

Betriebliche Informationssysteme

L9 – Data Model, Data Governance, Data Responsible, Data Quality

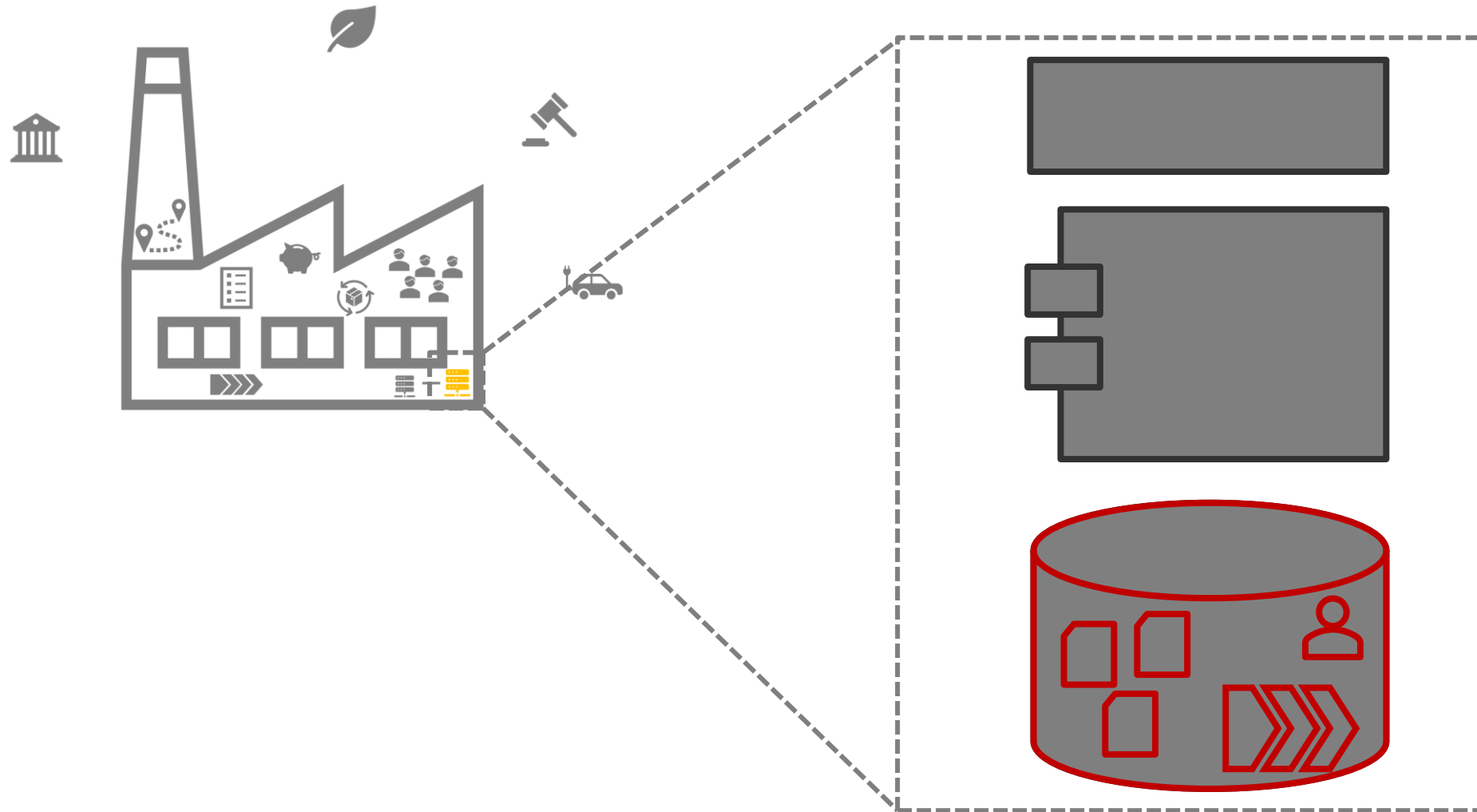


L9 – Data Model, Data Governance, Data Responsible, Data Quality

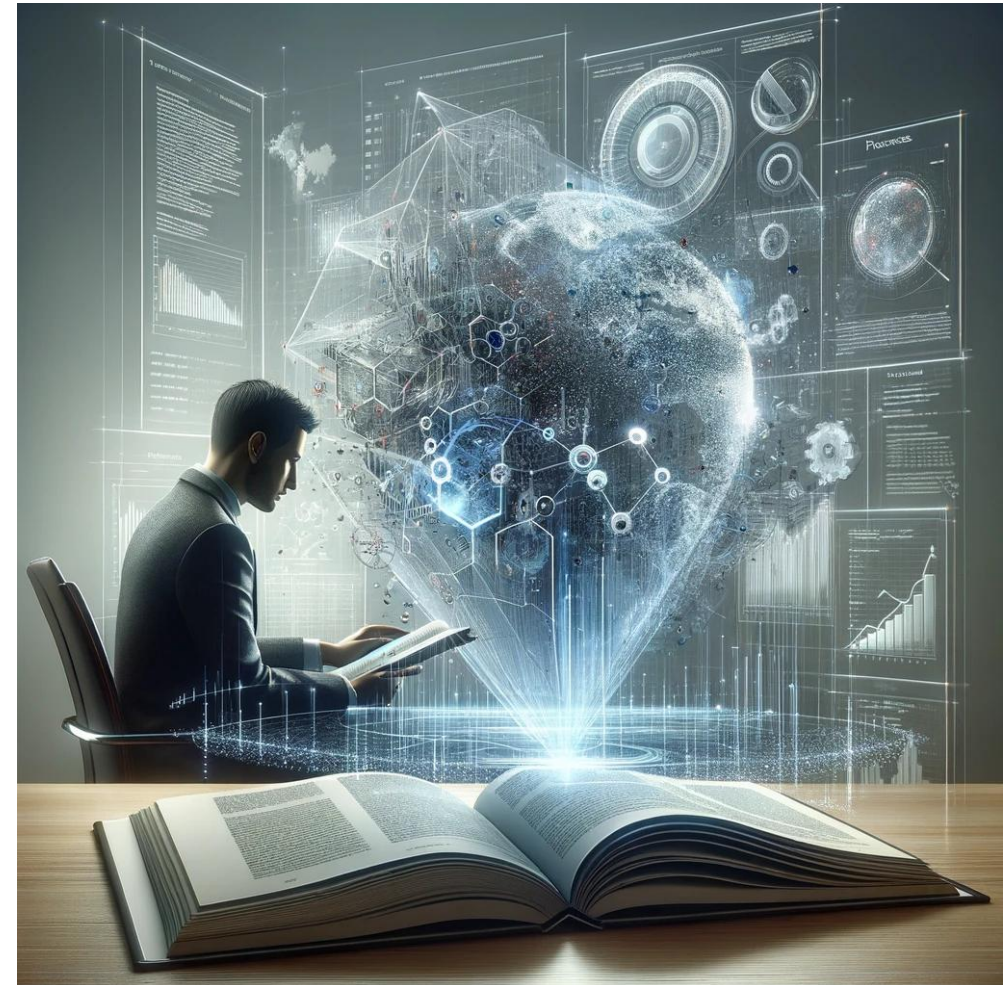
„Wie kann man Daten organisatorisch managen (Governance)?“

- Den Nutzen von Data Governance kennen
- Verstehen, wie ein Data Model mit Data Rollen, Data Prozessen und Weisungen & Standards zusammenhängen
- Die Zusammenhänge von Daten Governance und Datenqualität, sowie mit Herausforderungen mit Daten generell, verstehen
- *Weiterführende Literatur*
 - Gluchowski, Peter, Data Governance, 2020.
[Zugriff O'Reilly](#)

Big Picture



Data Governance



Quelle: OpenAI. (2024). ChatGPT (Version 4), Chatbot-Output vom 26.03.2024
(Prompt: «create an image with a person looking at data. beside the person a book about processes»)

Begriff „Data Governance“

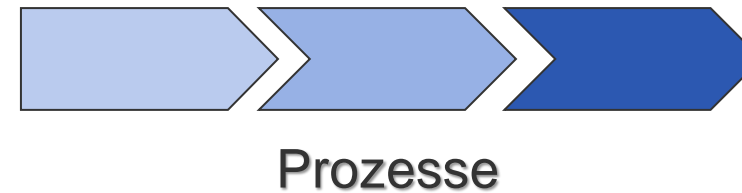
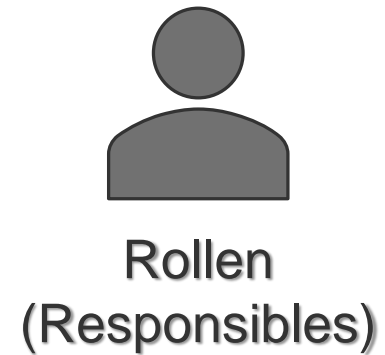
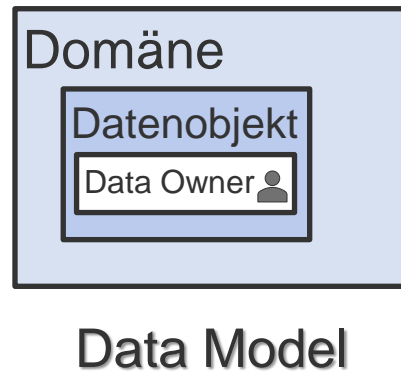
Strategischer Ansatz zur **Organisation, Steuerung und Kontrolle** der wachsenden Menge und Vielfalt an Daten. Gluchowski, Peter, Data Governance, Kapitel 1, 2020

Data Governance zielt auf ein **effektives Management von Daten** ab, da Daten von Unternehmen **als Vermögenswerte** betrachtet werden sollten. Khatri, V.; Brown, C. V.: Designing Data Governance. Communications of the ACM, 53 (1), 2010, S. 148-152;/ Ladley, J.: Data governance. How to design, deploy, and sustain an effective data governance program. Morgan Kaufmann, Amsterdam u.a., 2012

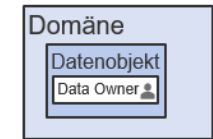
Die Einführung einer Data Governance ist **kein primär technisches Projekt**, sondern vielmehr ein **Programm zur Transformation** in eine digitale Organisation, in der Daten als vollwertiger Vermögenswert genutzt werden. Gluchowski, Peter, Data Governance, Kapitel 2, 2020

In Abgrenzung zur IT-Governance, in deren Verantwortungsbereich lediglich die IT-Systemlandschaft und Programme fallen [Khatri & Brown 2010], widmet sich die Data Governance den (digitalen) Daten und Informationen. Weill, P.; Ross, J. W.: IT Governance – How Top Performers Manage IT Decision Rights for Superior Results. Harvard Business Review Press, Brighton, 2004.

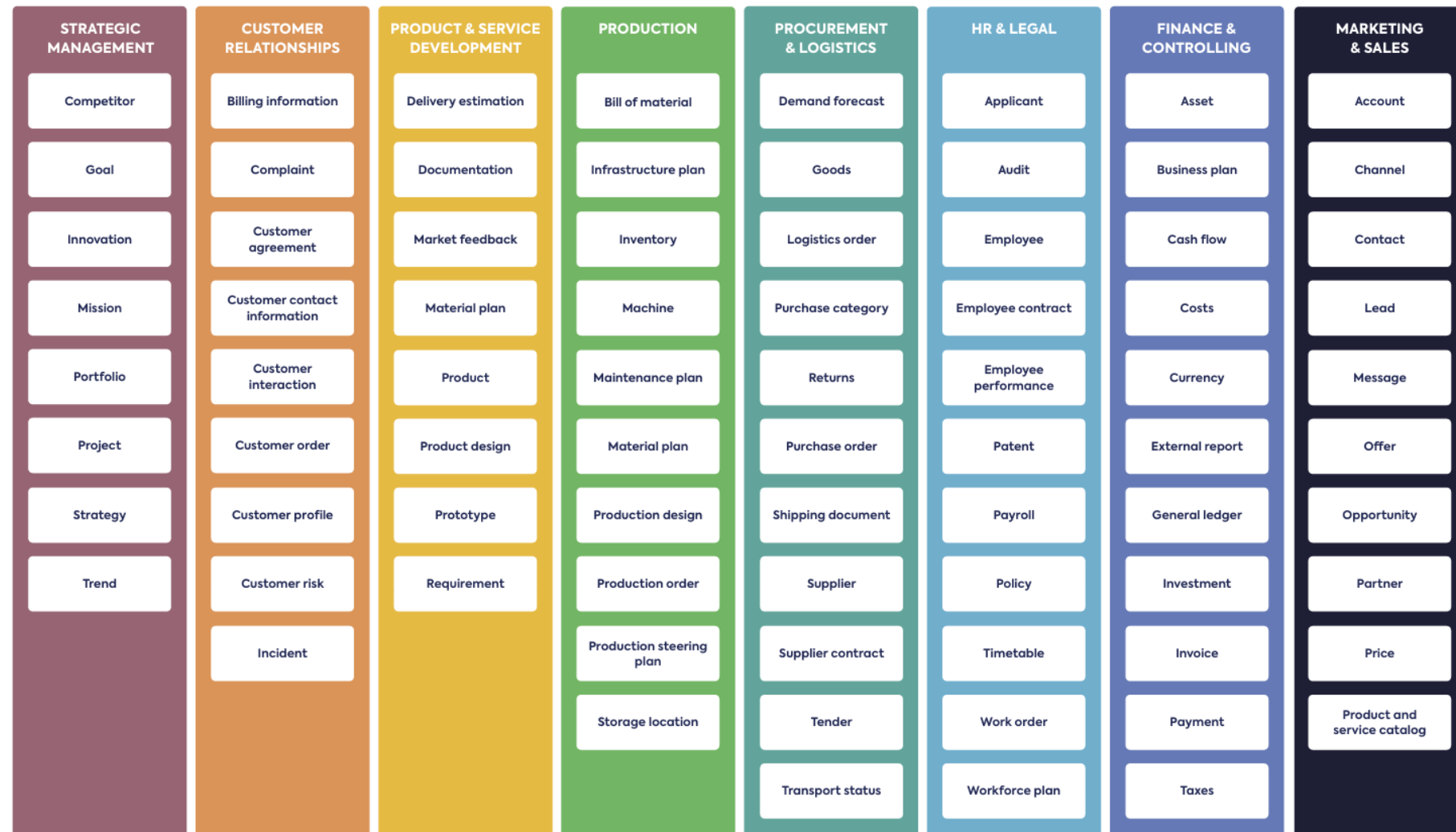
Wichtigste Elemente einer Data Governance



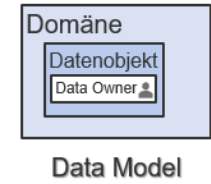
Data Model



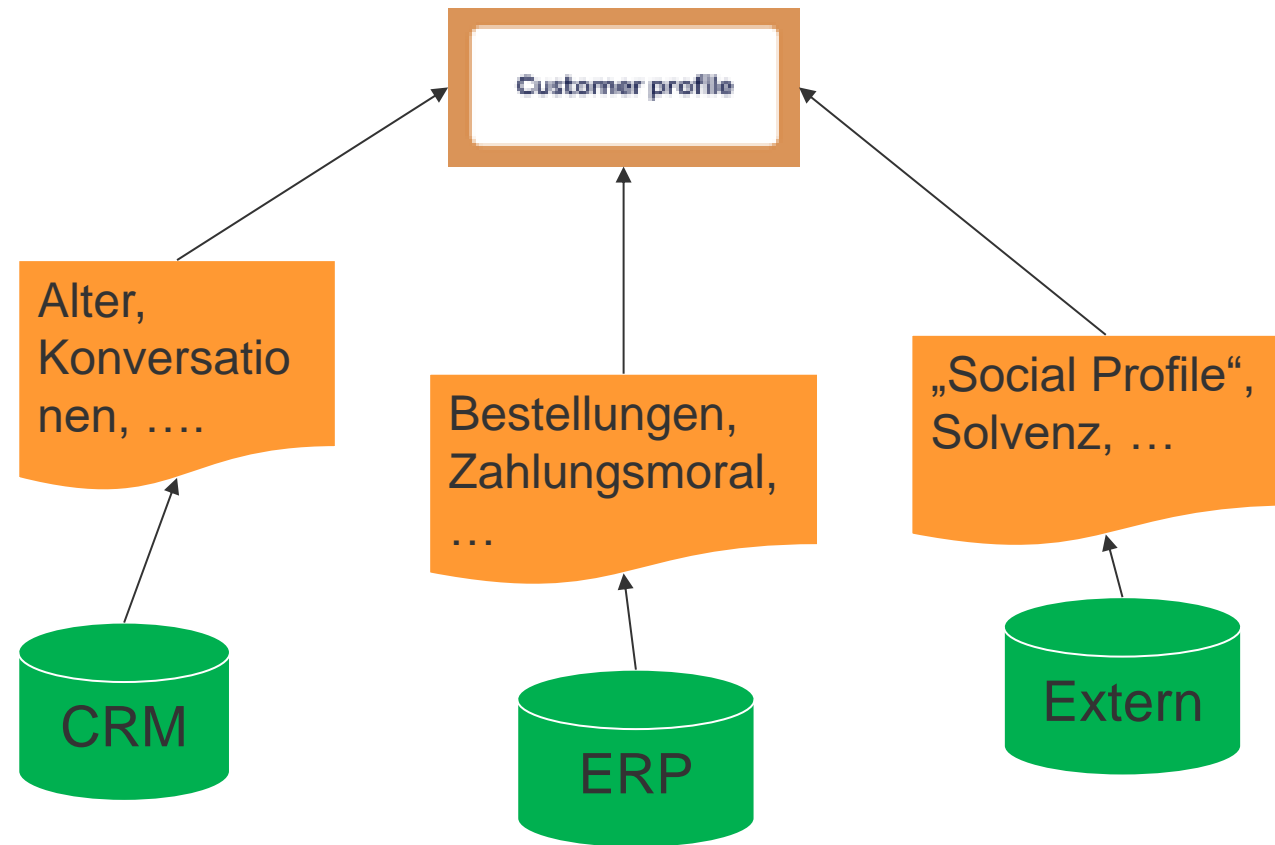
Data Model



Quelle: <https://www.leanix.net/en/wiki/it-architecture/enterprise-data-model#form>, Abgerufen 14.04.2024



Am Beispiel „Customer Profile“



Legende

Daten

Quellsystem

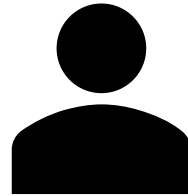


Rollen

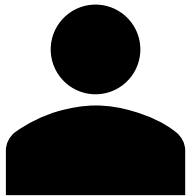
- **Rollen, welche für die Data Governance benötigt werden, festlegen.**
- Dabei die **Entscheidungsbefugnisse, Verantwortlichkeiten** sowie **Zuständigkeiten definieren.**
- Je nach Unternehmensgrösse zur Unterstützung ein **Data Governance Office** etablieren



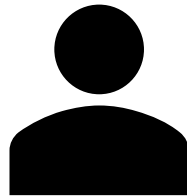
Mögliche typische Rollen



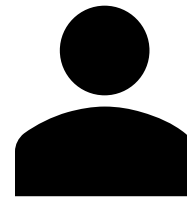
Chief Data Officer (CDO)



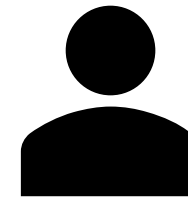
Technical Data Owner



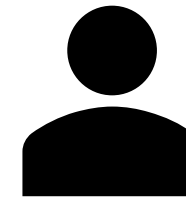
Fachlicher Data Owner



Data Engineer



Data Scientist



Data Analyst



Mögliche typische Rollen - Beschreibung

Chief Data Officer (CDO)

- Verantwortlich für die Entwicklung und Umsetzung der Datenstrategie des Unternehmens.
- Führungskraft, die die datenbezogenen Aktivitäten des Unternehmens übergreifend koordiniert.
- Fördert die datengesteuerte Entscheidungsfindung in allen Bereichen des Unternehmens.

Fachlicher Data Owner

- Verantwortet die Datenobjekte in seinem Bereich
- Versteht die Daten, kann diese interpretieren und ist für deren Datenqualität zuständig

Technischer Data Owner

- Unterstützt bei der Speicherung, Abfrage und Verwaltung der Datenbestände
- Datenbankadministrator (DBA), Data Warehouse Engineer, Data Lake Engineer, Cloud Data Engineer, ...

Data Engineer

- Entwickelt und verwaltet die Dateninfrastruktur des Unternehmens.
- Stellt sicher, dass Daten korrekt und effizient gespeichert, verarbeitet und abgerufen werden können.
- Arbeitet mit Data Scientists zusammen, um die Datenzugänglichkeit und -integration zu gewährleisten.

Data Scientist

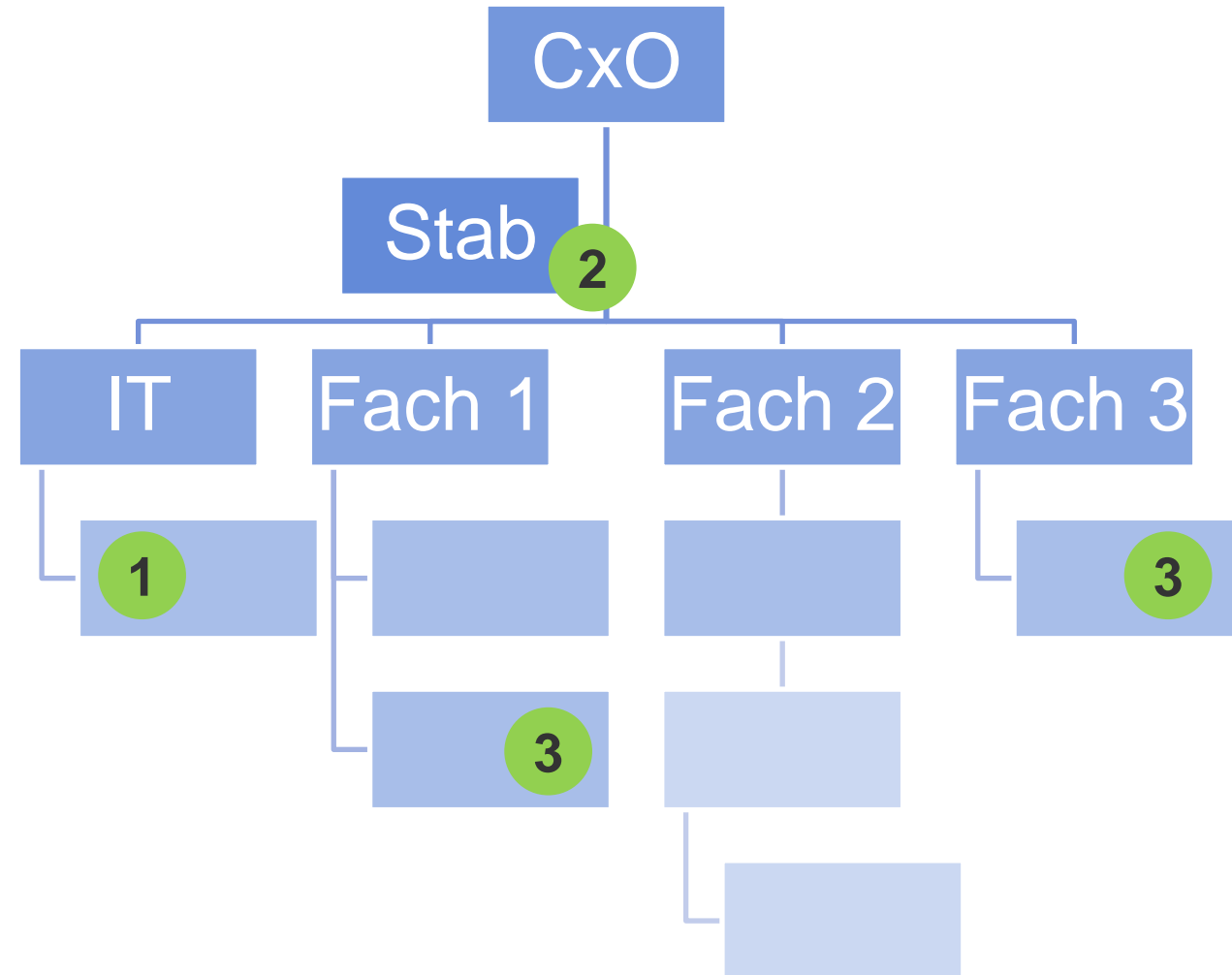
- Wendet Data-Mining-, Machine-Learning- und Statistiktechniken an, um wertvolle Erkenntnisse aus Daten zu gewinnen.
- Entwickelt Modelle und Algorithmen zur Analyse von Daten und zur Vorhersage von zukünftigen Trends.
- Arbeitet mit anderen Teams zusammen, um die Ergebnisse der Datenanalyse in die Praxis umzusetzen.

Data Analyst

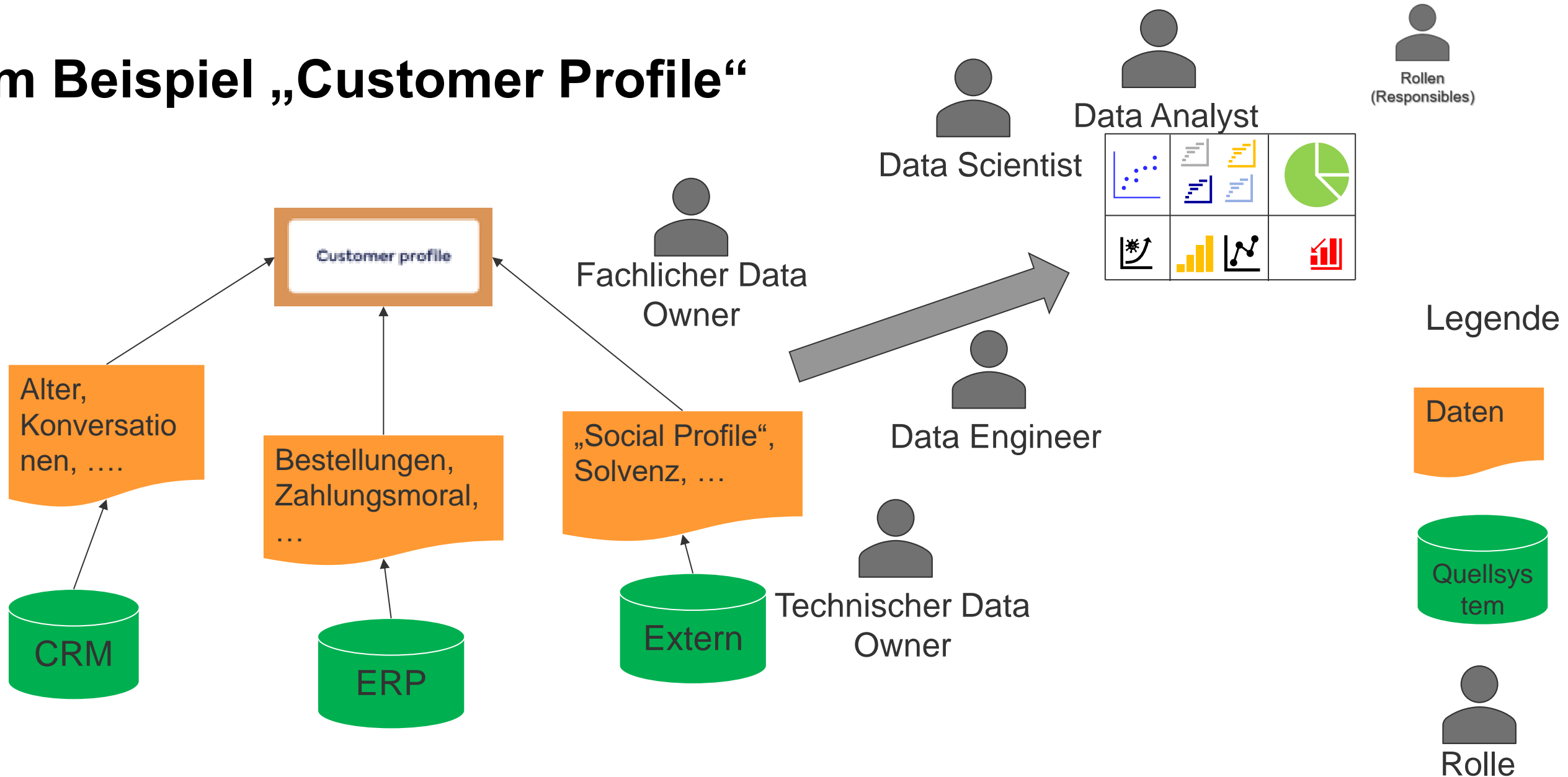
- Sammelt, bereinigt und analysiert Daten aus verschiedenen Quellen.
- Erstellt Berichte und Visualisierungen, um die Ergebnisse der Datenanalyse zu kommunizieren.
- Unterstützt andere Teams bei der Nutzung von Daten zur Verbesserung ihrer Entscheidungsfindung.



Eingliederung in der Organisation



Am Beispiel „Customer Profile“





Weisungen & Standards

- **Richtlinien und Standards, abgeleitet von der Strategie** (warum und mit welcher Zielvorstellung sich die Organisation mit dem Thema auseinandersetzt)
- **Standards für Bewertung und Qualität von Datenbeständen**
- **Weisung zur Einhaltung interner und externer Regeln und Richtlinien:**
 - CH-Datenschutzgesetz (DSG) bzw. Datenschutz-Grundverordnung (DSGVO)
 - Branchenspezifische Anforderungen, bspw. Bankaufsichtliche Anforderungen (FinMA)
- **Weisungen bezüglich Sicherheit:**
 - Definition von Sicherheitsregeln im Zusammenhang mit den genutzten Daten
 - Festlegung von Zugriffsrechten (Berechtigungskonzept)
 - Vorgehensweisen bei Sicherheitsverstößen



Quelle: OpenAI. (2024). ChatGPT (Version 4o), Chatbot-Output vom 07.11.2024 (Prompt: «Person neben einem riesigen Stapel Papier.»)

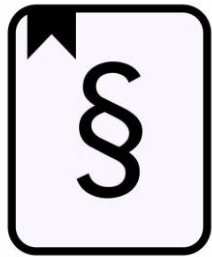


Prozesse

- ✓ Verbindliche Verfahren zu den Bereichen **Datensicherheit, Datenmodellierung oder interner sowie externer Datenaustausch.**
- ✓ Stichprobenartige Überprüfung der **Datenqualität**, u.a. mittels **Data Cleansing**: Identifizierung, Korrektur oder Entfernung von fehlerhaften, inkonsistenten oder unvollständigen Daten
- ✓ **Überwachung und Steuerung** interner und externer Datenverarbeitungsprozesse.
 - ✓ Data Lineage zeigt, verfolgt und analysiert, wie Daten über alle Anwendungen und Datenquellen verteilt sind. Zu wissen, woher die Daten stammen, ist nicht nur für Compliance-Berichte, sondern auch für den Aufbau vertrauenswürdiger und erklärbarer KI-Modelle relevant.
 - ✓ Ergänzung: Evidenzerhebung bezüglich des Teilens und der Abfrage von Daten. Ziel ist die Herstellung der notwendigen Transparenz darüber, wer wann welche Daten nutzt.
- ✓ Kontinuierliche Überprüfung & Aktualisierung des Data Model.

Data Governance unterstützt uns folgendes zu erreichen...

*Umsetzung von Gesetzen
und Regulatorien*



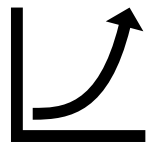
*Zeitnahe und exakte Aussagen zu
Dienstleistungen, Produkten, Kunden, ...
→ Entscheidungsfindung*



*Basis einer guten Datenqualität &
Vermeidung von
Datensilos*



Exponentiell wachsende Datenvolumen

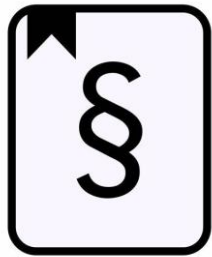


*Definition von Rollen
mit Zuständigkeit*



... und folgendes zu verhindern:

*Umsetzung von Gesetzen
und Regulatorien*



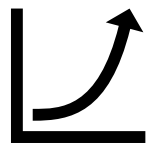
*Zeitnahe und exakte Aussagen zu
Dienstleistungen, Produkten, Kunden, ...
→ Entscheidungsfindung*



*Basis einer guten Datenqualität &
Vermeidung von
Datensilos*



Exponentiell wachsende Datenvolumen

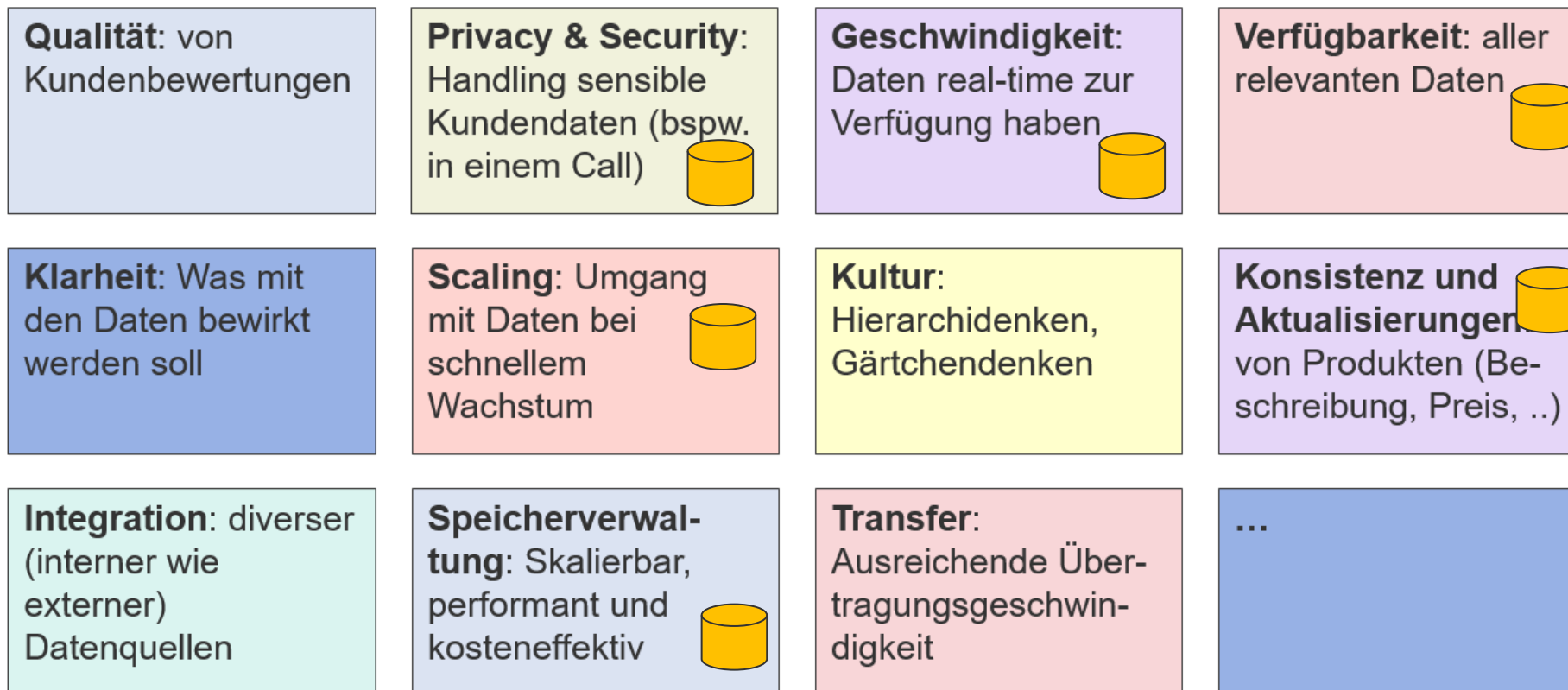


*Definition von Rollen
mit Zuständigkeit*



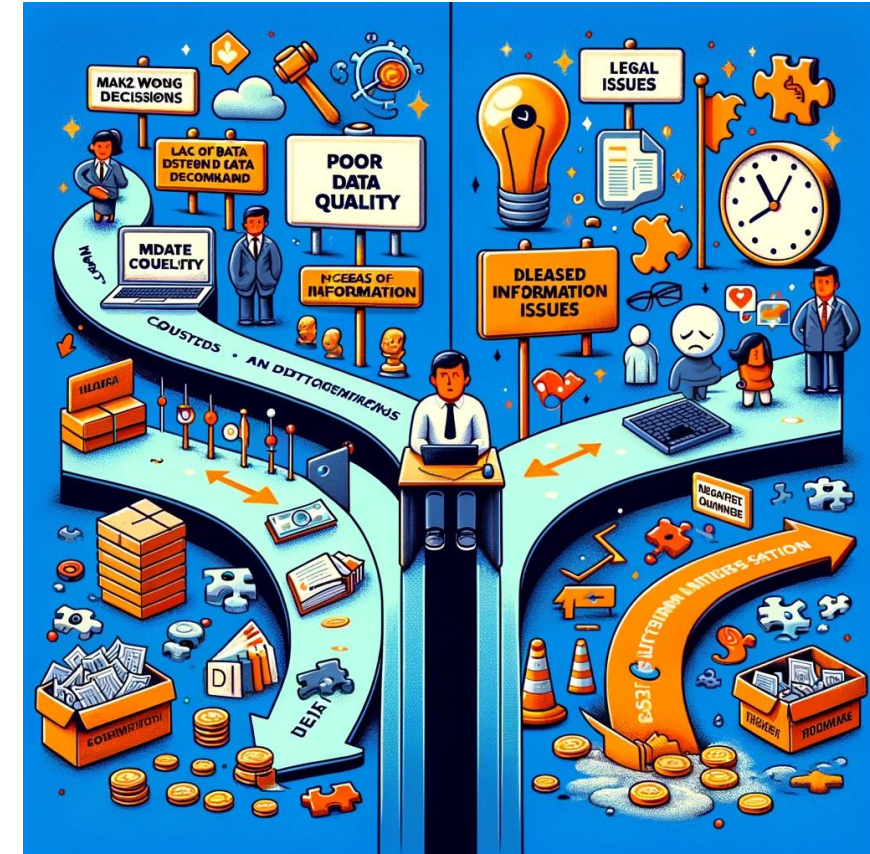


Wo kann Data Governance bei den Herausforderungen mit Daten unterstützen?



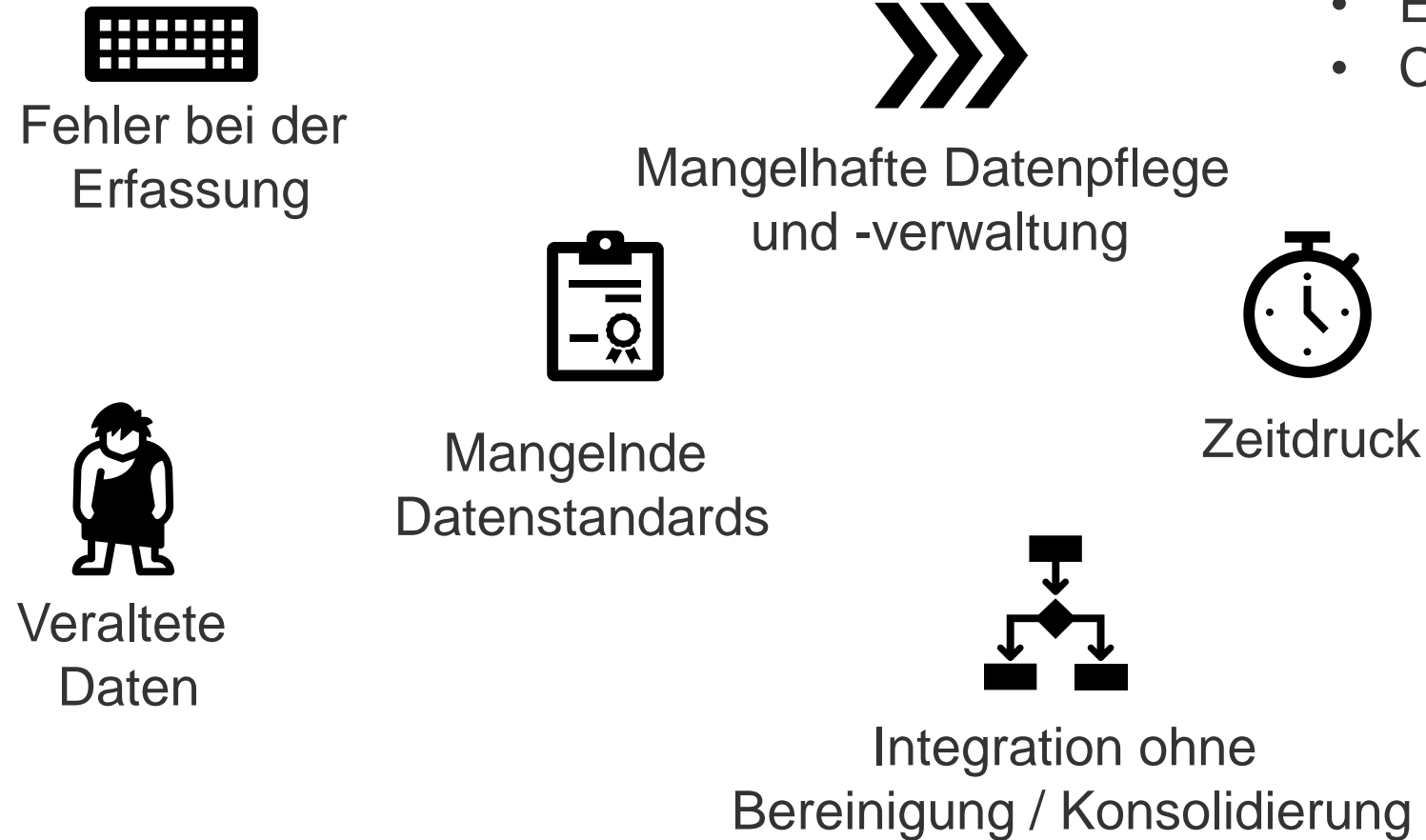
Datenqualität

Qualität: von
Kundenbewertungen



Quelle: OpenAI. (2024). ChatGPT (Version 4), Chatbot-Output vom 26.03.2024 (Prompt: «image for Ursache & Resultat schlechter Datenqualität»)

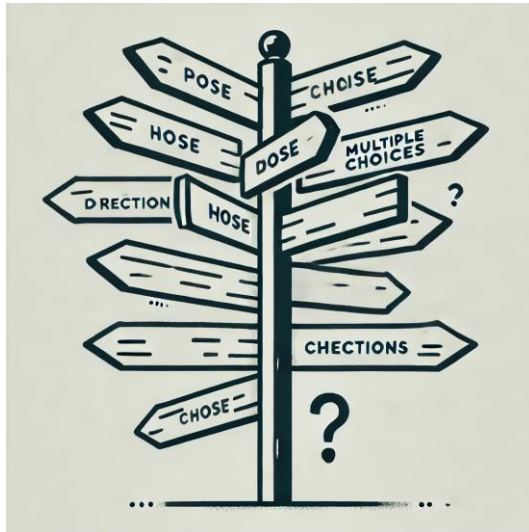
Ursachen mangelnder Datenqualität



Die Auswirkungen sind unterschiedlich

- Elektronische Gesundheitsakte vs.
- Cumulus Sammelprogramm bei der Migros

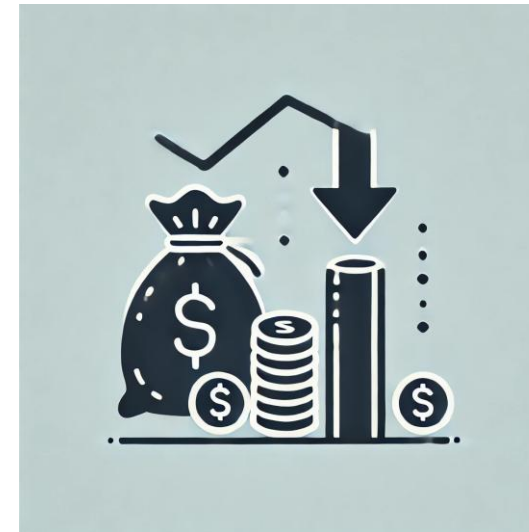
Resultate mangelnder Datenqualität



Quelle: OpenAI. (2024). ChatGPT (Version 4o), Chatbot-Output vom 07.11.2024 (Prompt: «Bild mit den Wegweisern, die in verschiedene Richtungen zeigen.»)



Quelle: OpenAI. (2024). ChatGPT (Version 4o), Chatbot-Output vom 07.11.2024 (Prompt: «Bild eines unzufriedenen Kunden mit Krone.»)



Quelle: OpenAI. (2024). ChatGPT (Version 4o), Chatbot-Output vom 07.11.2024 (Prompt: «Grafik, die erhöhte Kosten mit einem negativen, nach unten zeigenden Pfeil neben einem Geldsack oder Münzen darstellt»)



Quelle: OpenAI. (2024). ChatGPT (Version 4o), Chatbot-Output vom 07.11.2024 (Prompt: «Grafik mit einer Waage der Justitia neben einem Gerichtsgebäude, das rechtliche und regulatorische Risiken symbolisiert.»)



Datenqualitätsprobleme

Datenqualitätsprobleme	Beispiel
Duplikate	
Unvollständigkeit	
Fehlende Werte	
Falsche Formatierung	
Gültigkeit Wertebereich	
Widerspruch in den Daten	



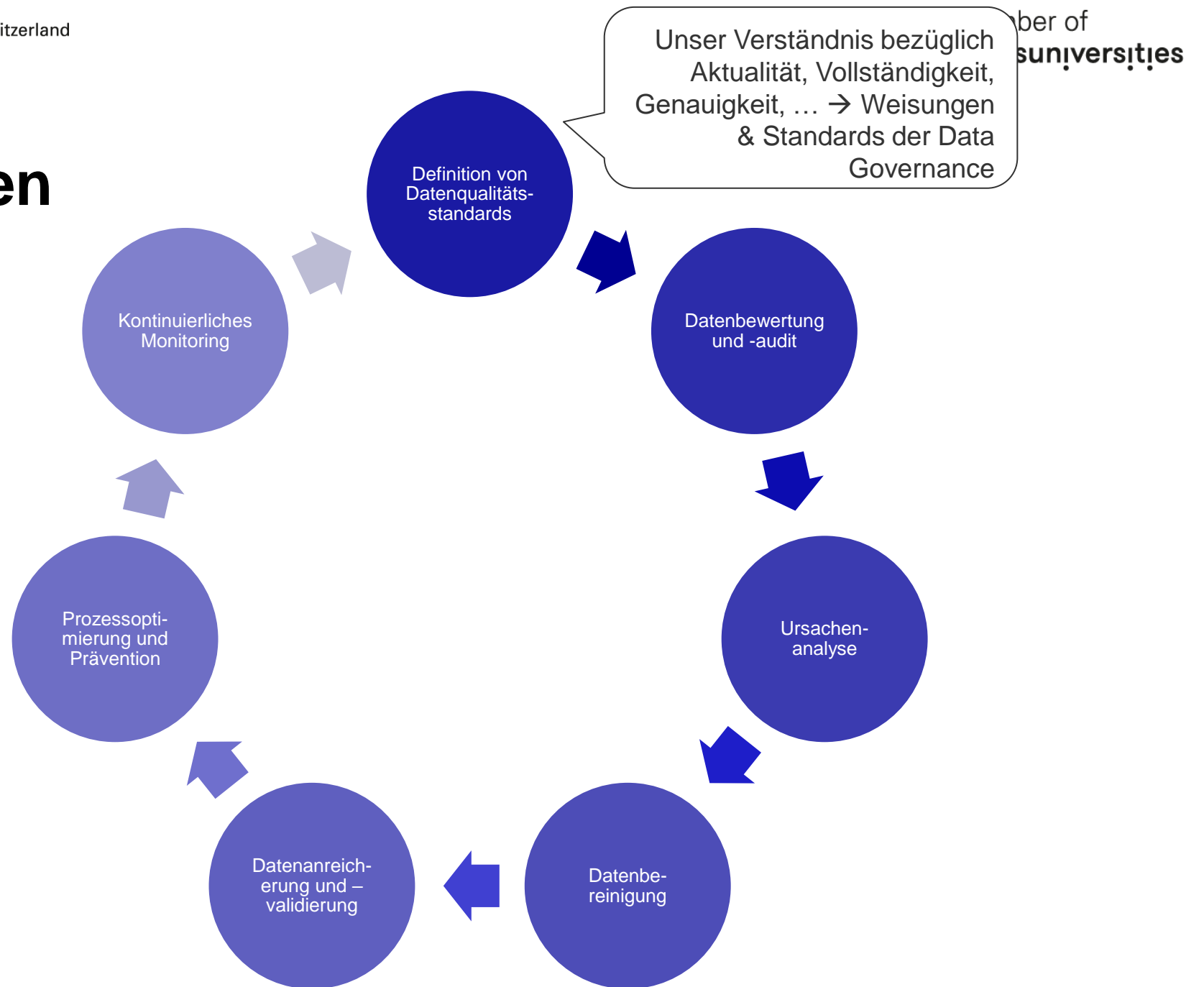
Übung: Liste mit Datenqualitätsprobleme

Überprüft den Verkaufsdatensatz eines Onlineshops (Übung Datenqualitätsprobleme.xlsx) auf folgende Datenqualitäts-Probleme:

1. **Duplikate**
2. **Unvollständigkeit**
3. **Fehlende Werte**
4. **Falsche Formatierung**
5. **Gültigkeit Wertebereich**
6. **Widerspruch in den Daten**

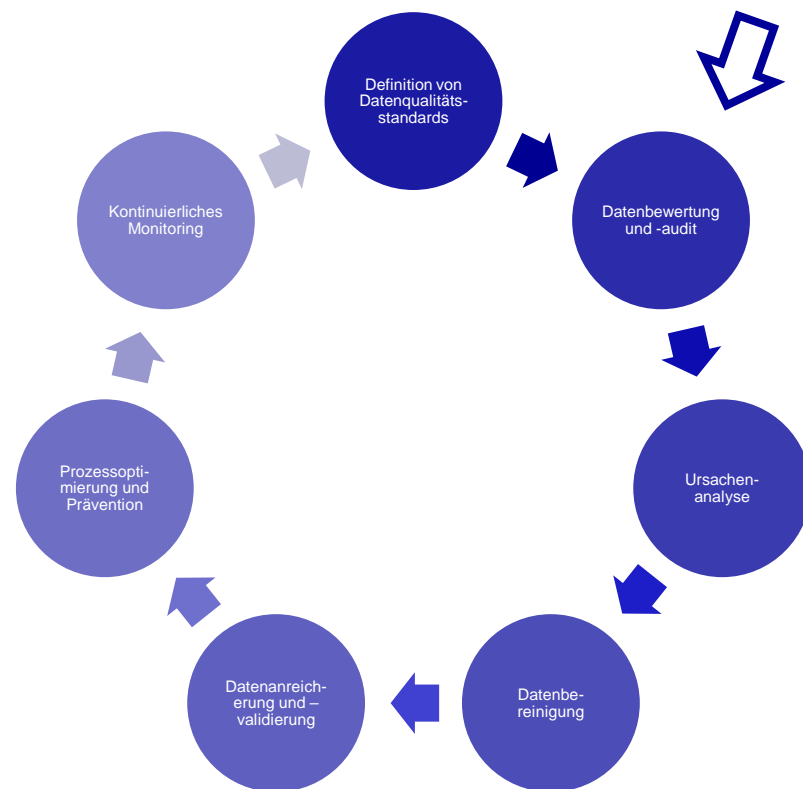
- *2-3er Gruppen*
- *Zeit: 10 Minuten*
- *Plenum: 10 Minuten*

Datenqualität: Erkennen & Beheben



Basierend auf: OpenAI. (2024). ChatGPT (Version 4), Chatbot-Output vom 26.03.2024 (Prompt: «wie können datenqualitätsprobleme erkannt und behoben werden?»)

Datenqualität: Erkennen & Beheben



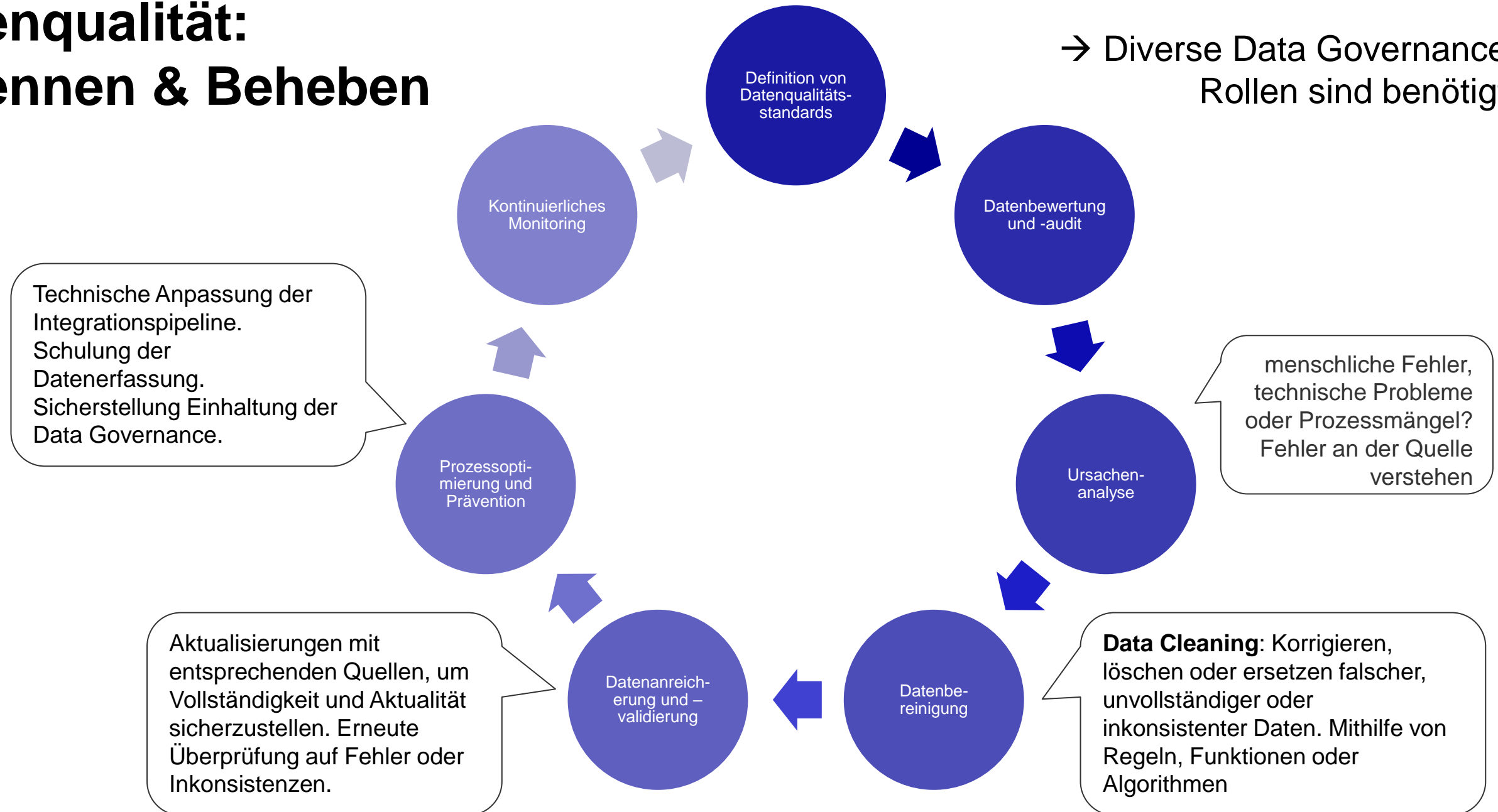
Datenbewertung und -audit durch Data Profiling: inkonsistente oder inkompatible Datentypen und -formate, gültige Werte und Bereiche, Vollständigkeit und Korrektheit

→ Know-How von Data Owner

Kunden-ID	Vorname	Nachname	Alter	Geschlecht	E-Mail	Telefon	Geburtsdatum	Letzter Besuch	Letzter Kauf	Gesamtausgaben
1001	Max	Mustermann	25	M	max@example.com	044 232 32 23	20.05.1999	15.10.2023	20.01.2024	100.5
1002	Maria	Müller		F	maria.com		30.12.1988	01.07.2022	05.12.2023	75.25
1003	Max	Mustermann	25	M	max@example.com	044 232 32 23		28.02.2024	10.11.2023	
1004	John	Doe	30	M	john@example.com	091 9292 929	17.08.1994		15.02.2024	150
1005	Eva	Schmidt	40	F		079 878 99 99	12.03.1983	20.06.2022	30.09.2023	
1006	Peter	Müller	35.5	M	peter@example.com	044 555 33 22	30.12.1988	05.07.2023	25.12.2023	80.75
1007	Lisa	Meier	28	F	lisa@example.com	023 224 22 22	28.02.1996	10.09.2023	10.03.2024	200
1008	Max	Mustermann	42	M	max@example.com	044 232 32 23	20.05.1999	01.03.2024	20.12.2023	120.5
1009	Emma	Mayer	29	F	emma@example.com	056 600 39 56	25.09.1995	15.05.2023	05.11.2023	180
1010	Hans	Schmidt	45	M	hans@example.com		10.06.1979	20.11.2023	28.02.2024	90.25
1011	Lisa	Meier	28	M	lisa@example.com		31.02.1996	10.09.2023	10.03.2024	200
1012	Sarah		35	F	sarah@example.com		20.07.1989		15.01.2024	150
1013	Chris	Peterson	45	M		568278389	05.09.1987	25.04.2023	30.10.2023	110.75
1014	Anne			F	anne@example.com	011 111 11 11		30.12.2023	20.02.2024	
1015	David	Brown	42	M	david@example.com	051 444 29 52	1975-15-30	15.08.2023	05.03.2024	250
1016	Emily	Johnson		F	emily@example.com		29.02.1980	20.10.2023	31.12.2023	180.25
1017	Mike		28	M	mike@example.com	073 292 92 00	10.12.1996		10.02.2024	300
1018	Lucy	Williams	33	F	lucy@example.com	081 923 292 9	25.04.1991	02.07.2023	22.12.2023	140.75
1019		Jackson	35	M	jackson@example.com	081 480 49 22	20.09.1989	10.11.2023		200
1020		Johnson	50	M			05.06.1974	10.04.2023	05.10.2023	350

Datenqualität: Erkennen & Beheben

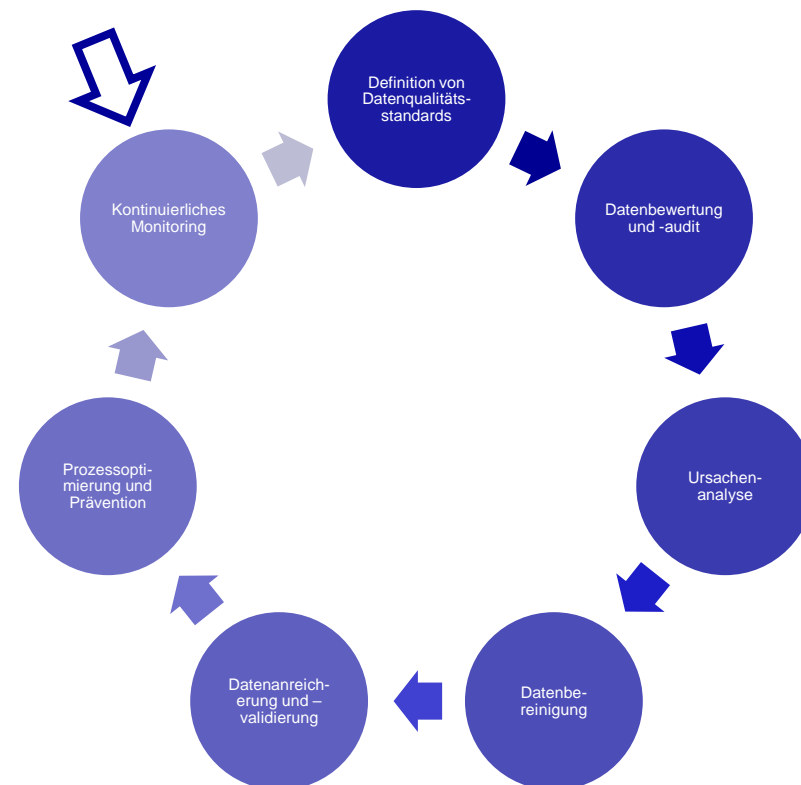
→ Diverse Data Governance
Rollen sind benötigt



Datenqualität: Datenbewertung und -audit

Kontinuierliches Monitoring: Qualität
kontinuierlich verfolgen, messen und bewerten

→ Know-How von Data Owner



Kunden-ID	Vorname	Nachname	Alter	Geschlecht	E-Mail	Telefon	Geburtsdatum	Letzter Besuch	Letzter Kauf	Gesamtausgaben
1001	Max	Mustermann	25	M	max@example.com	044 232 32 23	20.05.1999	15.10.2023	20.01.2024	100.5
1002	Maria	Müller		F	maria.com		30.12.1988	01.07.2022	05.12.2023	75.25
1003	Max	Mustermann	25	M	max@example.com	044 232 32 23		28.02.2024	10.11.2023	
1004	John	Doe	30	M	john@example.com	091 9292 929	17.08.1994		15.02.2024	150
1005	Eva	Schmidt	40	F		079 878 99 99	12.03.1983	20.06.2022	30.09.2023	
1006	Peter	Müller	35.5	M	peter@example.com	044 555 33 22	30.12.1988	05.07.2023	25.12.2023	80.75
1007	Lisa	Meier	28	F	lisa@example.com	023 224 22 22	28.02.1996	10.09.2023	10.03.2024	200
1008	Max	Mustermann	42	M	max@example.com	044 232 32 23	20.05.1999	01.03.2024	20.12.2023	120.5
1009	Emma	Mayer	29	F	emma@example.com	056 600 39 56	25.09.1995	15.05.2023	05.11.2023	180
1010	Hans	Schmidt	45	M	hans@example.com		10.06.1979	20.11.2023	28.02.2024	90.25
1011	Lisa	Meier	28	M	lisa@example.com		31.02.1996	10.09.2023	10.03.2024	200
1012	Sarah		35	F	sarah@example.com		20.07.1989		15.01.2024	150
1013	Chris	Peterson	45	M		568278389	05.09.1987	25.04.2023	30.10.2023	110.75
1014	Anne			F	anne@example.com	011 111 11 11		30.12.2023	20.02.2024	
1015	David	Brown	42	M	david@example.com	051 444 29 52	1975-15-30	15.08.2023	05.03.2024	250
1016	Emily	Johnson		F	emily@example.com		29.02.1980	20.10.2023	31.12.2023	180.25
1017	Mike		28	M	mike@example.com	073 292 92 00	10.12.1996		10.02.2024	300
1018	Lucy	Williams	33	F	lucy@example.com	081 923 292 9	25.04.1991	02.07.2023	22.12.2023	140.75
1019		Jackson	35	M	jackson@example.com	081 480 49 22	20.09.1989	10.11.2023		200
1020		Johnson	50	M			05.06.1974	10.04.2023	05.10.2023	350

Data Governance hat Impact auf „mein“ BIS, weil

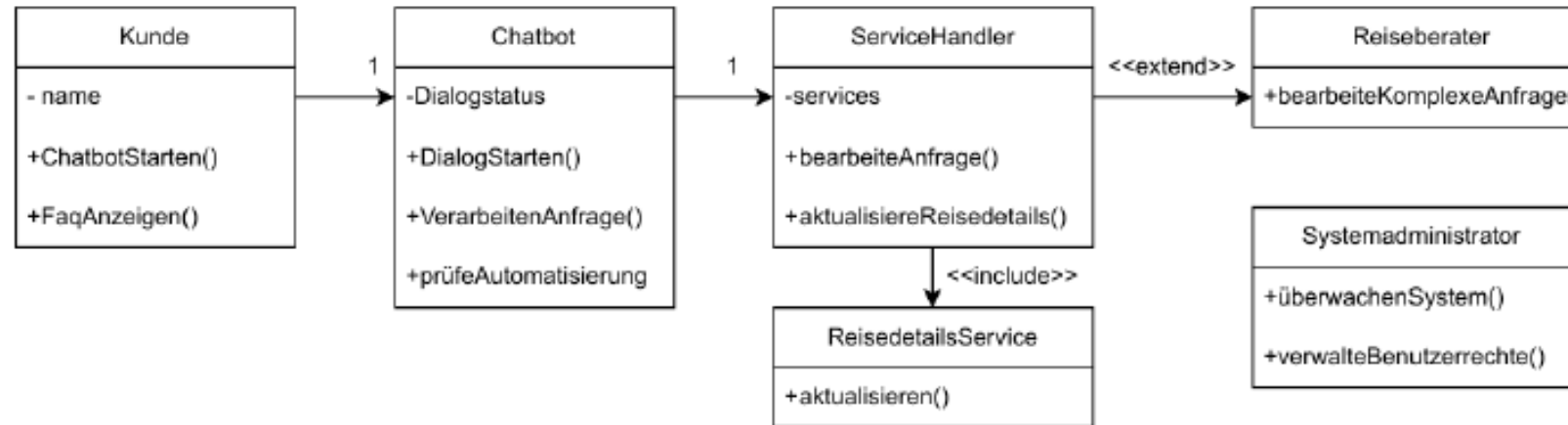
- ich die **Daten**, welche im BIS verwendet werden, **verstehen** muss. Dazu benötige ich **Data Owner**.
- ich wissen muss, aus **welchen Systemen, in welchem Format, über welche Schnittstelle, etc.** die Daten, welche mein BIS benötigt, bezogen werden können. Dazu wird ein Technischer Data Owner, ein Data Engineer benötigt.
- **Weisungen & Standards** auch für mein BIS gelten und diese eingehalten werden müssen.
- für eine **gute Datenqualität** im BIS eine vorhandene Data Governance einen positiven Einfluss hat.

Gruppenarbeit

Auftrag 8 & Auftrag 9



Auftrag 7: Klassenfeedback



Kunde

- Attribute: Name
- Methoden: +chatbotStarten(), +kundenanfrageStellen(), +rechnungAbrufen(), +faqAnzeigen()
- Beziehung: nutzt Chatbot (1 Kunde -> 1 Chatbot)

Das Klassendiagramm modelliert den Aufbau eines Systems zur Kundeninteraktion über einen Chatbot. Im Zentrum steht der Kunde, der den Chatbot nutzt, um Informationen wie Rechnungen, FAQs oder Anfragen abzurufen. Der Chatbot übernimmt dabei die Kommunikation und leitet Kun-

Auftrag 7: Feedback pro Gruppe

Vormittag

Gruppe 2

Gruppe 4 (kein Feedback)

Nachmittag

Gruppe 8

Gruppe 7

Gruppe 10

Gruppe 12

Gruppe 5

Gruppe 2 (kein Feedback)