



# AI-Governance by Design: Datenarchitekturen für vertrauenswürdige KI

*Wie Datenarchitektur, Semantik und  
Governance vertrauenswürdige KI  
ermöglichen*

MARKUS  
BEGEROW



#gerneperDu



**DATENBANKEN  
VERSTEHEN**

*für Anfänger und Profis*



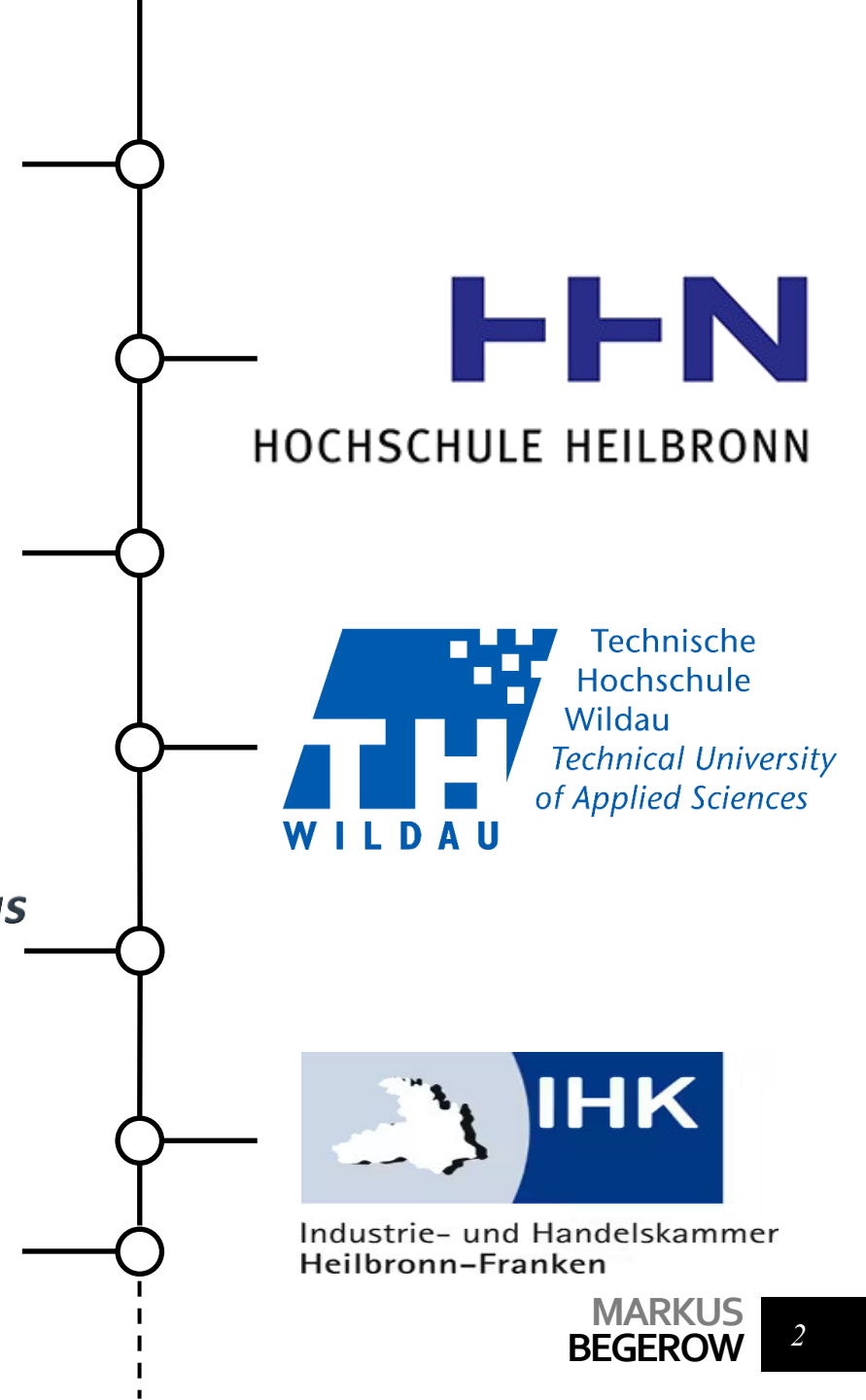
**Data & AI Campus**

*Your next level in Data & Analytics*



**wbh**

**WILHELM BÜCHNER  
HOCHSCHULE**



**MARKUS  
BEGEROW**





**SCAN MICH 😊**

# Wo stehen wir heute?

LEADERSHIP > LEADERSHIP STRATEGIES

## Why 95% Of AI Projects Fail And How Better Data Can Change That

By [Gary Drenik](#), Contributor. © Gary Drenik is a writer covering AI, analytics a...

[Follow Author](#)

Published Oct 15, 2025, 10:00am EDT



Source: [Forbes](#)

Gartner

[Who We Serve](#) [Our Solutions](#) [Conferences](#) [Webinars](#) [AI Hub](#) [New To Role](#)

[Newsroom](#) [Topics](#) [Media Contacts](#) [Media Resources](#) [Insights](#) [Archive](#)

[Newsroom](#) / [Information Technology](#) / [Q&A](#)

## Lack of AI-Ready Data Puts AI Projects at Risk

### Q&A with Roxane Edjlali

STAMFORD, Conn., February 26, 2025

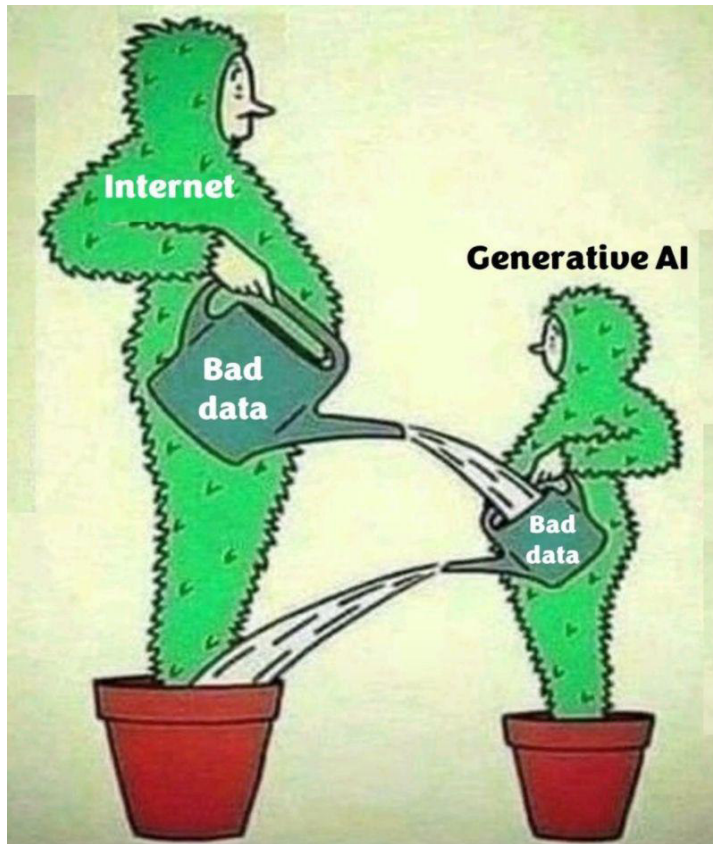
### Overview

Sixty-three percent of organizations either do not have or are unsure if they have the right **data** management practices for AI, according to a survey by Gartner. A survey of 1,203 data management leaders in July 2024 found that organizations that fail to realize the vast differences between AI-ready data requirements and traditional data management will endanger the success of their **AI efforts**.

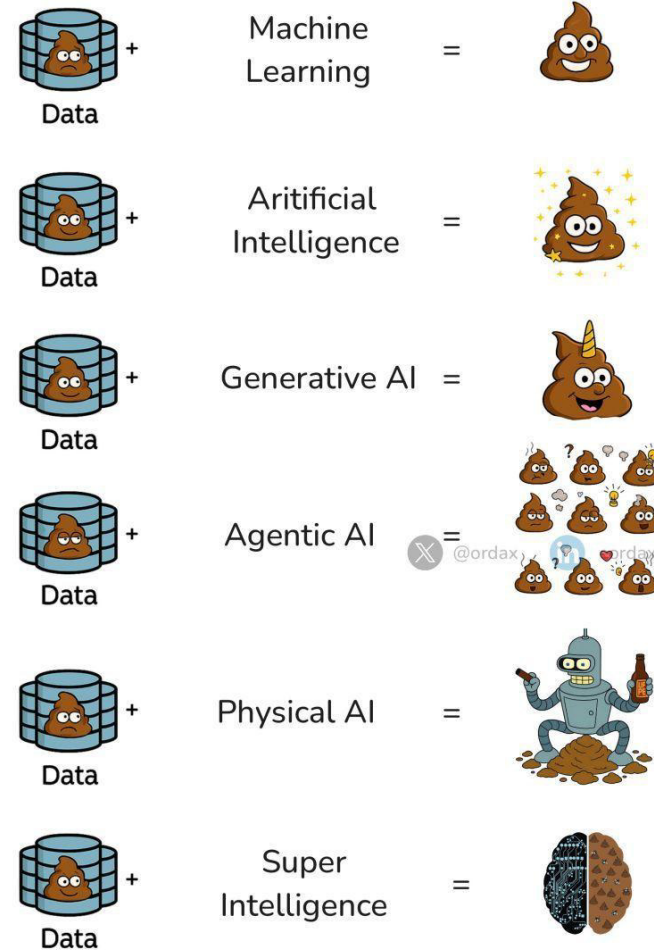
In fact, Gartner predicts that through 2026, organizations will abandon 60% of AI projects unsupported by **AI-ready data**.

Source: [Gartner](#)

# Wo stehen wir heute?

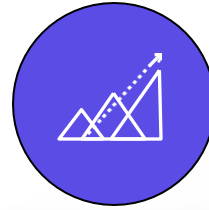


Source: [reddit](#)



Source: [LinkedIn](#)

# Aktuelle Herausforderungen in der Daten- und Analysewelt?



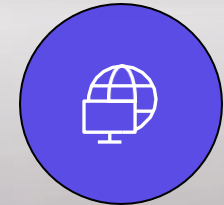
Den Bedarf an Datenzugriff und Künstlicher Intelligenz mit Governance in Einklang bringen



Wachsende Datenanforderungen treffen auf die begrenzte Skalierbarkeit klassischer Systeme



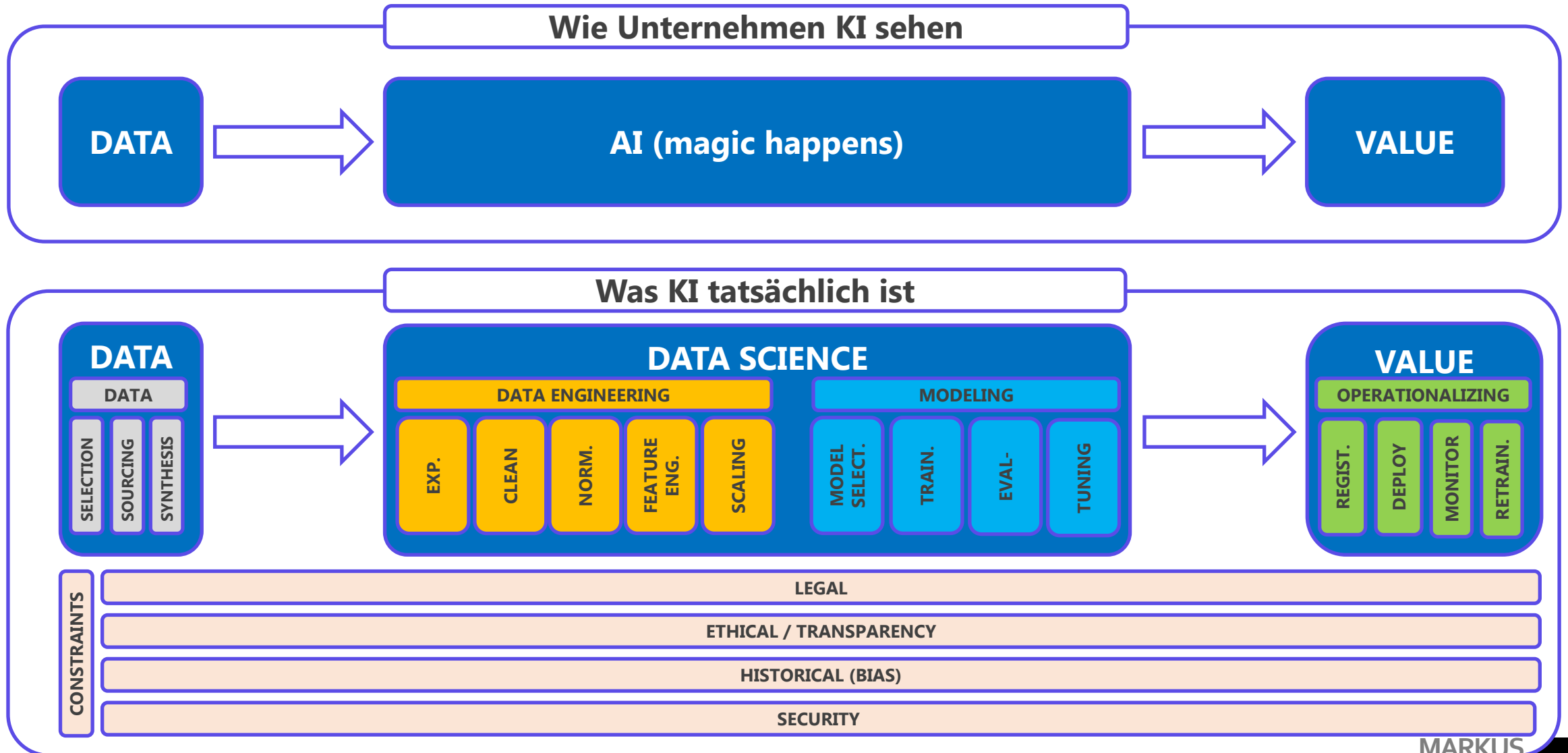
Datensilos abbauen und eine gemeinsame vertrauenswürdige Datenquelle etablieren



Das Versprechen von Künstlicher Intelligenz trotz begrenzter Ressourcen einlösen!, aber wie?



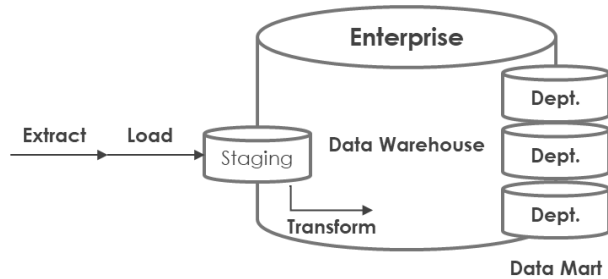
# Wahrnehmung von Künstlicher Intelligenz



# Architekturen schreiten voran...

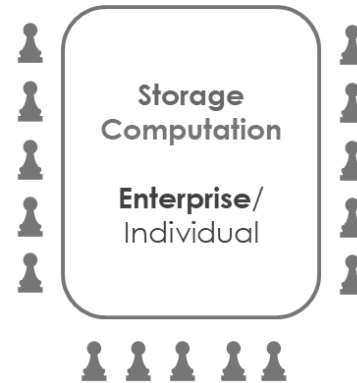
**Data Mesh?!?**

## Ende 1980er Data Warehouse



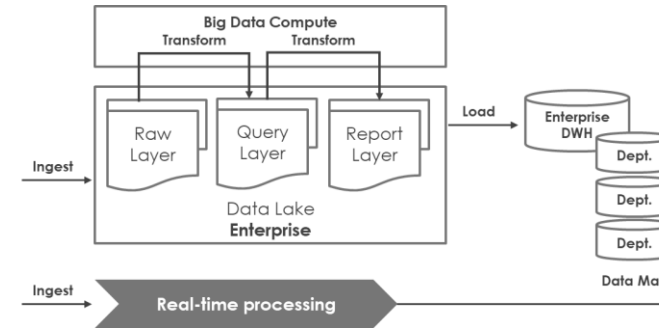
**Entlastung der Produktionssysteme**  
**Datenintegration und -qualität**  
**Historische Daten und Flexibilität**  
**Unabhängigkeit der Benutzer**  
**Erleichterte Geschäftsentscheidungen**

## Ende 2000er Data Lake



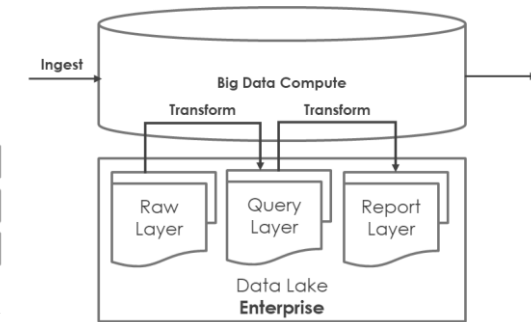
**Schema on read**  
**Entlastung teurer Ressourcen**  
**Schneller Datenzugriff**  
**Datenexploration**  
**HL für Transformationen**

## Mitte 2010 CDP



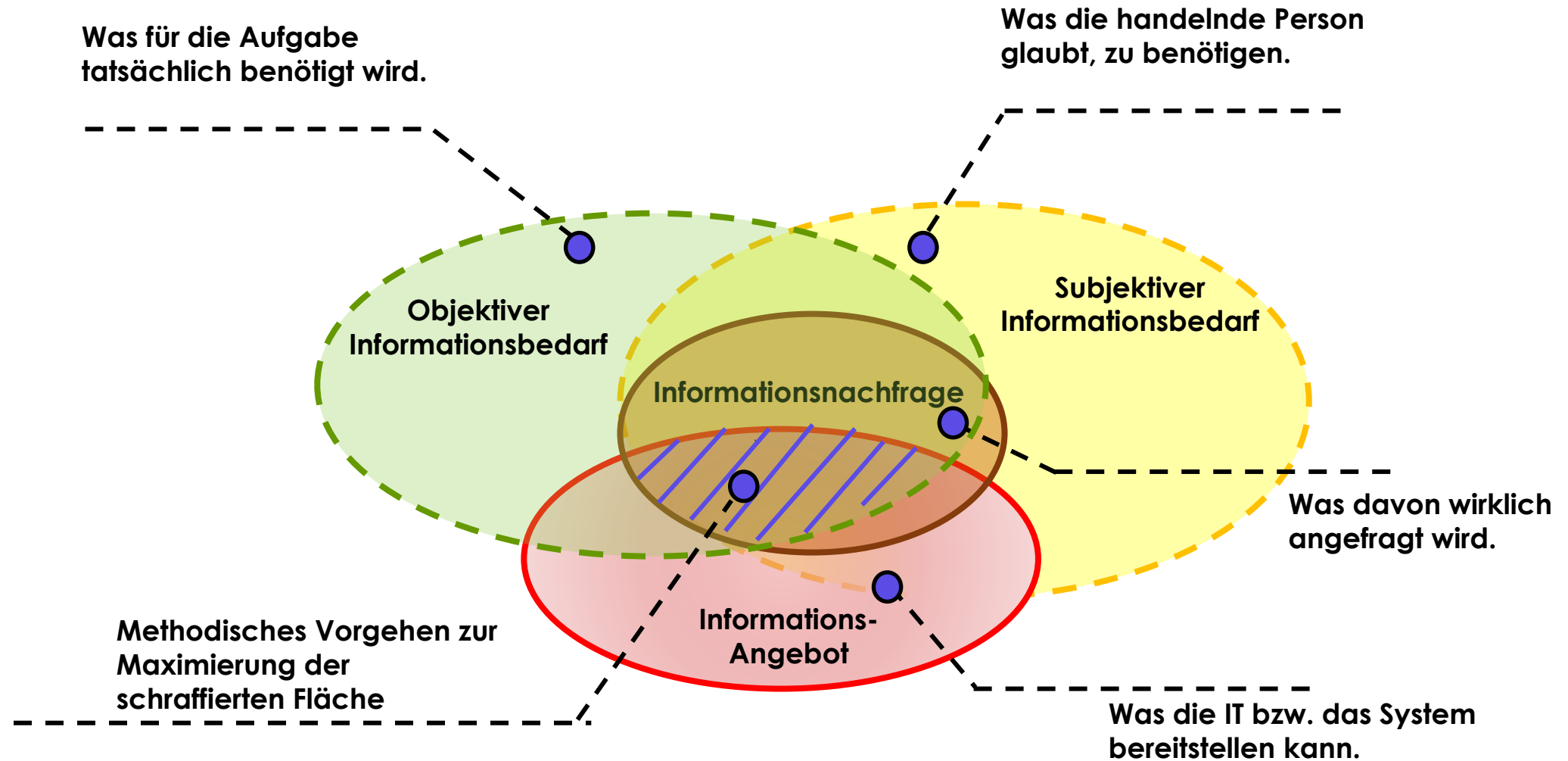
**Skalierbarkeit und Flexibilität**  
**Kostenersparnis**  
**Schneller Datenzugriff und hohe Verfügbarkeit**  
**Sicherheit und Compliance**  
**Integrierte Analyse- und KI-Dienste**

## Ab 2020 Data Lakehouse



**ACID-Transaktionen**  
**Datenversionierung**  
**Streaming und Batch**  
**Upserts and deletes**  
**Performance!!!**





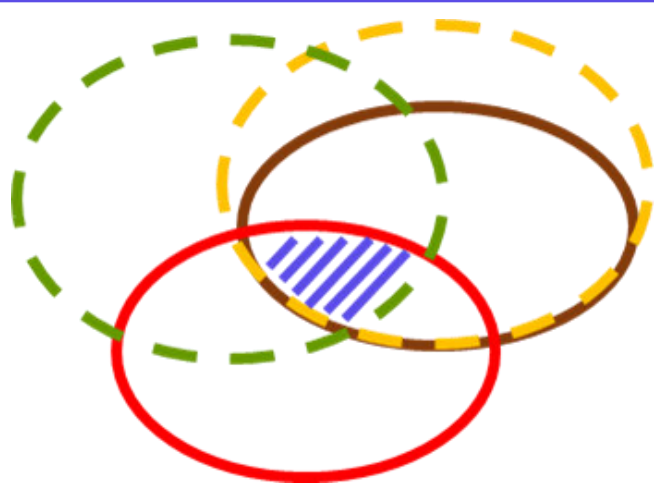
# Übersicht: Ergebnisse im Überblick

| KPI                                     | Dimension | Abteilung         | Kennzahlen Gruppe | Data Owner  | Datenherkunft (System)  | Prio | Erklärung                        | Zeit | Artikel | Händler | Vertriebskanal | Marketing-Kanal | Zahlart | Kunde | Versandkosten Cluster |
|-----------------------------------------|-----------|-------------------|-------------------|-------------|-------------------------|------|----------------------------------|------|---------|---------|----------------|-----------------|---------|-------|-----------------------|
| Anzahl Kundenbestellungen= Conversions  |           | (bestehendes) BI- | Marketing-        | Marketing   | dzt BI-System           | 1    | = Alle Kundenbestellungen        | x    | x       | x       | x              | x               | x       | x     |                       |
| Anzahl Bestellungen                     |           | (bestehendes) BI- | Marketing-        | Marketing   | dzt BI-System           | 1    | = Alle Bestellungen auf          | x    | x       | x       | x              | x               | x       | x     |                       |
| Anteil von registrierten und Gastkunden |           | (bestehendes) BI- | Marketing-        | Marketing   | dzt BI-System           | 1    | Berechnung am Wert und           | x    | x       | x       | x              | x               | x       | x     |                       |
| Anteil von B2B und B2C                  |           | (bestehendes) BI- | Marketing-        | Marketing   | dzt BI-System           | 1    | Berechnung am Wert und           | x    | x       | x       | x              | x               | x       | x     |                       |
| Customer Lifetime Value                 |           | (bestehendes) BI- | Marketing-        | Marketing   | dzt BI-System           | 1    | diskontiert auf den              | x    | x       | x       | x              | x               | x       | x     |                       |
| Provisionseinnahmen                     |           | (bestehendes) BI- | Operations        | Operations  | dzt BI-System           | 1    | Provisionseinnahmen =            | x    | x       | x       | x              | x               | x       | x     |                       |
| Deckungsbeitrag 1                       |           | (Vertriebs-       | Management        | Controlling | berechnet               | 1    | DB1 =                            | x    | x       | x       |                |                 |         |       | x                     |
| Deckungsbeitrag 2                       |           | (Vertriebs-       | Management        | Controlling | Manuelle Pflege (Excel, | 1    | DB2 = DB 1 -Kosten               | x    | x       | x       |                |                 |         |       | x                     |
| Bestellwert                             |           | (Vertriebs-       | Management        | Controlling | Commerce System         | 1    | ebenfalls Ø Bestellwert je       | x    | x       | x       |                |                 |         |       | x                     |
| Umsatz                                  |           | (Vertriebs-       | Management        | Controlling | Commerce System         | 1    | Wert aller Bestellungen,         | x    | x       | x       |                |                 |         |       | x                     |
| Auslieferumsatz                         |           | (Vertriebs-       | Management        | Controlling | Commerce System         | 1    | ebenfalls Ø                      | x    | x       | x       |                |                 |         |       | x                     |
| Bestellwert in Process                  |           | (Vertriebs-       | Management        | Controlling | Commerce System         | 1    | = Bestellwert nach Storno        | x    | x       | x       |                |                 |         |       | x                     |
| Bestellwert nach Storno                 |           | (Vertriebs-       | Management        | Controlling | Commerce System         | 1    | = Bestellwert minus Wert         | x    | x       | x       |                |                 |         |       | x                     |
| Ausgelieferte Bestellungen              |           | (Vertriebs-       | Management        | Controlling | Commerce System         | 1    | = Alle Bestellungen auf          | x    | x       | x       |                |                 |         |       | x                     |
| Kosten Google Shopping                  |           | Category          | Marketing         | Marketing   | Werbekanäle (Google     | 1    | ... =Berechnung fehlt in         | x    | x       | x       |                |                 |         |       |                       |
| SEO Ranking PDP                         |           | Category          | Marketing         | Marketing   | SEO-Tools               | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| in Performance Marketing                |           | Category          | Marketing         | Marketing   | nicht bekannt           | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Preisindex                              |           | Category          | Marketing         | Marketing   | Price Intelligence      | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Conversion Rate PDP (SKU)               |           | Category          | Marketing         | Marketing   | Google Analytics        | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Absprungrate PDP                        |           | Category          | Marketing         | Marketing   | nicht bekannt           | 1    | Absprungrate PDP                 | x    | x       | x       |                |                 |         |       |                       |
| Abbruchrate PDP                         |           | Category          | Marketing         | Marketing   | nicht bekannt           | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Bestpreis am Markt                      |           | Category          | Marketing         | Marketing   | Price Intelligence      | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Median Bestpreis am Markt               |           | Category          | Marketing         | Marketing   | Price Intelligence      | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Abweichung Bestpreis                    |           | Category          | Marketing         | Marketing   | Price Intelligence      | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Abweichung Median Bestpreis             |           | Category          | Marketing         | Marketing   | Price Intelligence      | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Preisrank                               |           | Category          | Marketing         | Marketing   | Price Intelligence      | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Preisindex                              |           | Category          | Marketing         | Marketing   | Price Intelligence      | 1    | ... =Berechnung fehlt in         | x    | x       | x       |                |                 |         |       |                       |
| Umsatz                                  |           | Category          | Controlling       | Controlling | Commerce System         | 1    | ... = Bestellwert bereinigt um   | x    | x       | x       |                |                 |         |       |                       |
| Verkäufe Toolneo 6 Monate               |           | Category          | Controlling       | Controlling | Commerce System         | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Verkäufe Toolneo 12 Monate              |           | Category          | Controlling       | Controlling | Commerce System         | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Verkäufe Toolneo gesamt                 |           | Category          | Controlling       | Controlling | Commerce System         | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |
| Bestellwert                             |           | Category          | Controlling       | Controlling | Commerce System         | 1    | ... =Berechnung fehlt in         | x    | x       | x       |                |                 |         |       |                       |
| Lieferantenerträge                      |           | Category          | Lieferant         | Category    | Manuelle Pflege (Excel, | 1    | ... =WKZ in [] Auswertung nach   | x    | x       | x       |                |                 |         |       |                       |
| Retouren                                |           | Category          | Logistik          | Operations  | Commerce System         | 1    | ... =Berechnung fehlt in [Liste, | x    | x       | x       |                |                 |         |       |                       |

# Vom Konzept zum Modell

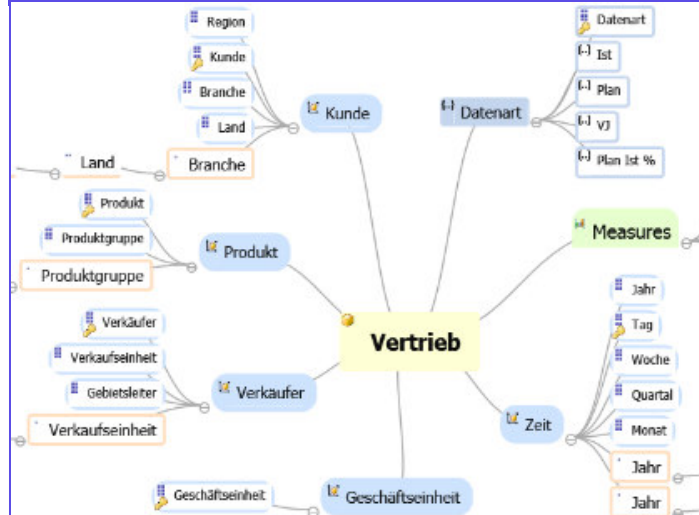
## Semantic Layer?

### Objektiver Informationsbedarf



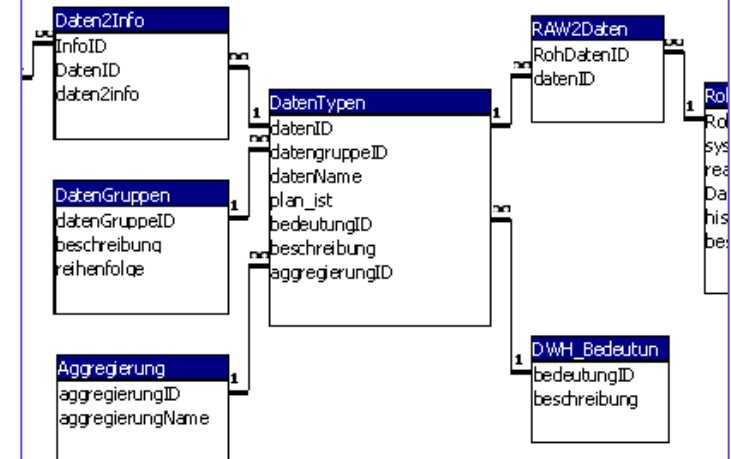
- Fragestellungen, Kennzahlen und Informationsbedarfe
- Perspektive des Fachbereichs / Managements
- Grundlage für die fachliche Struktur der Daten

### Konzeptionelles Datenmodell



- Fachliche Modellierung von Entitäten, Beziehungen und Kennzahlen
- Abstrakte Struktur ohne technische Details
- Übersetzt den Informationsbedarf in ein fachliches Datenmodell

### Technisches Datenmodell



- Physische Umsetzung im Datenbanksystem
- Definition von Tabellen, Schlüsseln und Beziehungen
- Grundlage für Implementierung und Datenverarbeitung

# Wieso brauchen wir die semantischen Zusammenhänge, es gibt doch „ChatGPT“?

---

## **Generische LLMs (z. B. ChatGPT)**

- basieren auf breitem Allgemeinwissen aus vielen Quellen
- können halluzinieren (plausible, aber falsche Aussagen)
- Quellen der Antworten sind nicht immer transparent
- Antwortqualität ist oft gut, aber nicht zwingend fachlich verlässlich für spezialisierte Domänen

## **LLMs mit domänenspezifischem Wissen**

- nutzen strukturierte Fachinformationen und semantische Modelle
- Beziehungen zwischen Fachbegriffen werden explizit abgebildet
- Anfragen können fachlich besser interpretiert werden
- Antworten werden präziser, konsistenter und kontextbezogener



# Semantik / Datenmodelle

---

- **Semantik** (altgriechisch) bedeutet „bezeichnen / ein Zeichen geben“
  - **Datenmodelle:** Daten sind eine **Sammlung von Fakten**, die nach einem Modell zusammengestellt werden.  
Das Datenmodell ist eine **Annäherung an die Realität**, gut genug für die anstehenden (analytischen) Aufgaben.  
vgl. Deghani Z.; (2022); „Data Mesh: Eine dezentrale Datenarchitektur entwerfen“
  - relationale Tabellen (SQL-Datenspeicher)
  - flache Beziehungen in verschachtelten Baumstrukturen (JSON-Strukturen)
  - Semantische Graphen
  - Wissensgraphen („Knowledge Graph“)
- = > Semantic Layer!!!

# Was genau ist der „Semantic Layer“?

---

Eine **semantische Schicht** stellt die **technische und organisatorische Grundlage** bereit, um Fachwissen strukturiert zu modellieren, zu verwalten und für KI nutzbar zu machen.

- **Regelwerk für Wissensmodellierung**

Definition, wie fachliche Beziehungen und Begriffe strukturiert beschrieben werden.

- **Abgleich mit externen Standards**

Harmonisierung eigener Regeln mit denen externer Institutionen, Normen oder Fachgremien.

- **Strukturierung von Informationen zu Wissen**

Verknüpfung von Daten und Fachbegriffen zu maschinenlesbarem Wissen.

- **Integration internen und externen Wissens**

Verbindung eigener Wissensbestände mit Wissen Dritter.

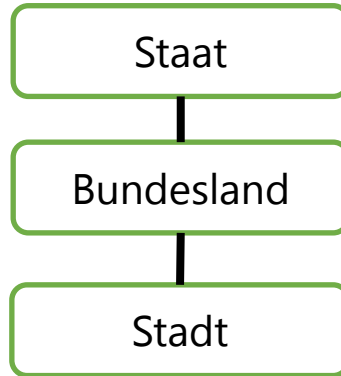
- **Automatisierte Konsistenzprüfungen**

Mechanismen zur Prüfung von Widerspruchsfreiheit und Qualität des Wissens.

- **Basis für spezialisierte KI-Systeme**

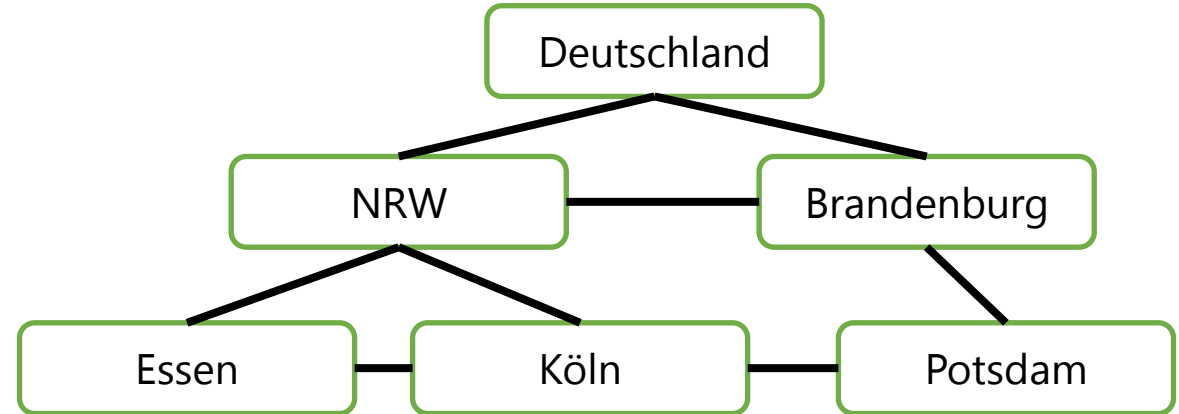
Bereitstellung hochwertiger Wissensstrukturen als Trainings- und Kontextbasis für domänenspezifische LLMs

# Semantischer Graph vs. Wissensgraph



## Semantischer Graph:

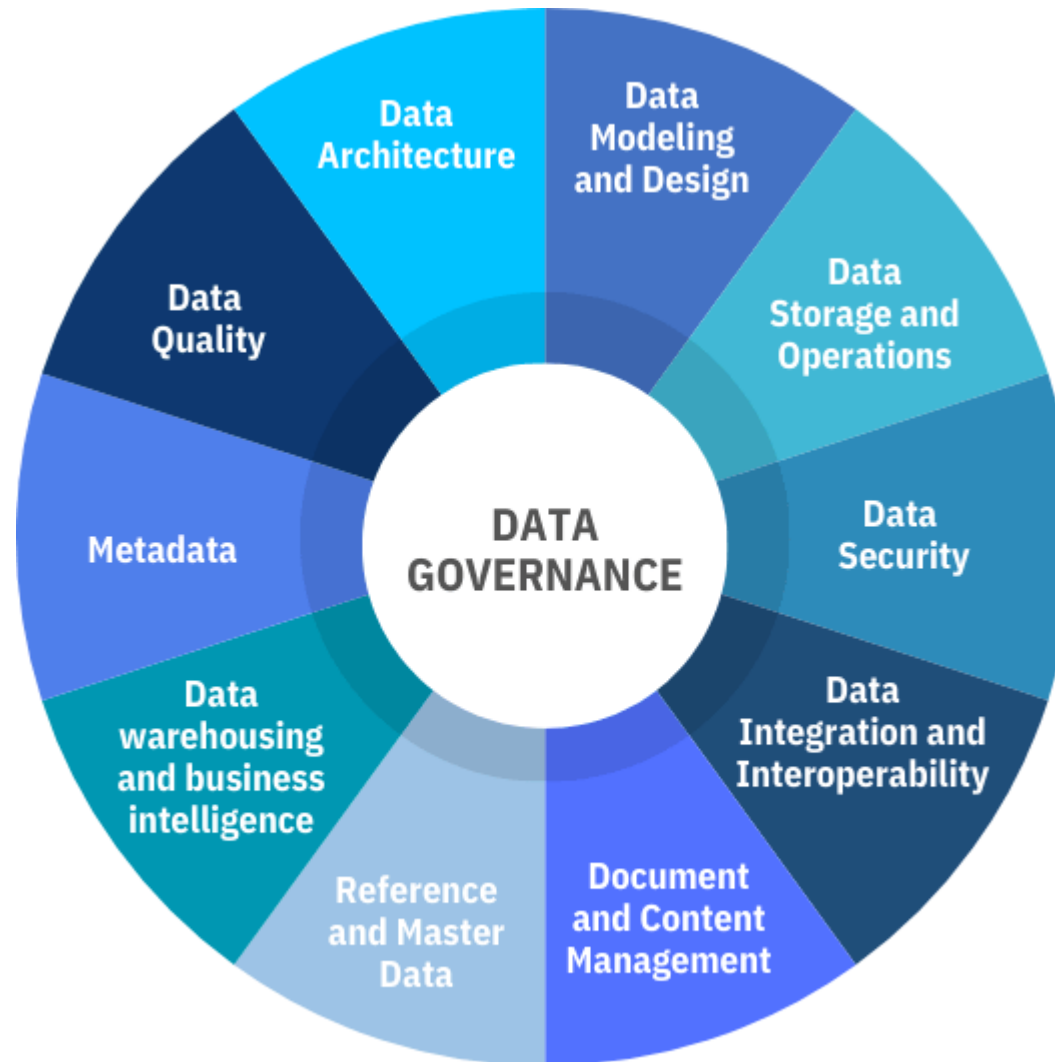
- Dient dazu, **Bedeutungen von Daten** in einem spezifischen Kontext durch Beziehungen und Attribute zu erfassen.
- Die Anwendungsbereiche liegen in der Forschung, in semantischen Webanwendungen, **Wissensmanagement in spezifischen Domänen**, ...



## Wissensgraph:

- Erweitert die Idee semantischer Graphen, um ein breites **Spektrum an Wissen** aus verschiedenen Quellen zu integrieren und zugänglich zu machen.
- Die Anwendungsbereiche liegen in wissensbasierten Systemen, **KI-Anwendungen**, Empfehlungssysteme, breite industrielle und akademische Nutzung, (...)
- **Wissensgraph** = viele semantische Graphen + Kontext

# Data Governance: Zentraler, fachlicher Baustein



Copyright © 2024 DAMA International

Source: [DAMA International](https://www.dama.org/)



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

Source: [DAMA International](https://www.dama.org/)



## „Lehre des Seins“

Eine **Ontologie** in der Informatik ist eine **formale Darstellung von Wissen**, die Konzepte in einem **hierarchischen Netzwerk mit Beziehungen verbindet**. Sie wird genutzt, um die Bedeutung von **Begriffen zu erfassen und zu modellieren**, und spielt eine **zentrale Rolle in Wissensmanagement**, semantischer Suche, **maschinellern Lernen, künstlicher Intelligenz** und anderen Wissensbereichen.

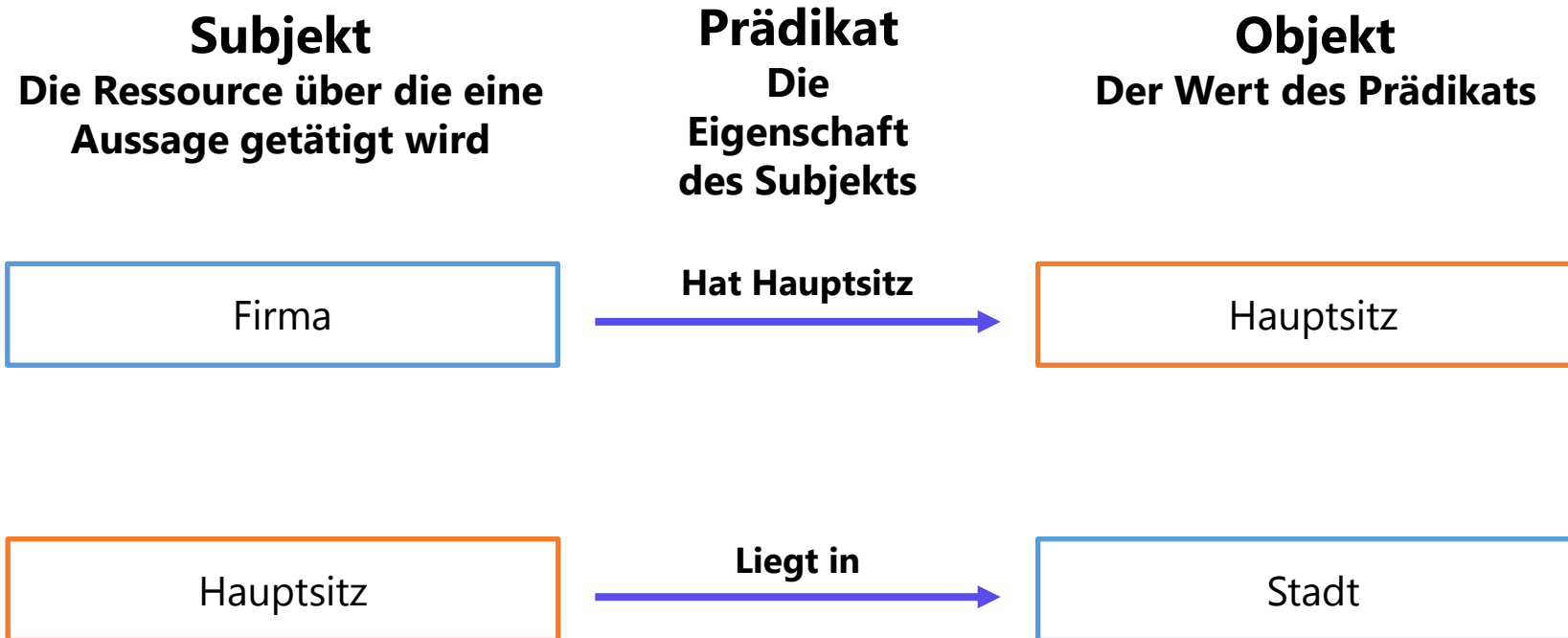
vgl. [Staab, Studer 2004]

Aber wie ist so eine Ontologie aufgebaut?

# Ressource Description Framework (RDF)

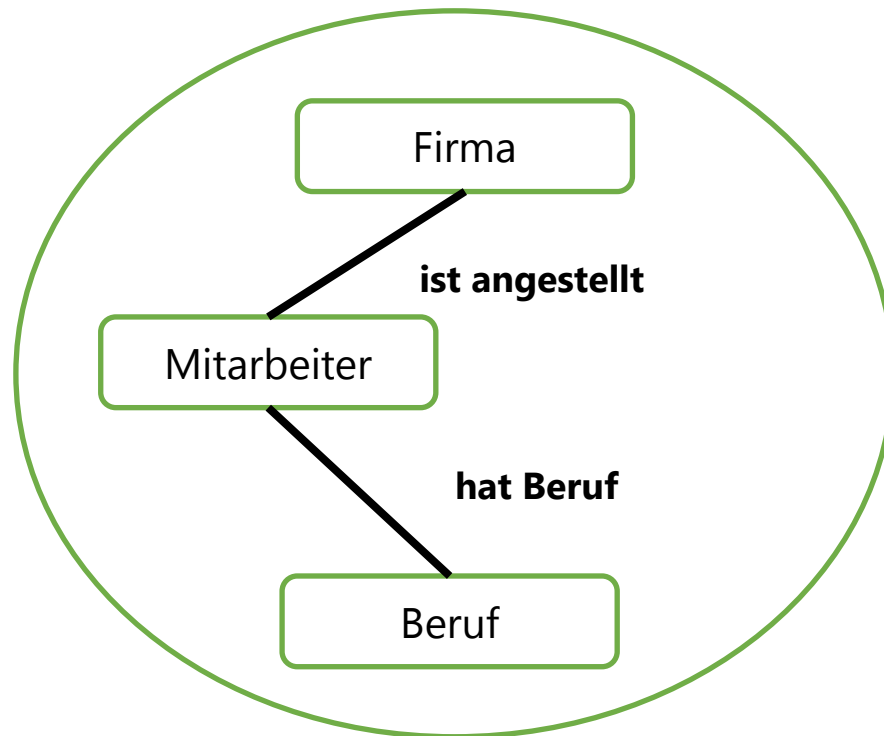
---

- RDF ist ein Standard für die Beschreibung von Informationen und deren Beziehung untereinander.

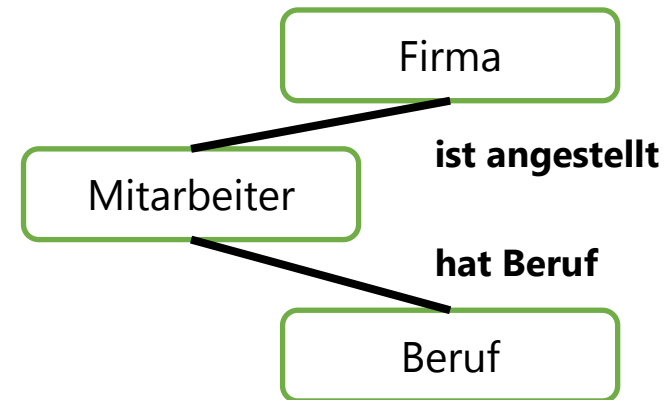


# Schema - und Instanzebene

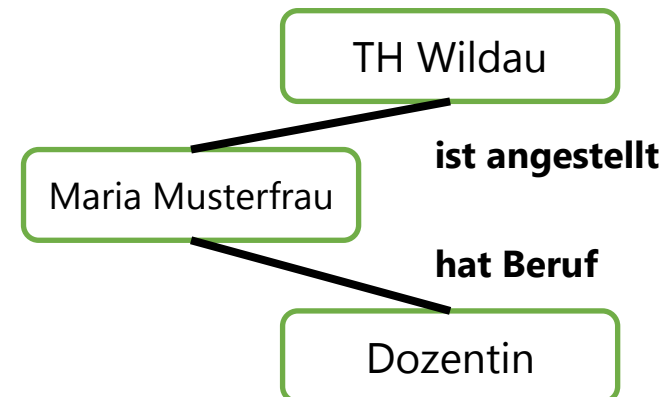
## Firmen-Ontologie



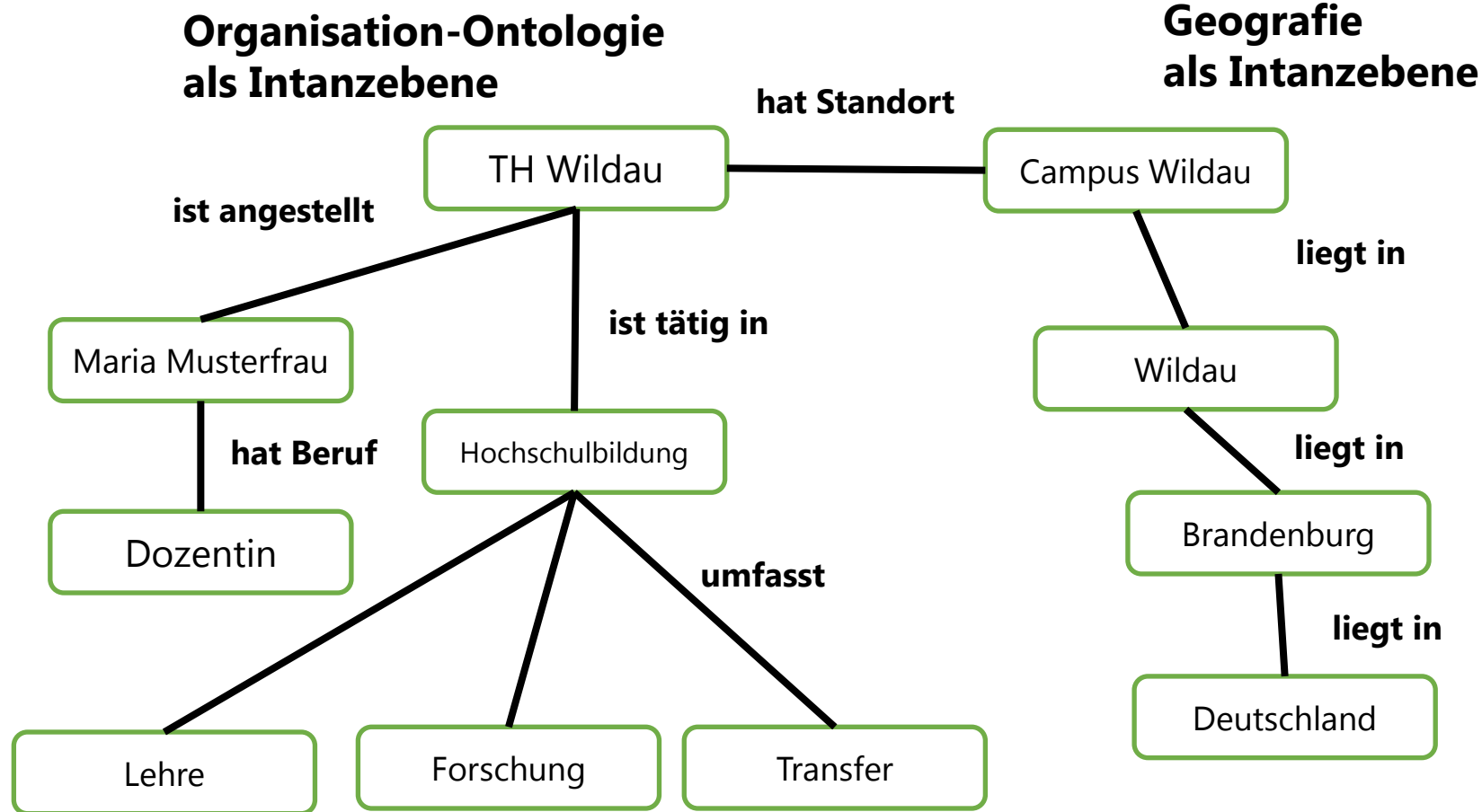
## Schemaebene



## Instanzebene



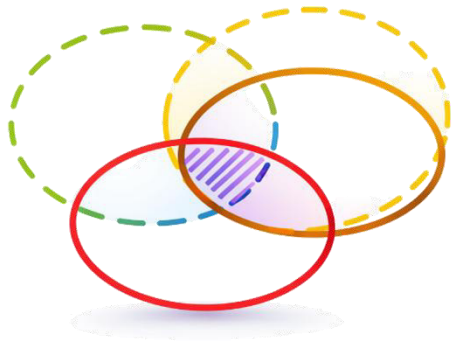
# RDF repräsentiert als Graph



```
{
  "entities": [
    {
      "id": "TH_Wildau",
      "type": "Organisation"
    },
    {
      "id": "Maria_Musterfrau",
      "type": "Person"
    },
    {
      "id": "Campus_Wildau",
      "type": "Ort"
    }
  ],
  "relations": [
    {
      "relation": "arbeitetBei",
      "source": "Maria_Musterfrau",
      "target": "TH_Wildau"
    },
    {
      "relation": "hatStandort",
      "source": "TH_Wildau",
      "target": "Campus_Wildau"
    },
    {
      "relation": "liegtIn",
      "source": "Campus_Wildau",
      "target": "Wildau"
    },
    {
      "relation": "liegtIn",
      "source": "Wildau",
      "target": "Brandenburg"
    },
    {
      "relation": "liegtIn",
      "source": "Brandenburg",
      "target": "Deutschland"
    }
  ]
}
```

# 3 Punkte zum Mitnehmen...

---



## KI beginnt beim Informationsbedarf

- KI braucht klare fachliche Konzepte
- nicht nur Datenpipelines
- Informationsbedarf → Datenmodell → Semantik

💡 **Gute KI beginnt nicht mit Modellen, sondern mit Fragen!**



## Semantik macht Daten für KI verständlich

- K.-Graphs strukturieren Fachwissen
- Ontologien definieren Begriffe und Beziehungen
- Semantic Layer verbindet Daten und Bedeutung

💡 **Daten werden erst durch Semantik zu Wissen!!**

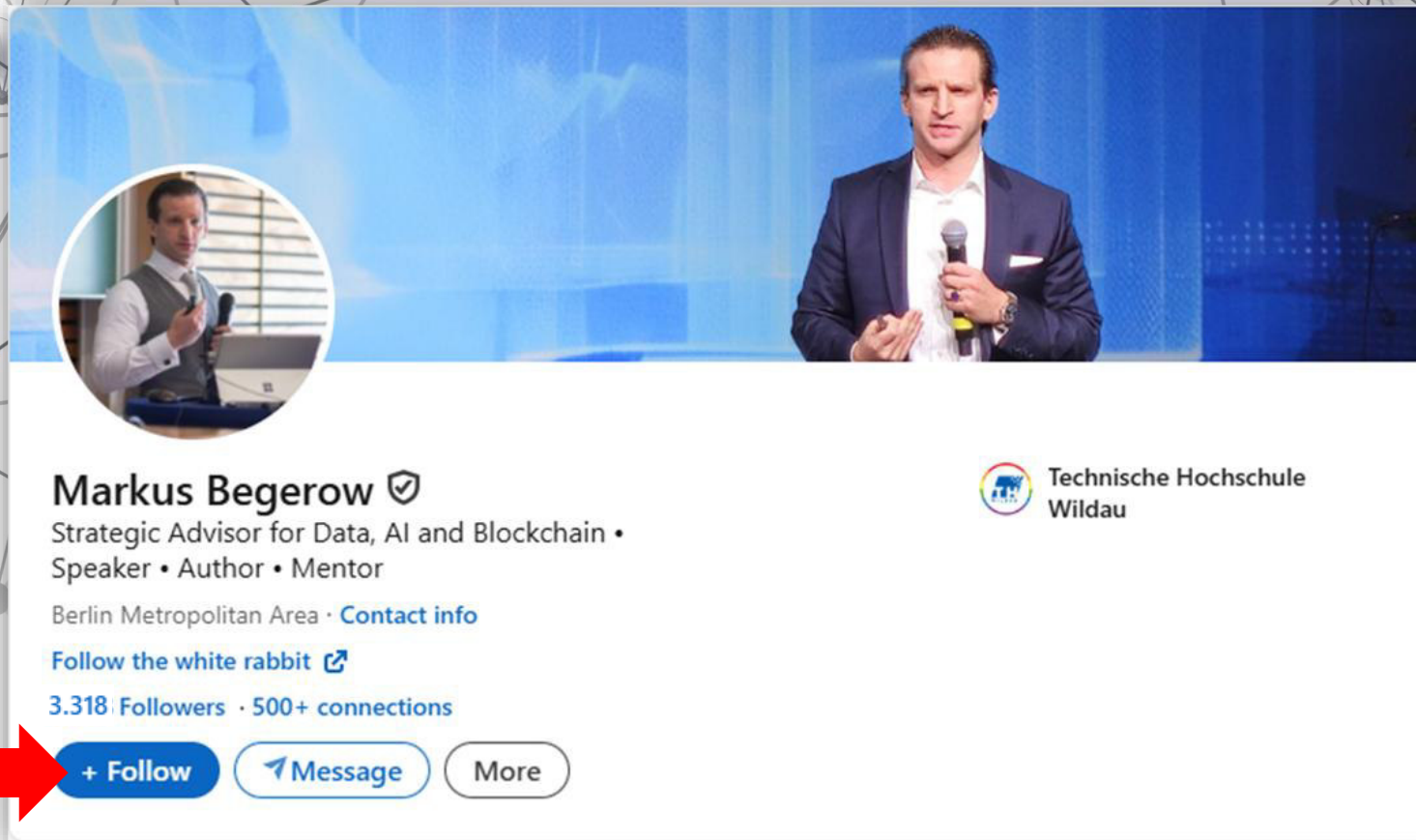


## Governance macht KI vertrauenswürdig

- Nachvollziehbare Datenquellen
- Versionierte Modelle & Prompts
- Prüfbarkeit von KI-Ergebnissen

💡 **Was nicht prüfbar ist, ist nicht produktionsreif!!!**





# THANK YOU!

Slides, Sources, Tools:  
[markus-begerow.de](https://markus-begerow.de)



📱 @markusbegerow

✉️ [mail@markus-begerow.de](mailto:mail@markus-begerow.de)

🌐 [linkedin.com/in/markusbegerow](https://linkedin.com/in/markusbegerow)