

# HeFDI-Forschungsdatentag 24. Juni 2022

Track	Session	Referent*in(nen)
1. Datenmodellierung & Coding	1.1. Einstieg in die Datenmodellierung	Andre Pietsch (JLU Gießen), Andreas Schieberle (Hochschule Darmstadt)



## Abstract:

Dieser Vortrag gibt einen Überblick über verschiedene Aspekte der Datenmodellierung und zeigt auf, warum die Modellierung von Daten beim Aufbau eines Systems eine zentrale Rolle einnehmen muss. Anhand eines einfachen Beispiels, welches sich durch den gesamten Vortrag zieht, wird gezeigt, welche unterschiedlichen Möglichkeiten es gibt, ein und dieselben Daten zu modellieren. Die Beispiele reichen von einer relationalen Datenbank hin zum JSON-Format. Vorkenntnisse sind nicht nötig, ein Grundverständnis, wie Computersysteme arbeiten, aber von Vorteil.

## Zu den Forschungsdatentagen:

Die HeFDI-Forschungsdatentage ermöglichen es Forschenden, Lehrenden und allen weiteren Interessierten, Themen und Angebote des Forschungsdatenmanagements (FDM) kennenzulernen und zu erproben. Zudem stellen verschiedene hessische Infrastruktur-Einrichtungen und Zentren ihre Dienste und Angebote vor. Die Forschungsdatentage sind ein Angebot der Landesinitiative HeFDI - Hessische Forschungsdateninfrastrukturen, welche vom Hessischen Ministerium für Wissenschaft und Kunst (HMWK) finanziert wird.

DOI: <https://doi.org/10.5281/zenodo.6982991>; Lizenzinformation: Creative Commons Attribution 4.0 International ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))



# Einstieg in die Datenmodellierung

Referenten:

Andre Pietsch (JLU Gießen)

Andreas Schieberle (Hochschule Darmstadt)



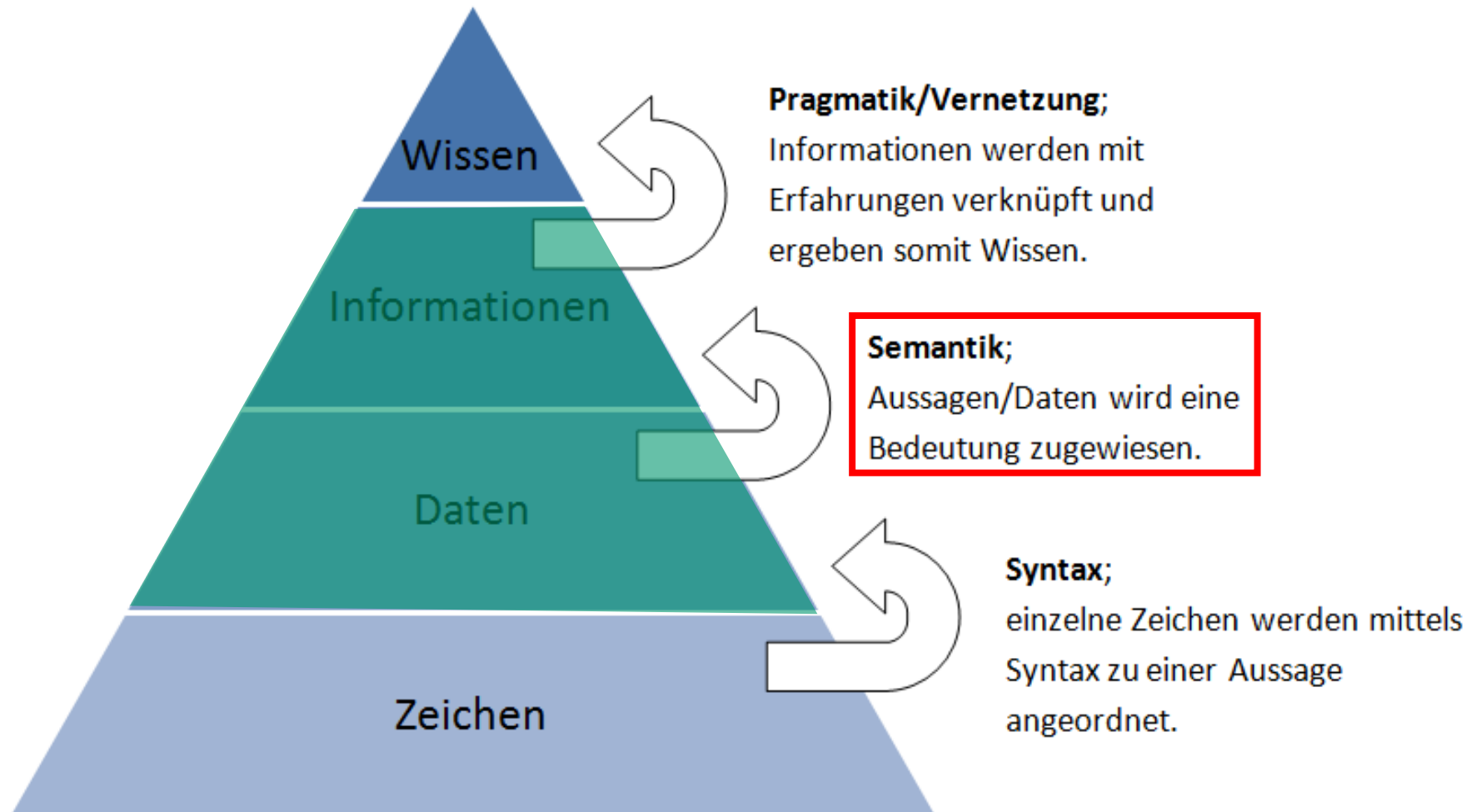
# Ein Beispiel für Datenmodellierung in Excel

Prof?	Wann?	Was?	
Max Musterforscher	05/1996-09/1999	Verwendungen von Abkürzungen im MA	Projekt
	03/2000-08/2003	Datenmodelle in den Sozialwissenschaften	Projekt
	WiSe 2003/2004	Einführungsveranstaltung	Lehrveranstaltung
	2019	Datenmodellierung sozialer Interaktionen	Publikation
Maximilia Musterforscherin (Wirtschaftswissenschaften)	WiSe 2015/2016	Einführung in die Wirtschaftswissenschaften	Lehrveranstaltung (Vorlesung)
	SoSe 2017	Externes Rechnungswesen	Lehrveranstaltung (Seminar)
	Okt. 2018 bis ????	Aufbau eines Infoportals für Neuunternehmer und Start-Ups	Sonstiges
	2021	Tracking als Teil mikro- und makroökonomischer Prozesse im digitalen Raum	Sammelwerk
Musterforscher / Musterforscherin	31.05.2022	hr	Interview

## Was ist hier schlecht umgesetzt?

- verbundene Zellen
- hierarchische Darstellung von Inhalten
- Nutzung von Kommentaren und Farbmarkierungen
- mehrere Informationen pro Zelle und Nutzung von unaufgelösten Abkürzungen





Quelle: Raffael Herrmann, 2012; [www.derwirtschaftsinformatiker.de](http://www.derwirtschaftsinformatiker.de)



# Zeichen – Daten – Information – Wissen

## Zeichen

Einzelne Zeichen ohne Beziehungen

1,2, ..., a, b, c, ..., X, Y,

## Daten

Sinnbehafteter Zusammenhang, allerdings ohne Semantik und Sinnzusammenhang

und, aber, 8:45, 24.06.2022

## Informationen

Kontextualisierung von einzelnen Daten in einen Sinnzusammenhang / Semantik etc.

HeFDI Forschungsdatentag am  
24.06.2022

## Wissen

Räumlicher, zeitlicher, kultureller Kontext und gekoppelt mit individuellen Kenntnissen, Fähigkeiten etc. Wissen ist individuelle Erkenntnis/Erfahrung

Wo, wann, wer, was ... inkl. aller  
weiteren Kontexte





# Wie kann man Informationen & Wissen beschreiben

## Beispiel „Bibliothek“

Formale Metadaten, wie

- Titel, Autor:innen
- Seitenzahlen
- Identifier

Inhaltliche Metadaten, wie

- Verschlagwortung
- Klassifikation
- Abstract

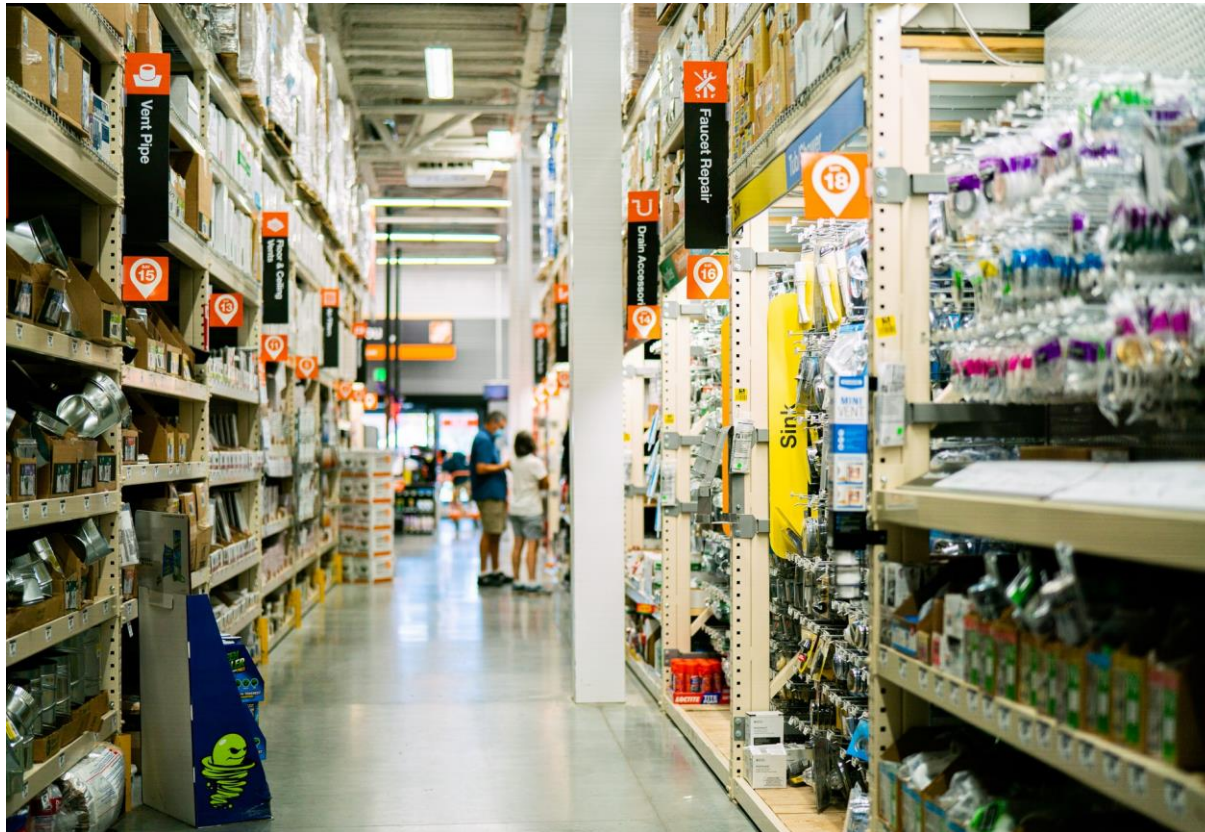


Quelle: <https://pixabay.com/de/photos/b%3c%bccher-regale-t%c3%bcr-eingang-1655783/>





# Wie kann man Informationen & Wissen beschreiben



Quelle: [https://unsplash.com/photos/loDPKt\\_srQ0](https://unsplash.com/photos/loDPKt_srQ0)

## Beispiel „Baumarkt“

- Gangbezeichnungen und Nummern
- Regal-Bezeichnungen

## Prinzip:

Thematisch ähnliches zusammen gruppieren.



# Wie kann man Informationen & Wissen beschreiben

## Filter Zusatzstoffe / Allergene




















Angebote ausschließen, die markierte Zusatzstoffe / Allergene beinhalten. [Filter leeren](#)

**Zusatzstoffe**

<input type="checkbox"/> (1) Farbstoff	<input type="checkbox"/> (2) Konservierungsstoff	<input type="checkbox"/> (3) Antioxidationsmittel
<input type="checkbox"/> (4) Geschmacksverstärker	<input type="checkbox"/> (5) Geschwefelt	<input type="checkbox"/> (6) Geschwärzt
<input type="checkbox"/> (7) Gewachst	<input type="checkbox"/> (8) Phosphat	<input type="checkbox"/> (9) Süßungsmittel
<input type="checkbox"/> (10) Alkohol	<input type="checkbox"/> (11) Mit einer Zuckerart und Süßungsmittel	<input type="checkbox"/> (12) Phenylalaninquelle
<input type="checkbox"/> (13) Koffeinhaltig		

**Allergene**

<input type="checkbox"/> (30) Glutenhaltiges Getreide
<input type="checkbox"/> (33) Fisch
<input type="checkbox"/> (36) Milch
<input type="checkbox"/> (39) Senf
<input type="checkbox"/> (42) Lupinen
<input type="checkbox"/> (30b) Roggen
<input type="checkbox"/> (30e) Dinkel
<input type="checkbox"/> (37b) Haselnüsse
<input type="checkbox"/> (37e) Pekannüsse
<input type="checkbox"/> (37h) Macadamia-/Queenslandnüsse

NUDELWELT I		Preise Studierende*   Bedienstete*   Gäste
	<b>Pasta (30,30a)</b> oder Vollkornpasta (30,30a) mit Sojabolognese (35,38)	Euro 2,30   4,30   4,80   
	<b>Pasta (30,30a)</b> oder Vollkornpasta (30,30a) mit Sauce Bolognese (2,38) Fleisch "Geprüfte Qualität - HESSEN"	Euro 2,40   4,40   4,90   
	<b>Pasta (30,30a)</b> oder Vollkornpasta (30,30a) mit veganem Petersilienpesto	Euro 2,20   4,20   4,70   
	<b>Zu unseren Pastagerichten bieten wir Ihnen zusätzlich an:</b> Grana Padano (2,32,36) Veganer Cashew-Crumble (37,41,37d)	 
	<b>Chinesische Mienudelsuppe (2,30,35,38,39,40,30a)</b> Preis pro 100 g	Euro 0,32   0,50   0,55   

## Beispiel „Speisekarte“

- Kategorien
- ggf. Nummern
- Bezeichnungen
- Hauptzutaten
- Allergiehinweise
- Preise





# Wie denkt der Mensch und wie „denkt“ eine Maschine?

## Mensch

- Denken in Bildern und Beziehungen
- Einfaches Erkennen von Bildern und Ähnlichkeiten
- Implizites Wissen
- soziokulturelle Einbettung

## Maschine

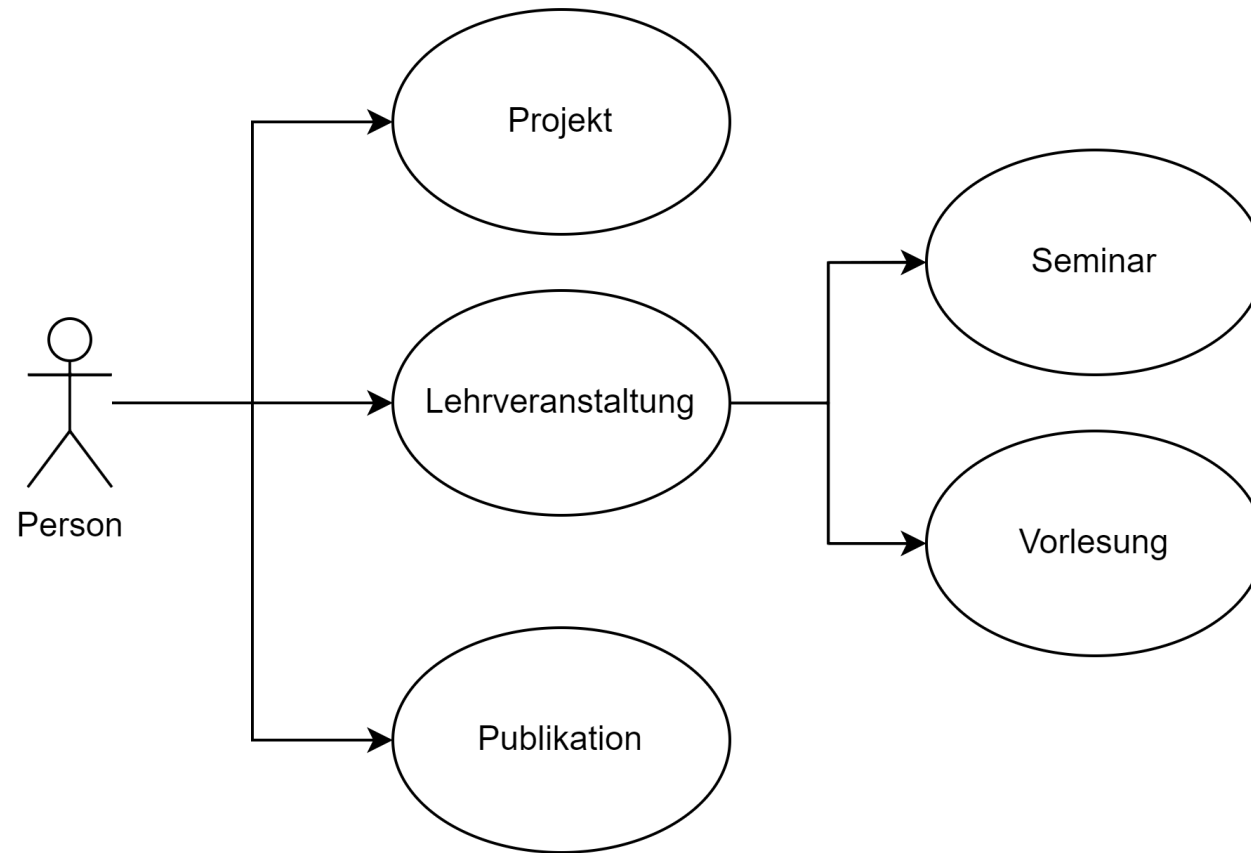
- bitweise Prozessierung von Informationen (in Nullen und Einsen)
- Speicheradressierung
- kein (implizites) Wissen vorhanden
- Simulation von Intelligenz durch Regel und Statistik



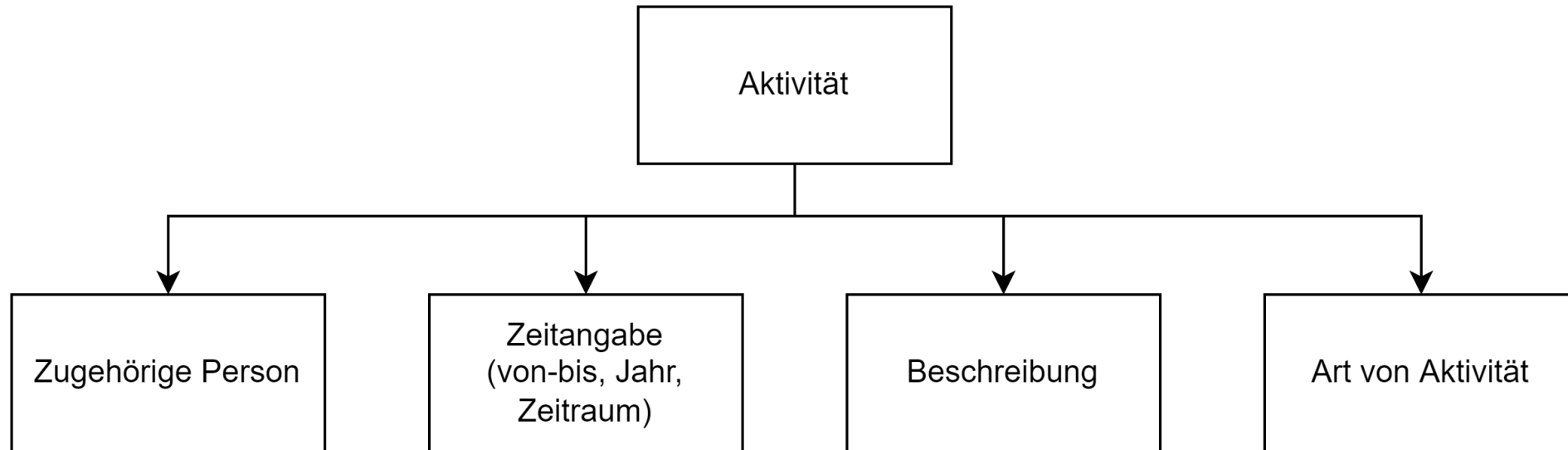
# Der Weg zum Datenmodell



# Modell skizzieren – Variante 1

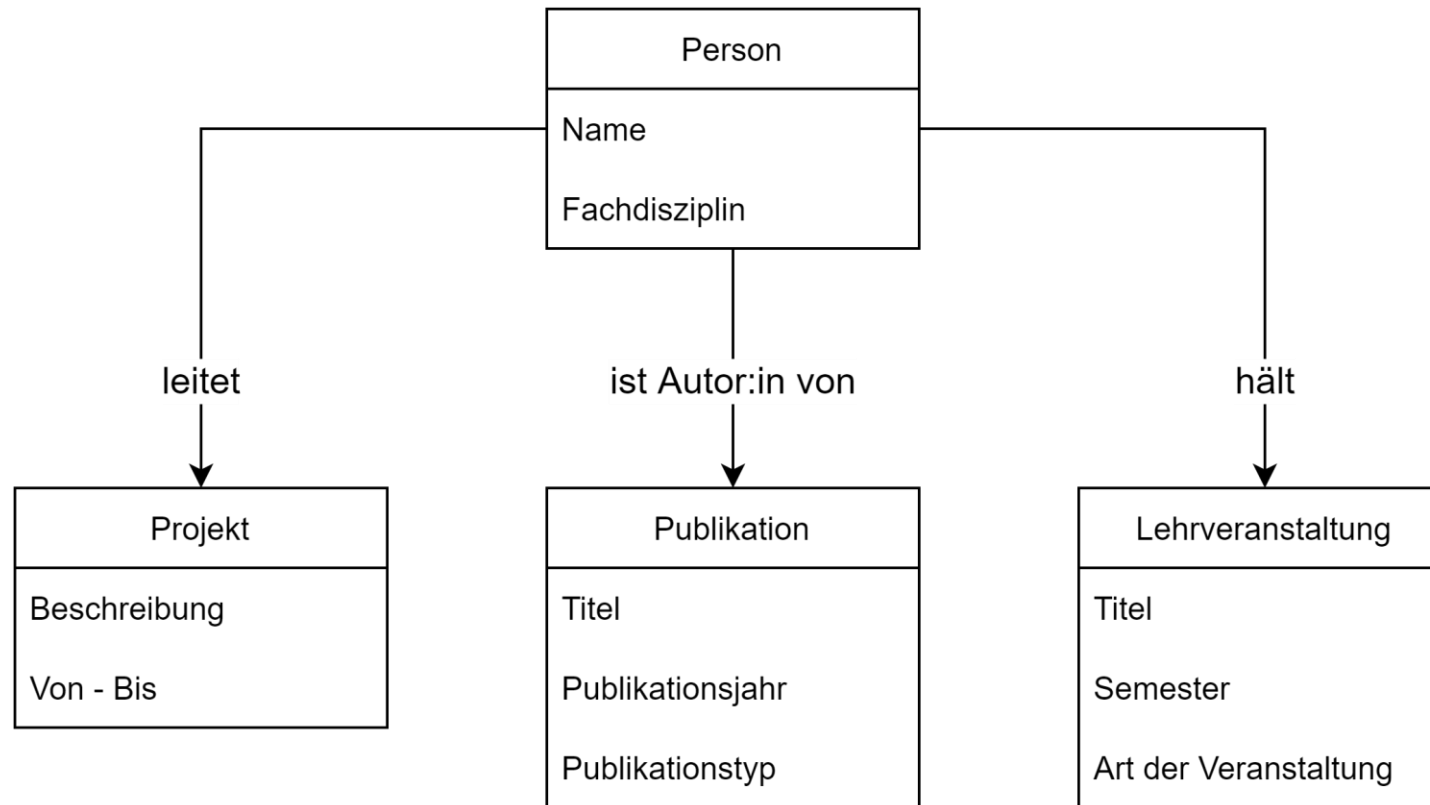


# Modell skizzieren – Variante 2





# Modellierung (angelehnt an UML – Unified Modeling Language)



# Unser Ausgangsbeispiel

Prof?	Wann?	Was?	
Max Musterforscher	05/1996-09/1999	Verwendungen von Abkürzungen im MA	Projekt
	03/2000-08/2003	Datenmodelle in den Sozialwissenschaften	Projekt
	WiSe 2003/2004	Einführungsveranstaltung	Lehrveranstaltung
	2019	Datenmodellierung sozialer Interaktionen	Publikation
Maximilia Musterforscherin (Wirtschaftswissenschaften)	WiSe 2015/2016	Einführung in die Wirtschaftswissenschaften	Lehrveranstaltung (Vorlesung)
	SoSe 2017	Externes Rechnungswesen	Lehrveranstaltung (Seminar)
	Okt. 2018 bis ????	Aufbau eines Infoportals für Neuunternehmer und Start-Ups	Sonstiges
	2021	Tracking als Teil mikro- und makroökonomischer Prozesse im digitalen Raum	Sammelwerk
Musterforscher / Musterforscherin	31.05.2022	hr	Interview



# Nullte Normalform (Modell)

Tabelle
Person : Text
Zeitraum : Text
Titel : Text
Kategorie : Text

**0NF:** Daten liegen tabellarisch vor.

**Weg:**

- verbundene Zellen auflösen
- sinnvolle Spaltenüberschriften definieren
- leere Zellen befüllen
- ggf. fehlende Daten ergänzen



# Nullte Normalform (Daten)

Person	Zeitraum	Titel	Kategorie
Max Musterforscher (Sozial-/Altertumswissenschaften)	05/1996-09/1999	Verwendungen von Abkürzungen im MA	Projekt
Max Musterforscher (Sozial-/Altertumswissenschaften)	03/2000-08/2003	Datenmodelle in den Sozialwissenschaften	Projekt
Max Musterforscher (Sozial-/Altertumswissenschaften)	WiSe 2003/2004	Einführungsveranstaltung (Seminar)	Lehrveranstaltung
Max Musterforscher (Sozial-/Altertumswissenschaften)	2019	Datenmodellierung sozialer Interaktionen (Monographie)	Publikation
Maximilia Musterforscherin (Wirtschaftswissenschaften)	WiSe 2015/2016	Einführung in die Wirtschaftswissenschaften	Lehrveranstaltung (Vorlesung)
Maximilia Musterforscherin (Wirtschaftswissenschaften)	SoSe 2017	Externes Rechnungswesen	Lehrveranstaltung (Seminar)
Maximilia Musterforscherin (Wirtschaftswissenschaften)	Okt. 2018 bis ????	Aufbau eines Infoportals für Neuunternehmer und Start-Ups	Projekt
Maximilia Musterforscherin (Wirtschaftswissenschaften)	2021	Tracking als Teil mikro- und makroökonomischer Prozesse im digitalen Raum	Sammelwerk
Musterforscher / Musterforscherin	31.05.2022	hr	Interview





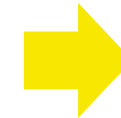
# Erste Normalform (Modell)

**1NF:** Attribute liegen atomar vor,  
Wiederholungsgruppen sind ausgelagert.

## Weg:

- Person -> Vorname, Nachname, Fachdisziplin
- Zeitraum -> Von, Bis, Semester, Publikationsjahr
- Kategorie -> Kategorie, Typ
- Mehrere Werte in einer Zelle in mehrere Zeilen auslagern.

Tabelle



Tabelle



# Erste Normalform (Daten)

Für Excel meist  
ausreichend!

Vorname	Nachname	Fachdisziplin	Von	Bis	Semester	Publikations jahr	Titel	Kategorie	Typ
Max	Musterforscher	Sozial-/Altertumswissenschaften	01.05.1996	30.09.1999			Verwendungen von Abkürzungen im MA	Projekt	
Max	Musterforscher	Sozial-/Altertumswissenschaften	01.03.2000	31.08.2003			Datenmodelle in den Sozialwissenschaften	Projekt	
Max	Musterforscher	Sozial-/Altertumswissenschaften	01.10.2003	31.03.2004	WiSe 2003/2004		Einführungsveranstaltung	Lehrveranstaltung	Seminar
Max	Musterforscher	Sozial-/Altertumswissenschaften				2019	Datenmodellierung sozialer Interaktionen	Publikation	Monographie
Maximilia	Musterforscherin	Wirtschaftswissenschaften	01.10.2015	31.03.2016	WiSe 2015/2016		Einführung in die Wirtschaftswissenschaften	Lehrveranstaltung	Vorlesung
Maximilia	Musterforscherin	Wirtschaftswissenschaften	01.04.2017	30.09.2017	SoSe 2017		Externes Rechnungswesen	Lehrveranstaltung	Seminar
Maximilia	Musterforscherin	Wirtschaftswissenschaften	01.10.2018				Aufbau eines Infoportals für Neuunternehmer und Start-Ups	Projekt	
Maximilia	Musterforscherin	Wirtschaftswissenschaften				2021	Tracking als Teil mikro- und makroökonomischer Prozesse im digitalen Raum	Publikation	Sammelwerk
	Musterforscher		31.05.2022	31.05.2022			hr	Aktivität	Interview
	Musterforscherin		31.05.2022	31.05.2022			hr	Aktivität	Interview



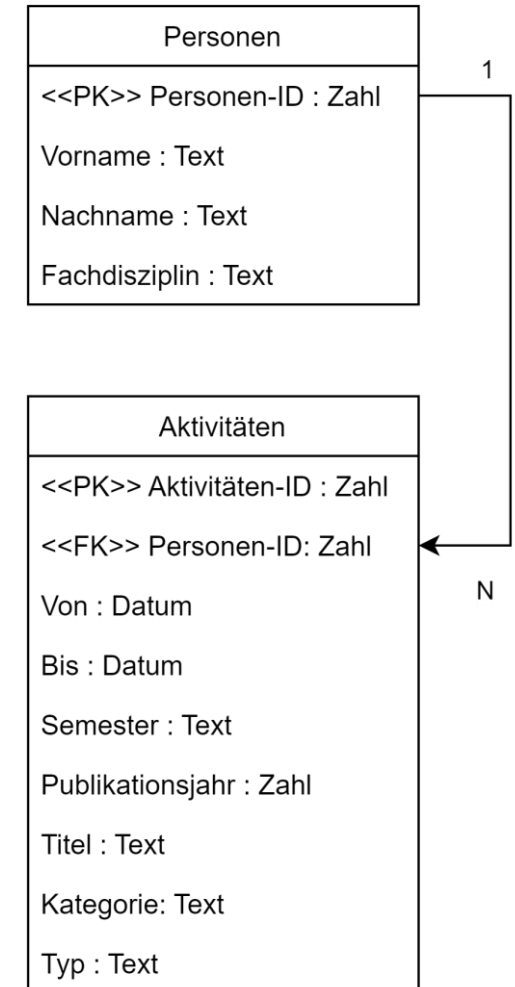
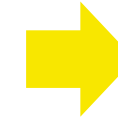
# Zweite Normalform (Modell)

**2NF:** Alle Nicht-Schlüsselattribute sind von allen **Schlüsselattributen** abhängig und nicht nur von einem Schlüsselteil.

**Weg:**

- ID als Schlüsselattribut (PK) einführen
- Spalten *Vorname*, *Nachname* und *Fachdisziplin* in eigene Tabelle auslagern
- ID der Person als Fremdschlüssel (FK) in der Aktivitäten-Tabelle eintragen

Tabelle
Vorname : Text
Nachname : Text
Fachdisziplin : Text
Von : Datum
Bis : Datum
Semester : Text
Publikationsjahr : Zahl
Titel : Text
Kategorie: Text
Typ : Text



# Zweite Normalform (Daten)

Personen			
Personen-ID	Vorname	Nachname	Fachdisziplin
1	Max	Musterforscher	Sozial-/Altetumswissenschaften
2	Maximilia	Musterforscherin	Wirtschaftswissenschaften

**Für viele  
Anwendungsfälle  
ausreichend!**

Aktivitäten								
Aktivitäten-ID	Personen-ID	Von	Bis	Semester	Publikationsjahr	Titel	Kategorie	Typ
1	1	01.05.1996	30.09.1999			Verwendungen von Abkürzungen im MA	Projekt	
2	1	01.03.2000	31.08.2003			Datenmodelle in den Sozialwissenschaften	Projekt	
3	1	01.10.2003	31.03.2004	WiSe 2003/2004		Einführungsveranstaltung	Lehrveranstaltung	Seminar
4	1				2019	Datenmodellierung sozialer Interaktionen	Publikation	Monographie
5	2	01.10.2015	31.03.2016	WiSe 2015/2016		Einführung in die Wirtschaftswissenschaften	Lehrveranstaltung	Vorlesung
6	2	01.04.2017	30.09.2017	SoSe 2017		Externes Rechnungswesen	Lehrveranstaltung	Seminar
7	2	01.10.2018				Aufbau eines Infoportals für Neuunternehmer und Start-Ups	Projekt	
8	2				2021	Tracking als Teil mikro- und makroökonomischer Prozesse im digitalen Raum	Publikation	Sammelwerk
9	1	31.05.2022	31.05.2022			hr	Aktivität	Interview
10	2	31.05.2022	31.05.2022			hr	Aktivität	Interview





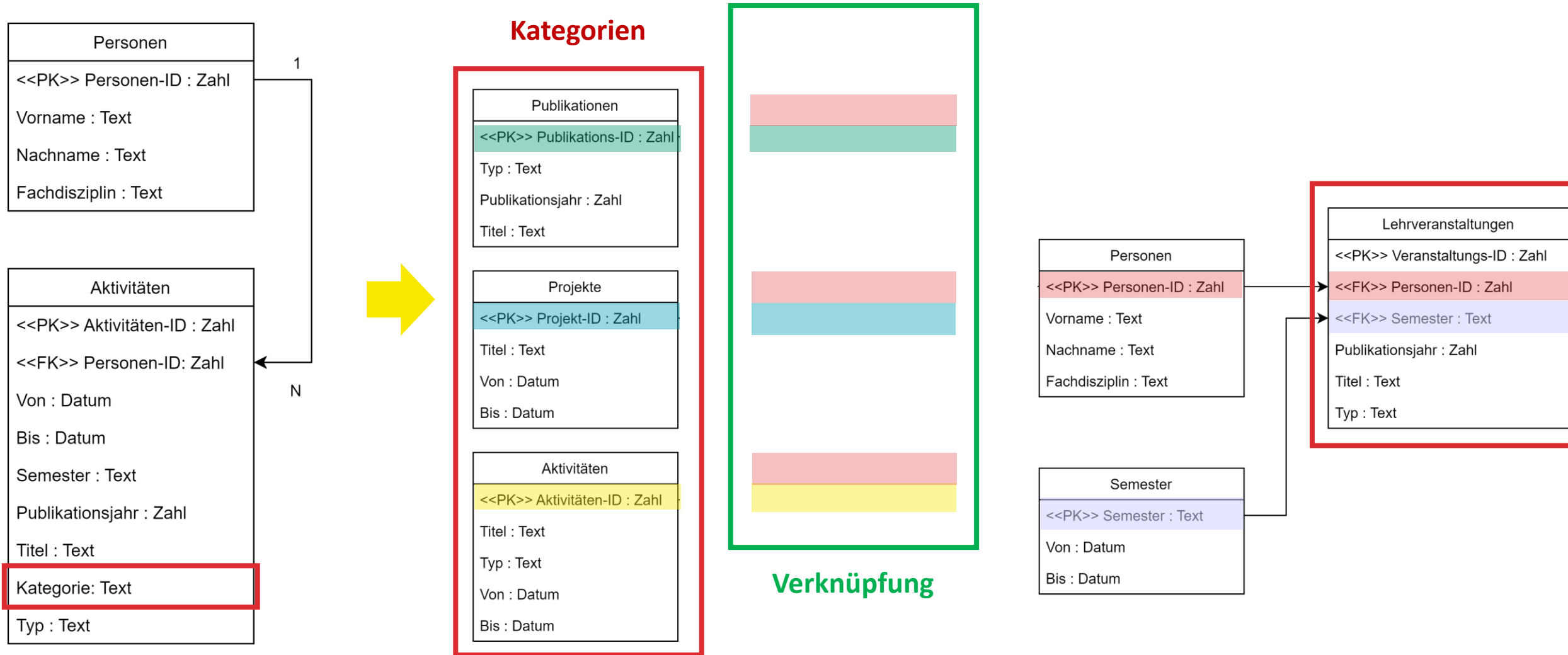
# Dritte Normalform

**3NF:** Kein Nicht-Schlüsselattribut ist von einem anderen Nicht-Schlüsselattribut abhängig.

Aktivitäten								
Aktivitäten-ID	Personen-ID	Von	Bis	Semester	Publikationsjahr	Titel	Kategorie	Typ
1	1	01.05.1996	30.09.1999			Verwendungen von Abkürzungen im MA	Projekt	
2	1	01.03.2000	31.08.2003			Datenmodelle in den Sozialwissenschaften	Projekt	
3	1	01.10.2003	31.03.2004	WiSe 2003/2004		Einführungsveranstaltung	Lehrveranstaltung	Seminar
4	1				2019	Datenmodellierung sozialer Interaktionen	Publikation	Monographie
5	2	01.10.2015	31.03.2016	WiSe 2015/2016		Einführung in die Wirtschaftswissenschaften	Lehrveranstaltung	Vorlesung
6	2	01.04.2017	30.09.2017	SoSe 2017		Externes Rechnungswesen	Lehrveranstaltung	Seminar
7	2	01.10.2018				Aufbau eines Infoportals für Neuunternehmer und Start-Ups	Projekt	
8	2				2021	Tracking als Teil mikro- und makroökonomischer Prozesse im digitalen Raum	Publikation	Sammelwerk
9	1	31.05.2022	31.05.2022			hr	Aktivität	Interview
10	2	31.05.2022	31.05.2022			hr	Aktivität	Interview



# Dritte Normalform (Modell)



# Dritte Normalform (Daten)

Relevant für SQL  
Datenbanken!

Person			
Personen-ID	Vorname	Nachname	Fachdisziplin
1	Max	Musterforscher	Sozial-/Altertumswissenschaften
2	Maximilia	Musterforscherin	Wirtschaftswissenschaften

Publikation			
Publikations-ID	Typ	Jahr	Titel
1	Monographie	2019	Datenmodellierung sozialer Interaktionen
2	Sammelwerk	2021	Tracking als Teil mikro- und makroökonomischer Prozesse im digitalen Raum

Lehrveranstaltung				
Veranstaltungs-ID	Personen-ID	Typ	Titel	Semester
1	1	Seminar	Einführungsveranstaltung	WiSe 2003/2004
2	2	Vorlesung	Einführung in die Wirtschaftswissenschaften	WiSe 2015/2016
3	2	Seminar	Externes Rechnungswesen	SoSe 2017

Projekt			
Projekt-ID	Titel	Von	Bis
1	Verwendungen von Abkürzungen im MA	01.05.1996	30.09.1999
2	Datenmodelle in den Sozialwissenschaften	01.03.2000	31.08.2003
3	Aufbau eines Infoportals für Neuunternehmer und Start-Ups	01.10.2018	

Aktivität				
Aktivitäts-ID	Titel	Typ	Von	Bis
1	hr	Interview	31.05.2022	31.05.2022

Person_Publikation			
PersPubl-ID	Personen-ID	Publikations-ID	Rolle
1	1	1	Autor
2	2	2	Herausgeber

Semesterzeiten		
Semester	Von	Bis
WiSe 2003/2004	01.10.2003	31.03.2004
WiSe 2015/2016	01.10.2015	31.03.2016
SoSe 2017	01.04.2017	30.09.2017

Person_Projekt		
PersProj-ID	Personen-ID	Projekt-ID
1	1	1
2	1	2
3	2	2

Person_Aktivität		
PersAkt-ID	Personen-ID	Aktivitäten-ID
1	1	1
2	2	1



# Wozu der Aufwand?

## Ziel: Konsistenzhöhung durch Redundanzvermeidung

- existierende Datensätze müssen **nicht erneut erfasst** werden
  - z.B. Semesterzeiten, Personennamen
  - gleichzeitig: Fehlervermeidung
- Änderungen an Daten müssen **nur an einer Stelle** vorgenommen werden
  - z.B. Änderung des Titels oder Umbenennung der Fachdisziplinen
- **Vermeidung inkonsistente Daten**
  - z.B. bei Löschung oder fehlerhaften Änderungen
- Reduktion des **Speicherbedarfs**





# Grenzen relationaler Datenmodelle

## Aufwand

- Datenmodell muss zu Beginn aufwändig definiert werden
- Spätere Änderungen sind aufwendig umzusetzen

## Eignung

- Gut geeignet für Datenarten, die sich in Tabellen abbilden lassen
- Schlecht geeignet für abstrakte oder komplexe Datenarten

## Segmentierung

- Zusammenhängende Daten werden in unterschiedlichen Tabellen gespeichert (Normalisierung)
- Komplexe Abfragen mit negativer Performance

## Skalierbarkeit

- Schlecht geeignet bei sehr großen Datenmengen
- Gut vertikal skalierbar (größerer Server) aber schwer horizontal skalierbar (mehr Server)



# Nicht Relationale Datenmodelle / Datenbanken

## Key-Value Stores

- Einfache Schlüssel-Wert Paare
- Fokus auf Lese-/Schreibgeschwindigkeit
- Geeignet z.B. für Caching oder Real-Time-Anwendungen

## Spaltenorientierte Datenbank

- Zugriff über einzelne Spalten statt Tabellen
- Fokus auf Lesegeschwindigkeit und Skalierbarkeit
- Geeignet z.B. für verteilte Datenhaltung in mehreren Rechenzentren

## Graphdatenbanken

- Daten werden als Graphen von Knoten und Kanten dargestellt
- Fokus auf Beziehung zwischen Daten
- Geeignet für stark vernetzte Daten, z.B. Wissensnetze, Linked Data

## Dokumentdatenbanken

- Daten werden als Dokument dargestellt (z.B. als JSON oder XML)
- Jedes Dokument kann andere Datenfelder enthalten („schema-less“)
- Fokus auf zusammengehörige Daten



# Ein Beispiel für Datenmodellierung in zwei verschiedenen Formaten



# XML – eXtensible Markup Language

- Auszeichnungssprache
- 1998 veröffentlicht durch das World Wide Web Consortium (W3C)
- Regelwerk / Grammatik zur Objektbeschreibung
- geeignet für die Darstellung hierarchisch strukturierter Daten
- Verarbeitung großer Mengen von XML-Daten eher langsam im Vergleich zu JSON

```
<?xml version="1.0"?>
<quiz>
  <frage>
    Wer war der fünfte
    deutsche Bundespräsident?
  </frage>
  <antwort>
    Karl Carstens
  </antwort>
  <!-- Anmerkung: Wir
    brauchen mehr Fragen -->
</quiz>
```

**XML**

Quelle: [de.wikipedia.org/wiki/Extensible\\_Markup\\_Language](https://de.wikipedia.org/wiki/Extensible_Markup_Language) von ✓ (CC-SA 2.5)



# JSON – JavaScript Object Notation

- von Programmiersprachen unabhängig
- Datenaustauschformat in Textform
- verwendet für den Datenaustausch zwischen verschiedenen Anwendungen
- kompakte Beschreibung der Daten, daher schnell Verarbeitung
- leicht lesbar

