur die theoretische Vollständigkeit beurteilen, sondern auch die praktische Anwendbarkeit und Nutzen.

#### Datenqualitätsprobleme

Aufgrund der großen Datenmengen und Vielzahl an unterschiedlichen Datensätzen in der AquaFlow GmbH ist die Erfassung von Datenqualität ein unübersichtliches Thema geworden. Nutzer von Daten, insbesondere bei denen, die nicht direkt aus ihrem Arbeitsumfeld stammen, stellen sich häufig Fragen wie *wie gut sind die vorliegenden Daten* oder *sind beide Datensätze an einem Speicherort von gleicher Qualität*. Diese Unsicherheiten führen zu Verzögerungen und Unsicherheiten bei der Nutzung und Analyse der Daten.

Konkrete Probleme in Bezug auf die Datenqualität sehen in der AquaFlow GmbH wie folgt aus:

- Offene Fragen über Soll-Zustände und Anforderungen an Daten
- Unterschiedliche Darstellungen und Formate für informationsgleiche Daten

Phasen	Bereiche von Data Governance	Entscheidungsdomänen				
		Datenprinzipien	Datenqualität	Metadata	Datenzugriff	Datenlebenszyklus
Definition	Governance Struktur					
	Daten Operationen					
	Datenstrategie & Technologien					
Implementierung	Governance Struktur					
	Daten Operationen					
	Datenstrategie & Technologien					
Überwachung	Governance Struktur					
	Daten Operationen					
	Datenstrategie & Technologien					

Abbildung 3.8: Domänenmatrix, angelehnt an [4]

- Überwachung der Datenqualität
- Prozesse bei neuen Datensätzen
- Duplikate von Datensätzen

#### **Fragen zu Datenzugriffen**

In Bezug auf Datenzugriffe bestehen in der AquaFlow GmbH zahlreiche Herausforderungen, die oft durch fehlende formelle Verantwortlichkeiten entstehen. Diese Unklarheiten führen dazu, dass Nutzer nicht wissen, an wen sie sich wenden müssen, um Zugriff auf bestimmte Daten zu erhalten. Darüber hinaus ist häufig unklar, wer für den Schutz der Daten, die Gewährleistung der Datenqualität und die Verwaltung anderer wichtiger Aspekte zuständig ist. Diese Unsicherheiten resultieren in Verzögerungen und Ineffizienzen bei der Nutzung und Analyse der Daten. Sensible Daten sind oft für alle Anwender zugänglich, was erhebliche Sicherheitsrisiken mit sich bringt.

Konkrete Probleme in Bezug auf Datenzugriffe sehen wie folgt aus:

- Unklarheit über Datenherkunft
- Keine Informationen über die Existenz vorhandenen Datenbestände
- Fehlende Zugriffsprotokolle und -kontrollen
- Wie erhalte ich Zugriff zu diesen Daten?
- Übersicht über aktuelle Zugriffsrechte
- Rollenwechsel eines Mitarbeiters

#### **Einhaltung neuer gesetzlicher Vorschriften**

Unsicherheiten bei der Einhaltung von gesetzlichen Vorschriften sind häufig ein wichtiger Treiber von Data Governance Aktivitäten. Die Nichteinhaltung kann weitreichende Konsequenzen nach sich ziehen. Neben direkten Folgen wie finanziellen Strafen und rechtlichen Konsequenzen können auch indirekte Folgen wie der Verlust von Kundenvertrauen und Schaden an der Reputation des Unternehmens eintreten. Anwender und Verantwortliche sind häufig mit der Komplexität der zu beachtenden Gesetze und Standards

konfrontiert. Dies betrifft insbesondere personenbezogene Daten, deren Schutz durch Gesetze wie die DSGVO geregelt ist. Branchenspezifische Vorschriften und Best-Practices können beim Umgang mit den Daten behilflich sein.

Konkrete Probleme in Bezug auf die Einhaltung gesetzlicher Vorschriften sehen in der AquaFlow GmbH wie folgt aus:

- Fehlende Übersicht über geltende gesetzliche Anforderungen und Vorschriften
- Unklare Verantwortlichkeiten für die Einhaltung und Überwachung der gesetzlichen Vorschriften
- Mangelnde Dokumentation und Nachverfolgbarkeit von Compliance-relevanten Aktivitäten
- Unzureichende Schulungen und Sensibilisierungsmaßnahmen für Mitarbeiter

## 4 Konzeption eines Frameworks

In Abschnitt 2.4 werden die Voraussetzungen und Herausforderungen, vor denen KMUs insbesondere im Bezug auf die Datenverwaltung und -nutzung stehen, erläutert. Durch zunehmende Datenmengen, der immer komplexer werdenden Datenlandschaft innerhalb eines Unternehmens und immer umfangreicheren Datenschutzbestimmungen müssen KMUs sicherstellen, dass ihre Daten zum einen sicher, geschützt und gesetzeskonform gespeichert und genutzt werden. Gleichzeitig ist es unerlässlich, den maximalen geschäftlichen Nutzen aus den verfügbaren Daten zu ziehen, um in der globalisierten Geschäftswelt konkurrenzfähig zu bleiben. In der Literatur und der Geschäftswelt gibt es aus diesem Grund in letzter Zeit verstärkte Bemühungen, geeignete Datenverwaltungsmaßnahmen und Vorgehensweisen zu entwickeln. Daraus sind eine Vielzahl an Data Governance Frameworks aus den verschiedenen Bereichen entstanden. Diese sind jedoch meist für große Unternehmen konzipiert, da in diesen die Herausforderungen der Datenverwaltung zuerst auftraten. Viele Autoren geben an, dass ihre Frameworks auch herunter skaliert werden können, jedoch fehlen in der Literatur dafür konkrete Beispiele. Aus diesem Grund wird in dieser Arbeit ein Data Governance Framework entwickelt, das die spezifischen Anforderungen und Herausforderungen von KMUs berücksichtigt und in der Praxis umsetzbar ist. Hierfür werden die in Abschnitt 2.3 vorgestellten Frameworks analysiert, um zu bestimmen, welche Elemente und Funktionen für KMUs geeignet sind und übernommen werden können. Ziel ist es, ein praxisnahes und skalierbares Data Governance Framework zu schaffen, das KMUs dabei unterstützt, ihre Daten effizient zu verwalten.

### 4.1 Analyse bestehender Frameworks

Im Folgenden werden die in Abschnitt 2.3 vorgestellten Frameworks analysiert. Dabei wird untersucht, welche Elemente auch für KMUs relevant sind und möglicherweise in das hier erarbeitete Framework einfließen können. Durch diese Analyse soll die Forschungsfrage 2 beantwortet werden.

Bei der Untersuchung werden zunächst die Kernaspekte und besonderen Stärken der Frameworks identifiziert, um zu erkennen, worin sich das jeweilige Framework hervorhebt und was es besonders gut kann. Im Anschluss daran werden die Aspekte und Merkmale aufgezeigt, die das Framework für KMUs ungeeignet machen bzw. für KMUs angepasst werden müssen. Zum Schluss werden die Elemente und Funktionen der vorgestellten Frameworks identifiziert, die für KMUs besonders relevant sind und in das in dieser Arbeit ausgearbeitete Konzept einfließen sollen.

### 4.1.1 DAMA

DAMA bietet eine umfassende und strukturierte Herangehensweise an Datenmanagement. Der Fokus ist dabei darauf ausgelegt, eine standardisierte Sprache und klare Richtlinien zu etablieren, die die Kommunikation und die Zusammenarbeit innerhalb eines Unternehmens sowie in der Forschung und Literatur ermöglichen. Mit den elf Schlüsselbereichen ist es sehr umfangreich und unterstützt Anwender in allen Gebieten des Datenmanagements. Data Governance ist dabei das zentrale Bindeglied.

Rollen und Verantwortlichkeiten spielen eine bedeutende Rolle im Framework von DAMA. Durch sie sind Aufgaben und Zuständigkeiten genau geregelt, was zu einer besseren Entscheidungsfindung führt. Die Zuteilung von Verantwortlichen für Datensätze, Entscheidungen und Prozesse fördert eine höhere Effizienz sowie Qualität der Abläufe und Daten. Außerdem werden die Abläufe transparenter und nachvollziehbarer und die Sicherheit der Daten wird erhöht, da spezifische Personen dafür zuständig sind.

Mit Prinzipien und Richtlinien legt DAMA grundlegende Regeln fest, wie Daten innerhalb des Unternehmens erhoben, gespeichert, verarbeitet und genutzt werden. Durch sie soll es in allen Bereichen des Daten Managements klare Vorgaben und Best-Practices geben, was die Konsistenz, Qualität und Sicherheit der Daten erhöht. Zudem unterstützen sie bei der Einhaltung gesetzlicher und regulatorischer Anforderungen, wodurch Risiken minimiert werden.

Datenqualität ist ein wichtiger Treiber der Data Governance Aktivitäten bei DAMA und spielt eine entscheidende Rolle für den Erfolg des Unternehmens. Eine hohe Datenqualität stellt sicher, dass die Daten des Unternehmens vollständig, konsistent und aktuell sind, wodurch verlässliche Analysen und Entscheidungen möglich sind. Die Implementierung von Prozessen und Maßnahmen für die Überwachung und Verbesserung der Qualität ist dabei ein wesentlicher Bestandteil. Dies umfasst die Erstellung von Datenstandards,

regelmäßige Prüfungen und Audits sowie Schulungen der Mitarbeiter im Umgang mit Daten.

Ein geeignetes Data Governance Reifegradmodell ist wichtig für den Erfolg der Data Governance Bemühungen. Dieses ermöglicht, den aktuellen Stand des Programmes zu bewerten und Bereiche zu identifizieren, die optimiert werden müssen. Dadurch ist es Unternehmen möglich, ihre Prozesse und ihr Verhalten schrittweise zu verbessern und so auch langfristige Ziele zu erreichen. Außerdem werden Fortschritte messbar und sichtbar gemacht, was sich wiederum positiv auf die Unternehmenskultur auswirken kann.

Das von DAMA in DMBOK vorgestellte Data Governance Framework kann für KMUs schwer umzusetzen sein, da es alleine durch den Umfang des Frameworks eine Komplexität bietet, die für kleine Unternehmen und Verantwortliche in diesen mit weniger Expertise überfordernd sein kann. Daher müssen KMUs vorher die Komplexität reduzieren und das Framework praxisorientierter gestalten. Zudem ist es für KMUs schwierig, Maßnahmen und Funktionen aus allen elf Bereichen umzusetzen. Hier müssen Domänen priorisiert und in eine Roadmap oder Reifegradmodell eingearbeitet werden, um Verbesserungen schrittweise zu ermöglichen. Generell ist das Framework generisch und allgemein gehalten, um für alle Anwender relevant zu sein. KMUs benötigen jedoch oft mehr Anleitungen und Vorgaben, da ihnen häufig die Expertise und Kapazität fehlen, sich intensiv mit den umfangreichen Inhalten des Frameworks zu beschäftigen.

DAMA bietet jedoch auch eine Reihe von Aspekten, die für KMUs geeignet sind. Die Schaffung grundlegender Prinzipien, die für alle Bereiche gelten, ist sinnvoll, um eine Basis für den Umgang mit Daten zu schaffen. Außerdem werden dadurch Best-Practices gefördert und einheitliche Standards etabliert. Ebenso wie in allen Frameworks ist eine klare Rollenverteilung und die Zuweisung von Verantwortlichkeiten essenziell, damit notwendige Aufgaben wie die Einhaltung und Überwachung der Datenqualität auch tatsächlich ausgeführt werden. Die Einführung und Betonung eines Reifegradmodells ist ebenfalls für KMUs von großem Nutzen, um zum einen den aktuellen Stand und derzeit noch auftretende Probleme zu erkennen, aber auch um die nächsten Verbesserungsmaßnahmen im Blick zu haben.

### 4.1.2 DGI

Das gesamte DGI Data Governance Framework ist wertbasiert. Der Fokus in allen Phasen und Prozessen liegt auf den Vorteilen und Ergebnissen des Programmes. Dafür müssen

klare, messbare und unternehmensspezifische Ziele und Metriken formuliert werden, welche die Grundlage für das gesamte Handeln bilden. Im Laufe des Programmes werden die Fortschritte kontinuierlich mit den formulierten Werten und Zielen abgeglichen, um das Programm gegebenenfalls anzupassen.

Kontrollen und Kontrollmechanismen sind ein wichtiger Baustein des DGI Frameworks. Mithilfen von Kontrollen wird nicht nur die Sicherheit der Daten kontinuierlich überprüft und sichergestellt, sondern auch überprüft, dass der größtmögliche Mehrwert aus den Daten gezogen wird. Wichtig dabei ist, dass die Kontrollen den gesamten Lebenszyklus abdecken. Kontrollen werden hauptsächlich mit Kontrollmechanismen umgesetzt. Diese können sowohl präventiv als auch investigativ sein.

Mit definierten Regeln werden klare Vorgaben für alle Prozesse und Verantwortlichkeiten gemacht. Dies umfasst sowohl Richtlinien, Definitionen und Glossare als auch Regeln für Entscheidungsrechte. Dort wird festgelegt, wer wann welche Entscheidung treffen darf oder muss. Durch die Vorgaben im Kontext der Entscheidungsrechte wird eine klarere Rechenschaftspflicht im Unternehmen und damit auch ein bewussterer Umgang mit Daten gefördert. Außerdem sorgen die Regeln für einen konsistenten Umgang mit den Daten und stellen sicher, dass branchenspezifische Best-Practices umgesetzt werden.

Menschen und Prozesse sind ebenfalls entscheidend für den Erfolg des Programmes. Laut DGI sollte ein spezialisiertes Team für die Förderung und das Vorantreiben der Data Governance Strategien verantwortlich sein. Sie sorgen unter anderem für die Unterstützung aus der Managementebene, die tägliche Umsetzung des Programmes und der Richtlinien und Standards sowie die Überwachung der Metriken und der Einhaltung der gesetzten Ziele. Gleichzeitig müssen dokumentierbare und wiederholbare Prozesse bestehen, die eine effiziente Durchführung ermöglichen. Für beide Aspekte ist eine klare Kommunikation wichtig.

Um das DGI Framework für KMUs nutzbar zu machen, wären einige Anpassungen erforderlich. Insgesamt ist das Framework mit zehn formulierten Bereichen ebenfalls sehr umfangreich, was es für kleinere Unternehmen oft schwer umsetzbar macht, da ihnen die personellen und finanziellen Mittel fehlen. DGI gibt selber an, dass ihr Framework für größere Unternehmen gedacht ist. Während bei kleineren Unternehmen oft noch ein informelles Data Governance System ausreichen kann, sei dies bei größeren Unternehmen nicht der Fall. Ebenfalls erfordert das Framework ein spezialisiertes Team aus Experten aus verschiedenen Bereichen, um die Data Governance Aktivitäten voranzutreiben. KMUs verfügen jedoch häufig nicht über die erforderlichen Fachkräfte, um so ein Team

zu etablieren. Ähnlich verhält es sich mit den Ressourcen für die Umsetzung der Kontroll- und Überwachungsmechanismen. Durch den Fokus von DGI auf größere Unternehmen gibt es keine spezifische Vorgehensweise für die Anpassung der Vorgänge auf die Bedürfnisse und Herausforderungen von KMUs. Aus den genannten Gründen ist das DGI Data Governance Framework nicht ohne Anpassungen für KMUs mit begrenzten Ressourcen und spezifischen Anforderungen anwendbar.

Gleichzeitig können aber auch einige der Aspekte und Maßnahmen für ein Framework für KMUs übernommen werden. Ein wertebasiertes Handeln, bei dem eine klare Mission, Ziele und messbare Metriken vorgegeben werden, hilft auch kleineren Unternehmen, Data Governance Aktivitäten in ihrem Alltag zu berücksichtigen. Ebenso ist ein Fokus auf die Ausübung der Data Governance Aktivitäten und der Einsatz von Kontrollen während des gesamten Lebenszyklus der Daten wichtig, um die Sicherheit und Integrität der Daten zu gewährleisten. Ein weiterer wichtiger Punkt ist die Benennung eines oder mehrerer Experten, die Data Governance vorantreiben. Im Vergleich zu DGI muss dieses Team jedoch nicht zwingend aus Fachkräften aus vielen unterschiedlichen Bereichen bestehen. Zudem sollte auch ein größerer Fokus auf Kommunikation und Dokumentation liegen, um alle Beteiligten stets informiert zu halten und einen Austausch zu fördern, sobald Probleme oder Verbesserungen erkannt werden. Ebenfalls sollten Gedanken über die Nutzung verschiedener Datenprodukte wie Data Warehouses, Datenkataloge, Dashboards und Metadaten-Management-Systeme frühzeitig angestellt werden, um zu gewährleisten, dass Daten vertrauensvoll, strukturiert und zugänglich sind. Diese Datenprodukte sollen vertrauenswürdige Informationsquellen sein. Schließlich kann eine Aufteilung der Aktivitäten in *little g* und *Big G* sinnvoll sein, um zwischen strategischen und operativen Aufgaben zu unterscheiden.

### 4.1.3 PwC

Das Framework von PwC umfasst eine Reihe von Kernaspekten. Die Strategie ist dabei ein wesentlicher Bestandteil und gibt die Richtung aller Aktivitäten vor. Dabei werden die beiden Haupttreiber, Kontrolle bzw. Sicherheit sowie die Monetarisierung berücksichtigt und damit die Data Governance Strategie genau mit der Unternehmensstrategie abgestimmt. Dies stellt sicher, dass Data Governance Initiativen direkt zur Erreichung der Geschäftsziele beitragen.

Ebenso bezieht PwC bestehende Datenmanagement Prozesse, Infrastrukturen und Konzepte in die Data Governance Überlegungen mit ein. Dadurch ist eine bessere und effizientere Integration neuer Konzepte und Aktivitäten möglich und vorhandene Ressourcen werden optimiert und die Effizienz der Datenverwaltung gesteigert. In Bezug auf bestehende Prozesse und die Denkweise der Mitarbeitenden spielt auch *Change Management* (dt. Veränderungsmanagement) eine Rolle. Neue Konzepte und Ideen stimmen möglicherweise nicht mit den bestehenden überein und es braucht Schulungen und Zeit, diese zu akzeptieren und umzusetzen.

Der Datenlebenszyklus wird ebenso als einer der Kernaspekte betrachtet. Dies umfasst die Verwaltung des gesamten Datenflusses. Dies reicht von der Erfassung und Speicherung sowie Nutzung bis hin zur Archivierung beziehungsweise Löschung. Dadurch wird gewährleistet, dass das Unternehmen über die gesamte Lebensspanne die Kontrolle über die Daten hat. Dadurch bleiben sie nutzbar und helfen dabei, gesetzliche Vorschriften einzuhalten. Außerdem wird so der Überblick über alle Daten behalten.

Eine zentrale Schicht des Frameworks sind die Data Governance Kernfunktionen, die sich aus verschiedenen Komponenten zusammensetzen. Dazu gehören die Datenverwaltung (Data Stewardship) mit Aspekten wie Datenqualität, Datenschutz und Besitzrechten sowie die kontinuierliche Verbesserung und technologische Weiterentwicklung (Innovation), welche die Erstellung und Pflege von Datenprodukten wie Glossaren und Metadaten-Management-Systemen umfasst.

PwC legt ebenfalls einen großen Fokus auf die sogenannten Enabler, die für die tatsächliche Umsetzung der gesamten Data Governance Strategie verantwortlich sind. Dazu gehören Mitarbeiter, Prozess- und Betriebsmodelle sowie geeignete technologische Werkzeuge. Diese schaffen die benötigten Voraussetzungen und unterstützen bei der Implementierung der Aktivitäten.

Eine eins zu eins Umsetzung des PwC Data Governance Frameworks kann für KMUs aufgrund verschiedener Komponenten herausfordernd sein. Zum einen gibt es im Framework keine klare Priorisierung der verschiedenen Maßnahmen und Initiativen, was dazu führen kann, dass KMUs versuchen, alle Aspekte gleichzeitig umzusetzen. Dies ist jedoch oft zu ressourcenintensiv für kleinere Unternehmen. Zum anderen setzt das Framework voraus, dass bestehende Datenmanagementprozesse bereits eine gewisse Reife haben, was bei KMUs nicht immer der Fall ist. Ohne eine ausgereifte Basis kann die Implementierung der vorgeschlagenen Maßnahmen komplex und ineffizient werden.

Insgesamt bietet das PwC Data Governance Framework eine gute Grundlage für ein geeignetes Framework für KMUs. Die Orientierung der Data Governance Ziele an der bisherigen Unternehmensstrategie stellt sicher, dass wirklich nur benötigte Governance Aktivitäten umgesetzt werden. Außerdem erleichtert dies das Veränderungsmanagement, da die Governance Ziele mit den bereits bekannten übereinstimmen und von allen Beteiligten akzeptiert werden, was die Einführung neuer Prozesse und Maßnahmen reibungsloser gestaltet. Ebenfalls ist es für KMUs sinnvoll, sich auf die tägliche Umsetzung zu fokussieren. Dies wird insbesondere durch ein geeignetes operatives Modell und weitere Enabler erreicht. Auch die Einbindung der bestehenden Prozesse im Bereich Datenmanagement in das neue Framework ermöglichen eine bessere Integration, da Data Governance nicht nur darauf aufgesetzt wird, sondern Teil davon wird. Der Fokus auf den gesamten Lebenszyklus wird ebenso wie bei DGI übernommen.

### 4.2 Bewertung der Frameworks

Im Folgenden werden die Erkenntnisse aus den Analysen der drei Data Governance Frameworks zusammengefasst und zusammengeführt. Die Kernelemente, die Punkte bei denen Anpassungsbedarf für die Umsetzung eines KMUs besteht und das Relevante für das auf KMU-spezialisierte Framework sind in Abbildung 4.1 zusammengefasst.

Basierend auf den durchgeführten Analysen und der besonderen Herausforderungen und Voraussetzungen von KMUs ergeben sich folgende Elemente, auf die bei der Entwicklung des spezialisierten Frameworks besonders Augenmerk gelegt wird.

- Übereinstimmung mit der Unternehmensstrategie (PwC)
- Fokus auf Enabler und operatives Modell (PwC)
- Unterscheidung little g und Big G (DGI)
- Wertebasiertes Handeln und klare Zielformulierung (DGI)
- Reifegradmodell (DAMA)
- Formulierung grundlegender und allgemeingültiger Prinzipien und Regeln (DAMA und DGI)
- Benennung eines Führungsteams oder Führungsperson (DAMA und DGI)

	DAMA	DGI	PwC
Kernelemente	<ul style="list-style-type: none"> <li>• ganzheitlicher Ansatz mit Data Management zusammen</li> <li>• Rollen &amp; Verantwortlichkeiten</li> <li>• Prinzipien und Richtlinien für jede Domäne</li> <li>• Datenqualität</li> <li>• Reifegradmodell</li> </ul>	<ul style="list-style-type: none"> <li>• wertebasiertes Handeln</li> <li>• Entscheidungsrechte &amp; Verantwortlichkeiten</li> <li>• Kontrollen &amp; Standards</li> <li>• Menschen und Prozesse</li> </ul>	<ul style="list-style-type: none"> <li>• Data Governance Strategie</li> <li>• Daten Management</li> <li>• Datenlebenszyklus</li> <li>• Verantwortlichkeiten für Daten</li> <li>• Data Governance Ermöglicher</li> </ul>
Anpassungsbedarf	<ul style="list-style-type: none"> <li>• Reduzierung der Komplexität</li> <li>• Priorisierung der wichtigsten Domänen</li> <li>• Konkretere Vorgaben für Implementierung und Überwachung</li> </ul>	<ul style="list-style-type: none"> <li>• Vereinfachung der Prozesse</li> <li>• Anpassung der Rollen</li> <li>• Reduzierung der Komplexität</li> </ul>	<ul style="list-style-type: none"> <li>• Priorisierung der wichtigsten Bestandteile</li> <li>• Unterstützung für Kulturwandel</li> </ul>
Relevantes für KMUs	<ul style="list-style-type: none"> <li>• Einhaltung grundlegender Prinzipien</li> <li>• Klare Rollenverteilung</li> <li>• Definieren von Richtlinien und Standards</li> <li>• Einführung eines Reifegradmodelles</li> </ul>	<ul style="list-style-type: none"> <li>• klare Ziele und Metriken</li> <li>• Definierte Entscheidungsrechte</li> <li>• Datenprodukte</li> <li>• Standards für Datenflüsse</li> <li>• Fokus auf Zusammenarbeit</li> </ul>	<ul style="list-style-type: none"> <li>• Übereinstimmung mit Unternehmensstrategie</li> <li>• Fokus auf operatives Modell</li> <li>• Einbindung in tägliche Arbeit</li> <li>• Förderung einer Datenkultur</li> <li>• DG nur mit Enablern möglich</li> </ul>

Abbildung 4.1: Vergleich der Kernelemente, relevanten Aspekte für KMUs und Anpassungsbedarf der Data Governance Frameworks von DAMA, DGI und PwC

- Datenprodukte (DGI)
- Einbindung bestehender Prozesse und Datenmanagement (PwC)
- Klare Kommunikation und Dokumentation (DGI)

### 4.3 CoreGovernance Framework

Aus den Erkenntnissen der vorherigen Abschnitte, in denen die etablierten Data Governance Frameworks von DAMA, DGI und PwC untersucht wurden, sowie den in Abschnitt 2.4 beschriebenen besonderen Herausforderungen und Voraussetzungen, ergibt sich das für KMUs spezialisierte *CoreGovernance* Framework. Der Name verdeutlicht, dass das Framework den Fokus auf die wesentlichen Kernaspekte von Data Governance legt. Das Modell versucht dabei, einerseits die wichtigsten Elemente der bewährten Frameworks zu integrieren und andererseits die Besonderheiten von KMUs zu berücksichtigen.

#### 4.3.1 Übersicht

In Abbildung 4.2 ist die grafische Übersicht des CoreGovernance Frameworks zu sehen. Das Framework ist in drei Dimensionen aufgebaut, die hauptsächlich zur besseren Verständlichkeit beitragen. Auf der linken Seite sind die drei Reihen in drei Anwendungsbereiche unterteilt. Zum einen gibt es die Geschäftsebene, die alle bestehenden Geschäftsaktivitäten wie die strategische Ausrichtung, BI und Datenmanagement umfasst. Ganz unten befindet sich die Data Governance Ebene, die alle spezifischen Data Governance Elemente enthält. Auf der mittleren Ebene geht es um die tägliche Umsetzung von Data Governance. Die Elemente werden wie bei PwC ebenfalls Enabler genannt und bilden das Rückgrat des Frameworks. Sie umfassen das operative Modell, das die täglichen Aufgaben, Meetings und Prozesse beschreibt. Technologien und eine entsprechende Dokumentation unterstützen das operative Modell dabei.

Die zweite Dimension auf der rechten Seite umfasst die gleichen Ebenen wie die Dimension Anwendungsbereiche. Hier wird der Fokus jedoch auf die Planungsphase zu Beginn der Data Governance Bemühungen gelegt. Diese Dimension verdeutlicht den Anwendern, was in den verschiedenen Ebenen getan werden muss. Strategie und Datenmanagement existieren bereits im Unternehmen und müssen überarbeitet und angepasst werden, um

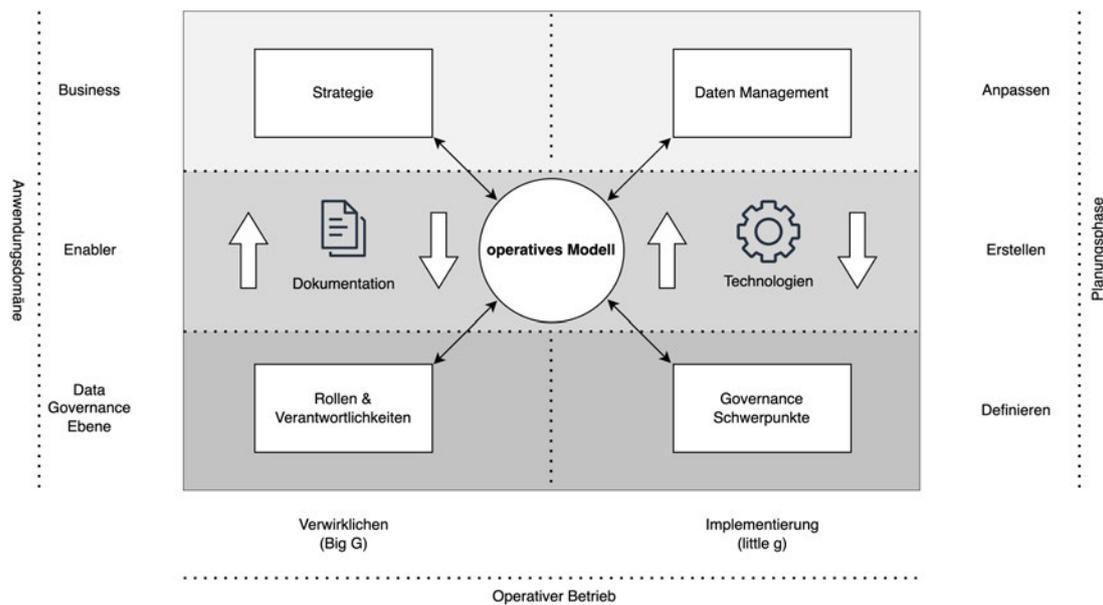


Abbildung 4.2: Übersicht des CoreGovernance Frameworks

mit den Data Governance Zielen im Einklang zu stehen. Rollen, Verantwortlichkeiten, Datenkonzepte, Regeln und Richtlinien in der untersten Ebene müssen definiert und formell festgelegt werden. Das bedeutet die Schaffung neuer Strukturen und Konzepte. Enabler wie das operative Modell und Technologien müssen erstellt und umgesetzt werden. Dies beinhaltet hauptsächlich die technische und organisatorische Entwicklung und Implementierung der Systeme, um als operative Werkzeuge der Data Governance dienen zu können.

Die dritte Dimension beschäftigt sich mit dem Einsatz und den Aufgaben der Bereiche im operativen Betrieb. Diese Dimension halbiert die Abbildung vertikal und ist wie bei DGI in große und kleine Data Governance Aktivitäten aufgeteilt (*little g und Big G*). Die großen Aktivitäten beinhalten langfristige strategische Aufgaben wie die Definition und Anpassung der Rollen, Anpassung der Unternehmensstrategie sowie die Dokumentation und Überwachung der Governance Schwerpunkte. Die kleinen Aktivitäten umfassen operative Aufgaben und Bereiche, die regelmäßig und im Alltag durchgeführt werden. Dazu gehören die Erstellung und Pflege von Datenprodukten sowie die Überwachung der Datenqualität und die Einhaltung der Datenschutzvorgaben. Diese Aktivitäten sorgen dafür, dass Data Governance im Geschäftsalltag integriert ist und kontinuierlich angepasst wird.

### 4.3.2 Kernkomponenten

Im Folgenden erfolgt eine detaillierte Betrachtung der Kernkomponenten. Dabei werden auch Komponenten berücksichtigt, die in der grafischen Übersicht nicht explizit auftauchen. Der Fokus liegt dabei insbesondere auf der Bedeutung der Komponenten für KMUs.

#### **Strategie/Vision**

Eine geeignete Strategie und Vision bilden die Grundlage für eine erfolgreiche Umsetzung von Data Governance. Strategie definiert dabei die allgemeinen Aktionsziele und -vorgänge und gibt die Richtung für jede Entscheidung, jeden Prozess und jede Maßnahme vor. Sie stellt somit die langfristige Verhaltensweise für das Erreichen langfristiger Ziele dar [53] und beinhaltet folglich nicht nur eine Sammlung von Zielen sondern auch konkrete Maßnahmen und Vorgehensweisen. Dabei ist wichtig, die Ziele und Strategie allen Beteiligten verständlich zu machen, damit alle stets die Mission vor Augen haben und den mit Data Governance verbundenen Mehraufwand akzeptieren und mittragen. Dafür sollten Ziele realistisch und greifbar formuliert sein mit Berücksichtigung der vorhandenen Expertise und Voraussetzungen des Unternehmens. Dadurch bleiben sie ambitioniert, sind jedoch auch erreichbar. Des Weiteren ist es entscheidend, dass die formulierten Ziele im Einklang mit den Geschäftszielen stehen. Das große Ziel von Data Governance bleibt, diese bei der Erreichung zu unterstützen. Dafür sollten die aktuell Wichtigsten priorisiert werden, um Ressourcen effizient einzusetzen. Für jedes Ziel sollten Metriken definiert werden, um kontinuierlich den Fortschritt und Erfolg der eingeleiteten Maßnahmen bewerten zu können. Dadurch kann das Programm gezielt überwacht und bei Bedarf Anpassungen vorgenommen werden.

#### **Datenmanagement**

Datenmanagement ist ebenfalls ein Element, das bereits im Unternehmen existiert und durch die Einführung von Data Governance angepasst wird. Im Fokus dabei sollten Verbesserungen bei Aspekten wie Datenqualität, Sicherheit und Verfügbarkeit der Daten stehen. Generell geht es bei den Anpassungen beim Datenmanagement hauptsächlich darum, die Elemente und Maßnahmen, die von den Governance Schwerpunkten definiert

werden, umzusetzen. Hierzu gehören unter anderem der Einsatz von Standards und Richtlinien. Auch die Einführung von Zugriffsrechten und die Datendemokratisierung spielen eine wichtige Rolle bei der Planung des Datenmanagements. Dadurch wird ebenfalls die Sicherheit sowie der Nutzen der Daten erhöht.

Ein weiterer zentraler Aspekt ist die Implementierung eines Lebenszyklusmodells für die Daten. Dadurch wird gewährleistet, dass zu jedem Zeitpunkt die Kontrolle über die Daten herrscht und fundierte Entscheidungen getroffen werden. Ohne dieses Modell können Daten im Unternehmen verweilen, bei denen die Herkunft, Relevanz und ursprünglicher Zweck unklar sind. Solche *Datenleichen* können zu potenziellen Risiken und Ineffizienzen führen.

Allgemeine Datenkonzepte und Datenarchitekturen sind ebenfalls wesentliche Bestandteile des Datenmanagements. Sie sorgen dafür, dass Datenstrukturen effizient und skalierbar sind, was auch für KMUs relevant ist. Aufgrund der Komplexität und des Umfangs dieser Themen werden sie in dieser Arbeit jedoch nicht im Detail behandelt, da der Schwerpunkt auf der Implementierung von Data Governance Maßnahmen liegt.

In Zusammenarbeit mit der Strategie und Verantwortlichen der Geschäftsdomänen ist die Priorisierung der Datendomänen ein wichtiger Schritt. Aufgrund der eingeschränkten Ressourcen sollten KMUs zunächst die wichtigsten Datendomänen identifizieren, deren Verwaltung optimieren und die Governance Schwerpunkte darauf fokussieren. Wichtige Domänen können zum einen besonders sensitive Daten sein sowie Datensätze, in denen großes Potenzial steckt und die besonders wichtig für Geschäftsprozesse sind. Nach der erfolgreichen Umsetzung der Maßnahmen in diesen Bereichen können die Prozesse verbessert und schrittweise auf weitere Datendomänen ausgeweitet werden.

Abschließend ist die kontinuierliche Überwachung und Verbesserung der Datenqualität der zentrale Bestandteil dieses Bereichs. Automatisierte Kontrollen stellen sicher, dass die Daten den definierten Standards und Ansprüchen entsprechen und bei Nicht-Einhaltung passende Maßnahmen eingeleitet werden.

#### **Rollen und Verantwortlichkeiten**

Auch wenn bei KMUs im Vergleich zu größeren Unternehmen die Ressourcen, Expertise und Personal limitiert sind, spielen Rollen und Verantwortlichkeiten ebenfalls eine entscheidende Rolle bei der Umsetzung und dem Erfolg der Data Governance Strategie. Die

Rollen und Verantwortlichkeiten bilden ein Rahmenwerk für Entscheidungsfindungen und strukturierte Arbeitsabläufe. Außerdem fördert es das Bewusstsein der Mitarbeiter für ihre Aufgabenbereiche, was Missverständnisse und ineffiziente Abläufe reduziert. Ebenso fördern klare Zuweisungen die Rechenschaftspflicht, wodurch Aufgaben gewissenhafter erledigt werden.

Im Vergleich zu großen Frameworks wie dem von DAMA ist es für KMUs vorteilhafter, mit weniger Rollen zu starten und diese bei Bedarf zu erweitern. Dadurch fokussiert sich das Framework zunächst auf das Wesentliche, kann aber skaliert werden. Folgende Rollen sollten jedoch auf jeden Fall verteilt werden:

Das **DGO** ist das zentrale Gremium aller Data Governance Aktivitäten, wird vom CDO geleitet und sollte auch in einem KMU vorkommen. Hier werden alle wesentlichen Entscheidungen zu Richtlinien, Standards, Prozessen und Normen getroffen sowie die grundlegenden Aspekte der Umsetzung definiert. Darüber hinaus ist das DGO für die Überwachung und Einhaltung der definierten Vorschriften sowie die kontinuierliche Überwachung der Datenqualität verantwortlich. Dadurch übernimmt es in einem KMU auch die Aufgaben des DGC. Es besteht in der Regel aus dem CDO, Data Ownern, Data Stewards sowie bei Bedarf Stakeholdern aus dem Geschäftsbereich. Für die Überwachung erwartet das DGO regelmäßige Berichte der Verantwortlichen. In einem KMU besteht das DGO in der Regel aus den verantwortlichen Personen der unterschiedlichen Datenquellen. Diese verfügen über die entsprechende Geschäftsexpertise und können beurteilen, wie die Daten im Unternehmenskontext strukturiert sein müssen und wie sie am besten eingesetzt werden.

Der **CDO** ist die wichtigste Person im Kontext von Data Governance in KMUs und zeichnet sich durch eine große Expertise aus. Er muss sowohl über technisches als auch über geschäftliches Verständnis verfügen, da er die Schnittstelle zwischen den Geschäftsinteressen der Business-Abteilungen und den technischen Teams, die für die Dateninfrastruktur verantwortlich sind, darstellt. In einem KMU tritt häufig der Fall ein, dass nicht alle Abteilungen umfassend über die spezifischen Anforderungen ihrer jeweiligen Datengebiete informiert sind. Daher muss der CDO ein breitgefächertes Wissen haben und sich kontinuierlich weiterbilden. Eine weitere Aufgabe ist die Identifikation von Risiken im Kontext von Daten und die Sicherstellung, dass sensible Daten stets geschützt und nur autorisierten Personen zugänglich sind. Außerdem sorgt der CDO dafür, dass alle Abteilungen effizient arbeiten und die definierten Maßnahmen des DGO in die Praxis umsetzen. Durch seine zentrale Rolle und Expertise ist er in der Lage, die benötigten

Ressourcen und Zeiten für die Umsetzung der Maßnahmen einzuschätzen und gibt entsprechende Empfehlungen ab.

In KMUs ist es wichtig, die Verantwortung für die Qualität und Sicherheit der Daten den passenden Personen zu übertragen. Diese **Data Owner** sollten über umfassendes geschäftliches Fachwissen verfügen und sich in ihrem jeweiligen Anwendungsgebiet gut auskennen. Häufig haben sie bereits Verantwortung für ihren Bereich und diese Verantwortung wird nun formal auf die Verwaltung der konkreten Daten und Datenprozesse erweitert. Aus diesem Grund ergibt es in KMUs Sinn, bestehende Führungskräfte oder Abteilungsleiter zu Data Ownern zu ernennen, da diese sich bereits mit den operativen Abläufen und geschäftlichen Anforderungen auskennen. Die Aufgaben der Data Owner bleiben dabei grundsätzlich die gleichen wie in allen Unternehmen und sind in Unterabschnitt 2.2.3 ausführlich ausgeführt.

**Data Stewards** sind normalerweise für die operative Verwaltung und Pflege der Daten in ihrem zugewiesenen Datenbereich verantwortlich. In KMUs sind die Teams jedoch häufig relativ klein und verfügen nicht über die Kapazitäten und Expertise, eine Person als Data Steward abzustellen. In solchen Fällen ist es sinnvoll, zunächst keinen speziellen Data Steward in den Teams einzusetzen, sondern den Data Owner zusammen mit dem Team für diese Aufgaben in die Verantwortung zu nehmen. Wächst das Team im Laufe der Zeit, kann skaliert werden und ein spezifischer Data Steward ernannt werden. Laut DAMA „*werden die besten Data Stewards gefunden, nicht gemacht*“ [34]. Zu dem Zeitpunkt der Skalierung gibt es eine Person im Team, die sich am meisten um Data Governance bezogene Aufgaben kümmert und diese sollte als Data Steward eingesetzt werden. Ebenso können zu Beginn der Data Governance Aktivitäten abteilungsübergreifend ein oder mehrere Personen zu Data Stewards ernannt werden, die als Ansprechpartner für datenbezogene Fragen dienen und bei der Implementierung von Richtlinien und Standards unterstützen.

#### **Governance Schwerpunkte**

Die Governance Schwerpunkte umfassen die wesentlichen Maßnahmen und Richtlinien, die neu in das Unternehmen integriert werden und für eine effektive Umsetzung von Data Governance sorgen. Viele dieser Schwerpunkte wurden bereits im Kontext von anderen Frameworks genannt. Im Folgenden werden die Schwerpunkte und Maßnahmen vorgestellt, die für KMUs relevant sind und insbesondere ihr Bezug dazu erläutert.

Die kontinuierliche Überwachung, Sicherstellung und Verbesserung der **Datenqualität** ist einer der zentralen Ziele von Data Governance. Für KMUs bedeutet das, zunächst eine grundlegende Bewertung der vorhandenen Datenbestände durchzuführen. Dabei müssen klare Soll-Zustände und Standards definiert werden, um die Datenqualität systematisch verbessern zu können. Die Einhaltung der gewünschten Datenqualität sollte in sämtliche Entscheidungen miteinbezogen und kontinuierlich überwacht werden. Insbesondere automatisierte Mechanismen zur Überwachung der Qualität sind in diesem Kontext essenziell, um die Datenqualität sicherzustellen. Dadurch können individuelle Fehler reduziert werden und die Überwachung hängt nicht von zeitintensiven manuellen Prüfungen ab.

Wie bereits von PwC und dem Punkt *Datenmanagement* dargelegt, stellt die **Einführung eines Lebenszyklus-Modells** für Daten eine entscheidende Maßnahme dar. Ein solches Modell ermöglicht es KMUs, Datenflüsse innerhalb des Unternehmens zu verstehen, und zu entscheiden, was mit Daten passieren soll, die bereits genutzt wurden oder sich länger im Datenökosystem befinden. Dadurch wird insbesondere das Risiko von Datenmissbrauch oder -verlust reduziert und die Sicherheit der Daten gefördert. Die Verantwortung für die Verwaltung des Lebenszyklus liegt bei den Data Ownern, da diese sich am besten mit der Datendomäne auskennen. Die Entwicklung der geeigneten Prozesse erfolgt in Zusammenarbeit zwischen den Data Ownern und den Data Stewards, da diese sich mit den technischen Möglichkeiten auskennen und diese umsetzen können. Der Prozess kann dabei iterativ mit regelmäßigen Überprüfungen erfolgen, um auf Herausforderungen und regulatorische Änderungen reagieren zu können. Mit automatisiertem Feedback und Audits wird sichergestellt, dass sich die Daten in dem passenden Lebenszyklus befinden.

Ein passendes Konzept für das **Stammdatenmanagement** sollte ebenfalls entwickelt werden. Durch die Etablierung einer *Single Source of Truth* (dt. einzige Quelle der Wahrheit) wird versucht, Datenduplikate zu vermeiden. Dies gewährleistet, dass auch Mitarbeiter aus unterschiedlichen Abteilungen auf korrekte, aktuelle und konsistente Daten zugreifen können.

Wie bereits in Unterabschnitt 2.4.2 erwähnt, stellt das Fehlen klarer Anweisungen zur **Überwachung** eine große Herausforderung für KMUs bei der erfolgreichen Implementierung eines Data Governance Programms dar. Dabei ist dieses wichtig, um sicherzustellen, dass definierte Standards und Richtlinien im Tagesgeschäft tatsächlich umgesetzt und vorhandene Ressourcen gezielt und effizient eingesetzt werden. Für das **Monitoring** werden die in der Strategie festgelegten Metriken und Ziele bedeutsam. Diese bilden die

Grundlage für die Überwachung und Bewertung der Fortschritte. Eine Berichterstattung sollte regelmäßig an eine zentrale Instanz wie den CDO oder das DGO erfolgen. Dies kann durch Berichte mit Einblicken aus den jeweiligen Abteilungen geschehen oder durch automatisierte Systeme, die relevante Daten und Kennzahlen sammeln und auswerten. Es ist sinnvoll, möglichst viele Überwachungsprozesse zu automatisieren und manuelle Kontrollen eher zur Überprüfung der automatisierten Prozesse einzusetzen. Die Ergebnisse dieser Überwachungen sollten den Beteiligten regelmäßig als Feedback mitgeteilt werden, um Transparenz zu schaffen und sicherzustellen, dass alle Akteure auf dem gleichen Informationsstand sind.

Neben Datenqualität ist die **Sicherheit** der Haupttreiber von Data Governance Aktivitäten. KMUs werden vermehrt von Hackern anvisiert, da sie im Vergleich zu größeren Unternehmen häufig schwächere Sicherheitsvorkehrungen haben [13]. KMUs verfügen über weniger Ressourcen und weniger Erfahrung im Umgang mit Hackern und IT-Sicherheitsvorkehrungen. Daher ist es umso wichtiger, dass sensible Daten im Falle eines Angriffs geschützt sind. Dafür ist die Erstellung eines Sicherheits- und Zugriffsmodells entscheidend und dass die Daten klassifiziert werden. Dabei werden sensible, personenbezogene Daten nur verwendet, wenn sie wirklich benötigt werden. Durch die Klassifizierung der Daten und Auflistung in einem Glossar werden **Zugriffsrechte** verteilt und nachvollziehbar.

**Richtlinien und Regeln** bilden die allgemeingültige Grundlage für alle Handlungen und Aktivitäten bezogen auf den Umgang mit Daten. Für KMUs ist es besonders wichtig, klare und präzise Richtlinien zu formulieren, da so der Einsatz von bewährten Verfahren in den Geschäftsalltag integriert werden. Sie helfen dabei, Unsicherheiten und Missverständnisse zu vermeiden und dienen als Orientierungshilfe für ein effiziente und sichere Datenverwaltung. Der Fokus bei der Formulierung der Richtlinien sollte dabei auf den essenziellen Aspekten liegen, um die Übersichtlichkeit zu erhalten. Andernfalls besteht die Gefahr, dass die Beteiligten durch die Vielzahl an Vorgaben überfordert werden und die Umsetzung der wirklich wesentlichen Aspekte vernachlässigt wird.

#### **Operatives Modell**

Das operative Modell nimmt eine zentrale Stellung innerhalb des Frameworks ein, da es die tatsächliche Umsetzung der Data Governance Maßnahmen koordiniert und überwacht. In bisherigen Arbeiten wird häufig auf ein operatives Modell hingewiesen, jedoch

## 4 Konzeption eines Frameworks

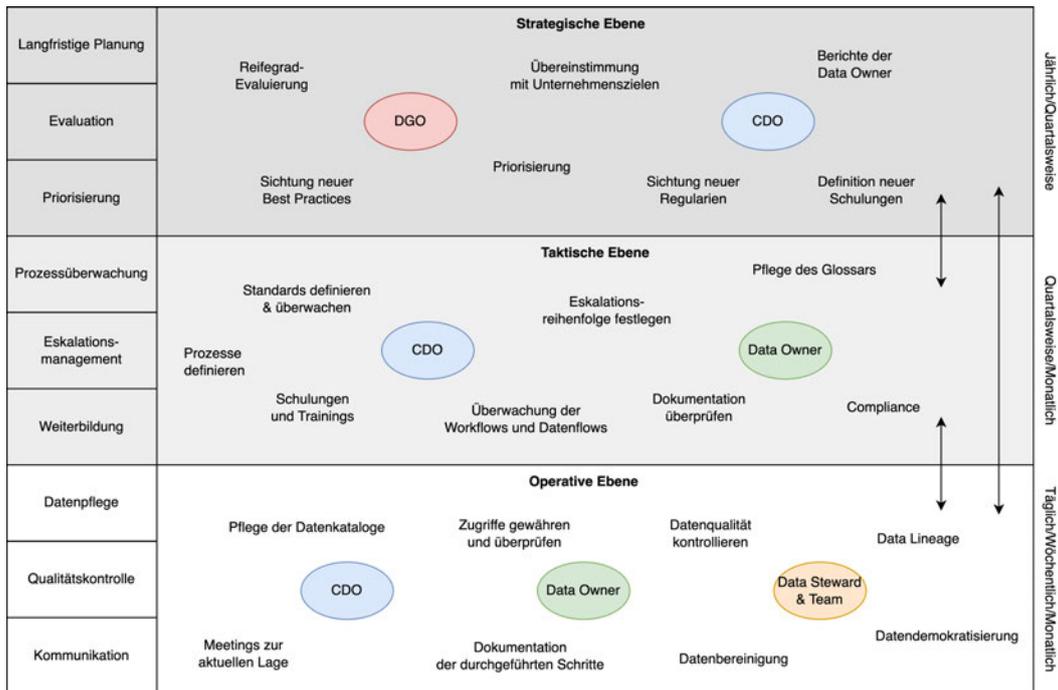


Abbildung 4.3: Übersicht über anfallende Aufgaben im operativem Modell des Core Governance Frameworks

ohne eine konkrete Ausgestaltung vorzunehmen, da diese zu stark vom jeweiligen Unternehmen abhängt. Dies ist korrekt, jedoch können präsentierte Modelle untersucht und an die individuellen Gegebenheiten angepasst werden. Dadurch wird eine solide Grundlage geschaffen, auf der Unternehmen aufbauen können. Abbildung 4.3 zeigt ein operatives Modell, das als Vorlage dienen kann.

Aufgrund des möglicherweise geringeren Fachwissens handelt es sich dabei um ein zentralisiertes Modell. Es gibt es eine zentrale Instanz, die die Regeln und die Richtung der Governance Bemühungen vorgibt. Dies ermöglicht es wenigen Experten, ihr Wissen auf alle Aspekte und Regeln zu übertragen und erleichtert die Koordination. Wenn die Größe der Teams und deren Expertise zunimmt, kann das Modell in ein föderatives Modell umgewandelt werden. Dabei bilden sich innerhalb der Abteilungen eigene Arbeitsgruppen, die spezifische Regeln und Richtlinien für ihren Datenbereich aufstellen und etablieren.

Das Modell ist ebenfalls in drei Ebenen unterteilt, in die strategische, die taktische und die operative Ebene. Diese Unterteilung gibt dabei jeweils den Umfang und Geltungsbereich der auszuführenden Aktivitäten an. Die Hauptaufgaben der Ebenen sind auf der linken

Seite zusammengefasst und bieten einen Überblick über die wesentlichen Tätigkeiten. Im großen Bereich auf der rechten Seite sind beispielhaft wesentliche Aufgaben aufgeführt, welche durch unternehmensspezifische Tätigkeiten ergänzt werden können.

Strategische Maßnahmen sind langfristige Unternehmungen, müssen wohl bedacht sein und werden jährlich bis quartalsweise ausgeführt. Kernelemente davon sind die langfristige Zielsetzung von Data Governance, die mit denen des Unternehmens übereinstimmen müssen, die Priorisierung dieser Ziele sowie die Evaluierung des Reifegrads des Programms. Hauptsächliche Akteure dabei sind das DGO und der CDO. Beispiele für Aktivitäten sind die genannten Hauptaufgaben sowie weitere Aufgaben wie die Definition und Erstellung neuer Schulungen, die Sichtung neuer Regularien und Best-Practices sowie die Überwachung der Berichte der Data Owner.

Die taktische Ebene umfasst die Maßnahmen, die zur Erreichung der langfristigen, strategischen Ziele notwendig sind und werden in der Regel monatlich bis quartalsweise ausgeführt. Der Fokus dabei liegt auf der Umsetzung und Überwachung der Standards und Prozesse. Ebenso wichtig dabei sind klare Vorgaben für Prozesse bei unterschiedlichsten Anwendungsfällen. Die Prozessüberwachung beinhaltet ebenfalls das Eskalationsmanagement, das regelt, welche Schritte eingeleitet und welche Instanzen und Rollen einbezogen werden müssen, wenn Probleme oder Unklarheiten auftreten. Außerdem ist die Weiterbildung der Beteiligten ein zentraler Bestandteil. Beispiele für konkrete Aktivitäten sind die Überwachung der Prozesse und Compliance in Meetings, Durchführung von Schulungen und Trainings sowie die Pflege des Glossars und der Dokumentation. Hauptakteure sind hier der CDO und die verschiedenen Data Owner.

Die operative Ebene konzentriert sich auf die täglich und wöchentlich anfallenden Aufgaben und Aktivitäten, um die langfristigen Ziele zu erreichen. Dabei stehen insbesondere die Datenpflege, die Qualitätskontrolle sowie die Kommunikation und Zusammenarbeit im Fokus. Während die anderen beiden Ebenen noch in vielen Unternehmen ähnlich sind, können sich die konkreten Aktivitäten auf dieser Ebene je nach Domäne und Größe des Unternehmens sowie Reifegrad des Modells unterscheiden. Beispiele für regelmäßig anfallende Aufgaben sind das Nachverfolgen der Datenherkunft, Pflege der Datenkataloge und Dokumentation, Vergabe von Zugriffsrechten sowie Treffen in den Teams zur Besprechung von Problemen und Fortschritten. Involviert sind alle Beteiligten: der CDO, Data Owner, Data Stewards und die Teams.

### Dokumentation

Die Dokumentation aller Aktivitäten ist ein essenzieller Enabler für Data Governance. Mündlich übertragene Informationen geraten schnell in Vergessenheit und auftretende Fragen können nicht immer von zentralen Akteuren wie dem CDO oder Data Ownern beantwortet werden. Außerdem fördert Dokumentation die Transparenz und Nachvollziehbarkeit von Prozessen und Entscheidungen im Unternehmen. Ebenso handelt es sich bei Dokumentation um einen wichtigen Kommunikationskanal, den andere Abteilungen beispielsweise nutzen, wenn sie auf Daten arbeiten, deren Verantwortungsbereich in anderen Abteilungen liegt. Auch für neue Mitarbeiter werden so Zusammenhänge klar und sie können sich leichter einarbeiten. Die Dokumentation sollte dabei gut strukturiert sein und die wesentlichen Punkte klar und übersichtlich darstellen. Dafür wird in dieser Arbeit zwischen operativer und struktureller Dokumentation unterschieden. Operative Dokumentation bezieht sich auf die Beschreibung von laufenden Aktivitäten und Prozessen im Rahmen von Data Governance. Dies umfasst die Beschreibung tatsächlich stattfindender Maßnahmen und Entscheidungen. Strukturelle Dokumentation hingegen umfasst die grundlegenden Richtlinien, Regeln und Konzepte, die allgemeingültig sind und als Vorlage für konkrete Umsetzungen dienen. Im Folgenden sind Beispiele für beide Arten von Dokumentation aufgelistet.

#### Operative Dokumentation:

- **Durchgeführte Aktivitäten und Entscheidungen:** Dokumentation der konkreten Maßnahmen, die im Rahmen von Data Governance umgesetzt werden, einschließlich der getroffenen Entscheidungen.
- **Protokolle von Meetings:** Mitschriften der Besprechungen, um Entscheidungen und Diskussionen nachvollziehbar zu machen.
- **Schulungsunterlagen:** Materialien und Ressourcen, die zur Weiterbildung der Mitarbeiter verwendet werden.
- **Reports und regelmäßige Berichte:** Regelmäßige Berichte zur Überwachung und Bewertung der Data Governance Maßnahmen und Fortschritte.

#### Strukturelle Dokumentation:

- **Richtlinien:** Allgemeine Anweisungen und Leitlinien für den Umgang mit Daten im Unternehmen.

- **Regeln im Umgang mit Daten:** Konkrete Vorschriften, wie bestimmte Daten zu behandeln sind.
- **Eskalationsprozesse:** Vorgehensweise und Hierarchie für das Melden und Bearbeiten von Problemen und Unklarheiten.
- **Ansprechpartner:** Liste der zuständigen Personen für alle Data Governance relevanten Themen.
- **Datenmanagementkonzepte:** Beschreibung grundlegender Datenstrukturen, -flüsse und -architekturen, die zur Erhöhung der Datenqualität beitragen.

Wichtig bei beiden Arten der Dokumentation ist, dass die Informationen leicht zugänglich sind und aktuell bleiben. Dafür sollte sie digital erfolgen und an einem zentralen Ort für jeden verfügbar sein. Dies kann beispielsweise eine Website im Intranet sein. Eine weitere Möglichkeit ist die Auflistung der Verantwortlichkeiten und Richtlinien sowie die Übersicht der Datenstrukturen im Datenkatalog. Dadurch sind die relevanten Informationen an einem Ort verfügbar und können direkt abgerufen werden.

Die Herausforderung bei der Dokumentation ist, dass sie beim Erstellen keinen direkten Mehrwert bringt und dadurch als zusätzliche Arbeit und Belastung wahrgenommen wird, wodurch sie oft vernachlässigt wird. Hierbei stehen der CDO und die Data Owner in der Verantwortung, die Einhaltung der Dokumentation sicherzustellen und Unterstützung sowie Hilfsmittel für die Mitarbeiter bereitzustellen. Beispiele hierfür sind Vorlagen und Checklisten, die Anweisungen und Strukturen für die Dokumentation vorgeben.

### **Technologien**

Die Auswahl und der Einsatz für das Unternehmen geeigneter Technologien und Werkzeuge sind essenziell für Data Governance. Sie unterstützen bei sich wiederholenden Prozessen, der Einhaltung von Standards und Richtlinien und erhöhen die Effizienz und Sicherheit beim Umgang mit Daten. Wichtig dabei ist, dass die ausgewählten Technologien den spezifischen Anforderungen und Bedürfnissen des Unternehmens entsprechen. Sie sollten ausreichend umfangreich sein, um die wichtigsten Funktionen und Anwendungen zu unterstützen und gleichzeitig nicht so umfangreich sein, dass sie unübersichtlich und schwer verständlich sind.

Ein **Datenkatalog** dient als zentrale Übersicht aller verfügbaren Datenquellen und Datenbestände. Diese sind dort mit ihren Metadaten gelistet. Die Metadaten liefern Informationen über Formate, Inhalte und Nutzung und erhöhen die Wiederverwendbarkeit und Effizienz der Daten. Zusätzlich ist die Nennung der verantwortlichen Person hilfreich. Die Extraktion der Metadaten sollte automatisch erfolgen und mit manuellen Kommentaren ergänzt werden. Das Werkzeug, mit dem der Datenkatalog umgesetzt wird, muss ebenfalls in der Lage sein, ein Geschäftsglossar zu erstellen und zu verwalten. Dadurch werden Begriffe innerhalb des Unternehmens vereinheitlicht und es kommt zu keinen Missverständnissen. Außerdem ist so die Integration von Sicherheitsklassifikationen für sensitive und nicht sensitive Daten möglich. Ein weiterer Punkt ist die Verfolgung der Daten und die Bestimmung ihrer Herkunft und Abhängigkeiten. Einige Datenkataloge bieten die Möglichkeiten an, diese anhand von Metadaten zu verfolgen. Der CDO ist für die Pflege und Aktualisierung des Datenkatalogs verantwortlich und stellt sicher, dass alle Einträge korrekt sind und regelmäßig überprüft werden.

**Werkzeuge zur Erfassung der Datenqualität** ermöglichen eine automatisierte Überwachung der Datenqualität. Dafür werden verschiedene Qualitätsmetriken wie Vollständigkeit, Genauigkeit, Konsistenz und Aktualität definiert und Soll-Werte bestimmt. Diese werden mit den vorhandenen und anfallenden Daten kontinuierlich und ohne manuellen Aufwand abgeglichen. Bei Abweichungen und Verstößen gegen diese Vorgaben ist die Einrichtung eines Benachrichtigungssystems hilfreich, um Verantwortliche direkt informieren zu können. Visualisierungen der Ergebnisse machen diese greifbarer und verständlicher. Dadurch können Trends abgeleitet und Domänen priorisiert werden.

Eine technologiegestützte **Zugriffsverwaltung** spielt ebenfalls eine wichtige Rolle für die Sicherheit der Daten und Einhaltung der Gesetze und Regularien. Um eine granulare Steuerung der Zugriffe auf sensitive Daten zu ermöglichen, müssen Rollen und Zugriffsrechte für diese definiert werden. Dabei ergibt es zu Beginn Sinn, die eingebauten Lösungen der Datenquellen dafür zu verwenden. Die Rollen können dabei über den Datenkatalog und Glossar eingeteilt werden. Verantwortlich für die Zugriffsverwaltung sind die entsprechenden Data Owner. Die Einrichtung von Audit-Trails hilft die Nachvollziehbarkeit von Datenzugriffen nachzuverfolgen und potenzielle Sicherheitsvorfälle aufzuklären.

**Visualisierungen** wie Dashboards helfen bei der Übersicht und dem Verständnis von Metriken, Fortschritten und der Einhaltung von Data Governance Aktivitäten. Abstrakte Zahlen sind auf den ersten Blick oft schwer verständlich. Durch Visualisierungen wie

Dashboards, Graphen und Diagramme können diese interpretierbar dargestellt werden, wodurch schneller reagiert werden kann und Entscheidungen präziser getroffen werden können.

**Automatisierungswerkzeuge** sind entscheidend für sich wiederholende Aktivitäten und Prozesse wie die Überwachung der Datenqualität, die Bereinigung der Daten, die Bestimmung der Datenherkunft, die Erstellung von Metadaten, die Erstellung von Dashboards, die Benachrichtigung bei Verstößen, das Erstellen von Backups und generell für Datenpipelines. Dadurch werden Kontrollen stetig ausgeführt und manuelle Fehler reduziert.

### 4.3.3 Überwachung und Reifegradmodell

Reife beschreibt in Unternehmen den Zustand der Vollständigkeit, Volkommenheit oder Bereitschaft [21]. Dazu werden sogenannte Reifegradmodelle entwickelt, die eine Anzahl von Reifegradstufen definieren [71]. Diese repräsentieren die unterschiedlichen Stufen der Vollständigkeit bis zu den definierten Zielen und geben den gewünschten Entwicklungspfad bis dorthin vor [10]. Die Stufen sind dabei diskret und können nicht übersprungen werden [10][34].

Im Bezug auf Data Governance bezeichnet Reifegrad das Ausmaß, in dem das Unternehmen die Prozesse, Richtlinien, Strukturen und Praktiken, die durch Data Governance definiert werden, umgesetzt hat [46]. Ein Reifegradmodell für Data Governance besteht ebenfalls aus Stufen und beschreibt verschiedene Prozesscharakteristiken, welche in jeder Stufe unterschiedlich stark ausgeprägt sind [34]. Untersuchungen von Marchildon et al. in [46] zeigen, dass eine erfolgreiche Umsetzung von Data Governance häufig nicht an falschen Definitionen scheitert, sondern daran, dass Unternehmen Schwierigkeiten haben, ihren aktuellen Reifegrad zu beurteilen und dadurch nicht in der Lage sind, konkrete Maßnahmen abzuleiten.

In der Literatur und Praxis existieren verschiedene etablierte Reifegradmodelle, die typischerweise fünf bis sechs Stufen umfassen. DAMA adaptiert ein Modell von *Select Business Solutions* und schlägt fünf Stufen vor, wobei zusätzlich noch eine Stufe null der Vollständigkeit wegen definiert wird und sehr selten in der Realität vorkommt [34]. Das Modell von Gartner umfasst auch sechs Stufen, wobei die erste Phase ebenfalls das Fehlen formaler Strukturen beschreibt [29]. Ähnlich verhält es sich bei Capability Maturity

Model Integration (CMMI), das ursprünglich für die Prozessoptimierung in der Softwareentwicklung entwickelt wurde [17]. Mittlerweile wird das Modell auch auf Prozesse der Datenverarbeitung übertragen, um diese in Unternehmen systematisch zu bewerten und zu verbessern, wobei die Anzahl der Stufen mit den anderen Modellen identisch ist.

Die Struktur der Reifegradmodelle hat sich bewährt und unterscheidet sich nur geringfügig untereinander. KMUs besitzen andere Voraussetzungen und Ressourcen als größere Unternehmen, das übergeordnete Ziel von Data Governance ist aber identisch. Mit konsistenten, vorhersehbaren, wiederholbaren und zuverlässigen Prozessen die Sicherheit, die Wiederverwendbarkeit und den Nutzen von Daten im Unternehmen zu erhöhen. Aus diesem Grund wird die grundlegende Struktur der Reifegradmodelle übernommen und im Folgenden kurz vorgestellt.

- **Stufe 0 Nicht bewusst:** Hier besitzt das Unternehmen keine formellen Prozesse und Richtlinien im Umgang mit den Daten. Datenqualität und Datensicherheit befinden sich nicht im Bewusstsein der Beteiligten.
- **Stufe 1 Bewusst:** Dies ist häufig die Phase, in der die Unternehmen starten. Prozesse werden in der Regel reaktiv ausgeführt. Sofern es formale Prozessstrukturen gibt, werden diese nicht in allen Bereichen und Teams angewandt. Rollen und Verantwortlichkeiten sind ebenfalls in begrenzten Bereichen umgesetzt. Es gibt jedoch keinen Standard und gemeinsamen Ansatz für die Verwaltung und gemeinsame Nutzung der Unternehmensdaten.
- **Stufe 2 Grundlagen:** Erste einheitliche Tools und Rollen bei der Prozessausführung sind vorhanden. Ebenso gibt es eine zentrale Überwachung der Datenverwaltung. Datenqualität ist im Bewusstsein der Beteiligten, ebenso sind Stamm- und Referenzdaten vertraute Konzepte.
- **Stufe 3 Erweitern und Anpassen:** Es beginnt mit der Einführung skalierender Prozesse, die unternehmensweit einheitlich sind. Es wird vermehrt Wert auf Datenqualität gelegt und Daten werden im Unternehmen geteilt. Die Prozesse sind weiter entwickelt und mittlerweile so automatisiert, dass manuelles Eingreifen nur noch selten auftritt. Dadurch sind die Ergebnisse der Prozesse beim Durchführen besser vorhersehbar.
- **Stufe 4 Skalieren und Optimieren:** Die Fortschritte aus den vorherigen Stufen ermöglicht dem Unternehmen, Ergebnisse neuer Projekte und Aufgaben bereits vorher vorherzusagen und die Risiken der Daten abzuschätzen. Insgesamt gibt es

mittlerweile für alles standardisierte Werkzeuge und Prozesse und ein gut organisiertes zentrales Planungsorgan. Datenqualität ist mittlerweile konsistent auf einem hohen Niveau und Daten können von Anfang bis Ende verfolgt werden.

- **Stufe 5 Experten:** In diesem Stadium gibt es für alle auftretenden Möglichkeiten klare Prozesse und Entscheidungsfindungen. Prozesse sind mittlerweile alle automatisiert und für die Messung der Datenqualität und der Prozessqualität gibt es verständliche Kenngrößen. Das Unternehmen fokussiert sich auf die kontinuierliche Verbesserung und Innovationen, um die Datenverwaltung weiter zu optimieren

Um den aktuellen und künftige Reifegrade bestimmen zu können, müssen spezifische Prozesscharakteristiken definiert werden. Diese Merkmale dienen als Indikatoren für den Fortschritt und die Effektivität des Data Governance Programms. In der Regel wird der Fokus auf verschiedene Schwerpunktbereiche gelegt wie beispielsweise die elf Bereiche von DAMA. CMMI legt den Fokus in seinem Datenmodell auf die sechs Bereiche Datenmanagement Strategie, Data Governance, Datenqualität, Plattform und Architektur sowie Datenoperationen [49]. Innerhalb dieser Bereiche werden gezielt Fragen und Charakteristiken formuliert, die den Fortschritt und Effektivität von Data Governance bewerten. Diese Fragen lassen sich häufig in folgende Unterbereiche einteilen:

- **Aktivitäten:** Hier wird untersucht, ob Prozesse und Aktivitäten in diesem Bereich für diese Charakteristik existieren und wie gut diese definiert sind. Weitere Beispiele für Charakteristiken könnte die Umsetzung von Best-Practices sein oder wie ausgereift die Prozesse sind und wie gut sie umgesetzt werden.
- **Werkzeuge:** Dieser Bereich bewertet beispielsweise, wie automatisiert Prozesse sind und inwieweit Technologien diese unterstützen. Charakteristiken zu Schulungen zu diesen Technologien, abhängig von den Rollen und wie stark sie diese verwenden und deren Verbreitung im Unternehmen sind weitere Möglichkeiten.
- **Standards:** Hier werden Fragen und Merkmale zur Existenz und Anwendung von Standards gestellt. Wie gut sind sie dokumentiert? Werden sie kontrolliert und umgesetzt?
- **Menschen und Ressourcen:** Dieser Bereich fokussiert sich auf die Beteiligten und die vorhandenen Strukturen und Infrastruktur. Dabei werden Charakteristiken zu Rollen, Entscheidungsrechten und Kompetenzen der Mitarbeiter definiert.

Die Fragen und Charakteristiken müssen dabei so konzipiert sein, dass sie je nach aktuellem Reifegrad weiterentwickelt werden können. Wenn beispielsweise die Reife des Datenqualitätsmanagements bewertet wird, wird in Stufe eins untersucht, ob überhaupt Verfahren dafür vorhanden sind. In höheren Stufen können diese Fragen sich ausweiten, wie konsistent sie in unterschiedlichen Abteilungen durchgeführt werden und wie regelmäßig Berichte erstellt werden. Innerhalb jedes Reifegrades erfolgt dann die Ordnung in eine Skala, die angibt, wie weit die Umsetzung für das Erreichen des Zieles dieses Reifegrades ist. DAMA schlägt eine Skalierung von eins bis vier vor [34]. Eine Bewertung von eins bedeutet, dass die Maßnahme noch nicht begonnen wurde. Bei einer Einstufung von zwei ist die Maßnahme in der Implementierungsphase. Eine Bewertung von drei zeigt, dass die Maßnahme implementiert und funktionsfähig ist, jedoch noch Optimierungspotenzial besteht. Eine Einstufung von vier bedeutet, dass die Maßnahme vollständig integriert ist und konsistent gute Ergebnisse liefert.

Eine Visualisierung des aktuellen Standes bei den Reifegraden kann dazu beitragen, Schwachstellen und Handlungsbedarf zu identifizieren und den Fortschritt nachvollziehbar darzustellen. Dies erleichtert den Beteiligten das Verständnis des aktuellen Reifegrades und fördert die gezielte Planung und Umsetzung von Verbesserungsmaßnahmen. Abbildung 4.4 zeigt eine mögliche Darstellung des aktuellen Standes. Dabei wird der Reifegrad der elf Kernbereiche des Data Managements dargestellt. Der äußere, blaue Ring zeigt den gewünschten Grad an, den das Unternehmen anstrebt, der innere orangene Ring stellt den aktuellen Reifegrad dar. Die Bereiche, die den größten Abstand haben, stellen das größte Risiko da und können beispielsweise priorisiert werden.

Häufig besteht das Problem bei der Umsetzung der Reifegrade nicht explizit beim passenden Modell, sondern vielmehr bei den Unklarheiten, wie die Bewertung erfolgen soll und welche Fragen dafür gestellt werden müssen. Dafür muss ein passender Fragenkatalog und passende Prozessmerkmale definiert werden, die auf das Unternehmen zugeschnitten sind. Marchildon et al. haben in ihrer Arbeit [46] 72 Fragen entwickelt, die als Grundlage dienen können. Diese können an die spezifischen Anforderungen des Unternehmens angepasst werden. Wichtig anzumerken dabei ist noch, dass bei der Bewertung das Ziel ist, aktuelle Stärken und Schwächen aufzudecken und nicht diese direkt zu lösen oder zu verharmlosen.

Die Informationen und Konzepte aus den verschiedenen Modellen werden zusammengeführt und daraus das in Abbildung 4.5 dargestellte Reifegradmodell vorgeschlagen. Das vorgeschlagene Modell kann KMUs als Grundlage dienen und umfasst ebenfalls sechs

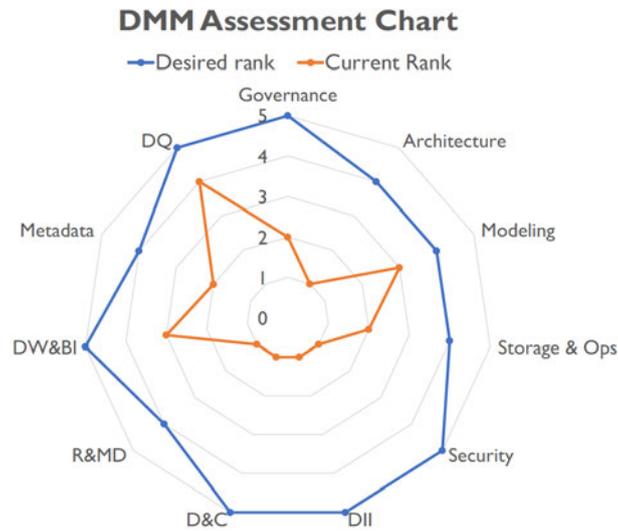


Abbildung 4.4: Visualisierung der Reifegrade in den verschiedenen Kernbereichen [34]

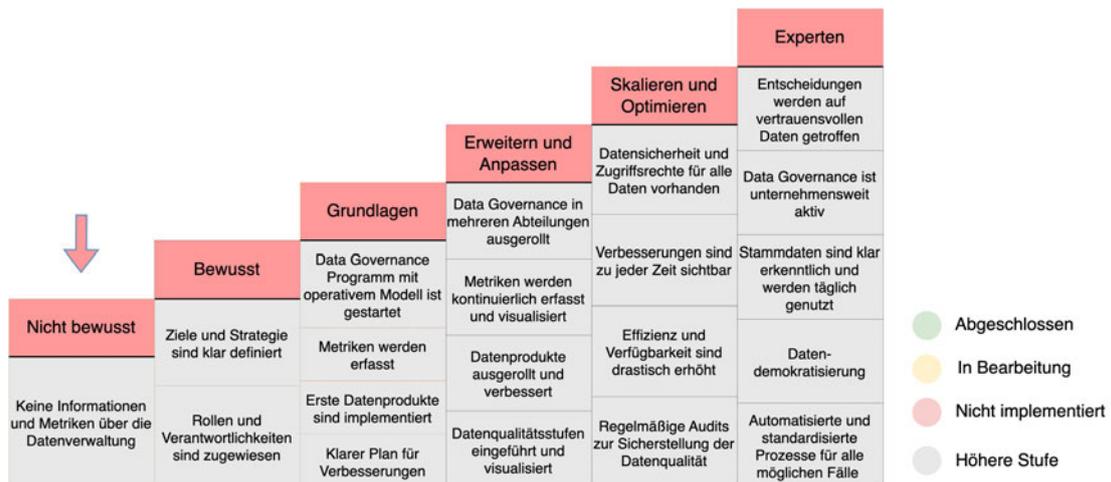


Abbildung 4.5: Visuelles Modell für die Erfassung des Reifegrads

Stufen, die den Fortschritte und die Reife des Data Governance Programms abbilden. Dadurch werden zum einen Schwachstellen und Bereiche, bei denen Handlungsbedarf besteht, identifiziert. Gleichzeitig wird der Weg bis hin zu den Zielen festgehalten. Innerhalb jeder Stufe können eigene Charakteristiken und Aufgaben definiert werden, wodurch eine Anpassung des Reifegradmodells an die individuellen Anforderungen möglich ist. Innerhalb jeder Reifegradstufe erfolgt für eine gute Übersicht eine Bewertung auf einer Skala von eins bis drei, die den Fortschritt der Umsetzung anzeigt. Die farbliche Darstellung der Bewertung ermöglicht eine Verdeutlichung des aktuellen Stands für die Beteiligten.

## 5 Case Study: Zusammenführen von Theorie und Praxis

Aufgrund fehlender Literatur und Beispielen wurde in dieser Arbeit in Kapitel 4 ein umfassendes Data Governance Framework entwickelt, das auf die spezifischen Bedürfnisse und Herausforderungen von KMUs zugeschnitten ist. Um eine weitere Lücke in der Literatur zu schließen, wird das entworfene Framework CoreGovernance im folgenden Kapitel exemplarisch an dem fiktiven Unternehmen AquaFlow GmbH umgesetzt. Damit soll begonnen werden, zu zeigen, wie die oft theoretischen Konzepte von Data Governance in die Praxis integriert werden können und welche Schritte notwendig sind, um ein effektives Data Governance Programm zu etablieren.

### 5.1 Vorgehensweise

Im Folgenden werden die notwendigen, strukturellen Schritte vorgestellt, um einen initialen Stand von Data Governance in der AquaFlow GmbH zu etablieren. Data Governance ist kein abzuschließendes Projekt, sondern bedarf dauerhafter Aktivitäten und Überwachung. Jedoch müssen insbesondere zu Beginn der Einführung einige Aufgaben, Aktivitäten und Prozesse definiert werden. Aus diesem Grund wird im Folgenden ein High-Level-Überblick über die benötigten Aktivitäten gegeben.

In der **Initialisierungsphase** wird der Grundstein für das Data Governance Programm gelegt. Dafür werden zunächst die wichtigsten Stakeholder vereinbart und mehrere Kick-off-Meetings mit diesen durchgeführt. In diesen werden Ziele und der Umfang abgesteckt sowie Erwartungen festgelegt. Im Anschluss daran erfolgt die Erfassung des aktuellen Stands im Bezug auf Datenverwaltung und -nutzung innerhalb des Unternehmens. Ebenso muss ein Überblick über alle vorhandenen Datenquellen und -bestände erstellt werden.

In der nächsten Phase müssen **die Ziele und die Strategie** konkretisiert und niedergeschrieben werden. Dies umfasst spezifische Data Governance Ziele wie Datenqualität, Datensicherheit, Verfügbarkeit, Compliance und Effizienzsteigerung sowie das Etablieren einer datengetriebenen Unternehmenskultur. Für das Erreichen der gesteckten Ziele wird eine High-Level Strategie formuliert, die die Definition von Prozessen, Richtlinien und Metriken beinhaltet.

Im Anschluss daran erfolgt **die Planung und das Design** der konkreten Maßnahmen. Dabei werden die verschiedenen Rollen und Verantwortlichkeiten zugewiesen und die Kommunikations- und Entscheidungsstrukturen definiert. Dafür wird zunächst das operative Modell und die konkreten Aktivitäten für die strategische, taktische und operative Ebene festgelegt und detailliert beschrieben. Ebenso werden Bedarf und Anforderungen an die benötigten technologischen Werkzeuge wie Datenkataloge, Datenqualitätskontrollen, Zugriffsmanagement, Dashboards und Automatisierungssysteme definiert.

Nach der Planung findet **die Umsetzung** der Aktivitäten statt. Dabei werden die im Rahmen des operativen Modells entwickelten Prozesse, Systeme und Maßnahmen in den täglichen Betrieb integriert. Dies umfasst ebenfalls die Implementierung und Konfiguration der erforderlichen Technologien. Zudem finden Schulungen und Trainings statt, um die Mitarbeiter und Beteiligten weiterzubilden.

Eine **kontinuierliche Überwachung** der Aktivitäten ist ebenso ein wichtiger Bestandteil wie deren Planung und Umsetzung. Dabei erfolgen regelmäßiges Monitoring und Berichterstattungen, die die Einhaltung der Richtlinien und der definierten Datenqualität überprüfen. Anhand der Ergebnisse der Überwachung der Prozesse und des Reifegrads können gezielt Maßnahmen zur kontinuierlichen Verbesserung beschlossen werden. Wichtig dabei ist, die Ergebnisse in geeigneten Formaten und Darstellungen an die Beteiligten weiterzugeben.

**Die Kommunikation und ein Kulturwandel** hin zu daten- und wertbasierten Entscheidungen ist für den langfristigen Erfolg des Data Governance Programms entscheidend. Mithilfe geeigneter Kommunikation und Dokumentation wird sichergestellt, dass der Informationsfluss innerhalb des Unternehmens intakt bleibt und Fortschritte, Erfahrungen und Erfolge weitergegeben werden. Führungskräfte und Verantwortliche müssen diese Denkweisen verinnerlichen und den Beteiligten vorleben.

## 5.2 Ziele und Strategie

Um die definierten Ziele zu erreichen, ist eine klare und umsetzbare Strategie notwendig. Eine Strategie besteht nicht nur aus Zielen, sondern gibt auch die Richtung vor und definiert die Maßnahmen, die unternommen werden müssen, um diese Ziele zu erreichen.

Die Hauptziele für die Data Governance bei AquaFlow Solutions GmbH sind:

- **Hochwertige Datenqualität**

- Ziel: Alle Daten sind zu jeder Zeit vollständig, korrekt, konsistent und aktuell.
- Ergebnis: Fundierte und qualitativ hochwertige Entscheidungen können auf Basis dieser Daten getroffen werden.
- Strategie: Einführung eines kontinuierlichen Datenqualitätsmanagements.
- Metriken: Datenfehlerquote, Anzahl durchgeführte Datenbereinigungen, Übereinstimmungsquote mit Soll-Zustand.
- Maßnahmen: Einführung von standardisierten Datenvalidierungsprozessen, Implementierung von Datenbereinigungs-Tools, regelmäßige Schulungen.

- **Sichere Daten in allen Lebenszyklen**

- Ziel: Alle Daten sind stets bekannt, und insbesondere sensible und personenbezogene Daten sind geschützt und vor unberechtigtem Zugriff sicher.
- Ergebnis: Reduzierung von Sicherheitsvorfällen und Schutz sensibler Daten.
- Strategie: Aufbau eines robusten Datenschutzes- und Sicherheitsrahmens.
- Metriken: Manuelle Prüfungen, Anzahl Sicherheitsmaßnahmen, Anzahl Alerts, Anzahl Sicherheitsvorfälle, Prozentwert, auf wie viele der Daten nur über Berechtigungen zugegriffen werden kann.
- Maßnahmen: Implementierung von Zugriffssystemen mit Berechtigungen, Sicherstellung, dass sensitive Daten nicht zugreifbar und lesbar sind, Einführung von Kontrollen.

- **100% verfügbare Daten**

- Ziel: Keine Ausfälle, auch nicht bei Infrastrukturausfall.

- Ergebnis: Stetige Verfügbarkeit von Daten für alle berechtigten Nutzer.
- Strategie: Sicherstellung einer hochverfügbaren Dateninfrastruktur.
- Metriken: Systemverfügbarkeit, Prozesszeiten bei Datensuche, Prozentwert von Replikas.
- Maßnahmen: Implementierung von Replikas und Partitionen, Erstellung von Recovery-Plänen, Einführung von Self-Service-Tools.

- **Datengetriebene Unternehmenskultur**

- Ziel: Mitarbeiter sind geschult und haben Data Governance stets im Hinterkopf bei allen Aktivitäten.
- Ergebnis: Lösungen und Entscheidungen basieren auf Datenanalysen.
- Strategie: Förderung einer datenorientierten Denkweise und Entscheidungsfindung.
- Metriken: Durchgeführte Schulungen und deren Erfolg, Mitarbeiterzufriedenheit, Anzahl datenbasierter Entscheidungen.
- Maßnahmen: Motivation der Mitarbeiter zur Umsetzung, regelmäßige Schulungen, klare Ziele vor Augen behalten.

- **Effizienz maximieren**

- Ziel: Prozesse sind, wo es möglich ist, automatisiert und standardisiert, Fehlerminimierung durch Reduzierung manueller Eingriffe.
- Ergebnis: Effizientere Prozesse und reduzierte Fehlerquoten.
- Strategie: Automatisierung und Standardisierung der Prozesse.
- Metriken: Zeitersparnis bei Prozessen, Anzahl standardisierter Prozesse, Fehlerquote.
- Maßnahmen: Automatisierung von Prozessen, Einführung von Workflow-Management-Systemen, regelmäßige Überwachung und Optimierung.

- **Compliance**

- Ziel: Erfüllung aller gesetzlichen Vorgaben und aller ausgewählten Normen und Standards.
- Ergebnis: Keine Verstöße gegen gesetzliche Vorgaben und interne Standards.
- Strategie: Compliance-Management und zentrales Repository aller geltenden Vorschriften.
- Metriken: Anzahl Verstöße, Anzahl Überprüfungen, prozentualer Wert von Nicht-Einhaltung.
- Maßnahmen: Sicherstellung des Zugriffs für alle relevanten Vorgaben, Implementierung von Überwachungsstrategien.

Diese strategischen Maßnahmen und Prozesse sind darauf ausgelegt, die definierten Ziele effizient zu erreichen und eine nachhaltige Verbesserung der Data-Governance-Strukturen innerhalb der AquaFlow Solutions GmbH zu gewährleisten.

### 5.3 Rollen und Verantwortlichkeiten

Eine geeignete Zuweisung der Rollen ist entscheidend für den Erfolg der Data Governance Aktivitäten für die AquaFlow GmbH. Die ausgewählten Personen tragen die Verantwortung, ihr Wissen im gesamten Unternehmen zu verbreiten und in den Geschäftsalltag zu integrieren. Gleichzeitig müssen die Rollen jedoch in das bestehende Gefüge passen, um reibungslose Abläufe zu gewährleisten. Im Folgenden werden die spezifischen Rollen und Verantwortlichkeiten in der AquaFlow GmbH zugewiesen.

Abbildung 5.1 zeigt die Struktur der Rollen des Data Governance innerhalb der AquaFlow GmbH. Im Zentrum der Governance Aktivitäten befindet sich das DGO. Dieses setzt sich aus den Data Ownern der verschiedenen Datenquellen und dem CDO zusammen. Es trifft sich in der Regel quartalsweise und übernimmt die in Unterabschnitt 4.3.2 beschriebenen Aufgaben. Der CDO formt neben dem DGO mit den zwei ernannten Data Stewards das DG-Kernteam. Dieses trifft sich monatlich, um operative Entscheidungen zu treffen und die Umsetzung der Data Governance Aktivitäten zu überwachen. Dabei geht es insbesondere um konkrete Maßnahmen, wie die taktischen und strategischen Ziele erreicht werden können. Die beiden Data Stewards arbeiten mit den Teams in den verschiedenen Abteilungen zusammen und unterstützen sie bei der Implementierung und Einhaltung der getroffenen Entscheidungen und Data Governance Maßnahmen. Der Data Owner

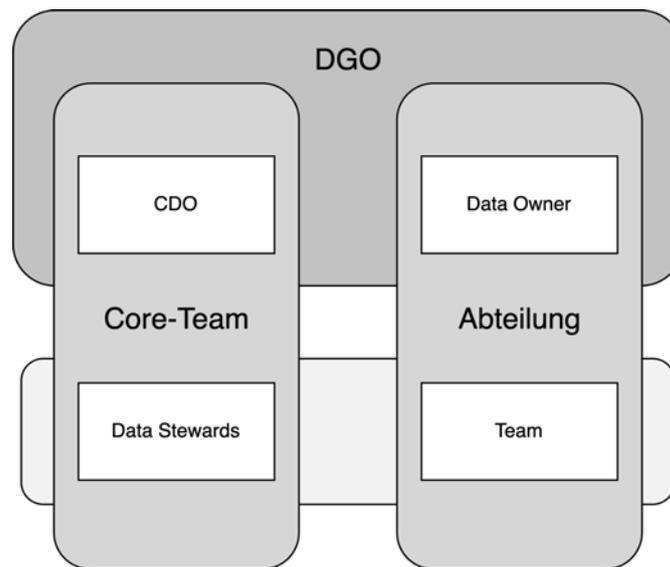


Abbildung 5.1: Übersicht der Data Governance Rollen in der AquaFlow GmbH

sorgt im täglichen Betrieb zusammen mit seinem Team dafür, dass die Sicherheit der Daten gewährleistet ist und die geschäftlichen Ziele erreicht werden.

## CDO

Die Rolle des CDO sollte idealerweise ein langjähriger und erfahrener Mitarbeiter übernehmen, der über eine umfangreiche Expertise in sämtlichen Unternehmensbereichen verfügt. Die Person muss ein tiefes Verständnis der Geschäftsinteressen des Unternehmens aufweisen und das große Ganze stets im Blick behalten. Die Rolle des CDO könnte auch durch ein Mitglied der Geschäftsführung ausgefüllt werden, da diese die strategischen Ziele einschätzen und umsetzen kann. Allerdings ist die zeitliche Belastung der Geschäftsführung oft nicht mit den zeitintensiven Aufgaben, die die Rolle des CDOs mit sich bringt, zu vereinbaren. Ebenso sollte die Person über ein grundlegendes technisches Verständnis verfügen, um die IT-Aspekte von Data Governance effektiv verwalten zu können. Aus diesem Grund wird in der AquaFlow GmbH der Vertriebsleiter, der seit Jahren dem Unternehmen zugehörig ist, zum CDO ernannt.

## **DGO**

Das DGO setzt sich aus Vertretern der verschiedenen Abteilungen zusammen und wird vom CDO geleitet. Diese Zusammensetzung gewährleistet, dass alle relevanten Perspektiven und Anforderungen berücksichtigt werden. Die IT-Abteilung bringt beispielweise technisches Verständnis mit ein und gibt Einschätzungen über die technische Machbarkeit von Datenprodukten ab. Der Vertreter aus dem Vertrieb verfügt über Kenntnisse über Kundenbedürfnisse und Marktentwicklungen und hat Anforderungen an die Qualität und Verfügbarkeit der Kundendaten. Die Vertreter der anderen Abteilungen, also der Entwicklung, dem Marketing und der Produktion haben ihre eigenen Anforderungen und Sichtweisen auf die Nutzung und Verwaltung der Daten und bringen diese im DGO ein. Bei den Vertretern der Abteilungen handelt es sich um die jeweiligen Abteilungsleiter, da sie das größte Verständnis für die geschäftlichen Anforderungen haben und die Verantwortung für ihren Bereich und die damit verbundenen Aktivitäten tragen.

Das DGO trifft sich in Regelmeetings, die durch das operative Modell koordiniert werden, und entwickelt und überwacht Richtlinien, Standards, Prozesse und Normen. Es ist die zentrale Stelle für Data Governance Maßnahmen und hat stets das große Ganze im Blick, um die Aktivitäten des Unternehmens strategisch zu lenken und weiterzuentwickeln.

## **Data Owner**

In der AquaFlow GmbH sind bereits etablierte Strukturen und Verantwortlichkeiten innerhalb der verschiedenen Abteilungen vorhanden. Um die Integration von Data Governance zu erleichtern, wird die Verantwortung der Datenbestände bzw. der Datenquellen explizit einer Abteilung zugewiesen, wobei die Abteilungsleiter die Verantwortung für diese Daten übernehmen. Die zugewiesenen Data Owner sind dabei verantwortlich für die Einhaltung der im DGO festgelegten Richtlinien, Standards und Prozesse. In kleineren Unternehmen arbeiten die Data Owner eng mit ihren Teams zusammen, um sicherzustellen, dass die Data Governance Maßnahmen effektiv umgesetzt werden.

## **Data Stewards**

Die Einführung von Data Stewards in kleinen Teams kann herausfordernd sein. In solchen Fällen kann der Data Owner, der Teil des Teams ist, die Funktion des Data Stewards

gemeinsam mit dem Team vorübergehend übernehmen. Mit dem Wachstum des Unternehmens und der Erhöhung der Reife der Data Governance Aktivitäten sollten jedoch Data Stewards benannt werden. Um den Übergang zu erleichtern, erfolgt zunächst die Benennung zweier Data Stewards für das gesamte Unternehmen. Diese arbeiten eng mit dem CDO, den Data Ownern und den Teams zusammen und bieten Unterstützung bei spezifischen Fragen bezüglich der Umsetzung von Data Governance Maßnahmen. Sie bilden in Abbildung 5.1 das *DG-Kernteam*.

In der AquaFlow GmbH werden die Rollen der Data Stewards von erfahrenen Mitarbeitern aus der IT-Abteilung und der Marketing-Abteilung übernommen. Der Vertreter der IT-Abteilung bringt tiefes technisches Verständnis und Erfahrung im Umgang mit Dateninfrastrukturen und der Datenverwaltung mit. Er bietet insbesondere Hilfestellung bei der Umsetzung der technischen Aktivitäten. Der Data Steward aus der Marketing-Abteilung bringt umfassendes Wissen über Kundenanalysen und die strategische Nutzung von Daten mit ein. Die Kombination von technischer und geschäftlicher Expertise stellt sicher, dass den Teams alle offenen Fragen beantwortet werden können.

### Zuweisung der Verantwortlichkeiten

Die Zuweisung der Datenquellen und der damit verbundenen Datensätze ist ein entscheidender Faktor für eine effektive Data Governance. Durch diese Zuweisung werden klare Verantwortlichkeiten für Aspekte wie Qualität, Sicherheit und Effizienz geschaffen, was auch eine Rechenschaftspflicht einführt. Die Datenquellen werden den jeweiligen Anwendungsgebieten zugeordnet, wobei die Data Owner die Verantwortung für die Einhaltung der Maßnahmen übernehmen. In der Tabelle 5.1 ist die erfolgte Zuweisung zu sehen.

Datenquelle	Abteilung	Verantwortlicher
Logs	IT-Abteilung	IT-Leiter
Sensordaten IoT	Entwicklung	Entwicklungsleiter
Web-Analytics	Marketing	Marketing-Leiter
ERP-System	Produktion	Produktionsleiter
CRM-System	Vertrieb	stellv. Vertriebsleiter
Produktionsanlagen	Produktion	Produktionsleiter
Gesetze und Standards	HR-Abteilung	HR-Leiter

Tabelle 5.1: Zuweisung der Datenquellen an die Abteilungen

## 5.4 Reifegradmodell

Ein passendes Reifegradmodell ist ebenfalls für die AquaFlow GmbH von entscheidender Bedeutung, um Fortschritte zu überwachen und diese sowohl den Stakeholdern als auch den beteiligten Mitarbeitern transparent darzustellen. Dadurch können Probleme identifiziert und entsprechende Verbesserungsmaßnahmen ergriffen werden, während gleichzeitig das Gesamtziel im Auge behalten wird. Dafür wird das in Unterabschnitt 4.3.3 für KMUs entwickelte und in Abbildung 4.5 dargestellte Modell übernommen.

Zunächst muss eine Ist-Analyse, oder auch Assessment, durchgeführt werden, um den aktuellen Stand der Prozesse und Aktivitäten im Kontext der Datenverwaltung zu erfassen. Hierfür werden die im Reifegradmodell definierten Prozessfortschritte und Charakteristiken verwendet. Zur Erfassung des aktuellen Standes werden Fragen aus der Fallstudie [46] von Marchildon et al. herangezogen. Beispielhaft sind elf Fragen und die dazugehörigen Antworten in Tabelle 5.2 aus den verschiedenen Bereichen der Fallstudie aufgeführt. Es ist wichtig zu beachten, dass sich die Fragen häufig auf das Data Governance Framework (DGF) beziehen. Beim erstmaligen Assessment sind damit die bereits bestehenden Strukturen und Prozesse gemeint, auch wenn diese noch nicht formal als DGF formuliert sind.

Tabelle 5.2: Assessment der Data Governance Charakteristiken

<b>Bereich</b>	<b>Frage</b>	<b>Antwort</b>
Datenrisikomanagement und Compliance	In ihrem Unternehmen, ist das DGF auf das Risikomanagement und die Einhaltung der Vorschriften abgestimmt? Es gibt eine ... Abstimmung zwischen dem DGF und dem Risikomanagement und der Einhaltung der Vorschriften.	schwache
Datenwertschöpfung	Bietet das DGF wichtige organisatorische Vorteile? Das DGF bietet ... wichtige organisatorische Vorteile.	wenige
Organisationsstruktur und -bewusstsein	Sind sich die Mitarbeiter des Unternehmens darüber im Klaren, wie man mit Daten als Vermögenswert umgeht? ... Mitarbeiter sind sich darüber im Klaren.	Einige

<b>Bereich</b>	<b>Frage</b>	<b>Antwort</b>
Datenrichtlinien und -regeln	Sind in Ihrem Unternehmen Data Governance Richtlinien und Regeln dokumentiert? Data Governance Richtlinien und Regeln sind ... dokumentiert.	gar nicht
Datenverwaltung	Haben die Geschäftsbereiche/Abteilungen in Ihrem Unternehmen einen eigenen Data Steward? ... Geschäftsbereiche/Abteilungen haben ihren eigenen Data Steward.	Keine
Datenqualität	Gibt es in Ihrem Unternehmen einen Prozess für den Umgang mit Datenqualitätsproblemen?	Es gibt keine Prozesse für den Umgang mit Datenqualitätsproblemen.
Datenlebenszyklusmanagement	Werden in Ihrem Unternehmen Metriken verwendet, um die Erhebung, Verwendung, Speicherung und Löschung von Daten aus allen Datenquellen zu bewerten? ... Metriken werden verwendet.	Wenige
Datensicherheit	Sind in Ihrem Unternehmen sensible Daten verschlüsselt? Sensible Daten sind ... verschlüsselt.	teilweise
Datenarchitektur	Basiert die Architektur der unstrukturierten Datensysteme auf etablierten Standards? Die Architektur der unstrukturierten Datensysteme ... auf etablierten Standards.	basiert nicht
Datenklassifizierung und Metadaten	Gibt es in Ihrem Unternehmen ein Business Glossar für die wichtigsten organisatorischen Begriffe? Das Glossar umfasst ... Schlüsselbegriffe.	keine
Archivierung und Audits	Werden in Ihrem Unternehmen Änderungen an kritischen Daten geprüft? Änderungen an kritischen Daten werden ... geprüft.	selten

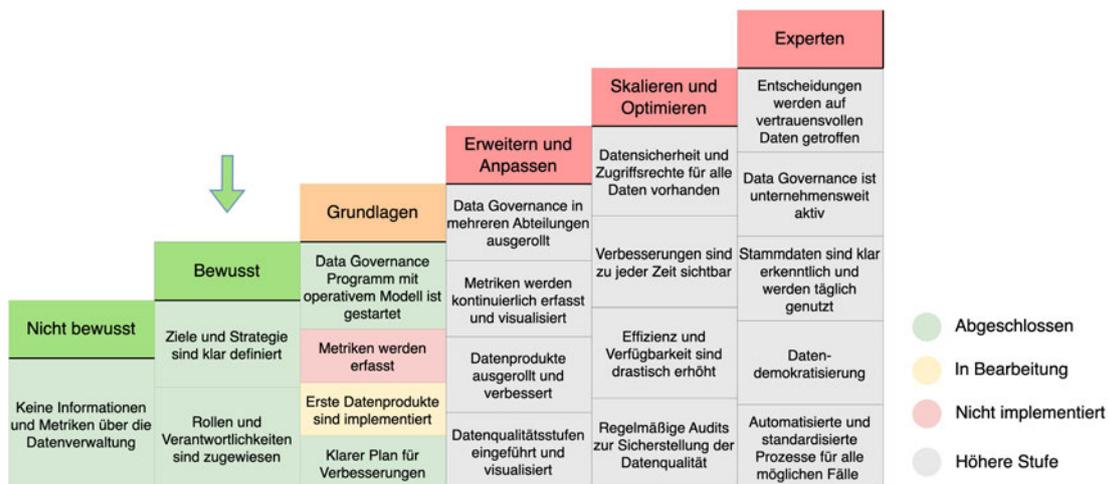


Abbildung 5.2: Aktueller Stand des Reifegrads der AquaFlow GmbH

Um ein umfassendes Bild des aktuellen Reifegrads zu erhalten, muss das Assessment um weitere Fragen ergänzt werden. Dabei ist wichtig, dass alle Fragen in unterschiedlichen Abstufungen beantwortet werden können, um den Fortschritt bestimmen zu können und Spielraum für weitere Verbesserungen zuzulassen. Die abgestuften Antworten lassen sich zu spezifischen Charakteristiken zusammenfassen, wie es in Abbildung 5.2 beispielhaft dargestellt ist.

Vor der Umsetzung der Schritte und Maßnahmen in diesem Kapitel befand sich die AquaFlow GmbH in der Stufe *Nicht bewusst*. Durch die Zuweisung der Rollen und Verantwortlichkeiten so wie die klare Zielsetzung und Strategiewaueisung hat das Unternehmen mittlerweile die Bedeutung von Data Governance erkannt und beginnt, seine Aktivitäten gezielt danach auszurichten. Es befindet sich in der Stufe *Bewusst* und hat begonnen, die Aktivitäten, die für das Erreichen der nächsten Stufe *Grundlagen* umgesetzt sein müssen, zu implementieren. Die bereits umgesetzten Punkte werden in Grün dargestellt, diejenigen, die sich gerade in Bearbeitung befinden oder optimiert werden, in Gelb, und die noch nicht implementierten Punkte in Rot. Diese farbliche Darstellung bietet eine klare Übersicht darüber, welche Maßnahmen noch fehlen und worauf der Fokus gelegt werden sollte. Nach der Umsetzung weiterer Maßnahmen wird der Reifegrad des Unternehmens zu einem späteren Zeitpunkt der Arbeit neu untersucht, um den weiteren Fortschritt zu überprüfen.

## 5.5 Operatives Modell

Das in Unterabschnitt 4.3.2 definierte und in Abbildung 4.3 dargestellte operative Modell wird ebenfalls von der AquaFlow GmbH übernommen. Dieses Modell dient dazu, allen Beteiligten ihre regelmäßig anfallenden Aufgaben aufzuzeigen und sicherzustellen, dass nichts Wichtiges vergessen wird. Es definiert dabei nicht nur die einzelnen Aufgaben und Prozesse, sondern gibt auch die Kommunikationswege vor, indem es die Regelmeetings und andere Kommunikationskanäle festlegt. Es ist dabei die Aufgabe des CDO, das operative Modell zu pflegen und zusammen mit dem DGO regelmäßig zu überarbeiten. Der CDO ist in allen drei Ebenen präsent und überwacht die Einhaltung der Prozesse sowie die Durchführung der Maßnahmen durch die Beteiligten.

Neben den in Abbildung 4.3 dargestellten Aufgaben und Aktivitäten ist es auch die Aufgabe des operativen Modells, Strukturen für Prozesse vorzugeben, um diese zu vereinheitlichen und optimiert ausführen zu lassen. Im Folgenden werden solche Vorgaben für drei zentrale Prozesse vorgestellt. Diese beinhalten den Umgang mit Datenqualitätsproblemen, die Integration neuer Datensätze sowie die Einführung und Durchführung von Datenkontrollen.

Das in Abbildung 5.3 dargestellte Ablaufdiagramm zeigt den standardisierten Prozess zur Einhaltung der Datenqualität in der AquaFlow GmbH. Werden bei der automatisierten oder manuellen Überprüfung Probleme bezüglich der Datenqualität erkannt, so werden entsprechende Schritte eingeleitet. Dadurch wird eine systematische, koordinierte und wiederholbare Vorgehensweise bei der Behebung von Datenqualitätsfehlern gewährleistet.

Ebenso ist die Einführung klarer Prozessvorgaben für die Integration neuer Datensätze sinnvoll, um sicherzustellen, dass nur qualitativ hochwertige Daten in das System aufgenommen werden und dass diese Daten allen relevanten Nutzern zur Verfügung stehen. Dafür durchläuft der Datensatz mehrere Prüfungen. Dazu gehören die Vorprüfung und die Sicherstellung der Übereinstimmung mit dem Business-Glossar. Durch das Hinzufügen zum Datenkatalog und dem Anreichern von Metadaten wird sichergestellt, dass der Datensatz leicht auffindbar ist und von verschiedenen Teams für unterschiedliche Anwendungen genutzt werden kann. Die Zuweisung eines klaren Verantwortlichen und die Vergabe von Zugriffsrechten gewährleistet die Sicherheit und Qualität der Daten. Mit diesen strukturierten Schritten wird eine effektive Integration neuer Daten ermöglicht. Der vollständige Prozessablauf ist in Abbildung A.2 dargestellt.

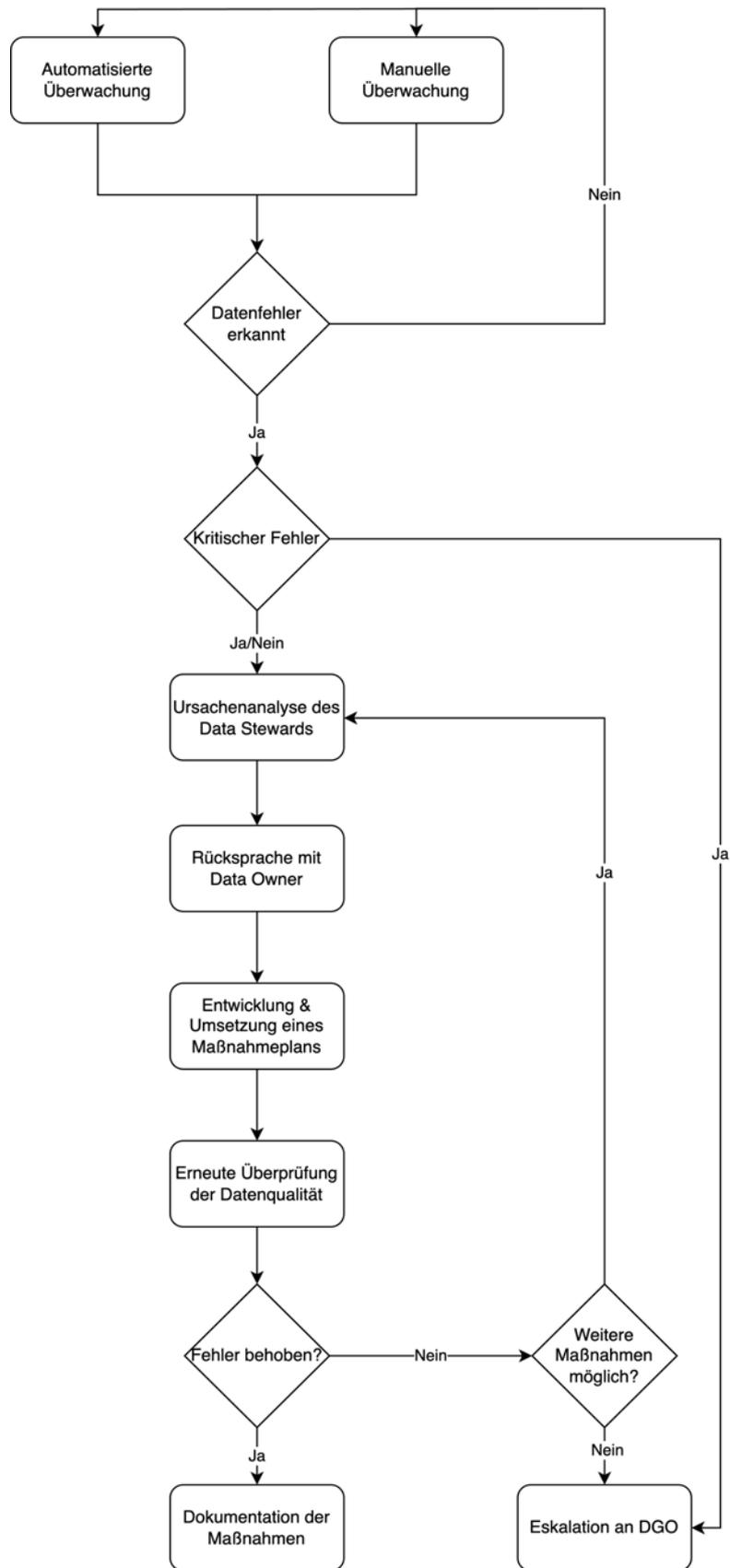


Abbildung 5.3: Prozessablauf für Probleme mit der Datenqualität

Datenkontrollen sind ein essenzieller Bestandteil für die Sicherung der Datenqualität. Das in Abbildung A.3 dargestellte Ablaufdiagramm zeigt die Prozessvorgaben zur Einführung solcher Kontrollen. Dabei werden zunächst die relevanten und zu überprüfenden Daten untersucht. In Absprache mit verschiedenen Stakeholdern werden klare Ziele für die Datenkontrolle sowie Ziel- und Schwellwerte, die verschiedene Aktionen auslösen, definiert. Im Anschluss daran werden geeignete Kontrollmechanismen entwickelt, die die gesetzten Ziele erreichen und die Qualität der Daten überprüfen. Ebenso werden Darstellungsoptionen wie Visualisierungen, Dashboards und Berichte definiert und implementiert, die eine Überwachung der Datenqualität und der Ergebnisse durch die Verantwortlichen ermöglichen. Mit einem Prüfplan wird die Häufigkeit sowie die Art und Weise der Datenkontrollen festgelegt. Ebenso werden darin Maßnahmen für Fälle formuliert, in denen Zielwerte nicht erreicht werden. Dies umfasst beispielsweise die Definition von Eskalationsreihenfolgen, die Durchführung zusätzlicher Überprüfungen und etwaige Anpassungen der Verarbeitungsschritte und Kontrollmechanismen. Im Anschluss daran erfolgt eine initiale Datenkontrolle, bei der die implementierten Schritte überwacht werden und geprüft wird, ob die Mechanismen alle zuvor formulierten Ziele und Anforderungen erfüllen. Im Falle von Abweichungen werden notwendige Anpassungen vorgenommen. Nach Erfüllung der Anforderungen wird eine abschließende Überprüfung durch den Data Steward durchgeführt. Im Anschluss daran werden sämtliche Prozesse, Ziele und Abläufe dokumentiert, woraufhin die Freigabe durch den Data Owner und den Data Steward erfolgt.

## 5.6 Technologische Umsetzung

Die bisherigen Aktivitäten und Maßnahmen konzentrierten sich auf die organisatorischen Aspekte und die Definition zentraler Strategien und Rollen für die AquaFlow GmbH. Um die definierten Ziele von Data Governance auch im operativen Alltag zu verwirklichen, bedarf es konkreter Implementierungen, die diese Vorgaben in die Praxis umsetzen. Dazu gehören unter anderem ein Datenkatalog, ein Business Glossar, Technologien und Pipelines zur Sicherstellung der Datenqualität sowie Automatisierungswerkzeuge, die die einzelnen Schritte und Maßnahmen automatisiert ausführen.

## Datenkatalog

Ein Datenkatalog spielt ebenfalls für KMUs eine wesentliche Rolle und wird im Core-Governance Framework als zentraler Bestandteil für das Erreichen der Data Governance Ziele angesehen. In Unterabschnitt 4.3.2 werden die Anforderungen und Ziele an den Datenkatalog sowie die dahinter stehende Technologie definiert. Insgesamt dient ein Datenkatalog als zentrales Verzeichnis für alle verfügbaren Datenquellen und Datenbestände. Das Verzeichnis umfasst die Metadaten der verschiedenen Datenbeständen, die sowohl automatisiert als auch manuell hinzugefügt werden. Dazu gehören Informationen über Formate, Herkunft, bisherige Nutzungen sowie die verantwortliche Person für die Daten. Ebenso ist es wichtig, im Datenkatalog eine klare Rollenverteilung und Darstellung der Verantwortlichkeiten zu integrieren. Dies erhöht die Transparenz und Effizienz, indem es eindeutig festlegt, wer für die Verwaltung und den Zugriff auf bestimmte Daten zuständig ist. So wird sichergestellt, dass alle Beteiligten ihre Aufgaben kennen und effizient zusammenarbeiten können. Eng im Zusammenhang mit Datenkatalogen steht häufig das Business Glossar. Dieses bietet einheitliche Begriffsdefinitionen und Zugriffskonzepte, die im gesamten Unternehmen angewendet werden.

Aus diesen Kernfunktionen des Datenkatalogs ergeben sich mehrere Anforderungen an die Technologie, die den Datenkatalog umsetzt.

- **Einfache Einbindung von verschiedenen Datenquellen:** Durch die Vielzahl an Datenquellen muss die Technologie gängige Formate und Technologien unterstützen und diese problemlos einbinden.
- **Automatisierte Erfassung von Metadaten:** Durch die Menge an Datenbeständen ist die manuelle Erfassung aller Metadaten zu aufwendig. Daher muss die Technologie Metadaten automatisch aus den eingebundenen Datenquellen extrahieren können. Jedoch müssen auch manuelle Anpassungen vorgenommen werden können.
- **Benutzerfreundliche Oberfläche:** Eine intuitive und übersichtliche Benutzeroberfläche ist entscheidend, um die Akzeptanz des Datenkatalogs im Unternehmen zu erhöhen.
- **Möglichkeit zur Zuweisung von Rollen:** Die Zuweisung von Rollen und Verantwortlichkeiten müssen mit der Technologie möglich sein, um Zugriffsrechte flexibel

und skalierbar vergeben zu können und gleichzeitig die Sicherheit sensibler Daten zu gewährleisten.

- **Erweiterbare Glossarfunktion:** Über einen integrierten Business Glossar können Begriffsdefinitionen einheitlich verwaltet und bei Bedarf erweitert werden.
- **Skalierbarkeit und Flexibilität:** Die Lösung sollte skalierbar und flexibel sein, um den wachsenden Anforderungen des Unternehmens gerecht zu werden. Dies bedeutet, dass der Datenkatalog sowohl mit zunehmendem Datenvolumen als auch mit sich verändernden Geschäftsanforderungen mithalten kann.

Hamma hat in [32] Technologien für Data Governance im Kontext von KMUs untersucht. Basierend auf dem Abgleich der Anforderungen und den Ergebnissen dieser Untersuchung wird in der AquaFlow GmbH *Datahub* als Datenkatalog verwendet. Datahub ist eine Open-Source-Plattform, die von *LinkedIn* mit dem Ziel entwickelt wurde, die Verwaltung von Metadaten und die Prozesse der Datenerkennung zu optimieren. Erreicht wird dies durch ein zentrales Repository, in dem alle Metadaten gespeichert, organisiert und abgerufen werden können. Diese umfassen alle möglichen Arten von Metadaten wie Schema, Abhängigkeiten, Eigentümer und Anmerkungen. Im Hintergrund nutzt Datahub die Volltextsuche-Engine *Elasticsearch*, um eine schnelle und effiziente Durchsuchbarkeit der Metadaten zu ermöglichen. Durch die Indizierung werden die Datensätze durchsuchbar, so dass die Benutzer anhand von Namen, Tags, Datenquellen und vielem mehr nach bestimmten Datensätzen suchen können. In Abbildung A.4 ist die Startseite von Datahub mit Suchleiste zu sehen. Neben dem gezielten Suchen nach Datenquellen können Nutzer auch in den vorhandenen Datensätzen stöbern. In Abbildung A.5 ist die Übersicht über die existierenden Datensätze in der AquaFlow GmbH in Listenform zu sehen, wodurch sich Anwender ebenfalls einen Überblick über die vorhandenen Datenquellen und Datensätze verschaffen können.

Die Einbindung von Datenquellen erfolgt in Datahub mittels sogenannten *Rezepten*. In diesen wird konfiguriert, wie die Metadaten extrahiert, transformiert und verarbeitet werden. Dies gewährleistet ein flexibles und skalierbares System, das für verschiedene Datenquellen geeignet ist und sich an unterschiedliche Anforderungen anpassen lässt. Zusätzlich besitzt Datahub die Funktion des Profilings der Daten, das für die Identifizierung weiterer Merkmale genutzt werden kann. Ebenso besteht die Möglichkeit, die Datenherkunft zu verfolgen, um Verarbeitungsschritte und Abhängigkeiten zwischen Daten nachvollziehen zu können.

Datahub bietet die Möglichkeit, direkt in der Anwendung ein Business Glossar zu erstellen und zu verwalten. Dabei können verschiedene Begriffe und Begriffsgruppen angelegt und bestimmten Datensätzen zugeordnet werden. In Abbildung A.6 sind die für die AquaFlow GmbH definierten Begriffsgruppen dargestellt. Mit der Gruppe *Classification* werden den Daten Sicherheitsstufen zugewiesen, über die der Zugriff für Personen und Gruppen geregelt werden kann. So kann zwischen vertraulichen und nicht vertraulichen Daten unterschieden werden. Andere Begriffe und Begriffsgruppen können diese Klassifizierungen erben, so dass beispielsweise alle personenbezogenen Daten den sensiblen Daten zugeordnet werden. In den Begriffsgruppen *Financial Data* und *Personal Information* werden Terminologien festgelegt und definiert, um sicherzustellen, dass Begriffe eindeutige Bedeutungen haben und nicht für andere Inhalte verwendet werden. In der Begriffsgruppe *Legal-Regulatory-Guidelines* sind die für das Unternehmen geltenden Gesetze, Vorschriften und Normen aufgeführt. Die zentrale Sammlung an einem Ort erleichtert das Auffinden der relevanten Vorgaben für die Anwender, wodurch die Einhaltung von Compliance-Vorschriften erleichtert wird. In Abbildung A.7 sind die für die AquaFlow GmbH bisher relevanten Gesetze und Normen im Datenkatalog zu sehen. Falls erforderlich, können spezifische Gesetze und Normen direkt den entsprechenden Datensätzen zugewiesen werden. In vielen Fällen ist die zentrale Sammlung in der Begriffsgruppe ausreichend, um den Überblick zu behalten. Die Begriffsgruppe ist unterteilt in die Kategorien GDPR (dt. DSGVO) und ISO-Standards, um eine gezielte Zuordnung und Navigation durch die verschiedenen Anforderungen zu ermöglichen. Jeder Eintrag enthält eine prägnante Beschreibung, den zuständigen Ansprechpartner für eventuelle Rückfragen sowie Verweise auf weitere Informationen und Ressourcen.

Neben den standardmäßigen tabellarischen Informationen wie Schema und Format können in Datahub auch weitere, beschreibende Informationen hinzugefügt werden. Diese erweiterten Metadaten ermöglichen es, den Verwaltern der Daten, den Kontext und den Zweck der Daten umfassender zu dokumentieren. Dadurch sind die Daten für Anwender schneller auffindbar und besser verständlich. In Abbildung A.8 sind die zusätzlichen Informationen dargestellt. Hierzu zählen Beschreibungen über den Datensatz sowie deren etwaige Verwendungen. Ebenso werden die Profile der Besitzer und Verwalter der Datensätze angezeigt. Der Datensatz *customers* aus dem CRM-System ist Herrn Müller, dem stellvertretenden Vertriebsleiter, zugewiesen. Die technische Betreuung der Daten übernehmen die beiden Data Stewards. *Tags* sind eine Möglichkeit, relevante Themenbereiche und Beziehungen zwischen verschiedenen Datensätzen einfach darzustellen. Über Schlüsselbegriffe können Anwender so ähnliche Datensätze schnell identifizieren und mit-

einander verknüpfen. Im Feld *Glossary Terms* werden Begriffe aus dem Glossar angezeigt, die im Datensatz vorkommen. Dadurch wird beispielsweise auf einen Blick deutlich, dass der Datensatz persönliche und sensible Daten enthält und möglicherweise nicht allen Anwendern ohne Weiteres zur Verfügung steht.

### **Zugriffsrechte**

Für den Schutz sensibler Daten und die Einhaltung gesetzlicher Vorschriften ist die Verwaltung von Zugriffsrechten entscheidend. Durch die Zuweisung der Datensätze zu Abteilungen und Abteilungsleitern werden die Verantwortlichkeiten klar geregelt. Die Verantwortlichen müssen sicherstellen, dass die Daten korrekt klassifiziert sind und Zugriffsrechte bei Bedarf erteilt werden. Für die Vergabe der Zugriffsrechte werden die IAM-Systeme der zugrundeliegenden Technologien genutzt. Im Datenkatalog ist ersichtlich, wer für die Daten verantwortlich ist und die Zugriffsrechte vergeben kann. Über das verknüpfte Profil kann der Benutzer beispielsweise per E-Mail die Zugriffsrechte beantragen. So hat zum Beispiel die Abteilungsleiterin der Entwicklungsabteilung alle Rechte für die in MinIO gespeicherten IoT-Sensordaten. MinIO verwendet, wie bereits erwähnt, die S3-API, wodurch die Rechteverteilung wie bei Amazon Web Services (AWS) direkt über Code gesteuert werden kann. In Listing A.1 ist die Erstellung der Policy zu sehen, die alle Rechte für den Zugriff und die Verwaltung des Buckets der IoT-Sensordaten vergibt. Wenn eine Person Zugriff auf den Datensatz anfordert, sollten in den meisten Fällen nur Leserechte gewährt werden. Die Erstellung der Leserechte für diesen Datensatz ist in der Listing A.2 zu sehen. Über die Kommandozeilenschnittstelle oder die grafische Benutzeroberfläche können die Policies Gruppen und Benutzern zugewiesen werden.

Durch diese flexible Rechtevergabe können die Zugriffsrechte fein granular definiert und vergeben werden. Data Owner erhalten dabei alle Rechte für einen Datensatz, während die meisten Anwender und Analysten in der Regel nur Leserechte erhalten. Data Stewards wiederum erhalten erweiterbare Rechte für alle Datensätze. Durch diese differenzierte Zuweisung wird sichergestellt, dass jeder Benutzer lediglich die Berechtigungen erhält, die für seine Aufgaben benötigt werden.

### **Datenqualität**

Eine hohe Datenqualität ist ein zentrales Ziel aller Data Governance Aktivitäten. Um ein konsistent hohes Qualitätsniveau sicherzustellen, reichen manuelle Kontrollen nicht aus.

Stattdessen sind flexible Technologien erforderlich, die auch große Datenmengen effizient überprüfen und bei Abweichungen Rückmeldung liefern. Dadurch ergeben sich an die einzusetzenden Technologien eine Reihe spezifischer Anforderungen.

- **Automatisierung:** Aufgrund des kontinuierlichen Datenflusses und der zunehmenden Datenmengen müssen die Kontrollen automatisiert durchführbar und nahtlos in bestehende Prozesse integrierbar sein.
- **Skalierbarkeit:** Eine weitere Anforderung ist eine Skalierbarkeit, um sich an potenziell wachsende Datenmengen anpassen zu können.
- **Flexibilität:** Neben dem Durchführen von einfachen Standardmessungen muss die Technologie in der Lage sein, komplexere und anwendungsspezifische Abfragen zu ermöglichen. Dadurch können alle Geschäftsanforderungen an die Daten abgedeckt werden.
- **Benutzerfreundlichkeit und Integration:** Eine einfache Bedienung sowie eine unkomplizierte Einrichtung und Einbindung in bestehende Prozesse erhöhen die Akzeptanz und die Nutzung der Technologie. Möglicherweise gibt es die Möglichkeit, dass auch nicht-technische Nutzer Kontrollen entwickeln können.
- **Echtzeit-Messungen:** Die Fähigkeit zur Durchführung von Echtzeit-Messungen ist von großer Bedeutung, um sofortige Einblicke in die Datenqualität zu erhalten und schnell auf auftretende Probleme reagieren zu können.

Basierend auf den Ergebnissen von Hamma in [32] hat sich die AquaFlow GmbH für den Einsatz von zwei Technologien zur Sicherung der Datenqualität entschieden. Zum einen wird die Open-Source Python-Bibliothek *Soda Core* ausgewählt, da es sich nahtlos in bestehende Systeme integrieren lässt und durch eine eigene Kontrollsprache auch von nicht-technischen Anwendern genutzt werden kann. Neben einer großen Auswahl an Standardprüfungen wie die Erfassung von Metriken wie Maxima oder Durchschnitte, Gültigkeitsprüfungen und Vollständigkeitsprüfungen gibt es in Soda auch die Möglichkeit, eigene Kontrollen mit SQL-Abfragen zu schreiben.

Zum anderen fällt die Wahl auf die Open-Source Bibliothek *Deequ*, um große Datenmengen, wie IoT-Daten oder Meldungen aus den Produktionsanlagen, effizient zu überwachen und auch zukünftig die Kontrollen bei weiter wachsenden Datenmengen durchführen zu können. Dies wird durch den Einsatz von Apache Spark als zugrundeliegende Technologie erreicht. Da die AquaFlow GmbH Apache Spark bereits als Datenverarbeitungsprogramm

verwendet, ist eine nahtlose Integration in die bestehende Infrastruktur möglich. Ähnlich wie bei Soda Core gibt es in Deequ eine Reihe von Basisprüfungen, die durch individuelle Kontrollen erweitert werden können.

Im Folgenden wird exemplarisch die Qualitätsmessung eines Datensatzes vorgestellt. Dabei handelt es sich um die Kundendaten aus dem CRM-System, für die es bisher keine Informationen über die Qualität der Einträge gibt. Zur Ermittlung der Qualität dieser Daten wird eine automatische Prüfung erstellt. Aufgrund der relativ kleinen Größe des Datensatzes wird Soda Core für die Durchführung der Kontrollen verwendet. Bei ersten manuellen Überprüfungen wird ersichtlich, dass die Fehler größtenteils bei falsch formatierten E-Mailadressen und fehlerhaften Telefonnummern liegen. Aus diesem Grund werden mit Soda Core Kontrollen implementiert, die sowohl die Validität als auch das korrekte Format der Einträge überprüfen. Abbildung A.9 zeigt die Ergebnisse der ersten Kontrollen. Hierbei kann ein Schwellenwert definiert werden, der festlegt, ab wann ein Check als *nicht bestanden*, *Warnung* oder *bestanden* markiert wird. In diesem Fall zeigt die Prüfung, dass knapp 12% der E-Mail-Adressen und 16% der Telefonnummern im Datensatz falsch formatiert sind und der Check dadurch nicht bestanden wird. Nach Erkennung des Problems können gezielt entsprechende Maßnahmen eingeführt werden. Zu den Maßnahmen in der AquaFlow GmbH zählen unter anderem die Sensibilisierung der Mitarbeiter für die korrekte Eingabe von Daten, um Fehler künftig zu vermeiden. Zusätzlich werden automatisierte Bereinigungsverfahren implementiert, wie die Korrektur von Formatfehlern anhand vordefinierter Muster. Durch diese Maßnahmen konnten die Fehler auf 5% bei E-Mailadressen und 6% bei Telefonnummern reduziert werden. Ein zusätzlicher Ansatz für eine weitere Verbesserung könnte die Einführung von Validierungsregeln direkt bei der Dateneingabe sein, um fehlerhafte Daten bereits im Erfassungsprozess zu verhindern.

### **Automatisierungswerkzeug**

Mit Hilfe von Automatisierungswerkzeugen kann sichergestellt werden, dass Maßnahmen auch langfristig und konsistent umgesetzt werden. So werden sowohl akute Prozesse wie die Überprüfung der Datenqualität und die Datenbereinigung als auch langfristige Initiativen wie die Archivierung und andere Lebenszyklusaufgaben zuverlässig durchgeführt. Insgesamt ergeben sich dadurch eine Reihe an Anforderungen an die Technologie.

- **Kontinuierliche Überwachung und Anpassung:** Die Technologie muss die Prozesse zuverlässig und kontinuierlich ausführen und bei veränderten Anforderungen anpassbar sein.
- **Compliance und Dokumentation:** Alle ausgeführten Schritte müssen dokumentiert werden und nachvollziehbar sein, um die Einhaltung von Compliance-Vorgaben sicherzustellen und bei Verstößen die Ursachen identifizieren zu können.
- **Skalierbarkeit:** Aufgrund der wachsenden Datenmengen muss die Technologie skalierbar sein, um auch in Zukunft bei steigender Komplexität alle Prozesse lückenlos ausführen zu können.
- **Benutzerfreundlichkeit:** Die Technologie sollte neben der Möglichkeit der Konfiguration mittels Code auch eine benutzerfreundliche grafische Oberfläche bieten. In dieser Oberfläche sollten Anwender in der Lage sein, kleine Änderungen wie die Anpassung von Parametern vorzunehmen und Pipelines und Prozesse zu aktivieren und zu deaktivieren. Darüber hinaus muss die Oberfläche die Möglichkeit bieten, laufende Prozesse in Echtzeit zu überwachen und bereits durchgeführte Vorgänge detailliert nachzuvollziehen.

Auf Grundlage der Untersuchungen von Hamma in [32] wird in der AquaFlow GmbH *Apache Airflow* für die Automatisierung und Orchestrierung der Prozesse und Abläufe verwendet. Apache Airflow ist eine Open-Source Plattform für die Entwicklung und Überwachung von Workflows. Die Workflows werden dabei über Code erstellt und können dadurch beliebig komplex gestaltet werden. Pipelines und Workflows können dank der flexiblen Planungssemantik von Apache Airflow in regelmäßigen Abständen automatisiert ausgeführt werden. Darüber hinaus ermöglicht Airflow die verteilte Ausführung von Workflows auf verschiedenen Knoten, wodurch die Automatisierung skalierbar wird. Individuelle Integratoren bieten zudem die Möglichkeit, Apache Airflow nahtlos in bestehende Cluster- und Cloud-Umgebungen zu integrieren. Abbildung A.10 zeigt die Darstellung eines Workflows zur Prüfung der Datenqualität in Form eines Graphen. Dabei werden die Daten aus der Datenquelle vorgeladen und mit Soda Core geprüft. Anschließend werden die Ergebnisse unabhängig vom Resultat in Elasticsearch geschrieben. Je nach Ausgang der Prüfung werden die verantwortlichen Personen mittels E-Mail und Microsoft Teams Nachricht informiert und weitere Schritte durchgeführt oder die Daten in die Datenbank geladen.

## 6 Ergebnisse & Analyse

In diesem Kapitel werden das theoretische Framework CoreGovernance sowie dessen Implementierung und Anwendung in der Testumgebung des fiktiven Unternehmens AquaFlow GmbH untersucht. Zunächst wird CoreGovernance qualitativ analysiert, indem verschiedene Aspekte und Domänen bewertet werden. Im Anschluss daran folgt eine quantitative Untersuchung, die sich auf die praktische Umsetzung des Frameworks in der Testumgebung konzentriert und die Auswirkungen davon untersucht. Abschließend werden die Ergebnisse und Erkenntnisse in einer allgemeinen Diskussion zusammengeführt.

### 6.1 Qualitative Untersuchung

Im folgenden Abschnitt wird die qualitative Analyse des CoreGovernance Frameworks durchgeführt. Der Fokus liegt dabei auf der Untersuchung, ob das Framework alle relevanten Bereiche abdeckt, die für Data Governance und insbesondere für die Anforderungen von KMUs entscheidend sind. Dafür wird die entwickelte Domänenmatrix verwendet, um die Abdeckung und Umsetzung des Frameworks systematisch zu bewerten. Anschließend wird das entwickelte CoreGovernance Framework mit den anderen vorgestellten Frameworks verglichen. Dabei wird untersucht, ob die Aspekte, die bei diesen Frameworks für KMUs problematisch waren, im CoreGovernance Framework berücksichtigt und gelöst werden konnten.

#### 6.1.1 Domänenmatrix

Die Domänenmatrix dient dazu, zu prüfen, ob das CoreGovernance Framework alle relevanten Bereiche und Aspekte der Data Governance abdeckt. Dafür werden die in drei Kategorien zusammengefassten, wesentlichen Maßnahmen und Strukturen, die nach Alhassan et al. für eine erfolgreiche Umsetzung von Data Governance notwendig sind, in den

Phasen	Bereiche von Data Governance	Entscheidungsdomänen				
		Datenprinzipien	Datenqualität	Metadata	Datenzugriff	Datenlebenszyklus
Definition	Governance Struktur	Grün	Grün	Grün	Grün	Grün
	Daten Operationen	Grün	Grün	Grün	Grün	Gelb
	Datenstrategie & Technologien	Grün	Grün	Grün	Grün	Grün
Implementierung	Governance Struktur	Grün	Grün	Grün	Grün	Grün
	Daten Operationen	Grün	Grün	Grün	Grün	Gelb
	Datenstrategie & Technologien	Grün	Grün	Grün	Grün	Grün
Überwachung	Governance Struktur	Grün	Grün	Grün	Grün	Gelb
	Daten Operationen	Gelb	Grün	Grün	Gelb	Grün
	Datenstrategie & Technologien	Gelb	Gelb	Grün	Gelb	Gelb

Abbildung 6.1: Domänenmatrix des CoreGovernance Frameworks

drei Bereichen Definition, Implementierung und Überwachung untersucht. Diese Untersuchung erfolgt für jeden der fünf Domänen, die nach Khatri und Brown für datenbasierte Entscheidungen entscheidend sind. In Abbildung 6.1 ist die ausgefüllte Domänenmatrix für das theoretische Framework zu sehen. Grün bedeutet dabei, dass das Framework für diese Kernmaßnahme bzw. -aktivität in der bestimmten Phase in der Datendomäne klare Anweisungen gibt und das Thema umfangreich behandelt. Ist das Feld gelb markiert, so sind teilweise Anweisungen vorgegeben. Rot bedeutet, dass die Maßnahme nicht ausreichend thematisiert wird und es wenige bis keine Anweisungen dafür gibt.

Die Thematik **Datenprinzipien** wird in CoreGovernance grundsätzlich behandelt und ist ein wesentlicher Bestandteil des Strategie- und Visionsteils. In diesem Kontext wird klar definiert, worauf es bei der Festlegung von Datenprinzipien ankommt wie beispielsweise die Übereinstimmung mit übergeordneten Unternehmenszielen. Die Verantwortung

für die Definition, Einhaltung und Pflege der Prinzipien wird dem DGO zugewiesen. Durch die Vorgabe von Prozessen werden alle drei Kategorien in der Definitionsphase als erfüllt angesehen.

In der Implementierungsphase regelt das operative Modell die Einhaltung der Verantwortlichkeiten und Umsetzung der Datenprinzipien. Das Modell definiert regelmäßige Meetings, legt Prozesse fest und führt die Aufgaben auf, die sicherstellen, dass die Datenprinzipien umgesetzt werden und mit der Unternehmensstrategie im Einklang stehen.

Die Überwachung der Datenprinzipien wird ebenfalls über das operative Modell geregelt. Im Modell sind ebenfalls klare Termine und Verantwortlichkeiten für die Überwachung durch das DGO auf der strategischen Ebene festgelegt. Etwas unklar bleibt, wann und wie die definierten Prozesse und Standards konkret überwacht werden sollen, weshalb hier das Feld mit gelb markiert ist.

Die Sicherstellung der **Datenqualität** ist ein wesentlicher Bestandteil des Frameworks. Zu Beginn der Maßnahmen müssen klare Standards und Dokumentationen für Prozesse, Abläufe und Datenflüsse erstellt werden. Das DGO trägt die Hauptverantwortung für die Datenqualität im Unternehmen. Die Data Owner sind verantwortlich für ihre Domäne und müssen in regelmäßigen Abständen Rechenschaft vor dem DGO ablegen. Ebenso werden Anforderungen an Technologien genannt wie etwa eine automatisierte Überwachung und Rückmelden bei Nicht-Einhaltung.

In der Implementierungsphase sieht das CoreGovernance Framework vor, dass Soll-Werte für die Datenqualität in den verschiedenen Domänen definiert und kontinuierlich mit den aktuellen Werten abgeglichen werden. Die Umsetzung der Maßnahmen tragen dabei alle Anwender mit, wobei die Data Owner und Data Stewards die Einhaltung sicherstellen. Die Prozesse für die Umsetzung sind im operativen Modell verankert. Die technologischen Anforderungen, die zuvor formuliert wurden, werden praktisch umgesetzt.

Die kontinuierliche Überwachung der Maßnahmen zur Sicherstellung der Datenqualität wird im Framework ebenfalls stark betont. Die Rechenschaftspflicht liegt beim DGO, das sicherstellen muss, dass die festgelegten Standards eingehalten werden. Visualisierungen sollen dabei helfen, die Datenqualität optisch darzustellen und Abweichungen sofort erkennbar zu machen. Allerdings bleibt unklar, wann und ob die festgelegten Soll-Werte überarbeitet werden müssen.

**Metadaten** spielen im CoreGovernance Framework eine wesentliche Rolle. Durch die Forderung nach Datendemokratisierung bestehen von Anfang an Anforderungen an die

Metadaten. Durch die Zuweisung der Datensätze zu Abteilungen sind auch die Verantwortlichkeiten für die Metadaten geregelt. Die definierten Prozesse und Anforderungen spiegeln sich in den Anforderungen an den Datenkatalog wider.

In der Implementierungsphase ist die Wahl und Umsetzung eines geeigneten Datenkatalogs, der die zuvor definierten Governance-Strukturen und Datenoperationen unterstützt, entscheidend. Dieser stellt sicher, dass Verantwortlichkeiten und Strukturen sichtbar und zugänglich in den Datensätzen dokumentiert sind. Außerdem wird im Framework die Bedeutung automatisierter Methoden zur Extraktion und Verwaltung von Metadaten hervorgehoben.

Die Überwachung und Pflege der Metadaten ist ebenfalls im operativen Modell verankert. Der CDO trägt die Hauptverantwortung für den Datenkatalog und stellt sicher, dass die Überwachung der Technologie und die Pflege der strategischen Ausrichtung kontinuierlich erfolgen. Dies gewährleistet, dass die Metadaten stets aktuell und relevant bleiben und die strategischen Ziele des Unternehmens unterstützen.

Die Bedeutung von **Zugriffsrechten** zum Schutz sensibler Daten wird im Framework mehrfach betont. Die Erstellung eines Sicherheits- und Zugriffsmodells durch das DGO ist dabei von Beginn an wichtig. Anforderungen und die Umsetzung der Technologien und Strategien wird ebenfalls angesprochen und thematisiert, um sicherzustellen, dass nur berechnigte Personen Zugriff auf die Daten haben.

Für die Implementierung der Sicherheitsvorkehrungen sind die einzelnen Data Owner verantwortlich. Diese sollen dabei die in der Technologie eingebauten Sicherheitsvorkehrungen nutzen. Zusätzlich muss eine übergreifende Struktur aus Rollen und Rechten etabliert werden, um sicherzustellen, dass der Zugriff auf Daten kontrolliert und autorisiert erfolgt.

Die Verantwortlichkeiten für die Überwachung der Zugriffsrechte sind durch das operative Modell geregelt. Allerdings bleibt die Frage offen, wie genau diese Überwachung durchgeführt werden soll. Ebenso könnte das Framework stärker auf mögliche Verbesserungsmaßnahmen wie die Einführung von Audit-Trails eingehen. Dabei stellt sich jedoch die Frage, wie umsetzbar und relevant solche Maßnahmen für KMUs sind.

Allgemeine Strategien und Richtlinien für ein **Lebenszyklusmodell** müssen zu Beginn der Data Governance Bemühungen vom DGO erarbeitet werden. Dies legt den Grundstein für die Schaffung geeigneter Lebenszyklusmodelle im Unternehmen, muss jedoch

mit den bisherigen Bemühungen vom Datenmanagement abgestimmt werden. Da die Anforderungen an das konkrete Lebenszyklusmodell stark von der Art und Nutzung der Daten abhängen, werden im CoreGovernance Framework noch keine konkreten Prozesse definiert.

Bei der Implementierung werden die allgemeinen Richtlinien und die domänenspezifischen Anforderungen in konkrete Prozesse und Technologien überführt. Pipelines und Datenflüsse stellen den kontrollierten Fluss von Daten durch die verschiedenen Phasen des Lebenszyklus sicher. Für die Umsetzung dieser Pipelines sind die Data Owner verantwortlich, die in enger Zusammenarbeit mit den Data Stewards arbeiten, um eine effektive Datenflusstruktur zu entwickeln.

Durch die iterative Umsetzung der Prozesse wird eine fortlaufende Überwachung der einzelnen Standards und Vorgaben gewährleistet. Prüfungen und Audits stellen sicher, dass die definierten Prozesse eingehalten werden und etwaige Abweichungen rechtzeitig erkannt werden. Wie und von wem diese technischen Überwachungsmaßnahmen umgesetzt werden, ist situationsabhängig und wird im Rahmen des Frameworks nicht näher beschrieben.

### 6.1.2 Vergleich mit den vorgestellten Frameworks

In Abbildung 4.1 werden die Ergebnisse der Analyse der drei vorgestellten Frameworks dargestellt. Dabei wird unter anderem aufgezeigt, in welchen Bereichen die einzelnen Konzepte für den Einsatz in KMUs angepasst werden müssen. Die drei analysierten Frameworks von DAMA, DGI und PwC weisen insgesamt viele Stärken auf, besitzen jedoch auch einige Merkmale, die ihre Anwendbarkeit in KMUs einschränken können.

Insgesamt sind die bisher existierenden Frameworks für Data Governance oft zu komplex und umfangreich, um den spezifischen Anforderungen von KMUs gerecht zu werden. Der Fokus der Frameworks richtet sich auf große Unternehmen, was sich in der weitreichenden und komplexen Gestaltung der Rollen und Prozesse zeigt. Insbesondere für weniger erfahrende Anwender erschwert dies die Identifizierung der für sie relevanten Bereiche. Darüber hinaus fehlen in den Frameworks aufgrund der konzipierten Allgemeingültigkeit konkrete Vorgaben für die Implementierung und Überwachung der vorgestellten Maßnahmen und Aktivitäten.

Ein weiterer kritischer Punkt ist die mangelnde Flexibilität der Frameworks, da von den Autoren oft nicht klar adressiert wird, welche die relevantesten Maßnahmen und Aktivitäten sind. KMUs besitzen andere Voraussetzungen und Ressourcen und können nicht alle Vorgaben und Prozesse gleichermaßen umsetzen. Insbesondere die begrenzten technischen und finanziellen Mittel sowie die kleinen personellen Kapazitäten machen es für KMUs schwierig, die umfangreichen Frameworks vollständig zu implementieren.

Das CoreGovernance Framework wurde daraufhin so entwickelt, dass es sowohl die wichtigsten Aspekte von Data Governance beinhaltet, aber auch die spezifischen Bedürfnisse und Anforderungen von KMUs berücksichtigt. Der Fokus liegt dabei auf den wesentlichen Bereichen von Data Governance, um diese dafür detaillierter und praxisnaher beschreiben zu können. Das Framework ist dabei flexibel gestaltet, sodass es an die individuellen Gegebenheiten kleinerer Unternehmen angepasst werden kann. Zum Beispiel können Data Owner in kleineren Organisationen ebenfalls andere, täglich anfallende Aufgaben übernehmen. Ebenso kann die Anzahl der Data Stewards schrittweise erhöht werden, und das gesamte System kann bei steigenden Mitarbeiterzahlen vom zentralisierten hin zu einem föderativen Modell transformiert werden.

Darüber hinaus legt das Framework einen starken Fokus auf die technische Machbarkeit im täglichen Betrieb. Mit dem operativen Modell und die technischen Enablern werden konkrete und adaptive Vorschläge für die Umsetzung der Maßnahmen gemacht. Dies ermöglicht es den Unternehmen, die Aktivitäten an ihre Bedürfnisse anzupassen und in ihre Strukturen zu integrieren. Damit bietet das Framework nicht nur theoretische Konzepte, sondern unterstützt Organisationen und KMUs durch Vorgaben und Anleitungen bei der Umsetzung der Konzepte in die Praxis.

Insgesamt bietet das CoreGovernance Framework eine praxisorientierte und adaptive Lösung, die die spezifischen Bedürfnisse von KMUs berücksichtigt und es ihnen ermöglicht, eine effektive Data Governance zu integrieren, ohne durch unnötige Komplexität oder starre Strukturen behindert zu werden.

## 6.2 Quantitative Untersuchung

Die quantitative Analyse fokussiert sich auf die Untersuchung der Wirksamkeit des CoreGovernance Frameworks unter realen Bedingungen. Dabei wird betrachtet, inwieweit die Strukturen und Vorgaben in der Lage sind, vorher auftretende Unsicherheiten in der

Entscheidungsfindung zu beseitigen und Antworten auf offene Fragen in den Evaluierungsszenarien zu geben.

### 6.2.1 Bewertung der Entscheidungsfindung

Vor der Einführung des CoreGovernance Frameworks gab es in den täglichen Abläufen und Entscheidungsfindungen eine Reihe von Unsicherheiten, die vor allem auf unklare Verantwortlichkeiten und fehlende Prozessvorgaben zurückzuführen waren. Diese Unsicherheiten führten häufig zu ineffizienten Prozessen, Verzögerungen bei Entscheidungen und einer Erhöhung der Fehleranfälligkeit. Oft war nicht klar, wer für bestimmte Aufgaben zuständig ist, was zu doppelter Arbeit oder zur Vernachlässigung wesentlicher Aufgaben führen kann. Ohne standardisierte Prozesse sind Entscheidungsfindungen nicht einheitlich und häufig nicht nachvollziehbar.

Das CoreGovernance Framework bietet Lösungen für diese Probleme, indem es klare Zuweisungen von Verantwortlichkeiten für Datensätze, Prozesse und deren Überwachung vornimmt. Abbildung 6.2 zeigt den Einfluss des CoreGovernance Frameworks auf die Einflussfaktoren in der AquaFlow GmbH. Unsicherheiten bei grün hinterlegten Faktoren werden von dem Framework direkt adressiert und beseitigt. Gelb markierte Bereiche werden von dem Framework unterstützt, können aber nicht vollständig beeinflusst werden. Graue Faktoren liegen außerhalb des Einflussbereichs des Frameworks.

Zu den gelb markierten Bereichen gehören die Regularien und technische Limitierungen. Das CoreGovernance Framework kann beim Umgang mit den Regularien lediglich Hilfestellung bei deren Umsetzung und Überwachung geben. Einfluss auf die geltenden Vorschriften hat es nicht. Ebenso unterstützt das Framework das Unternehmen, die bestehende Infrastruktur und verfügbare Ressourcen effektiv zu nutzen. Es bietet jedoch keine direkten Verbesserungen oder Erweiterungen der vorhandenen technischen Infrastruktur.

### 6.2.2 Evaluierungsszenarien

Um zu prüfen, wie das Framework mit realen Herausforderungen und Problemen umgeht, wurden in Unterabschnitt 3.3.2 Evaluierungsszenarien entworfen, die praxisnahe Situationen abbilden. Durch die Beantwortung dieser zeigt sich, wie das Framework in der Praxis angewendet wird.

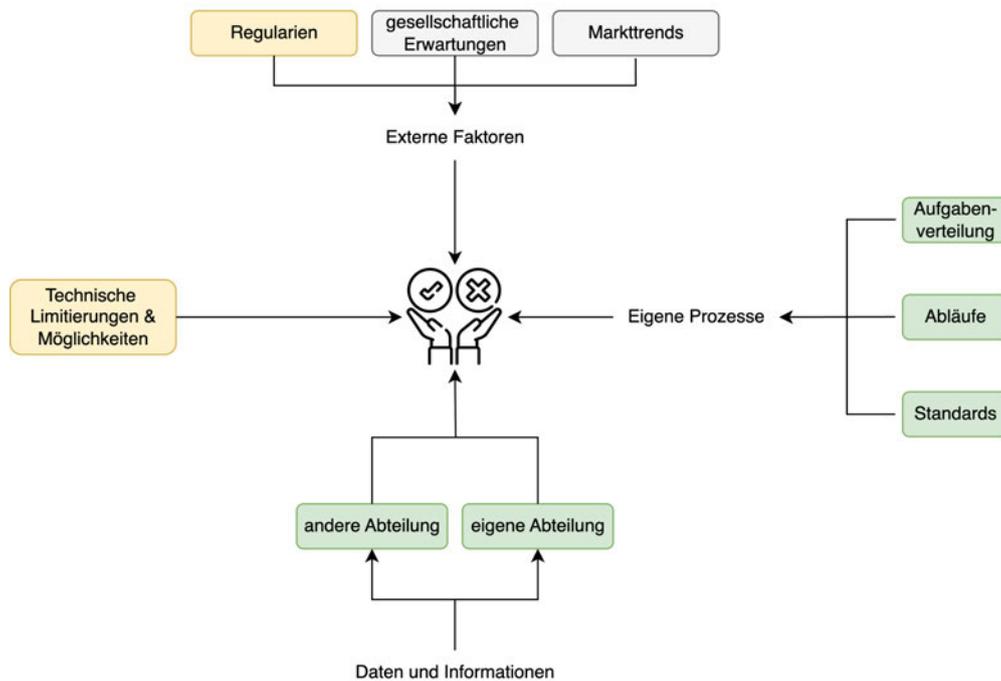


Abbildung 6.2: Einflüsse auf die Entscheidungsfindung mit Data Governance

## Datenqualität

Das Framework adressiert die auftretenden Fragen zur Datenqualität durch klare Zuständigkeiten und standardisierte Prozesse. Im Folgenden werden die zuvor beschriebenen Fragen und Herausforderungen beantwortet.

- **Offene Fragen zu Soll-Zuständen und Anforderungen:** Wenn Unklarheiten über die Soll-Zustände oder Anforderungen an die Daten bestehen, können die Antworten direkt beim Data Owner oder den Data Stewards eingeholt werden. Diese Verantwortlichen sowie ihre Kontaktdaten sind im Datenkatalog leicht auffindbar.
- **Unterschiedliche Darstellungen und Formate:** Unterschiedliche Darstellungen und Formate für identische Informationen weisen auf ein Qualitätsproblem hin. Die korrekte Version kann im Glossar nachgeschlagen werden. Wenn ein Begriff im Glossar fehlt, sollte der Data Owner oder ein Data Steward informiert werden. Gegebenenfalls kann das DGO zur Klärung der Situation eingeschaltet werden.
- **Überwachung der Datenqualität:** Die Entwicklung und Überwachung der Datenqualität kann ebenfalls bei den Data Ownern angefordert werden. Zusätzlich

können sich Anwender für die Visualisierung der Datenqualitätsergebnisse freischalten lassen, um kontinuierlich die Ergebnisse überwachen zu können.

- **Vorgehen bei neuen Datensätzen:** Die korrekte Vorgehensweise bei der Einführung neuer Datensätze, einschließlich der Bestimmung der Datenqualität, ist in den Prozessen des operativen Modells beschrieben. Dieses Modell gibt klare Anweisungen, wie die Qualität neuer Daten anfänglich zu bewerten ist und auch in Zukunft sichergestellt werden kann.
- **Doppelte Datensätze:** Wenn zwei identische Datensätze entdeckt werden, sollte der Data Owner oder das DGO informiert werden. Diese Instanzen sind dafür verantwortlich, die offene Frage zu untersuchen und zu beantworten, um die Konsistenz der Daten zu gewährleisten.

### Fragen zu Datenzugriffen

Das Framework stellt sicher, dass alle Aspekte des Datenzugriffs klar geregelt und dokumentiert sind. Im Folgenden werden die zentralen Herausforderungen und ihre Lösungen beschrieben.

- **Herkunft der Daten:** Im Datenkatalog sollten die Informationen zur Herkunft der Daten detailliert beschrieben sein. Falls diese Informationen fehlen, muss sich der Nutzer an den im Datenkatalog gelisteten Data Owner wenden, um die erforderlichen Auskünfte zu erhalten. Dieser muss Rechenschaft über die Vollständigkeit beim CDO ablegen.
- **Fehlende Metadaten:** Sollten Informationen oder Metadaten zu vorhandenen Datenbeständen fehlen, ist der CDO zu kontaktieren. Als verantwortliche Person für den Datenkatalog muss der CDO sicherstellen, dass alle aufgeführten Datensätze die erforderlichen Informationen bereitstellen.
- **Zugriffsprotokolle:** Fehlende Zugriffsprotokolle dürfen nicht existieren, da das DGO speziell Systeme entwickelt hat, um diese zu gewährleisten. Fehlen solche Kontrollen, muss umgehend der Data Owner oder das DGO informiert werden.
- **Zugriffsanfragen:** Der Zugriff auf Daten kann über den Data Owner beantragt werden. Die Kontaktdaten des Data Owners sind im Datenkatalog aufgeführt. Eine

Freischaltung erfolgt bei Vorliegen einer passenden Begründung zur Nutzung der Daten.

- **Überwachung der Zugriffsrechte:** Der Data Owner hat die Möglichkeit, durch die Nutzung von Rollen und Gruppen einzusehen, wer auf die Daten zugreift. Diese Rollen- und Gruppenstrukturen ermöglichen eine transparente und kontrollierte Vergabe von Zugriffsrechten.
- **Rollenwechsel:** Bei einem internen Rollenwechsel eines Mitarbeiters werden ihm die Zugriffsrechte der alten Rolle entzogen und die neuen Rechte entsprechend zugewiesen. Dies gewährleistet, dass die Zugriffe stets aktuell und rollenbasiert geregelt sind.

### **Einhaltung neuer gesetzlicher Vorschriften**

Das Framework unterstützt das Unternehmen bei der Implementierung und Überwachung gesetzlicher Anforderungen. Im Folgenden sind die Antworten des Frameworks auf gängige Fragen und Herausforderungen aufgeführt.

- **Zentralisierte Vorgaben:** Alle für das Unternehmen geltenden Anforderungen und Vorschriften sind zentral im Glossar des Datenkatalogs einsehbar. Ebenso sind dort weiterführende Informationen sowie Ansprechpartner aufgelistet, die bei Rückfragen zur Verfügung stehen.
- **Geregelte Verantwortlichkeiten:** Die Verantwortlichkeiten für die Einhaltung und Überwachung der Vorschriften sind über die Rollenverteilung geregelt und ebenfalls im Datenkatalog abrufbar. Bei Rückfragen steht der CDO und das DGO als Ansprechpartner zur Verfügung.
- **Dokumentation und Nachvollziehbarkeit:** Durch den Enabler Dokumentation und die im operativen Modell festgelegten Prozesse müssen alle Entscheidungen dokumentiert und nachvollziehbar sein. Sollten Unklarheiten über vergangene Prozesse und Entscheidungen auftreten, so können deren Dokumentationen zur Nachvollziehbarkeit herangezogen werden.
- **Schulungen und Sensibilisierungen:** Das operative Modell integriert Schulungen und Sensibilisierungsmaßnahmen fest in den Arbeitsalltag, um sicherzustellen,

dass alle Mitarbeiter über die aktuellen gesetzlichen Vorschriften informiert sind und diese einhalten.

### 6.2.3 Fortschritt im Reifegradmodell

Durch die Implementierung der Maßnahmen haben sich, wie in Unterabschnitt 6.2.1 und Unterabschnitt 6.2.2 gezeigt, bereits einige wesentliche Aspekte in der AquaFlow GmbH verbessert. In den ersten Schritten wurden grundlegende Strukturen und Zuständigkeiten definiert sowie erste Standards für Prozesse und Daten festgelegt. Mit der Verbesserung der Prozesse und der technischen Einführung von Werkzeugen wie dem Datenkatalog, Automatisierungswerkzeugen und Technologien zur Bestimmung der Datenqualität werden datenbasierte Entscheidungen kontinuierlich fundierter und nachvollziehbarer. Dies spiegelt sich auch im aktualisierten Reifegradmodell wider, das in Abbildung 6.3 dargestellt ist.

Im Vergleich zur letzten Evaluierung in Abbildung 5.2 wurden alle offenen Bereiche der Grundlagenstufe abgeschlossen. Die Datenprodukte wie der Datenkatalog, Zugriffsmodelle und Automatisierungswerkzeuge sind in allen Bereichen im Einsatz. Nach der ersten Iteration wurden aufgrund von Rückmeldungen bereits Anpassungen an diesen vorgenommen. Für das Erreichen der nächsten Stufe fehlt das Abschließen der flächendeckenden Visualisierung der Metriken sowie die Einführung und Umsetzung von Datenqualitätsstufen. Diese Stufen ermöglichen es, Datensätzen und Datenquellen Zertifikate zu vergeben, um deren Qualität und Integrität zu dokumentieren. Dadurch wird den Anwendern in Zukunft ein verlässliches System zur Erkennung von qualitativ hochwertigen Daten zur Verfügung gestellt.

## 6.3 Diskussion der Ergebnisse

In dieser Arbeit wurde das auf KMUs ausgerichtete CoreGovernance Framework entwickelt und in der Testumgebung der AquaFlow GmbH getestet. Untersuchungen ergaben, dass existierende Frameworks aus der Literatur und der Praxis so konzipiert sind, dass sie allgemeingültig sind und von einer breiten Audienz verwendet werden können. Diese Frameworks basieren ebenfalls häufig auf der Annahme, dass Data Governance hauptsächlich für größere Unternehmen relevant ist. Dadurch setzen die Autoren gewisse Organisationsstrukturen und umfangreiche Kapazitäten in der Infrastruktur voraus, was zu



angewendet. So wurden Rollen und Verantwortlichkeiten basierend auf der bestehenden Organisationsstruktur zugewiesen, und das operative Modell durch passende Prozessvorgaben erweitert. Ebenso wurden anhand der definierten Anforderungen die zentralen Datenprodukte wie der Datenkatalog, Automatisierungswerkzeuge und Datenqualitätsprüfungen implementiert. Dies veranschaulicht, wie die Anwendung der Konzepte in der Praxis erfolgreich umgesetzt werden kann.

Um die tatsächliche Wirksamkeit von Frameworks in der Praxis zu überprüfen, wurden spezifische Evaluierungsszenarien entwickelt. Diese Szenarien umfassen verschiedene alltägliche Situationen, Probleme und Herausforderungen, die im Unternehmen auftreten können. Die Lösung dieser realistischen Herausforderungen durch ein Framework bietet wertvolle Einblicke in die praktische Anwendbarkeit und zeigt die Verbesserungen, die das Framework mit sich bringt. Das CoreGovernance Framework hat dabei für alle Situationen Lösungen aufgezeigt und bewiesen, dass es auch in der Praxis die relevanten Aspekte abdeckt. Ebenso wurden die zuvor identifizierten Unsicherheiten bei der Entscheidungsfindung der einzelnen Abteilungen, die größtenteils aus unklaren Verantwortlichkeiten, fehlenden Prozessvorgaben und unzureichender Kommunikation entstanden, durch das Framework erheblich reduziert.

Zusammenfassend zeigt die Anwendung des CoreGovernance Frameworks auf die AquaFlow GmbH, dass das Framework nicht nur den spezifischen Anforderungen von KMUs gerecht wird, sondern auch die Lücke zwischen theoretischen Konzepten in der Literatur und dem Bedarf nach praktischen Umsetzungen schließt. Die Ergebnisse aus den qualitativen und der quantitativen Untersuchungen zeigen, dass der Fokus auf die wesentlichen Aspekte von Data Governance, gepaart mit Vorgaben für die praktische Umsetzung, die Effizienz und Sicherheit der Datenverwaltung von Unternehmen erhöhen kann. Dies führt zu optimierten und standardisierten Prozessen, einer besseren Nachvollziehbarkeit von Entscheidungen und einer höheren Datenqualität. Diese Punkte unterstützen das Unternehmen dabei, seine langfristigen Ziele zu erreichen.

# 7 Fazit

Die vorliegende Arbeit ging der Frage nach, wie ein praxisorientiertes Data Governance Framework speziell für die Bedürfnisse von KMUs entwickelt und implementiert werden kann. Dafür wurden bestehende Frameworks analysiert und deren Stärken identifiziert. Außerdem wurden die Voraussetzungen und Herausforderungen der KMUs untersucht, um anhand dessen ein adaptives Framework zu erstellen. Um die praktische Anwendbarkeit und Effektivität bei realistischen Bedingungen zu zeigen, wurde das CoreGovernance Framework in einer entwickelten Testumgebung eingesetzt.

## 7.1 Zusammenfassung

Für das Verständnis der Thematik wurden in Kapitel 2 die notwendigen theoretischen Grundlagen geschaffen. Zunächst wurde der Begriff Data Governance anhand verschiedener Quellen definiert und Einblicke in die wichtigsten Komponenten gegeben. Anschließend wurden die bestehenden Frameworks von DAMA, DGI und PwC vorgestellt, um die unterschiedlichen Ansätze aus verschiedenen Fachrichtungen zu untersuchen. Im Anschluss daran wurde die spezifische Rolle von Data Governance für KMUs untersucht, was zur Formulierung der Forschungsfragen führte.

Um das entwickelte CoreGovernance Framework und künftige, auf KMUs fokussierte Frameworks testen und validieren zu können, wurde in Kapitel 3 eine Testumgebung konzipiert. Dazu wird das fiktive KMU AquaFlow GmbH erstellt, und die auftretenden Probleme und Herausforderungen hervorgehoben. Die Beantwortung der Forschungsfrage 1 erfolgt durch die Analyse von Unsicherheiten bei der Entscheidungsfindung, die Entwicklung von Evaluierungsszenarien und die Erstellung einer Domänenmatrix. Gegenstand der Forschungsfrage ist die Gestaltung einer geeigneten Testumgebung, um Frameworks für KMUs zu untersuchen und zu validieren.

Kapitel 4 befasst sich mit der Entwicklung des CoreGovernance Frameworks, das speziell auf die Bedürfnisse von KMUs ausgelegt ist. Dafür wurden die drei in Kapitel 2 vorgestellten Frameworks auf ihre Stärken und Aspekte, die sie ungeeignet für KMUs machen, untersucht. Ebenso wurden im Rahmen der Analyse Elemente identifiziert, die in das neu erarbeitete Framework einfließen sollen. Damit wurde zugleich die Forschungsfrage 2 beantwortet. Anhand der Erkenntnisse aus der Analyse und den zuvor identifizierten Anforderungen von KMUs wurde das Data Governance Framework CoreGovernance entworfen, wodurch die Forschungsfrage 3 beantwortet wurde. Dabei wurden die Kernkomponenten näher vorgestellt und ihre besondere Bedeutung für KMUs hervorgehoben. Um die Fortschritte evaluieren und bei Bedarf Anpassungen vornehmen zu können, wurde ein geeignetes Reifegradmodell erstellt.

In Kapitel 5 wurde das CoreGovernance Framework in dem KMU AquaFlow GmbH exemplarisch eingeführt, um die praktische Machbarkeit zu demonstrieren. Dabei wurde die zu Beginn festgelegte Vorgehensweise nach und nach durchgeführt. Dies beinhaltet die Konkretisierung der Ziele und der Strategie, die Zuweisung der Verantwortlichkeiten an reale Personen, die Ausarbeitung und Einführung des operativen Modells sowie die technische Umsetzung der Datenprodukte. Diese umfassen einen Datenkatalog, ein Business Glossar, Technologien zur Datenqualitätsprüfung sowie Automatisierungswerkzeuge. Dabei wurden zunächst Anforderungen an die Technologien formuliert, und diese im Anschluss in die Praxis umgesetzt. Durch die unternehmensspezifische Umsetzung der Konzepte auf die Testumgebung wurde gleichzeitig die Forschungsfrage 4 beantwortet, die den Bedarf nach einer praxisgerechten Umsetzung eines Datenmanagementprogramms aufwirft, um die Konzepte des Data Governance Frameworks in einem Unternehmensalltag integrieren zu können.

In Kapitel 6 wurde das CoreGovernance Framework sowohl qualitativ als auch quantitativ untersucht. Dabei besteht neben dem Fokus auf die Bedürfnisse von KMUs auch der Anspruch, ein vollwertiges Data Governance Framework zu entwickeln. Die Analyse mithilfe der modifizierten Domänenmatrix ergab, dass das Framework alle relevanten Aspekte von Data Governance abdeckt. Außerdem zeigte die Untersuchung, dass die Aspekte und Elemente, die bei den existierenden Frameworks für KMUs problematisch waren, im CoreGovernance Framework erfolgreich adressiert wurden. Darüber hinaus differenziert sich die Arbeit von der bisherigen Literatur, indem sie die Anwendbarkeit des Frameworks unter realen Bedingungen untersucht. Diese quantitative Analyse ergab, dass das Framework Unternehmen nicht nur theoretisch unterstützen kann, sondern auch bei praktischen Herausforderungen und Problem wirksam hilft. So werden beispielswei-

se Unsicherheiten bei der Entscheidungsfindung beseitigt, offene Fragen bei alltäglichen, datenbasierten Fragen beantwortet und Probleme in den Evaluierungsszenarien gelöst.

## 7.2 Ausblick

Diese Arbeit hat gezeigt, wie die Einführung des CoreGovernance Frameworks bereits starke positive Effekte für ein Unternehmen haben kann. Eine Weiterführung der Untersuchung würde wertvolle Erkenntnisse darüber liefern, wie das Framework vollständig integriert werden kann und ob sich die positiven Effekte verstärken. Ebenso könnten die Rückschlüsse aus der Implementierung im Unternehmen dazu verwendet werden, Anpassungen am CoreGovernance Framework vorzunehmen. Dadurch könnte das Framework in verschiedenen Iterationen optimiert werden.

In weiterführenden Arbeiten sollte das CoreGovernance Framework in einem realen KMU erprobt werden. Eine dort durchgeführte Langzeitstudie würde wertvolle Erkenntnisse über die Praxisrelevanz und Wirkungsweise des Frameworks unter realen Bedingungen liefern. Die Umsetzung der Rollenverteilung in einem echten Unternehmenskontext sowie die Implementierung des operativen Modells könnten aufzeigen, ob die entwickelten Strukturen in der Praxis realistisch sind und inwieweit sie die antizipierten Unsicherheiten beseitigen können. Weitere Erkenntnisse werden sich in dem flächendeckenden Einsatz der vorgeschlagenen Technologien erhofft, ob diese die erwarteten Effekte bezüglich Datenqualität, Effizienz und Sicherheit liefern. Ebenso interessant ist es, wie die Akzeptanz der Mitarbeiter gegenüber den neuen Prozessen und Technologien ausfällt und wie das Change-Management und Schulungen gestaltet werden müssen, um die reibungslose Integration des CoreGovernance Frameworks sicherzustellen. Die gewonnenen Erkenntnisse könnten ebenfalls dazu beitragen, das Framework weiter zu optimieren und den spezifischen Anforderungen und Voraussetzungen von KMUs anzupassen.

# Literaturverzeichnis

- [1] ABRAHAM, Rene ; SCHNEIDER, Johannes ; VOM BROCKE, Jan: Data governance: A conceptual framework, structured review, and research agenda. In: *International Journal of Information Management* 49 (2019), S. 424–438. – URL <https://www.sciencedirect.com/science/article/pii/S0268401219300787>. – ISSN 0268-4012
- [2] AL-RUITHE, Majid ; BENKHELIFA, Elhadj ; HAMEED, Khawar: A systematic literature review of data governance and cloud data governance. In: *Personal Ubiquitous Comput.* 23 (2018), jan, Nr. 5–6, S. 839–859. – URL <https://doi.org/10.1007/s00779-017-1104-3>. – ISSN 1617-4909
- [3] ALDOSERI, Abdulaziz ; AL-KHALIFA, Khalifa N. ; HAMOUDA, Abdel M.: Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. In: *Applied Sciences* 13 (2023), Nr. 12. – URL <https://www.mdpi.com/2076-3417/13/12/7082>. – ISSN 2076-3417
- [4] ALHASSAN, Ibrahim ; SAMMON, David ; DALY, Mary: Data governance activities: a comparison between scientific and practice-oriented literature. In: *Journal of Enterprise Information Management* 31 (2018), 01, S. 00–00
- [5] AMAZON WEB SERVICES, INC.: *Was ist Containerisierung?* 2023. – URL <https://aws.amazon.com/de/what-is/containerization/>. – Stand: 28.05.2024
- [6] APACHE SPARK: *Apache Spark Wins Daytona GraySort 100TB Benchmark.* 2014. – URL <https://spark.apache.org/news/spark-wins-daytona-gray-sort-100tb-benchmark.html>. – Stand: 29.05.2024
- [7] APACHE SPARK: *Cluster Mode Overview - Spark 3.1.1 Documentation.* 2024. – URL <https://spark.apache.org/docs/latest/cluster-overview.html>. – Stand: 30.05.2024

- [8] B, Stein ; A, Morrison: The enterprise data lake: Better integration and deeper analytics. In: *PwC Technology Forecast: Rethinking integration* (2014)
- [9] BALTASSIS, Elias ; GOUREVITCH, Antoine ; QUARTA, Lucas: Good data starts with great governance / Boston Consulting Group. 2019. – Forschungsbericht
- [10] BECKER, Jörg ; KNACKSTEDT, Ralf ; PÖPPELBUSS, Jens: Developing Maturity Models for IT Management. In: *Business & Information Systems Engineering: The International Journal of WIRTSCHAFTSINFORMATIK* 1 (2009), Nr. 3, S. 213–222. – URL <https://EconPapers.repec.org/RePEc:spr:binfse:v:1:y:2009:i:3:p:213-222>
- [11] BEGG, Carolyn ; CAIRA, Thomas: Exploring the SME quandary: data governance in practise in the small to medium-sized enterprise sector. In: *The Electronic Journal of Information Systems Evaluation* 15 (2012), Nr. 1, S. 3–13. – ISSN 1566-6379
- [12] BROUS, Paul ; JANSSEN, Marijn ; VILMINKO-HEIKKINEN, Riikka: Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles, 09 2016, S. 115–125. – ISBN 978-3-319-44420-8
- [13] BUNDESAMT FÜR SICHERHEIT IN DER INFORMATIONSTECHNIK (BSI): *Informationen und Hilfestellungen für KMU*. 2024. – URL [https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/KMU/KMU\\_node.html](https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/KMU/KMU_node.html). – Stand: 12.07.2024
- [14] BUZZELLI, B.: *Data Quality Engineering in Financial Services*. O'Reilly Media, 2022. – URL [https://books.google.de/books?id=k\\_-VEAAAQBAJ](https://books.google.de/books?id=k_-VEAAAQBAJ). – ISBN 9781098136901
- [15] CHENG, Guangming ; LI, Yao ; GAO, Zhiwei ; LIU, Xiaoyin: Cloud data governance maturity model. In: *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2017, S. 517–520
- [16] CHEONG, Lai ; CHANG, Vanessa: The need for data governance: A case study. In: *ACIS 2007 Proceedings - 18th Australasian Conference on Information Systems* (2007), 01
- [17] CMMI INSTITUTE: *CMMI Appraisal Levels*. 2024. – URL <https://cmmi.institute.com/learning/appraisals/levels>. – Accessed: 2024-08-28

- [18] CNBC: *Fines for breaches of EU privacy law spike sevenfold to \$1.2 billion, as Big Tech bears the brunt*. Januar 2022. – URL <https://www.cnn.com/2022/01/18/fines-for-breaches-of-eu-gdpr-privacy-law-spike-sevenfold.html>
- [19] COMMISSION, European ; INTERNAL MARKET, Entrepreneurship Directorate-General for ; SMES: *User guide to the SME definition*. Publications Office, 2020
- [20] COMMUNICATION, European Commission. Directorate-General: *Unleashing the Full Potential of European SMEs*. Publications Office of the European Union, 2020. – URL <https://books.google.de/books?id=F159zQEACAAJ>. – ISBN 9789276169123
- [21] CROSBY, P.B.: *Quality is Free: The Art of Making Quality Certain*. McGraw-Hill, 1979 (Mentor book). – URL <https://books.google.de/books?id=n4IubCcpm0EC>. – ISBN 9780070145122
- [22] DATA GOVERNANCE INSTITUTE: *DGI Data Governance Framework*. – URL <https://datagovernance.com/the-dgi-data-governance-framework/>. – Stand 05.02.2024
- [23] DE HERT, P.J.A. ; PAPAKONSTANTINOY, V.: Three scenarios for international governance of data privacy: Towards an international data privacy organization, preferably a UN agency. In: *Journal of Law and Policy* 9 (2013), Nr. 2, S. 271–324. – ISSN 1467-9930
- [24] DIBOWSKI, Henrik ; SCHMID, Stefan: Using Knowledge Graphs to Manage a Data Lake. In: REUSSNER, Ralf H. (Hrsg.) ; KOZIOLEK, Anne (Hrsg.) ; HEINRICH, Robert (Hrsg.): *INFORMATIK 2020*, Gesellschaft für Informatik, Bonn, 2021, S. 41–50
- [25] DOCKER INC.: *Docker Overview*. <https://docs.docker.com/get-started/overview/>. – Stand: 28.05.2024
- [26] DOCKER INC.: *Overview of Docker Compose*. 2024. – URL <https://docs.docker.com/compose/>. – Stand: 29.05.2024
- [27] ENGELS, Barbara: Data Governance as the Enabler of the Data Economy. In: *Intereconomics* 54 (2019), Jul, Nr. 4, S. 216–222. – URL <https://doi.org/10.1007/s10272-019-0827-y>. – ISSN 1613-964X

- [28] EUROPEAN PARLIAMENT: *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)*. May 2016
- [29] FIRICAN, George: *Data Governance Maturity Models (Gartner)*. 2020. – URL <https://www.lightsondata.com/data-governance-maturity-models-gartner/>. – Stand: 17.07.2024
- [30] GORELIK, A.: *The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science*. O'Reilly Media, 2019. – URL <https://books.google.de/books?id=6SuGjgEACAAJ>. – ISBN 9781491931554
- [31] HAI, Rihan ; QUIX, Christoph ; JARKE, Matthias: Data lake concept and systems: a survey. In: *ArXiv abs/2106.09592* (2021)
- [32] HAMMA, Simon: *Evaluation of Tools for Data Governance Solutions*. 2024
- [33] IBRAHIM ALHASSAN, David S. ; DALY, Mary: Data governance activities: an analysis of the literature. In: *Journal of Decision Systems* 25 (2016), Nr. sup1, S. 64–75. – URL <https://doi.org/10.1080/12460125.2016.1187397>
- [34] INTERNATIONAL, Dama: *DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)*. Denville, NJ, USA : Technics Publications, LLC, 2017. – ISBN 1634622340
- [35] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO): *ISO 9000 Introduction and Support Package: Guidance on the Concept and Use of the Process Approach for management systems*. ISO/TC 176/SC 2/N 544R3. 2008
- [36] IQBAL, Muhammad ; KAZMI, Syed Hasnain A. ; MANZOOR, Amir ; SOOMRANI, Abdul R. ; BUTT, Shujaat H. ; SHAIKH, Khurram A.: A study of big data for business growth in SMEs: Opportunities & challenges. In: *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2018, S. 1–7
- [37] JANSSEN, Marijn ; BROUS, Paul ; ESTEVEZ, Elsa ; BARBOSA, Luis S. ; JANOWSKI, Tomasz: Data governance: Organizing data for trustworthy Artificial Intelligence. In: *Government Information Quarterly* 37 (2020), Nr. 3, S. 101493. – URL <https://www.sciencedirect.com/science/article/pii/S0740624X20302719>. – ISSN 0740-624X

- [38] JARMUL, K.: *Practical Data Privacy: Enhancing Privacy and Security in Data*. O'Reilly Media, 2023. – URL <https://books.google.de/books?id=HnSSzwEACAAJ>. – ISBN 9781098129460
- [39] JOHANNES SCHNEIDER, Christian M. ; BROCKE, Jan V.: Artificial Intelligence Governance For Businesses. In: *Information Systems Management* 40 (2023), Nr. 3, S. 229–249. – URL <https://doi.org/10.1080/10580530.2022.2085825>
- [40] KARKOŠKOVÁ, Soňa: Data Governance Model To Enhance Data Quality In Financial Institutions. In: *Information Systems Management* 40 (2022), 03, S. 1–21
- [41] KHATRI, Vijay ; BROWN, Carol: Designing data governance. In: *Commun. ACM* 53 (2010), 01, S. 148–152
- [42] KHINE, PWINT PHYU ; WANG, ZHAO SHUN: Data lake: a new ideology in big data era. In: *ITM Web Conf.* 17 (2018), S. 03025. – URL <https://doi.org/10.1051/itmconf/20181703025>
- [43] KILKENNY, Monique F. ; ROBINSON, Kerin M.: Data quality: “Garbage in – garbage out”. In: *Health Information Management Journal* 47 (2018), Nr. 3, S. 103–105. – URL <https://doi.org/10.1177/1833358318774357>. – PMID: 29719995
- [44] KPMG/IMPACT: Information as an Asset: the Board Agenda. 1994. – Forschungsbericht
- [45] LEBOW COLLEGE OF BUSINESS’ CENTER FOR BUSINESS ANALYTICS REPORT: Trends in Data Governance and Data Quality. (2021). – URL <https://www.lebow.drexel.edu/sites/default/files/legacy/1639152057-lebow-precisely-report.pdf>
- [46] MARCHILDON, Philippe ; BOURDEAU, Simon ; HADAYA, Pierre ; LABISSIÈRE, Aldrin: Data governance maturity assessment tool: A design science approach. In: *Projectics / Proyéctica / Projectique* 20 (2018), Nr. 2, S. 155–193. – URL [https://www.cairn.info/load\\_pdf.php?ID\\_ARTICLE=PROJ\\_020\\_0155](https://www.cairn.info/load_pdf.php?ID_ARTICLE=PROJ_020_0155). ISBN 9782807392373
- [47] MCAFEE, Andrew ; BRYNJOLFSSON, Erik: Big data: the management revolution. In: *Harvard business review* 90 (2012), Nr. 10, S. 60–68
- [48] MCGILVRAY, D.: *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM)*. Elsevier Science, 2021. – URL <https://books.google.de/books?id=peT7DwAAQBAJ>. – ISBN 9780128180167

- [49] MEDYCKYJ-SCOTT, David ; STOCK, Kristin ; GIBB, Robert ; GAHEGAN, Mark ; DZIERZON, Helge ; SCHMIDT, Jochen ; COLLINS, Alison: *Our Land and Water National Science Challenge - A Data Ecosystem for Land and Water Data to Achieve the Challenge Mission*. 09 2016
- [50] MICROSOFT CORPORATE: *A Guide to Data Governance*. 1 Microsoft Way, Redmond, WA: , 2020. – URL <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4GETu>
- [51] MINDSQUARE: *Was ist ein Data Lake?* <https://mindsquare.de/knowhow/data-lake/>. – Stand: 27.05.2024
- [52] MINIO INC.: *MinIO: Overview*. 2024. – URL <https://min.io/product/overview>. – Stand: 29.05.2024
- [53] MÜLLER-STEWENS, Günter ; GILLENKIRCH, Robert: *Strategie*. 2024. – URL <https://wirtschaftslexikon.gabler.de/definition/strategie-43591>. – Gabler Wirtschaftslexikon, Stand: 09.07.2024
- [54] NAMBIAR, Athira ; MUNDRA, Divyansh: An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. In: *Big Data and Cognitive Computing* 6 (2022), 11, S. 132
- [55] NIELSEN, Olivia B. ; PERSSON, John S. ; MADSEN, Sabine: Why Governing Data Is Difficult: Findings from Danish Local Government. In: ELBANNA, Amany (Hrsg.) ; DWIVEDI, Yogesh K. (Hrsg.) ; BUNKER, Deborah (Hrsg.) ; WASTELL, David (Hrsg.): *Smart Working, Living and Organising*. Cham : Springer International Publishing, 2019, S. 15–29. – ISBN 978-3-030-04315-5
- [56] OPPENHEIM, Charles ; STENSON, Joan ; WILSON, Richard M. S.: Studies on Information as an Asset I: Definitions. In: *Journal of Information Science* 29 (2003), Nr. 3, S. 159–166. – URL <https://doi.org/10.1177/01655515030293003>
- [57] ORACLE CORPORATION: *Enterprise Information Management: Best Practices in Data Governance*. 2011. – Forschungsbericht
- [58] ORACLE CORPORATION: *What is MySQL?* 2024. – URL <https://www.oracle.com/mysql/what-is-mysql/>. – Stand: 29.05.2024
- [59] OTTO, Boris: Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. In: *Communications of the Association for Information Systems* 29 (2011), 08, S. Article 3

- [60] PANIAN, Z.: Some practical experiences in data governance. 38 (2010), 02, S. 150–157
- [61] PANSARA, Ronak: Unraveling the Complexities of Data Governance with Strategies, Challenges, and Future Directions. In: *Transactions on Latest Trends in IoT* 6 (2023), Nr. 6, S. 46–56. – URL <https://ijsdcs.com/index.php/TLIoT/article/view/345>. – ISSN 3246-547X
- [62] PAUL P. TALLON, Ronald V. R. ; SHORT, James E.: The Information Artifact in IT Governance: Toward a Theory of Information Governance. In: *Journal of Management Information Systems* 30 (2013), Nr. 3, S. 141–178. – URL <https://doi.org/10.2753/MIS0742-1222300306>
- [63] PETZOLD, Bryan ; ROGGENDORF, Matthias ; ROWSHANKISH, Kayvaun ; SPORLEDER, Christoph: Designing data governance that delivers value / McKinsey Technology. 2020. – Forschungsbericht
- [64] POSTGRESQL GLOBAL DEVELOPMENT GROUP: *About PostgreSQL*. 2024. – URL <https://www.postgresql.org/about/>. – Stand: 29.05.2024
- [65] PWC: Global and industry frameworks for data governance. 2019. – Forschungsbericht
- [66] RILEY, Jenn: Understanding metadata. In: *Washington DC, United States: National Information Standards Organization (https://www.niso.org/publications/understanding-metadata-2017)* 23 (2017), S. 7–10
- [67] SALLOUM, Salman ; DAUTOV, Ruslan ; CHEN, Xiaojun ; PENG, Patrick X. ; HUANG, Joshua Z.: Big data analytics on Apache Spark. In: *International Journal of Data Science and Analytics* 1 (2016), Nov, Nr. 3, S. 145–164. – URL <https://doi.org/10.1007/s41060-016-0027-9>. – ISSN 2364-4168
- [68] SAMUNDISWARY, S. ; DONGRE, Nilma M.: Object storage architecture in cloud for unstructured data. In: *2017 International Conference on Inventive Systems and Control (ICISC)*, 2017, S. 1–6
- [69] SANTOSO, Leo W. ; YULIA: Data Warehouse with Big Data Technology for Higher Education. In: *Procedia Computer Science* 124 (2017), S. 93–99. – URL <https://www.sciencedirect.com/science/article/pii/S1877050917329>

022. – 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia. – ISSN 1877-0509
- [70] TROM, Lomso ; CRONJE, Johannes: *Analysis of Data Governance Implications on Big Data*. S. 645–654, 01 2020. – ISBN 978-981-13-3623-2
- [71] VAN LOOY, Amy ; DE BACKER, Manu ; POELS, Geert ; SNOECK, Monique: Choosing the right business process maturity model. In: *Information & Management* 50 (2013), Nr. 7, S. 466–488. – URL <https://www.sciencedirect.com/science/article/pii/S0378720613000621>. – ISSN 0378-7206
- [72] WEBER, Kristin ; OTTO, Boris ; ÖSTERLE, Hubert: One Size Does Not Fit All—A Contingency Approach to Data Governance. In: *J. Data and Information Quality* 1 (2009), jun, Nr. 1. – URL <https://doi.org/10.1145/1515693.1515696>. – ISSN 1936-1955
- [73] YANG, Chao-Tung ; CHEN, Tzu-Yang ; KRISTIANI, Endah ; WU, Shyhtsun F.: The implementation of data storage and analytics platform for big data lake of electricity usage with spark. In: *The Journal of Supercomputing* 77 (2021), Jun, Nr. 6, S. 5934–5959. – URL <https://doi.org/10.1007/s11227-020-03505-6>. – ISSN 1573-0484

# A Anhang

## A.1 Abbildungen

### A.1.1 Data Governance Modell

## Data Governance and Stewardship



(P) Planning, (C) Control, (D) Development, (O) Operations

Abbildung A.1: Übersicht über Data Governance Modell [34]

## **A.1.2 Prozessabläufe**

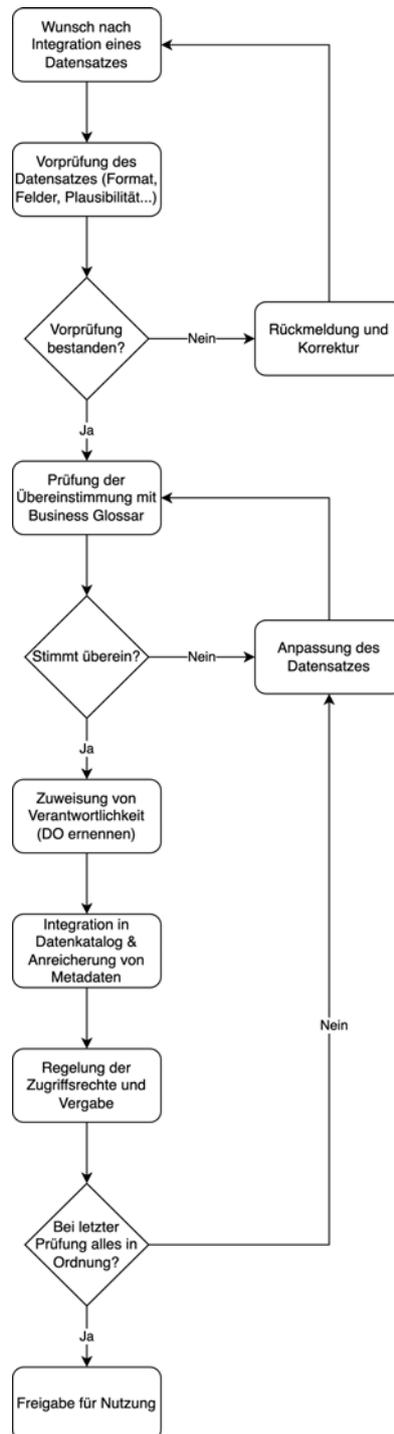


Abbildung A.2: Prozessablauf für Integration neuer Datensätze

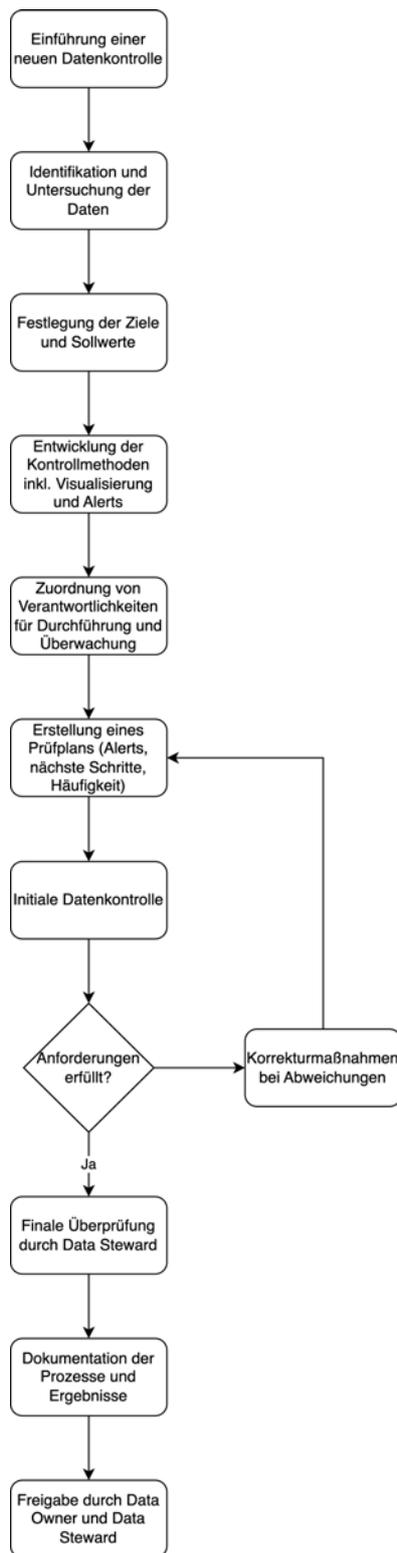


Abbildung A.3: Prozessablauf für die Einführung neuer Datenkontrollen

### A.1.3 Implementierung in DataHub

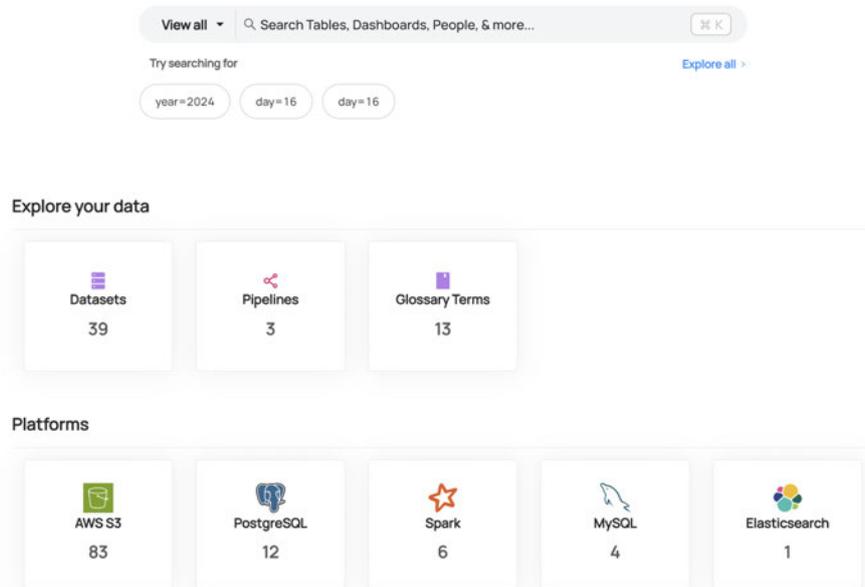
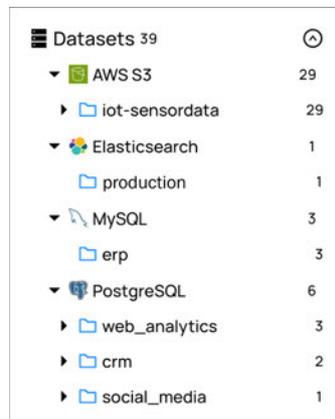


Abbildung A.4: Ausschnitt der Startseite von Datahub mit Suchleiste



Datasets 39	
▼ AWS S3	29
▶ iot-sensordata	29
▼ Elasticsearch	1
▶ production	1
▼ MySQL	3
▶ erp	3
▼ PostgreSQL	6
▶ web_analytics	3
▶ crm	2
▶ social_media	1

Abbildung A.5: Übersicht über vorhandene Datenquellen in Listenform



Business Glossary	
Classification	3
FinancialData	6
Legal-Regulatory-Guidelines	2
PersonalInformation	9

Abbildung A.6: Begriffsgruppen des erstellten Business Glossars

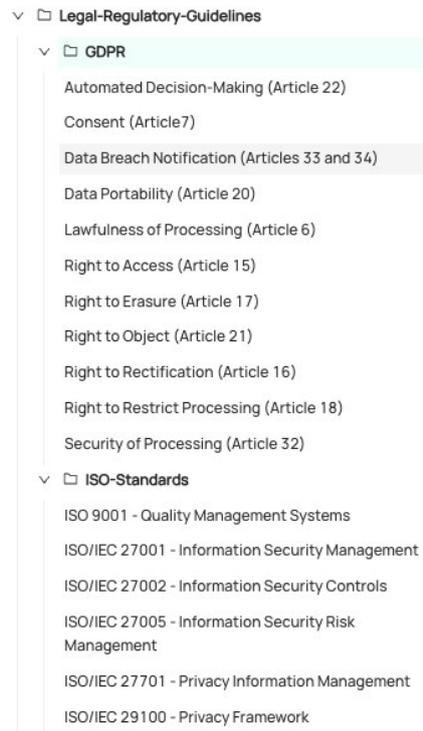


Abbildung A.7: Übersicht der rechtlichen Richtlinien und Normen

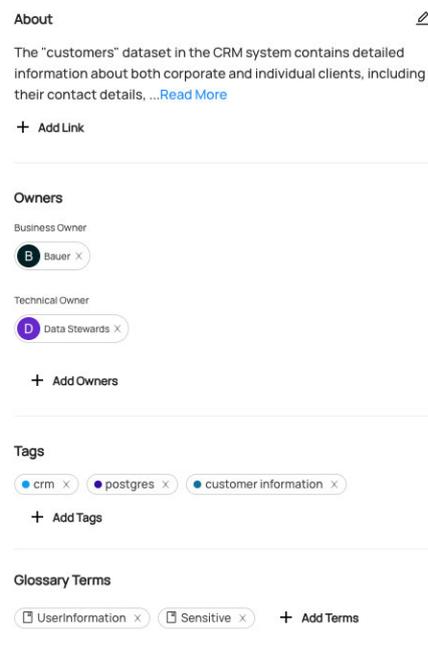


Abbildung A.8: Detaillierte Informationen eines Datensatzes

### A.1.4 Soda Core Ergebnis

```
2/2 checks FAILED:
  personal_data in social_source
    Ensure values are formatted as email addresses [FAILED]
      check_value: 11.69
      row_count: 231
      invalid_count: 27
    Ensure values are formatted as phone numbers [FAILED]
      check_value: 16.45
      row_count: 231
      invalid_count: 38
Oops! 2 failures. 0 warnings. 0 errors. 0 pass.
```

Abbildung A.9: Ergebnisse der ersten Qualitätsmessung

### A.1.5 Apache Airflow Workflow

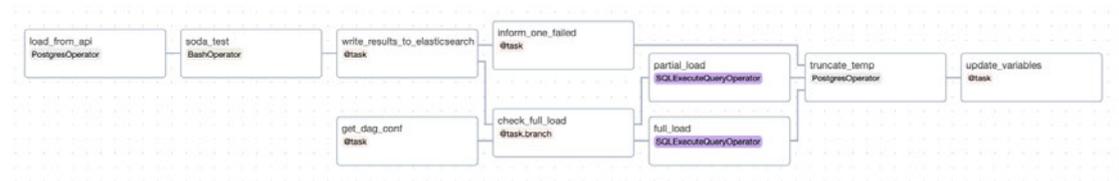


Abbildung A.10: Graphische Visualisierung eines Datenqualitäts-Workflows

## A.2 Listings

### A.2.1 Rechtezuweisung in MinIO

Listing A.1: Erstellung der Rechte für den Data Owner der IoT-Sensordaten in MinIO

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "s3:*",
      "Resource": [
        "arn:aws:s3:::iot-sensordata",
        "arn:aws:s3:::iot-sensordata/*"
      ]
    }
  ]
}
```

Listing A.2: Erstellung der Leserechte für die IoT-Sensordaten in MinIO

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
        "s3:ListBucket"
      ],
      "Resource": [
        "arn:aws:s3:::iot-sensordata",
        "arn:aws:s3:::iot-sensordata/*"
      ]
    }
  ]
}
```

## Erklärung zur selbständigen Bearbeitung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

\_\_\_\_\_

Ort	Datum	Unterschrift im Original
-----	-------	--------------------------