



Peter Gluchowski · Frank Leisten  
Gero Presser (Hrsg.)

# Architekturen für BI & Analytics

Konzepte, Technologien  
und Anwendungen



**Prof. Dr. Peter Gluchowski** leitet den Lehrstuhl für Wirtschaftsinformatik, insb. Systementwicklung und Anwendungssysteme, an der Technischen Universität in Chemnitz und konzentriert sich dort mit seinen Forschungsaktivitäten auf das Themengebiet Business Intelligence & Analytics. Er beschäftigt sich seit mehr als 25 Jahren mit Fragestellungen, die den praktischen Aufbau dispositiver bzw. analytischer Systeme zur Entscheidungsunterstützung betreffen. Seine Erfahrungen aus unterschiedlichsten Praxisprojekten sind in zahlreichen Veröffentlichungen zu diesem Themenkreis dokumentiert.



**Frank Leisten** ist als Cloud Solution Architect bei Microsoft tätig und Teil des internationalen Data Governance Rangers Team. Mit diesem verantwortet er die Umsetzung und Implementierung übergreifender Data-Governance-Strategien auf Basis der Microsoft Azure Cloud-Technologie. Offenheit steht hierbei im Vordergrund, das bedeutet sowohl On-Premise-, Hybrid- und Multi-Cloud-Ansätze werden berücksichtigt wie auch die Integration von Drittanbieter-Lösungen. Frank verfügt über umfassende Expertise in mehreren Data-Management-Disziplinen und nutzt diese Erfahrung zur Schaffung nachhaltiger Data-Governance-Lösungen.



**Dr. Gero Presser** ist Mitgründer und Geschäftsführer bei der QuinScape GmbH, einem Dortmunder IT-Dienstleistungsunternehmen mit 170 Mitarbeitern und dem Fokus auf Data & Analytics. Er organisiert die Meetup-Gruppe »Business Intelligence & Analytics Dortmund« mit über 1.000 Mitgliedern und ist Vorsitzender des TDWI Roundtable Ruhrgebiet.

Peter Gluchowski • Frank Leisten • Gero Presser (Hrsg.)

# Architekturen für BI & Analytics

Konzepte, Technologien und Anwendung

Edition TDWI



Peter Gluchowski  
[peter.gluchowski@wirtschaft.tu-chemnitz.de](mailto:peter.gluchowski@wirtschaft.tu-chemnitz.de)

Frank Leisten  
[frankleisten@microsoft.com](mailto:frankleisten@microsoft.com)

Gero Presser  
[Gero.Presser@QuinScape.de](mailto:Gero.Presser@QuinScape.de)

Lektorat: Julia Griebel, Christa Preisendanz  
Copy-Editing: Ursula Zimpfer, Herrenberg  
Satz: Ill-satz, [www.drei-satz.de](http://www.drei-satz.de)  
Herstellung: Stefanie Weidner, Frank Heidt  
Umschlaggestaltung: Anna Diechtierow  
Druck und Bindung: mediaprint solutions GmbH, 33100 Paderborn

Fachliche Beratung und Herausgabe von dpunkt.büchern in der Edition TDWI:  
Prof. Dr. Peter Gluchowski · [peter.gluchowski@wirtschaft.tu-chemnitz.de](mailto:peter.gluchowski@wirtschaft.tu-chemnitz.de)

Bibliografische Information der Deutschen Nationalbibliothek  
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;  
detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN:  
Print 978-3-86490-864-4  
PDF 978-3-96910-579-5  
ePub 978-3-96910-580-1  
mobi 978-3-96910-581-8

1. Auflage 2022  
Copyright © 2022 dpunkt.verlag GmbH  
Wiebinger Weg 17  
69123 Heidelberg

#### *Hinweis:*

Aus Gründen der besseren Lesbarkeit wird im weiteren Inhalt teilweise auf die gleichzeitige Verwendung der Sprachformen männlich, weiblich und divers (m/w/d) verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für alle Geschlechter.

#### *Hinweis:*

Dieses Buch wurde auf PEFC-zertifiziertem Papier aus nachhaltiger  
Waldwirtschaft gedruckt. Der Umwelt zuliebe verzichten wir  
zusätzlich auf die Einschweißfolie.



#### *Schreiben Sie uns:*

Falls Sie Anregungen, Wünsche und Kommentare haben, lassen Sie es uns wissen: [hallo@dpunkt.de](mailto:hallo@dpunkt.de).

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen.

Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autor noch Verlag noch Herausgeber können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

# Vorwort

Die Wahl einer geeigneten Architektur für die eigene Organisation stellt nach wie vor einen zentralen Erfolgsfaktor für das Gelingen von Business-Intelligence- und Analytics-(BIA-)Aktivitäten dar [DalleMule & Davenport 2017]. Nur durch ein stabiles architektonisches Fundament lassen sich die unterschiedlichen Anforderungen und Wünsche der Anspruchsgruppen in angemessener Zeit und Qualität umsetzen. Empirische Untersuchungen belegen, dass sich grundlegende Architekturbausteine in einigen Branchen (vor allem im Banken- und Versicherungsbereich) über Jahrzehnte nicht ändern, wohl aber stetigen Anpassungen unterliegen. Laut einer Studie aus dem Jahr 2015 nutzen zwei von drei Großunternehmen Data-Warehouse-Systeme, mit deren Aufbau vor zehn oder mehr Jahren begonnen wurde [Purwins 2015].

Insofern erweist sich der Aufbau eines umfassenden BIA-Ökosystems als langfristige Investition, zumal sich der Wechsel auf eine andere technologische Plattform als aufwendiges und langwieriges Unterfangen darstellen kann. Demzufolge sind alle Organisationen gut beraten, sich intensiv mit der Auswahl von Komponenten für eine ganzheitliche BIA-Landschaft auseinanderzusetzen.

Der Sammelband nähert sich dem Thema der BIA-Architekturen vom Allgemeinen kommend zum Speziellen und nimmt dabei eine Strukturierung entlang möglichst abgrenzbarer Teilgebiete bzw. Konzepte vor. Folgerichtig wurden die Beiträge des Bandes in drei Teile aufgegliedert, beginnend mit einer Einführung und der Vorstellung grundlegender Konzepte. Darauf aufbauend erfolgt die Präsentation von Plattformen und Ökosystemen großer Lösungsanbieter, bevor konkrete Architekturbeispiele die Ausführungen komplettieren.

Einleitend bieten Peter Gluchowski, Frank Leisten und Gero Presser in Teil 1 eine »Einführung in die BIA-Architekturen« und erörtern dabei ausgehend von den Anforderungen an eine ganzheitliche BIA-Architektur vor allem die klassischen Architekturen für BIA-Ökosysteme. Anschließend widmen sich Carsten Dittmar und Peter Schulz den »Architekturen und Technologien für Data Lakes« und gehen dabei auf die Abgrenzung zu artverwandten Begrifflichkeiten wie Data Puddle, Data Pond und Data Ocean ein. Der Beitrag zeigt, dass sich auch im Data Lake unterschiedliche Datenbereiche bzw. Zonen finden lassen. Michael Daum konzentriert sich mit seinen Ausführungen auf »Datenzugriffsstrategien für Ana-

lytics bei beschränktem Datenquellenzugriff« und bietet Lösungsansätze an – ein Thema mit weitreichender praktischer Bedeutung. Der zunehmenden Bedeutung einer Real-Time-Verarbeitung von Daten trägt das Kapitel zum Thema »Enterprise Application Integration: aktuelle Ansätze« von Martin Janssen Rechnung. Ausführlich erfolgt hier die Erläuterung der Funktionsweise des Enterprises Service Bus auch vor dem Hintergrund aktueller Cloud-Konzepte.

Im zweiten Teil des Sammelbandes wird der Reigen der unterschiedlichen Plattformen und Ökosysteme verschiedener Produkthanbieter eröffnet durch Christian Schneider und Gero Presser mit dem Thema »Cloud Data Platform für die Logistikbranche: eine Lösung auf Basis von AWS«. Mittels eines konkreten Einsatzbeispiels werden hier die einzelnen Komponenten von Amazon Web Service in ihrem Zusammenspiel beschrieben. Stefan Ebener präsentiert im Beitrag »Organise the world's data – like Google« zusammen mit seinen Co-Autoren das Google BIA-Ökosystem. Moderne Ansätze und neue Konzepte, wie sie sich in aktuellen Cloud-Architekturen finden, stehen hier im Vordergrund. Ebenfalls im Cloud-Bereich lässt sich der Ansatz von Microsoft verorten, der im Kapitel von Fabian Jogschies mit dem Titel »Die Modern-Data-Warehouse-Architektur von Microsoft« erläutert wird. Als nicht nur in die Zukunft, sondern ebenso in die Vergangenheit gerichtet erweist sich der Text von Daniel Eiduzzis unter der Überschrift »SAP Business Warehouse von gestern bis morgen«, der die Entwicklung des SAP BIA-Ökosystems über mehr als zwei Jahrzehnte beleuchtet.

Der dritte Teil des Sammelbandes beschreibt unterschiedliche Architekturbeispiele und verdeutlicht, dass sich spezifische Ausformungen zumindest teilweise aus den Besonderheiten einzelner Branchen und Rahmenbedingungen ergeben. Der erste Beitrag in diesem Teil von Thomas Müller, Lisa Anne Schiborr und Stefan Seyfert ist im Bankenbereich angesiedelt und arbeitet unter der Überschrift »Aus der Theorie in die Praxis – der Einfluss regulatorischer Anforderungen auf eine moderne Referenzarchitektur« heraus, wie sich vor allem regulatorische Anforderungen auf moderne Architekturen auswirken. Als ebenfalls im Finanzdienstleistungssektor verortet erweisen sich die Ausführungen von Nick Golovin und Don Seur mit der »Case Study: Crédit Agricole Consumer Finance Netherlands«, die sich insbesondere auf die Möglichkeiten und Grenzen einer Datenvirtualisierung fokussiert. Den gleichen Schwerpunkt bedienen Daniel Rapp, Thomas Niewel und Jörg Meiners mit dem Titel »Datenvirtualisierung« und stellen die eingesetzten Technologien am Beispiel der Festo Gruppe vor.

Ein sehr spezielles Einsatzgebiet bildet die Pharmabranche, wie der Beitrag von Jörg Krempien, Jörg Frank und Philipp Kazzner unter der Überschrift »BIA-Architekturen für klinische Studien« zeigt. Auch hier finden sich zahlreiche regulatorische Vorgaben und sonstige Anforderungen, die es technologisch abzudecken gilt. Ein speziell auf die Belange von Data Science und Machine Learning ausgerichtetes Konzept beinhaltet der Text von Gerhard Brückl und Timo Klerx zu »BIA-Architekturen in der Versicherungsbranche«, wobei auch hier ein Data

Lake als zentrale Speicherkomponente Verwendung findet. Dass tragfähige Architekturkonzepte nicht nur bei Großunternehmen zentrale Bedeutung haben, belegt Markus Begerow im Kapitel »BIA-Architekturen für kleine und mittlere Unternehmen«. Aus aktuellen, innovationstreibenden Themen für KMU werden hier Anforderungen an die Gestaltung von BIA-Lösungen abgeleitet. Stärker auf einzelne betriebswirtschaftliche Einsatzbereiche geht der Beitrag von Christian Fürstenberg, Oliver Zimmer und Björn Beuter »Integrierte Planung und Reporting im Business-Analytics-gestützten Controlling« ein und stellt dabei das Corporate Performance Management sowie Self-Service in den Vordergrund der Betrachtung.

Aufgrund des Umgangs und der Dynamik des Gesamtthemas ist es leider nicht möglich, alle Teilaspekte in der gleichen Tiefe zu durchdringen und diese überschneidungsfrei zu präsentieren. Allerdings vermittelt der Sammelband durch die Vielzahl der eingenommenen Perspektiven einen breiten Überblick über das Thema mit hinreichender Würdigung der wichtigsten Teilaspekte.

Das vorliegende Werk wendet sich schwerpunktmäßig an betriebliche Anwender und Entscheider aus den IT-Abteilungen und den Fachbereichen, aber auch an Mitarbeiter aus Beratungshäusern, IT-Dienstleistungsunternehmen und Hochschulangehörige sowie Studierende in den immer vielfältigeren Disziplinen rund um Data Management und Analytics. Die Herausgeber hoffen, dass die Leserinnen und Leser wertvolle Anregungen und Hinweise für die Konzeptionierung und Realisierung von Business-Intelligence- und Analytics-Architekturen in eigenen Projekten erhalten.

Ein herzlicher Dank gilt den einzelnen Autoren, die trotz ihrer anderen Verpflichtungen, voller Terminkalender und speziell in einem durch die Covid-19-Pandemie geprägten Zeitintervall dennoch fristgerecht ihre jeweiligen Beiträge einbringen konnten. Wie gewohnt äußerst angenehm und konstruktiv war die Zusammenarbeit mit dem dpunkt.verlag; vor allem gilt hier Christa Preisendanz und dem Verlagsteam ein besonderer Dank.

Abschließend bleibt uns nur zu wünschen übrig, dass die Leserinnen und Leser dieses Sammelbandes interessante und hilfreiche Impulse für ihre eigene Arbeit finden und sich ihr Blickwinkel auf das Thema BIA-Architekturen insgesamt erweitert. Für kritische oder bestätigende Anmerkungen stehen wir unter den nachfolgenden E-Mail-Adressen gerne zur Verfügung:

*Peter.Gluchowski@wirtschaft.tu-chemnitz.de*

*frankleisten@microsoft.com*

*Gero.Presser@QuinScape.de*

Chemnitz, Merzenich, Dortmund im August 2021

Peter Gluchowski, Frank Leisten, Gero Presser





# Inhaltsübersicht

<b>Teil I</b>	<b>Grundlagen</b>	<b>1</b>
<b>1</b>	<b>Einführung in die BIA-Architekturen</b> Peter Gluchowski • Frank Leisten • Gero Presser	<b>3</b>
<b>2</b>	<b>Architekturen und Technologien für Data Lakes</b> Carsten Dittmar • Peter Schulz	<b>23</b>
<b>3</b>	<b>Datenzugriffsstrategien für Analytics bei beschränktem Datenquellenzugriff</b> Michael Daum	<b>39</b>
<b>4</b>	<b>Enterprise Application Integration: aktuelle Ansätze</b> Martin Janssen	<b>53</b>
<b>Teil II</b>	<b>Plattformen und Ökosysteme</b>	<b>73</b>
<b>5</b>	<b>Cloud Data Platform für die Logistikbranche: eine Lösung auf Basis von AWS</b> Christian Schneider • Gero Presser	<b>75</b>
<b>6</b>	<b>Organise the world's data – like Google</b> Stefan Ebener • Stiv Sterjo • Sascha Kerbler • Andreas Ribbrock • Alex Osterloh • Diana Nanova • Christine Schulze • Lukas Grubwieser	<b>89</b>
<b>7</b>	<b>Die Modern-Data-Warehouse-Architektur von Microsoft</b> Fabian Jogschies	<b>121</b>
<b>8</b>	<b>SAP Business Warehouse von gestern bis morgen</b> Daniel Eiduzzis	<b>145</b>
<b>9</b>	<b>Aus der Theorie in die Praxis – der Einfluss regulatorischer Anforderungen auf eine moderne Referenzarchitektur</b> Thomas Müller • Lisa Anne Schiborr • Stefan Seyfert	<b>159</b>

<b>10</b>	<b>Case Study: Crédit Agricole Consumer Finance Netherlands</b>	<b>191</b>
	Nick Golovin • Don Seur	
<b>11</b>	<b>Datenvirtualisierung</b>	<b>201</b>
	Daniel Rapp • Thomas Niewel • Jörg Meiners	
<b>Teil III Architekturbeispiele</b>		<b>217</b>
<b>12</b>	<b>BIA-Architekturen für klinische Studien</b>	<b>219</b>
	Jörg Krempien • Jörg Frank • Philipp Kazzner	
<b>13</b>	<b>BIA-Architekturen in der Versicherungsbranche</b>	<b>235</b>
	Gerhard Brückl • Timo Klerx	
<b>14</b>	<b>BIA-Architekturen für kleine und mittlere Unternehmen</b>	<b>253</b>
	Markus Begerow	
<b>15</b>	<b>Integrierte Planung und Reporting im Business-Analytics-gestützten Controlling</b>	<b>275</b>
	Christian Fürstenberg • Oliver Zimmer • Björn Beuter	
<b>Anhang</b>		<b>291</b>
<b>A</b>	<b>Autoren</b>	<b>293</b>
<b>B</b>	<b>Abkürzungen</b>	<b>305</b>
<b>C</b>	<b>Literaturverzeichnis</b>	<b>311</b>
	<b>Index</b>	<b>319</b>

# Inhaltsverzeichnis

<b>Teil I</b>	<b>Grundlagen</b>	<b>1</b>
<b>1</b>	<b>Einführung in die BIA-Architekturen</b>	<b>3</b>
	Peter Gluchowski • Frank Leisten • Gero Presser	
1.1	BIA-Trends und -Entwicklungen . . . . .	3
1.2	Architekturkonzepte und -facetten . . . . .	5
1.3	Datenbezogene Rahmenbedingungen . . . . .	8
1.3.1	Datenstrategie . . . . .	9
1.3.2	Data Valuation . . . . .	11
1.3.3	Data Management . . . . .	13
1.4	Anforderungen an eine ganzheitliche BIA-Architektur . . . . .	17
1.5	Klassische Architekturen für BIA-Ökosysteme . . . . .	19
<b>2</b>	<b>Architekturen und Technologien für Data Lakes</b>	<b>23</b>
	Carsten Dittmar • Peter Schulz	
2.1	Historie der dispositiven Datenplattformen . . . . .	23
2.2	Das Data-Lake-Konzept . . . . .	24
2.3	Architektur eines Data Lake . . . . .	27
2.4	Datenarchitektur eines Data Lake . . . . .	31
2.5	Technologien für einen Data Lake . . . . .	32
2.6	Herausforderungen in der Umsetzung eines Data Lake . . . . .	36
<b>3</b>	<b>Datenzugriffsstrategien für Analytics bei beschränktem Datenquellenzugriff</b>	<b>39</b>
	Michael Daum	
3.1	Ursachen von Einschränkungen auf Datenquellen . . . . .	40
3.2	BIA-Anforderungen an Datenquellen . . . . .	44

3.3	Datenstrategische Überlegungen .....	44
3.3.1	Trennung von Problemstellung und technischer Lösung ....	45
3.3.2	Skalierbarkeit .....	46
3.3.3	Cloud-Strategie und Datenschutz .....	46
3.3.4	Data Management .....	47
3.4	Lösungsansätze bei unterschiedlichen Einschränkungen .....	47
3.4.1	Technische Probleme der Connectivity .....	47
3.4.2	(Firmen-)»Politische« Themen und Lizenzen .....	48
3.4.3	Datenschutzanforderungen beim Zugriff .....	49
3.5	Entkoppeln von Systemen und Datenvirtualisierung .....	49
3.6	Abgrenzung und weiterführende Themen .....	51
<b>4</b>	<b>Enterprise Application Integration: aktuelle Ansätze</b>	<b>53</b>
	Martin Janssen	
4.1	Ein altbekanntes Thema vor immer neuen Herausforderungen .....	53
4.2	Der Unterschied zwischen Theorie und Praxis ist in der Praxis weit höher als in der Theorie .....	55
4.3	Die Zeit des ESB .....	57
4.4	Neue Anforderungen durch die Clouds .....	59
4.5	Drei aktuelle Lösungsansätze .....	60
4.5.1	iPaaS – Integration mittels Low Code und als Turnkey-Lösung .....	60
4.5.2	Kafka – der neue ESB? .....	65
4.5.3	Serverless Integration – alles in der Cloud .....	68
4.6	Fazit .....	70

---

## **Teil II Plattformen und Ökosysteme** **73**

---

<b>5</b>	<b>Cloud Data Platform für die Logistikbranche: eine Lösung auf Basis von AWS</b>	<b>75</b>
	Christian Schneider • Gero Presser	
5.1	Herausforderung .....	76
5.2	Grundlegende Architektur .....	77
5.3	Technische Architektur mit AWS .....	78
5.4	Data Lake: AWS S3 und AWS Lake Formation .....	81

5.5	ETL und mehr: AWS Glue .....	82
5.6	Data Warehouse: AWS Redshift .....	83
5.7	Query Engine: AWS Redshift Spectrum .....	83
5.8	Visualisierung: AWS QuickSight .....	84
5.9	Flexibilität in der Architektur .....	85
5.10	Betrieb und Wartung .....	86
5.11	Ergebnis und Resümee .....	86
<b>6</b>	<b>Organise the world's data – like Google</b> .....	<b>89</b>
	Stefan Ebener • Stiv Sterjo • Sascha Kerbler • Andreas Ribbrock • Alex Osterloh • Diana Nanova • Christine Schulze • Lukas Grubwieser	
6.1	Einführung .....	89
6.1.1	Herausforderungen für eine erfolgreiche BI-Landschaft ....	91
6.1.2	Der Nutzen einer erfolgreichen BI-Landschaft .....	92
6.2	BI in der Public Cloud vs. On-Premises BI .....	93
6.2.1	Vom Budgetprozess hin zum aktiven Kostenmonitoring ...	94
6.2.2	Neue Unternehmensstrukturen rund um BIA in der Cloud .....	94
6.2.3	Trennung von Datenspeicherung und Rechenleistung .....	95
6.2.4	Elastizität und Skalierbarkeit .....	96
6.2.5	Infrastruktur als Code – IaC .....	97
6.2.6	Konvergenz von SQL und KI/ML .....	97
6.2.7	Vom Prototyp zur Applikation .....	98
6.3	Moderne Ansätze und neue Konzepte für BIA .....	98
6.3.1	Data Mesh aka Enterprise Data Evolution .....	98
6.3.2	Lake House als nächste Generation des Data Lake .....	102
6.4	Business Intelligence mit Google Cloud .....	106
6.4.1	Einführung einer serverlosen Architektur .....	108
6.4.2	Einführung innovativer KI/ML-Technologien .....	109
6.4.3	KI/ML im produktiven Einsatz .....	112
6.4.4	Ende-zu-Ende-Anwendung von einer mit KI/ML integrierten Datenplattform .....	116
6.4.5	Das »Big Picture« als Referenzarchitektur für ein modernes Lake House .....	116
6.4.6	Betrieb produktiver Anwendungen mit Google .....	118
6.5	Fazit und Ausblick .....	119

<b>7</b>	<b>Die Modern-Data-Warehouse-Architektur von Microsoft</b>	<b>121</b>
	Fabian Jogschies	
7.1	Datenablage mit Azure Data Lake Storage Gen2	121
7.2	Data Ingest und Orchestrierung	125
7.3	Transformation, Serving und ML mit Azure Synapse Analytics	127
7.4	Transformation, Serving und ML mit Azure Databricks	132
7.5	Data Lab Toolbox – Machine Learning	135
7.5.1	Azure Machine Learning (AML)	135
7.5.2	Azure Cognitive Services	138
7.6	Visualisierung mit Power BI	140
7.7	Data Governance mit Azure Purview	141
7.8	Azure DevOps	142
<b>8</b>	<b>SAP Business Warehouse von gestern bis morgen</b>	<b>145</b>
	Daniel Eiduzzis	
8.1	Business Intelligence made in Walldorf	145
8.1.1	SAP Business Warehouse – Wie alles begann	145
8.1.2	Probleme, Kritik und Herausforderungen im SAP BW-Kontext	147
8.2	Entwicklung des SAP BW	148
8.2.1	Die Zeit vergeht, das SAP BW bleibt	148
8.2.2	Der große Wurf bleibt aus	150
8.3	SAP Business Intelligence – heute und morgen	152
8.3.1	HANA und Cloud geben die Strategie vor	152
8.3.2	Features und Werkzeuge für Data Management und Data Integration	155
8.3.3	Reporting und Analyse dort, wo die Daten generiert werden	156
8.3.4	Hybride Konzepte als State-of-the-Art-Architektur	156
8.4	Ausblick und Fazit	158

<b>9</b>	<b>Aus der Theorie in die Praxis – der Einfluss regulatorischer Anforderungen auf eine moderne Referenzarchitektur</b>	<b>159</b>
	Thomas Müller • Lisa Anne Schiborr • Stefan Seyfert	
9.1	Aktuelle Herausforderungen .....	159
9.2	Historisierung .....	163
9.2.1	Bitemporale Historisierung .....	164
9.2.2	Best Practice .....	165
9.3	Datenschichtenarchitektur .....	166
9.3.1	Datenschichten der Referenzarchitektur .....	166
9.3.2	Archivierung und Housekeeping .....	169
9.4	Integrationsarchitektur .....	171
9.4.1	Verfahren und Werkzeuge .....	172
9.4.2	Anbindung Metadatenmanagement .....	173
9.4.3	Anbindung Datenqualitätsmanagement .....	174
9.5	Metadatenmanagement (MDM) .....	174
9.5.1	MDM – Kernanforderungen .....	176
9.5.2	Die Metamodelllandkarte (Modellsichten) .....	178
9.5.3	Data Lineage .....	179
9.5.4	Best Practice MDM – Technologie .....	181
9.5.5	Best Practice MDM – Architektur .....	182
9.6	Datenqualitätsmanagement (DQM) .....	183
9.6.1	DQM – Kernanforderungen .....	184
9.6.2	Prüfregeln .....	184
9.6.3	Korrekturen .....	185
9.6.4	Best Practice DQM – Architektur .....	186
9.7	Fazit und Handlungsempfehlungen .....	187
<b>10</b>	<b>Case Study: Crédit Agricole Consumer Finance Netherlands</b>	<b>191</b>
	Nick Golovin • Don Seur	
10.1	Herausforderungen .....	191
10.1.1	Lange Time-to-Market .....	192
10.1.2	Zugang zu Echtzeitdaten .....	193
10.1.3	Daten für operative Zwecke .....	193
10.1.4	DSGVO-Konformität .....	193
10.1.5	Anbindung von modernen Datenquellen .....	193

10.2	Anforderungen an die neue Lösung	194
10.3	Moderne Datenarchitektur	195
10.4	Use Cases	197
10.4.1	360°-Blick auf Kunden	198
10.4.2	Echtzeit-Sales-Monitoring	198
10.4.3	Marketinganalysen	199
10.4.4	Risk Management	199
10.4.5	Data Preparation für regulatorische Reportings	199
10.5	Schlussfolgerung: Datenvirtualisierung das Allheilmittel?	200
<b>11</b>	<b>Datenvirtualisierung</b>	<b>201</b>
	Daniel Rapp • Thomas Niewel • Jörg Meiners	
11.1	Moderne Datenarchitekturen für das Zeitalter der Digitalisierung	201
11.2	Datenvirtualisierung – ein Überblick	202
11.2.1	Anwendungsfälle der Datenvirtualisierung	204
11.3	Die Technologie der Datenvirtualisierung	206
11.3.1	Zugriff auf das Datenmodell	207
11.3.2	Datenschutz und Sicherheit	208
11.3.3	Query-Optimierung	209
11.3.4	Daten-Caching	210
11.3.5	Datenkatalog	211
11.4	Abgrenzung zu anderen Integrationstechnologien	211
11.5	Kundenbeispiel: Die Festo Gruppe	212
11.5.1	Unternehmensprofil der Festo Gruppe	212
11.5.2	Geschäftsanforderungen	212
11.5.3	Die Lösung	213
11.5.4	Die Mehrwerte	214
11.6	Zusammenfassung	215
11.6.1	Die Anwenderperspektive	216
11.6.2	Die Data-Governance-Perspektive	216
11.6.3	Die IT-Perspektive	216



<b>Teil III</b>	<b>Architekturbeispiele</b>	<b>217</b>
<b>12</b>	<b>BIA-Architekturen für klinische Studien</b>	<b>219</b>
	Jörg Krempien • Jörg Frank • Philipp Kazzner	
12.1	Über klinische Studien	219
12.2	Anforderungen an die BI-Architektur	221
12.3	Architekturdetails	222
12.3.1	Architekturüberblick	222
12.3.2	Staging	223
12.3.3	Core	225
12.3.4	Publish	226
12.3.5	Virtualisierung (Domänen durch Konfiguration)	227
12.4	Entwicklungsgeschichte und Ausblick	229
12.5	Use Cases	230
12.5.1	Virtual Cut Off	230
12.5.2	Subject Status	231
12.5.3	Baseline Flags	231
12.5.4	Clean Patient Tracker	232
12.5.5	Fraud Detection in Clinical Trials	232
12.5.6	Testautomatisierung (TAT)	232
<b>13</b>	<b>BIA-Architekturen in der Versicherungsbranche</b>	<b>235</b>
	Gerhard Brückl • Timo Klerx	
13.1	Ausgangssituation	235
13.2	Zielsetzung	236
13.3	Zielarchitektur	237
13.4	Data Lake	241
13.5	Datenverarbeitung	245
13.6	Ablaufsteuerung	247
13.7	Data-Science-Labor	249
13.8	Reporting	250

<b>14</b>	<b>BIA-Architekturen für kleine und mittlere Unternehmen</b>	<b>253</b>
	Markus Begerow	
14.1	Ausgangssituation . . . . .	253
14.2	Neue Themen als Treiber für Innovationen . . . . .	254
14.2.1	Stammdaten- und Datenqualitätsmanagement . . . . .	254
14.2.2	Cloud-Infrastrukturen . . . . .	255
14.2.3	Data Science im Mittelstand . . . . .	260
14.3	Konsequenzen für kleine und mittlere Unternehmen . . . . .	262
14.3.1	Tabellen- und Textdateien ersetzen keine Datenbank . . . . .	262
14.3.2	Cloud-Servicemodelle verstehen . . . . .	263
14.3.3	Data Science light einführen . . . . .	268
14.4	Fazit . . . . .	272
<b>15</b>	<b>Integrierte Planung und Reporting im Business-Analytics-gestützten Controlling</b>	<b>275</b>
	Christian Fürstenberg • Oliver Zimmer • Björn Beuter	
15.1	Die Entwicklung der Finanzplanung und -analyse . . . . .	276
15.2	Der Weg zur datengetriebenen Unternehmenssteuerung . . . . .	278
15.3	CCH® Tagetik – eine Lösung für alle Corporate-Performance-Management-Bereiche . . . . .	280
15.4	Data Literacy – Aufbau von Datenkompetenz im Controlling . . . . .	282
15.5	Einsatz und Nutzen von Self-Service im Controlling . . . . .	284
15.6	Power BI als Self-Service-Reporting- und Analyse-Architektur . . . . .	286
15.7	Einsatz von Power BI im Umfeld von CCH® Tagetik . . . . .	287
	<b>Anhang</b>	<b>291</b>
<b>A</b>	<b>Autoren</b>	<b>293</b>
<b>B</b>	<b>Abkürzungen</b>	<b>305</b>
<b>C</b>	<b>Literaturverzeichnis</b>	<b>311</b>
	<b>Index</b>	<b>319</b>

Teil I

---

# Grundlagen



# 1 Einführung in die BIA-Architekturen

Peter Gluchowski • Frank Leisten • Gero Presser

*Der vorliegende Beitrag setzt sich das Ziel, die Rahmenbedingungen für komplexe Business Intelligence & Analytics-(BIA-)Landschaften zu beleuchten. Den Ausgangspunkt für die Betrachtungen bildet der folgende Abschnitt, der BIA-Trends und -Entwicklungen in der letzten Dekade punktuell aufgreift und die Bedeutung für die zugehörigen dispositiven Ökosysteme herausarbeitet. Danach erfolgen eine Abgrenzung und Einordnung der BIA-Architektur zu verwandten Themen wie Unternehmensarchitektur, IT-Architektur, Anwendungsarchitektur und Infrastruktur (Abschnitt 1.2). Anschließend nähert sich Abschnitt 1.3 dem Architekturthema aus einer Datenperspektive, indem die Datenstrategie, die Wertermittlung von Daten und das Datenmanagement im Vordergrund der Betrachtung stehen. Der anschließende Abschnitt 1.4 beleuchtet die Anforderungen an eine ganzheitliche BIA-Architektur aus der Perspektive unterschiedlicher Anspruchsgruppen und macht deutlich, dass sich die Vorstellungen und Ziele erheblich voneinander unterscheiden können. Schließlich greift Abschnitt 1.5 die klassische Hub-and-Spoke-Architektur und die Schichtenarchitektur für BIA-Ökosysteme auf und verweist auf die zugehörigen Defizite.*

## 1.1 BIA-Trends und -Entwicklungen

In der letzten Dekade lässt sich eine zunehmende Komplexität analytischer Architekturen feststellen. Waren es noch vor zehn Jahren die klassischen Data-Warehouse-zentrierten Architekturkonzepte, die fast flächendeckend und ausschließlich Verwendung fanden, haben in der Zwischenzeit vielfältige zusätzliche Komponenten und Technologien Einzug in die BIA-Landschaften der Unternehmen gehalten.

Unterstützt wurde diese Entwicklung nicht zuletzt durch die intensive Diskussion um Big Data, die durch die Hypothese geleitet ist, dass die herkömmlichen Konzepte und Technologien nicht dazu in der Lage sind, alle aktuellen Anforderungen in geeigneter Form zu erfüllen. So greifen einige Veröffentlichungen zu dem Thema auf eine Negativabgrenzung zurück und stellen heraus, dass Big Data

dann gegeben ist, wenn die Kapazitäten und Funktionalitäten der klassischen Datenhaltung, -aufbereitung und -auswertung sich als nicht ausreichend erweisen [Dittmar et al. 2016, S. 3]. Zumeist wird Big Data heute durch die charakteristischen Eigenschaften beschrieben. Dann zeichnet sich Big Data nicht allein durch das immense Datenvolumen (Volume) aus, sondern ebenso durch die erhebliche Vielfalt an Datenformaten (Variety) sowie durch die Geschwindigkeit (Velocity), mit der neue Daten entstehen sowie verfügbar und damit analysierbar sind [Eaton et al. 2012, S. 5].

Allerdings lassen sich zahlreiche weitere Begrifflichkeiten mit dem Anfangsbuchstaben »V« und somit weitere Dimensionen identifizieren, mit denen Big Data umschrieben wird. Beispielsweise adressiert Veracity als Wahrhaftigkeit oder Richtigkeit der Daten eine weitere Eigenschaft von Big Data, zumal Auswertungen und die damit verbundenen Entscheidungen hierauf beruhen und falsche Daten zu fehlerhaften Analyseergebnissen führen können. Aufgrund der Datenvielfalt und des Datenvolumens erweist sich eine Überprüfung der Daten jedoch häufig als schwierig [Klein et al. 2013, S. 321]. Als weitere Begrifflichkeiten mit »V« finden sich Validity, Volatility, Variability und vor allem Value, auf die hier allerdings nicht weiter eingegangen wird [Gandomi & Haider 2015, S. 139; Khan et al. 2014, S. 3]. Es liegt auf der Hand, dass hieraus gänzlich neue Bedarfe resultieren, die es zu erfüllen gilt.

Auch seitens der Datenanalyse haben sich in den letzten Jahren bemerkenswerte Veränderungen eingestellt, die sich in einer verstärkten Hinwendung zu anspruchsvollen statistisch-mathematischen Verfahren unter Oberbegriffen wie künstliche Intelligenz, Machine Learning oder Data Science zeigen. Derzeit erweisen sich vor allem komplexe künstliche neuronale Netze (Deep Learning) als leistungsfähig, mit denen die Erforschung von Strukturzusammenhängen (Datenmustern) in Datenbeständen eine neue Qualität erreicht [Dorer 2019, S. 119 ff.].

Als Konsequenz aus diesen Entwicklungen erfolgte in zahlreichen Unternehmen eine zumindest teilweise Abkehr beispielsweise von den klassischen, festplattenorientierten relationalen Datenbanksystemen hin zur schemalosen und verteilten Ablage des Datenmaterials, mit der sich auch große und polystrukturierte Inhalte organisieren lassen. Daneben mündet die Forderung nach hoher Verarbeitungsgeschwindigkeit in neuen Herausforderungen, die eine Erweiterung oder Ergänzung der bislang üblichen Batch-orientierten Aufbereitung des Datenmaterials für analytische Zwecke zur Folge hat – spätestens dann, wenn Datenströme zu verarbeiten sind.

Begünstigt wird die Veränderung durch zahlreiche neue Technologien. Bezogen auf die Speicherung von Daten sei hier etwa auf In-Memory-Konzepte, NoSQL-Datenbanksysteme (z.B. als Key-Value Store) sowie auf Cloud-Technologien verwiesen. Im Frontend-Sektor dagegen haben Self-Service-Werkzeuge breiten Raum eingenommen.

Auch aus organisatorischen Gründen haben sich im letzten Jahrzehnt die Voraussetzungen für die Gestaltung von BIA-Architekturen geändert. So erfor-

dert die zunehmende Hinwendung zu agilen Gestaltungsmethodiken mit kurzen Entwicklungszyklen, dass sich inkrementelle und iterative Veränderungen im Systemaufbau auch mit den vorhandenen Landschaften realisieren lassen. Aufgrund des engen zeitlichen Rahmens erweist es sich dabei teilweise als unumgänglich, dass einzelne Entwicklungsschritte durch Automatisierungsverfahren und -komponenten beschleunigt werden. Aber auch aus dem Betrieb von BIA-Lösungen ergeben sich Beschleunigungsbedarfe, die oftmals unter dem Begriffsgebilde DataOps diskutiert werden [Detemple 2020]. Gefordert wird hier sowohl eine Datenpipeline als auch eine Analytics-Pipeline zur möglichst zeitnahen Zurverfügungstellung von Berichten, Dashboards und Analytics-Modellen für den Endanwender.

Weitere Rahmenbedingungen für die BIA-Landschaft ergeben sich aus externen, regulatorischen, aber auch internen Vorgaben, die es zu erfüllen gilt. Als wichtige regulatorische Vorgabe lässt sich die Datenschutz-Grundverordnung (DSGVO) anführen, aus der sich die Notwendigkeit eines besonderen Umgangs mit personenbezogenen Daten und der architektonischen Umsetzung ableiten lässt. In einzelnen Branchen existieren darüber hinaus spezielle Regularien, die weit über den einfachen gesetzlichen Standard hinausreichen. So kann für den Finanzdienstleistungssektor das Regelwerk der BCBS 239 angeführt werden, aus dem sich weitreichende Anforderungen an die Transparenz und Nachverfolgbarkeit der Verarbeitung von Daten ergeben.

Vor diesem Hintergrund wird deutlich, dass eine einfache Architektur mit wenigen Komponenten heute kaum ausreichen kann, um allen Anforderungen gerecht zu werden. Vielmehr stellt sich die Aufgabe, ein analytisches Ökosystem zu gestalten, in dem jeder Baustein definierte Funktionen übernimmt und dabei seine spezifischen Stärken einbringt. Naturgemäß ergibt sich hieraus die steigende Komplexität der Gesamtlandschaft, zumal das reibungslose Zusammenspiel der einzelnen Komponenten eine Herausforderung darstellt.

## 1.2 Architekturkonzepte und -facetten

Der Begriff Architektur findet in zahlreichen Wissensdisziplinen und thematischen Bereichen Verwendung. Allgemein repräsentiert eine Architektur die Gesamtheit aller beschreibenden Darstellungen (Entwurfsartefakte) der erkenntnisrelevanten Objekte derart, dass diese den Anforderungen entsprechend produziert und betrieben werden können (Qualität). Idealerweise bleiben die grundlegenden Teile der Beschreibung möglichst unverändert über die Nutzungsdauer erhalten [Zachman 1997], können aber an geänderte Bedingungen angepasst werden. Die Artefakte bilden neben der Repräsentation von Objekten auch deren Funktionen, Schnittstellen und Beziehungen sowie dynamische Aspekte ab, wie den zeitlichen Ablauf von Austauschbeziehungen [Krcmar 2015, S. 280 f.].

Im Kontext von Informationssystemen umfasst dies die modellhafte Beschreibung der grundsätzlichen Struktur eines Systems mit seinen Elementen, der Beziehungen zwischen den Elementen sowie den Beziehungen des Systems zur Umwelt

[ISO 2000; Knoll 2018, S. 889]. Neben der Spezifikation seiner Komponenten und ihrer Beziehungen unter allen relevanten Blickwinkeln lassen sich auch die Konstruktionsregeln zur Erstellung des Bauplans [Sinz 2019] sowie die Prinzipien zur Konstruktion, Weiterentwicklung und Nutzung des Systems zu einer Informationssystem-Architektur zählen [IEEE 2000].

Durch die umfassende, globale Sicht auf ein Informationssystem, die alle relevanten Komponenten beinhaltet, unterscheidet sich die Architektur von eingeschränkteren Ansätzen (z.B. der unternehmensweiten Datenmodellierung). Zudem erfolgt die Konzentration auf eher aggregierte Elemente und Beziehungen, um die Ganzheitlichkeit der Betrachtung zu ermöglichen, ohne den Überblick zu verlieren [Winter & Aier 2019].

Als Teil einer Informationssystem-Architektur beschreibt die Datenarchitektur eines Informationssystems auf Fachkonzept- oder Entwurfsebene die grundlegenden Datenstrukturen und bildet dabei die Datenarchitektur eines ganzen Unternehmens ab oder konzentriert sich als Datenarchitektur eines Anwendungssystems auf einen Ausschnitt des Unternehmens [Winter & Aier 2019]. Demgegenüber repräsentiert die IT-Infrastruktur die technischen Komponenten, bestehend aus Hardware, (System-)Software sowie baulichen Einrichtungen für den Betrieb von (Anwendungs-)Software [Patig et al. 2019].

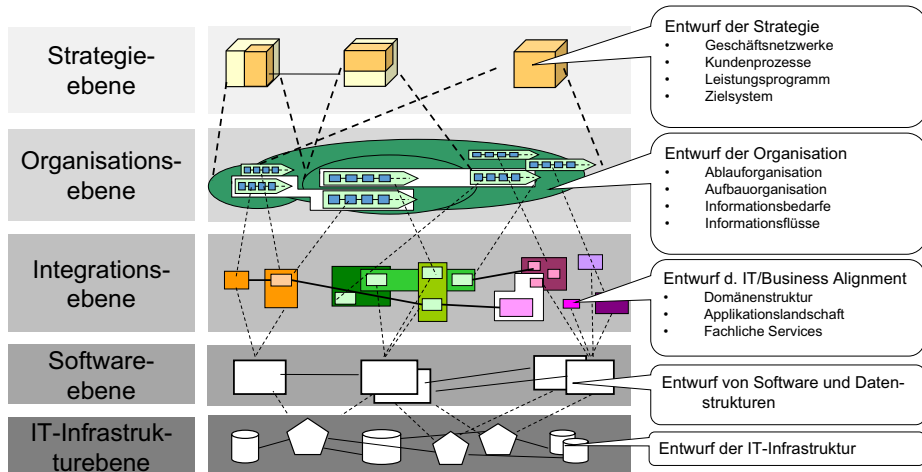
Um den Bezug zu geschäftlichen bzw. fachlichen Sichtweisen auf die Architekturen und damit ein gutes Business-IT-Alignment zu wahren, sind über die technische Perspektive hinaus weitere Aspekte zu berücksichtigen [Knoll 2018]. So lassen sich strategische und organisationale Ebenen abbilden, die auf den technischen Layern aufsetzen und diese ergänzen.

Auf jeder der betrachteten Architekturebenen finden sich unterschiedliche Objekte, deren Ausgestaltung und Zusammenwirken den Aufbau des Gesamtgebildes bestimmen (vgl. Abb. 1–1). Zur Gestaltung sind verschiedene Modelltypen verwendbar, die beim Entwurf der spezifischen Strukturen unterstützen. So finden sich auf der strategischen Ebene beispielsweise Modelle zur Abbildung von Geschäftsbeziehungen zu Kunden und Lieferanten. Bei der Beschreibung der Organisationsebene gelangen neben Prozesslandkarten und -modellen auch Organigramme sowie (fachliche) Informationslandkarten zur Anwendung. Auf der untersten Ebene, der IT-Infrastrukturebene, finden sich Beschreibungen über das Zusammenspiel (hardwarenaher) technischer Komponenten wie Modelle der Netzwerkinfrastruktur. Die Softwareebene darüber bildet neben den relevanten Datenstrukturen auch den Aufbau der Softwarekomponenten ab, beispielsweise auf Basis von Softwaremodulen oder auch -services.

Eine besondere Rolle spielt bei diesem Konzept die Integrationsebene, die sich als Mittler zwischen betriebswirtschaftlich-fachlicher und technischer Perspektive erweist. Hier werden einzelne Softwarebestandteile zu Anwendungen und Datenstrukturen zu Domänen verknüpft, um einzelne fachliche Prozesse



unterstützen zu können. Infolgedessen lassen sich hier Modelle der Applikationslandschaft und Domänenmodelle verwenden. Neben der Verknüpfungsfunktion erwies sich hier in der Vergangenheit die Entkopplung von fachlichen und technischen Komponenten als hilfreich, um die langsam sich ändernden technischen Gegebenheiten (mit Zykluszeiten von 6 bis 10 Jahren) mit den relativ schnell sich ändernden fachlichen Ebenen (von 3 bis 6 Monaten auf der Organisations-ebene bis zu 1–2 Jahren auf der Strategieebene) zu synchronisieren [Winter 2008, S. 24 ff.].



**Abb. 1–1** Architekturebenen des Business Engineering [Winter 2010, S. 90]

Vor dem Hintergrund von sich stetig schneller entwickelnden technologischen Innovationen und dem fast flächendeckenden Einzug von agilen Entwicklungsmethoden erweist es sich als fraglich, ob die unterschiedlichen Änderungsgeschwindigkeiten heute noch in dieser Form gegeben sind. Vielmehr scheint es oftmals so, dass die hohe technische Entwicklungsdynamik als Enabler Druck auf die fachlichen Strukturen und Prozesse ausübt. Als Indikator hierfür mag die oft mühselige Suche nach passenden Business Cases gelten, wenn neue, beispielsweise unstrukturierte Datenbestände in den Unternehmen verfügbar sind.

Der vorliegende Sammelband konzentriert sich mit den zugehörigen Beiträgen auf die Integrations- und die Softwareebene. Hierfür sollen unterschiedliche Architekturkonzepte vorgestellt und diskutiert werden, wie sie sich in einzelnen BIA-Ökosystemen präsentieren.

Bevor jedoch auf die konkreten Ausgestaltungen von BIA-Ökosystemen eingegangen wird, sind zunächst die Rahmenbedingungen für eine geeignete Architektur zu beleuchten, die sich sowohl aus strategischen Überlegungen zum Umgang mit Daten als wichtige Ressource als auch aus den konkreten Anforderungen der Stakeholder ergeben.

### 1.3 Datenbezogene Rahmenbedingungen

Ein Rahmen für die Verarbeitung von Daten in einem Unternehmen besteht im Wesentlichen aus folgenden Teilbereichen: Strategie, Management, Funktion, Prozess und Technologie. Der vorliegende Abschnitt skizziert und positioniert diese Teilbereiche, indem die Handlungsfelder beschrieben und die Wechselwirkungen untereinander aufgezeigt werden. Die folgende Abbildung 1–2 ordnet die Handlungsfelder den entsprechenden Themengebieten zu.

Die strategische Ebene definiert, wie sich Daten im Sinne des Geschäftsmodells nutzen lassen und eine geeignete Datenstrategie (vgl. Abschnitt 1.3.1) abgeleitet werden kann. Die Managementebene schafft ein geeignetes Rahmenwerk zum Umgang mit diesen Daten mittels der Funktionen des Datenmanagements (vgl. Abschnitt 1.3.3). Hierbei werden die Daten zuvor im Rahmen der Data Valuation bewertet und entsprechend ihrer strategischen, wirtschaftlichen und regulatorischen Bedeutung priorisiert und kategorisiert (vgl. Abschnitt 1.3.2). Die Funktionen des Datenmanagements unterstützen oder ermöglichen den Prozess der datenbasierenden Wertschöpfung, die von den Geschäftsfunktionen ausgeführt wird. Datenstrategie und Data Governance regulieren, steuern, verwalten und überwachen die Funktionen des Datenmanagements. Die Management- und Funktionsebenen werden hierbei auf den gesamten Lebenszyklus von Daten angewandt. Die Verwaltung des Lebenszyklus von Daten und die Wertschöpfung auf deren Basis finden im Habitat der Architekturen und ihrer Systemkomponenten statt.

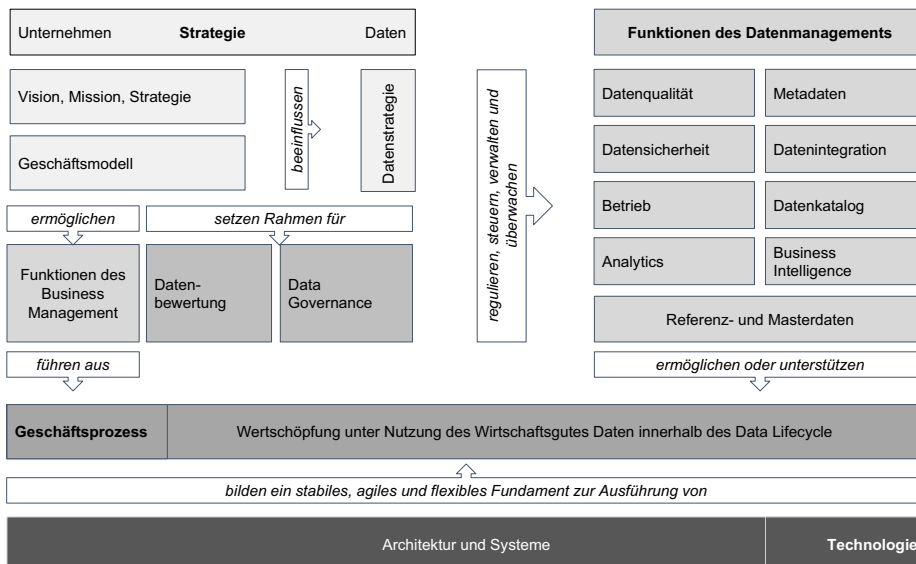


Abb. 1–2 Daten-Ökosystem in einem Unternehmen mit Ebenen und Handlungsfeldern

Dementsprechend fundamental sind die Architekturen und damit die Gesamtheit der Systeme zu gestalten, um eine dauerhafte und nachhaltige Basis zur Wertschöpfung aus Daten als Wirtschaftsgut und deren Verwaltung über den gesamten Lebenszyklus zu gewährleisten. Dabei sind vor allem Stabilität, Agilität und Flexibilität der Architekturen und Systeme sicherzustellen. Nachfolgend wird die Herleitung der Architekturanforderungen, von der Strategieebene ausgehend, beschrieben und in Abschnitt 1.4 als Anforderungen an eine BIA-Architektur zusammengefasst.

1.3.1 Datenstrategie

Allgemein wird auf der strategischen Ebene von der Unternehmensleitung (bzw. von den verantwortlichen Entscheidungsträgern) festgelegt, wie Daten im Sinne der Unternehmung einzusetzen sind. Im Rahmen der Strategiefindung muss die Denkweise über die Bedeutung von Daten an die jeweiligen spezifischen Bedingungen angepasst werden. Zahlreiche Unternehmen setzen Daten nach wie vor ausschließlich zur Unterstützung und Verbesserung bestehender Prozesse im Sinne von Messen und Verwalten ein [Rogers 2017]. Allerdings erfolgt – in immer mehr Organisationen und vor allem in den letzten Jahren – ein Umdenken in Bezug auf die Bedeutung und den Umgang mit Daten, wie in Tabelle 1–1 gegenübergestellt ist.

Früher	Heute
Die Datengenerierung innerhalb eines Unternehmens ist teuer	Daten werden ständig und überall generiert
Das Speichern und Verwalten von Daten stellt eine Herausforderung dar	Die Herausforderung besteht darin, Daten in wertvolle Informationen zu transformieren
Unternehmen nutzen nur strukturierte Daten	Unstrukturierte und semi-strukturierte Daten sind zunehmend nutzbar und stellen einen großen Wert dar
Daten werden in operativen Silos verwaltet	Wert generieren Daten insbesondere durch übergreifende Verbindungen
Daten sind ein Mittel zur Verbesserung von Prozessen	Daten sind ein immaterieller Vermögenswert und dienen damit der Wertschöpfung

Tab. 1–1 Anpassung der strategischen Denkweise [Rogers 2017, S. 139]

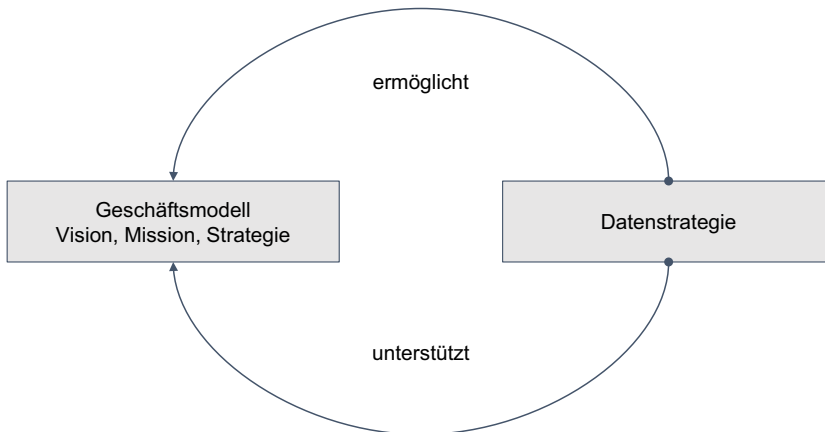
Zusammenfassend beinhaltet Tabelle 1–1 folgende Kernaussagen:

- Die Generierung von Daten erfolgt heute ubiquitär von Menschen und Maschinen.
- Die Herausforderung besteht nicht mehr in der Beschaffung von Daten, sondern in der Informationsgewinnung.
- Durch Einsatz fortschrittlicher Technologien und Methoden lassen sich aus Informationen Werte generieren.

Somit stellt sich bei der Strategieentwicklung die Aufgabe, Daten als wesentliche Schlüsselressource und damit wichtiges immaterielles Wirtschaftsgut für die eigene Organisation zu verstehen und zu behandeln, indem ein geeigneter Rahmen zu deren Bewirtschaftung definiert wird (vgl. hier auch die Datenstrategie der UN, abrufbar unter <https://www.un.org/en/content/datastrategy/index.shtml>). Dabei ist nicht zuletzt die zentrale Frage zu beantworten, welche Daten für das jeweilige Geschäftsmodell von Bedeutung sind oder sein könnten. Als exemplarische Einsatzgebiete von Daten, die bei der Definition einer Datenstrategie Bedeutung erlangen können, lassen sich anführen:

- Sammlung heterogener Datenarten für unterschiedlichste Zwecke
- Nutzung von Daten zur Prognose im Rahmen der Entscheidungsfindung
- Nutzung von Daten zur Entwicklung von Produktinnovationen
- Beobachtung des Verhaltens von Kunden
- Kombination von Daten aus diversen Bereichen bzw. Domänen

Aus einer strategischen Perspektive können Daten sowohl eine Supporter-Rolle (Unterstützer) als auch eine Enabler-Rolle (Ermöglicher) einnehmen (vgl. Abb. 1–3).



**Abb. 1–3** Zusammenhang zwischen Datenstrategie und Geschäftsmodell

Die eingenommene Rolle wird durch den jeweiligen Datenstrategieansatz bestimmt, wobei sich hier defensive von offensiven Ausprägungen abgrenzen lassen. Der defensive Teil verfolgt das Ziel, nachteilige Datenrisiken zu minimieren, und widmet Themen wie Datenschutz, Datenintegrität, Identifizierung, Standardisierung sowie dem operativen Verwalten der Daten besondere Aufmerksamkeit. Im BIA-Kontext wird als Ziel die Bereitstellung einer »Single Source of Truth« bzw. eines »Single Point of Truth« verfolgt. Als Treiber für diese strategische Ausrichtung fungieren u.a. allgemeine Anforderungen an den Betrieb der Lösungen neben regulatorischen Vorgaben, was im Ergebnis zu ausgeprägter Stabilität führt.

Dagegen verfolgt der offensive Ansatz das Ziel, die Wirtschaftlichkeit zu steigern, Performance-Verbesserungen zu erzielen und die Kundenzufriedenheit zu erhöhen. Auf Basis der offensiven Strategie werden »Multiple Versions of Truth« erzeugt. Der offensive Ansatz eröffnet Chancen und erreicht dies durch Anreicherung der Daten und Verwendung analytischer Anwendungen. Insgesamt wird die Agilität und somit auch die Resilienz bzw. Widerstandskraft des Unternehmens verbessert [DalleMule & Davenport 2017].

Bezogen auf die spezifische Situation einer Organisation sind beide Strategien zu beachten und in Betracht zu ziehen, um unter den jeweiligen Rahmenbedingungen (wie Branche und Geschäftsmodell) ein ausgewogenes Verhältnis zwischen Offensive und Defensive im eigenen Haus zu etablieren. Als Basisannahme muss der Datenstrategie die Einsicht zugrunde liegen, dass nur auf Basis einer defensiven Stabilität eine funktionierende Offensive erfolgreich sein kann.

Bei all dem gilt zu beachten, dass Daten als immaterielle Wirtschaftsgüter über andere Eigenschaften als materielle Wirtschaftsgüter verfügen und somit ein individuelles Umfeld zur Bewirtschaftung geschaffen werden muss. Tabelle 1–2 zeigt einige zentrale Unterscheidungskriterien auf.

Materielle Wirtschaftsgüter	Daten als Wirtschaftsgüter
Spezifische Distribution	Einfache Distribution (Internet etc.)
Einfache Ermittlung des Wertes – Bewertung anhand von Marktpreisen möglich	Komplexe und problematische Wertermittlung
Kosten einfach zu ermitteln	Kosten schwer zu ermitteln
Preisbildung bekannt	Preisbildung nahezu unbekannt
Individueller Besitz; Identifikation und Schutz leicht herstellbar	Vielfacher Besitz möglich; Identifikation und Schutz aufwendig
Hohe Vervielfältigungskosten	Geringe Vervielfältigungskosten
Gebrauch verursacht Wertverlust	Gebrauch generiert Wertgewinn durch Teilung

**Tab. 1–2** Gegenüberstellung von materiellen Gütern und Daten als Wirtschaftsgüter

Bekannte Methoden und Technologien zum Asset Management können dementsprechend nicht vollumfänglich – insbesondere im Hinblick auf die Ermittlung des Wertes von Daten – herangezogen werden. Der folgende Abschnitt widmet sich unter der Begrifflichkeit Data Valuation speziell dem Thema Wertermittlung von Daten.

1.3.2 Data Valuation

Die Standards zur Verwaltung von physischen Assets sind in ISO 55001 geregelt. Basierend auf diesem Standard entwickelte The Institute of Asset Management

(IAM) einen Leitfaden zur Bewertung von Wirtschaftsgütern, der als Voraussetzung ein Verständnis über Kosten und Risiken zu einem Vermögensgegenstand über den gesamten Lebenszyklus anführt [Fleckenstein & Fellows 2018, S. 15 ff.].

Als Wirtschaftsgut besitzen Daten einen Wert, der von den Unternehmen zwar erkannt wird, sich allerdings nur schwer bestimmen und quantifizieren lässt. Im International Accounting Standard (IAS) 38 sind immaterielle Wirtschaftsgüter definiert als identifizierbare nicht monetäre Vermögenswerte ohne physische Substanz. Diese Definition trifft auch auf Daten zu, dennoch fließen sie aber bislang nicht als Wirtschaftsgüter in die Bilanzen ein [Treder 2019, S. 43].

Eine Bewertung von Daten kann aus verschiedenen Blickrichtungen und in Abhängigkeit von ihrer spezifischen Rolle in einem Unternehmen durchgeführt werden. Eine gebräuchliche Einteilung unterscheidet zwischen den Bewertungskategorien Kosten, Nutzen und Marktwert. Signifikanten Einfluss auf den Wert von Daten übt die jeweilige Datenqualität aus [Krotova & Spiekermann 2020].

**Kosten** entstehen entlang des gesamten Data Lifecycle und weisen einen direkten Zusammenhang mit der Bewirtschaftung von Daten auf (z.B. Infrastruktur, Softwarelizenzen, Personal usw.). Eine naheliegende Option besteht darin, die Summe aller angefallenen Kosten für die Erzeugung bzw. Beschaffung und die Pflege der Daten für die Bewertung heranzuziehen. Falls Daten reproduzierbar oder ersetzbar sind, lassen sich alternativ die entsprechenden Reproduktionskosten oder Kosten für einen Datenersatz zugrunde legen [Rea & Sutton 2019, S. 6].

Bezüglich der Kosten seien an dieser Stelle auch Kosten für vermeintlich suboptimale Architekturansätze erwähnt. Lässt eine Architektur z.B. Datensilos zu, dann können Opportunitätskosten entstehen und – im ungünstigsten Fall – sogar die durch Daten gewonnenen Werte zerstören. Als mögliche Folgen von Datensilos lassen sich anführen [Treder 2019, S. 50]:

- Unterschiedliche Antworten auf die gleichen Fragestellungen
- Verzögerung bei der Umsetzung neuer Geschäftsmodelle
- Inkompatibilität bei der Zusammenführung von Silo-übergreifenden Daten
- Doppelte Arbeit
- Erschwerte Einhaltung regulatorischer Anforderungen

Der **Nutzwert** von Daten erweist sich als ungleich schwerer bestimmbar als die zugehörigen Kosten und lässt sich nicht immer exakt quantifizieren. Vor allem wenn neben dem tatsächlich generierten Nutzen (»finanzieller Nutzen«) auch der potenziell mögliche Nutzen (»finanzielle Chance«) erhoben werden soll [Glazer 1993], präsentiert sich die Erhebung als große Herausforderung und lässt sich nur zusammen mit Domänenexperten sowie mit erheblichem Aufwand näherungsweise ermitteln [Krotova & Spiekermann 2020]. Tatsächlich generierter Nutzen erwächst aus dem identifizierten Mehrwert, der sich aus der Datennut-