

E i n f ü h r u n g   G I S

BA AI Angewandte Informatik

# Geodaten II

T e c h n i s c h e   H o c h s c h u l e   D e g g e n d o r f



Prof. Dr. Roland Zink  
[roland.zink@th-deg.de](mailto:roland.zink@th-deg.de)

# Geodaten



## Definition Geodaten (Bill 2010, S. 263)

„Geodaten sind Daten über Gegenstände, Geländeformen und Infrastrukturen an der Erdoberfläche, wobei als wesentliches Element ein Raumbezug vorliegen muss. Sie beschreiben die einzelnen Objekte in der Landschaft und sind durch eine Position im Raum direkt (...) oder indirekt (...) referenzierbar.“



## Vektordaten – Sachlogik

Eine Vektordatei besteht aus zwei wesentlichen Elementen:

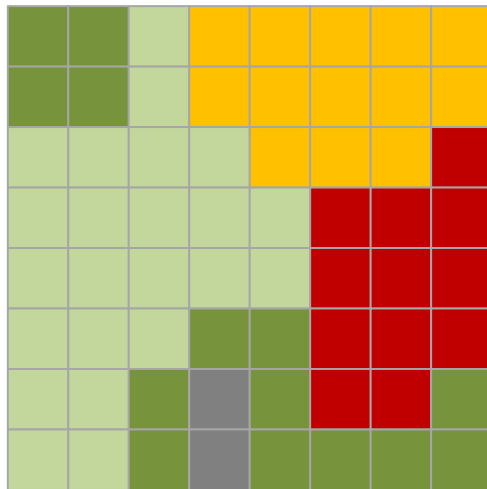
1. **Sachdatentabelle für Attributinformationen**

In dieser Tabelle werden sämtliche Informationen zu dem geometrischen Objekt gespeichert

2. **Speicherstruktur für die Geometriedaten**

Ebenfalls eine Tabelle, die sämtliche Knoten mit Koordinaten und bei Polylinien und Polygonen die Kanten bzw. Flächen dazwischen speichert

## Rasterdaten - Bild



R 128

G 128

B 128

Binärcode

100000001000000010000000

- In einer Fotodatei hat jedes Pixel eine Farbe, die mit einer Kennzahl aus dem Farbsystem gespeichert wird
  - Ein bekanntes Farbsystem ist RGB
  - Farben werden aus der Kombination der Grundfarben Rot, Grün und Blau mit der jeweiligen Intensität 0 bis 256 gebildet
  - Ein Farbpixel besteht folglich aus der Kombination von 3 Farbwerten
  - 3x8 Bit

## Rasterbilddaten – Worldfile



### Georeferenzierung

→ neben der Bilddatei gibt es eine Textdatei (ASCII) mit Zusatzinformationen

### Räumliche Referenz der Bilddatei

1. Es wird zu jeder Rasterzelle neben der Sachinformation zusätzlich noch die x- und y-Koordinate → großer Speicherplatzbedarf
2. Es wird eine Eckkoordinate definiert und gespeichert (meist links oben) und die Größe der Rasterzellen festgelegt → hoher Rechenaufwand

→ i.d.R. wird nach der Methode 2 vorgegangen

# Inhalt

1. Skalenniveaus
2. Primäre Metrik: XY-Daten
3. Sekundäre Metrik: Tabellen-Verknüpfungen
4. Qualität von Geodaten
5. Übung



# 1. Saklenniveaus

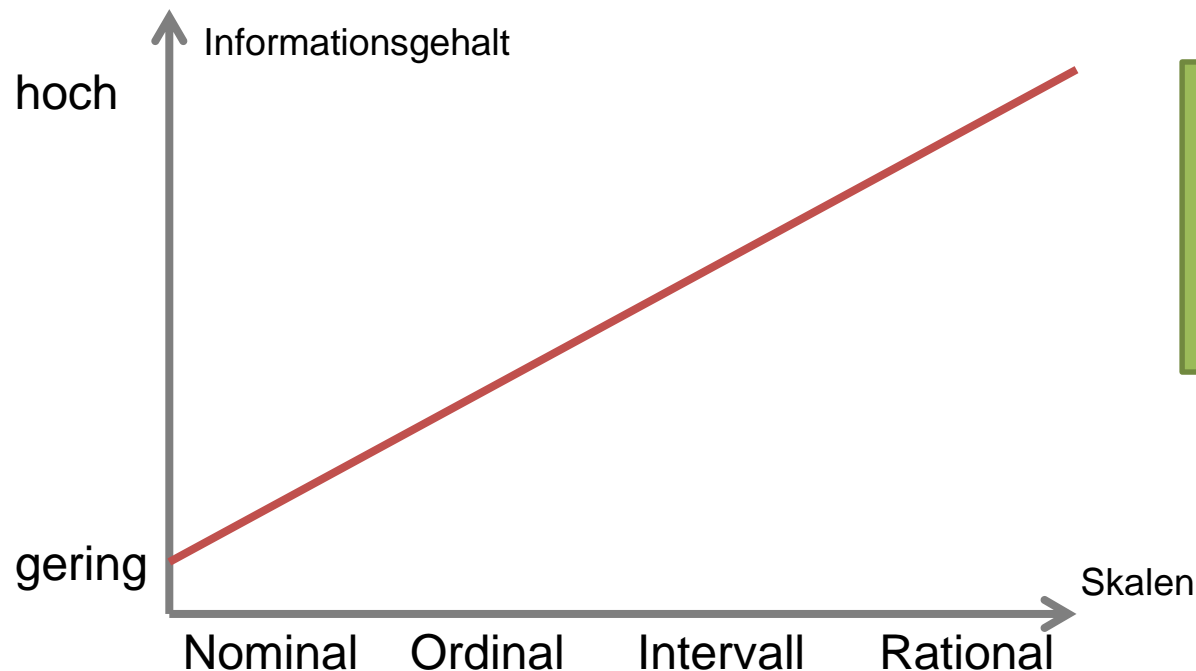


## Datenart und Skalenniveaus

Maßskalen sind für die Messung von Variablen und der Anwendung statistischer Methoden von zentraler Bedeutung.

Unterscheidung in

- **Nichtmetrische** Skalen: Nominal- und Ordinalskala
- **Metrische** Skalen: Intervall- und Rationalskala



Skalenniveaus sind hierarchisch geordnet: der Übergang zur niedrigeren Stufe ist nur durch einen Informationsverlust möglich!



## Datenart am Beispiel von Rasterdaten

### Nominaldaten

Beziehungen der Variablen ausgedrückt durch:

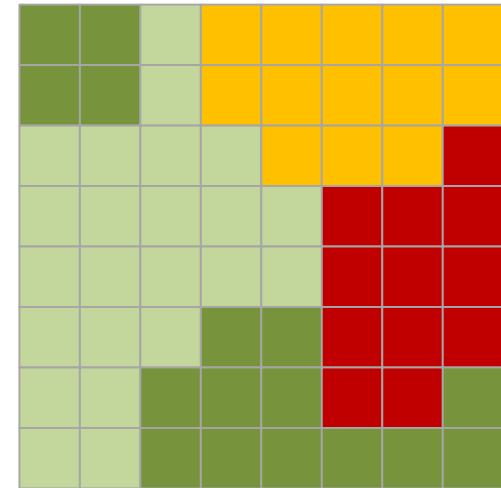
- Worte
- Buchstaben
- Zahlen (Code)

Keine Vergleichbarkeit im Sinne von Größer-Kleiner-Relation

### Mathematische Operationen

$$V_a = V_b$$

$$V_a \neq V_b$$



### Flächennutzung

	Siedlungsgebiet
	Ackerland
	Grünland
	Wald

## Datenart am Beispiel von Rasterdaten

### Ordinaldaten

Beziehungen der Variablen können in einer Rangordnung ausgedrückt werden.

Vergleichbarkeit im Sinne von Größer-Kleiner-Gleich-Relation möglich

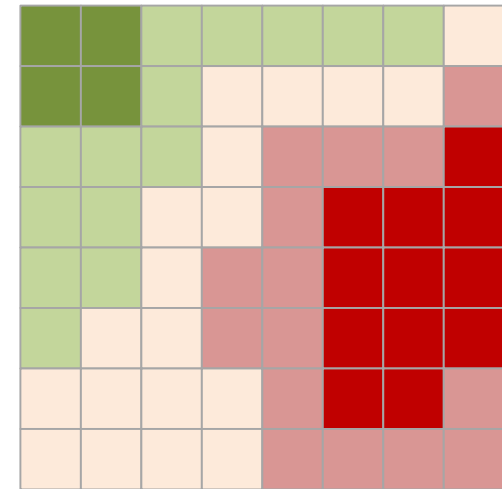
### Mathematische Operationen

$$V_a = V_b$$

$$V_a \neq V_b$$

$$V_a > V_b$$

$$V_a < V_b$$



### Lawinenrisiko



# Datenart am Beispiel von Rasterdaten

## Intervalldaten

Intervalldaten besitzen keinen Nullpunkt.

Dienen der Identifikation, Ordnung und Bewertung (additiv).

Keine Multiplikation oder Division!

## Mathematische Operationen

$$V_a = V_b$$

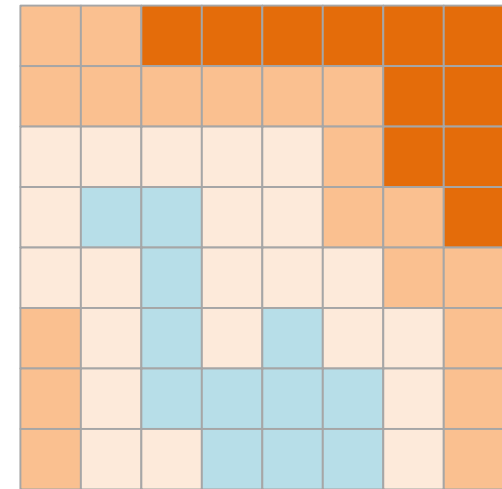
$$V_a \neq V_b$$

$$V_a > V_b$$

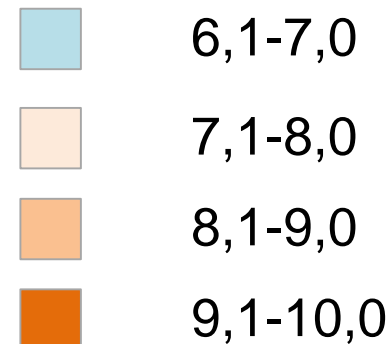
$$V_a < V_b$$

$$V_a = V_b + X$$

$$X = V_a - V_b$$



Durchschnittstemperatur  
°C



## Datenart am Beispiel von Rasterdaten

### Ratio-skalierte Daten

Rationaldaten besitzen einen Nullpunkt.

Dienen der Identifikation, Ordnung und Bewertung (additiv und multiplikativ).

→ Ein „Vielfaches“ kann angegeben werden.

### Mathematische Operationen

$$V_a = V_b$$

$$V_a \neq V_b$$

$$V_a > V_b$$

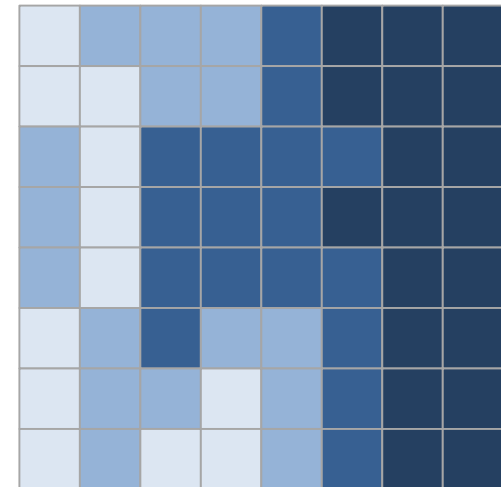
$$V_a < V_b$$

$$V_a = V_b + X$$

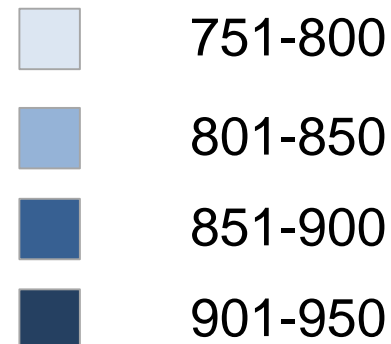
$$X = V_a - V_b$$

$$V_a = V_b * X$$

$$X = V_a / V_b$$



Niederschlag  
mm





Bearbeiten Sie die Aufgabe „Skalenniveaus“ im Übungsskript!



## 2. Primäre Metrik: XY-Daten

## Erstellung „neuer“ Geoinformationen und Geodaten

- 80-90% aller Informationen haben einen Raumbezug
- Der Raumbezug kann primär oder sekundär angegeben sein.
- Für die Verarbeitung in GIS ist es notwendig, die Daten in die richtige Form zu bringen!

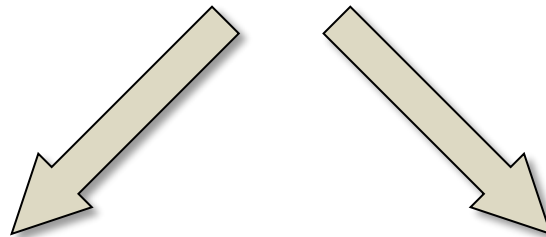


## Wiederholung

Grundlage von GIS ist der Raumbezug (= **Georeferenz**) aller Daten

Georeferenz beschreibt die Lage des Objektes/Information in einem Bezugssystem

### Georeferenz



#### **Primäre Metrik**

= direkter Raumbezug

ISO 19111 (Koordinatenreferenzsysteme)

#### **Sekundäre Metrik**

= indirekter Raumbezug

ISO 19112 (Raumbezug mit  
geographischen Identifikatoren)

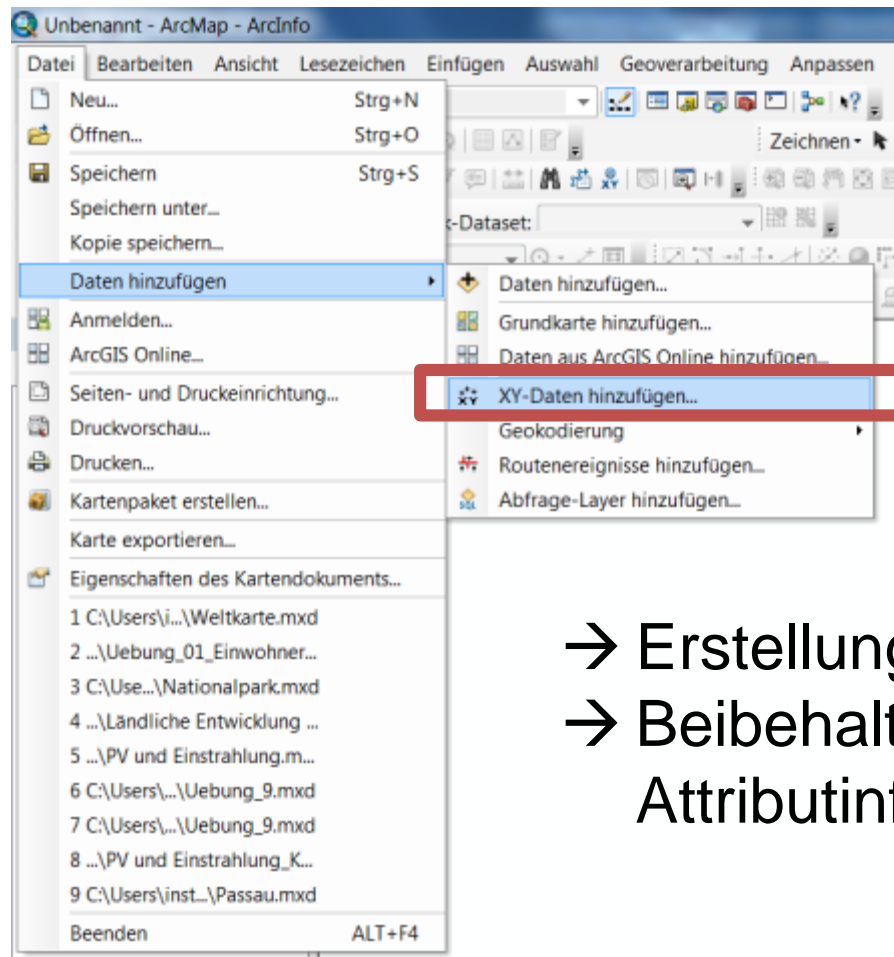




## Koordinatenangaben

- Informationen können auch direkt mit Koordinaten versehen sein
- Die Angaben müssen in Tabellenform gespeichert sein und die Koordinaten x,y in 2D und x,y,z in 3D umfassen  
→ Excel, CSV, TXT, DBF, ...
- Aufgaben beim Einbinden in GIS-Systeme
  - Richtiges Koordinatensystem
  - Richtige Koordinatenangabe
  - Tabellenform

# Koordinatenangaben



- Erstellung eines Vektorlayers (Punkt)
- Beibehaltung aller  
Attributinformationen aus der DB



# Koordinatenangaben

## Problem der räumlichen Referenz

- Räumliche Referenz ist im WGS84-System und wird mit Grad und Minuten (50°13') angegeben
- Für das Einfügen der Daten in GIS müssen Dezimalzahlen vorhanden sein

### Konvertierung von GMS-Daten (Grad, Minuten, Sekunden) in Dezimalgrad

$$\text{Dezimalgrad} = \text{Grad} + (\text{Minuten}/60) + (\text{Sekunden}/3600)$$

- Excel-Datenaufbereitung  
„Text in Spalten“-Werkzeug verwenden



Bearbeiten Sie die Aufgabe „Niederschlag“ im Übungsskript!



### 3. Sekundäre Metrik: Tabellen Verknüpfungen



# Datenbankschlüssel

## Primärschlüssel

- Identifiziert genau einen Datensatz
- Eindeutig identifizierbar, d.h. es darf in der Tabelle kein zweiter Datensatz mit demselben Primärschlüssel vorhanden sein.
- Erkennungsmerkmal: Feldname (Spaltenname) meist „...ID..“ genannt

ID	Vorname	Name	FK	...
0	Homer	Simpson	0	
1	Marge	Simpson	0	
2	Donald	Duck	1	
...				
n				

# Datenbankschlüssel

## Fremdschlüssel (foreign key = fk)

- sind dazu da einem Datensatz in Tabelle 1, einen oder mehrere Datensätze aus einer anderen Tabelle 2 zuzuordnen
- Der Fremdschlüssel in Tabelle 1 entspricht dabei dem eindeutigen Primärschlüssel Tabelle 2

ID	Straße	Hausnummer	Ort	...
0	Donat-Street	13	Springfield	
1	Daisy-Street	14	Entenhausen	
2	Edlmairstraße	6	Deggendorf	
...				
n				

# Datenbankschlüssel

Tabelle 1

ID	Vorname	Name	FK	...
0	Homer	Simpson	0	
1	Marge	Simpson	0	
2	Donald	Duck	1	

Tabelle 2

ID	Straße	Hausnummer	Ort	...
0	Donat-Street	13	Springfield	
1	Daisy-Street	14	Entenhausen	
2	Edlmairstraße	6	Deggendorf	

## Verknüpfung beider Tabellen

→ Homer und Marge Simpson wohnen gemeinsam in der Donat-Street 13 in Springfield





## Datenmodellierung – Relationenschema

- Die Beschreibung einer Basistabelle bezeichnet man in der Fachsprache auch als Relation.
- In einer Relation werden folgende Informationen gespeichert:
  - Name der Relation (Tabelle)
  - Liste von Merkmalen (Attribute)
  - Primärschlüsselfelder
  - Domänen (Einschränkungen des Wertebereiches)
  - Abhängigkeiten

Relation:  
Flächennutzung

ID	Geometrie	Größe	Umfang	Nutzung
0	Polygon	1.010	2.020	Wald
1	Polygon	...	...	...

Attribute:  
Flächennutzung  
(ID, Geometrie, Größe, Umfang, Nutzung)

Primärschlüssel (mit \*)  
Flächennutzung  
(ID\*, Geometrie, Größe, Umfang, Nutzung)

# Datenmodellierung – Relationenalgebra



## Relationenalgebra

Vereinigung

Durchschnitt

Differenz

Kartesisches  
Produkt

Projektion

Selektion

Joins



Innerer Join  
Outer Join  
Left Join  
Right Join



# Datenmodellierung – Relationenalgebra



## Relationenalgebra

Vereinigung

Durchschnitt

Differenz

Kartesisches  
Produkt

Projektion

Selektion

Joins

Tabelle 1

Vorname	Nachname
Homer	Simpson
Marge	Simpson
Bart	Simpson

Tabelle 2

Vorname	Nachname
Donald	Duck
Daisy	Duck
Bart	Simpson



Vorname	Nachname
Homer	Simpson
Marge	Simpson
Bart	Simpson
Donald	Duck
Daisy	Duck

# Datenmodellierung – Relationenalgebra

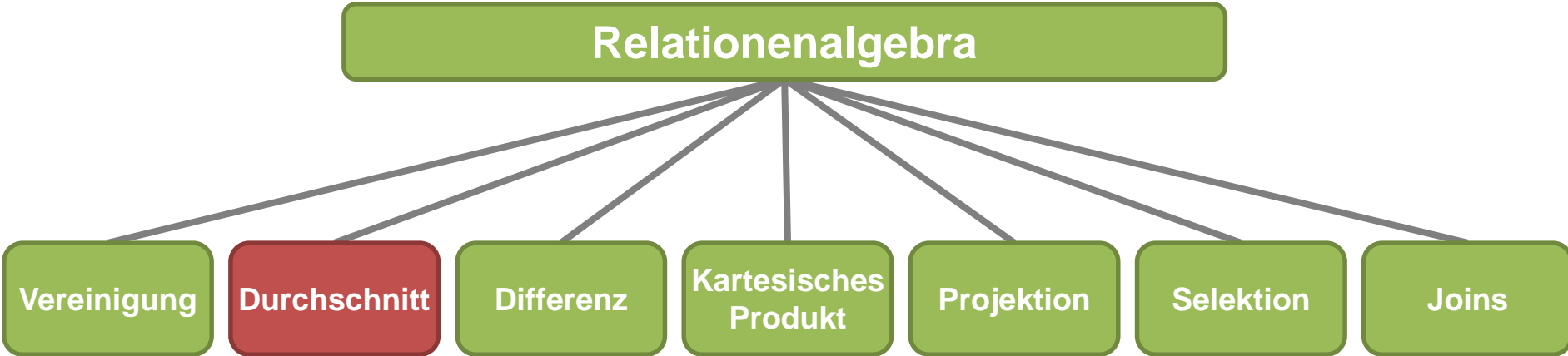
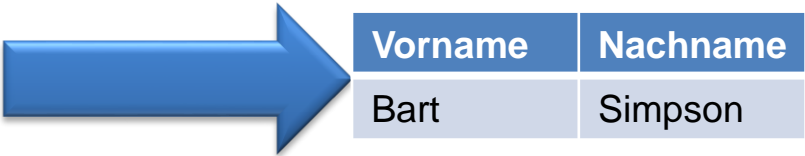


Tabelle 1

Vorname	Nachname
Homer	Simpson
Marge	Simpson
Bart	Simpson

Tabelle 2

Vorname	Nachname
Donald	Duck
Daisy	Duck
Bart	Simpson



# Datenmodellierung – Relationenalgebra

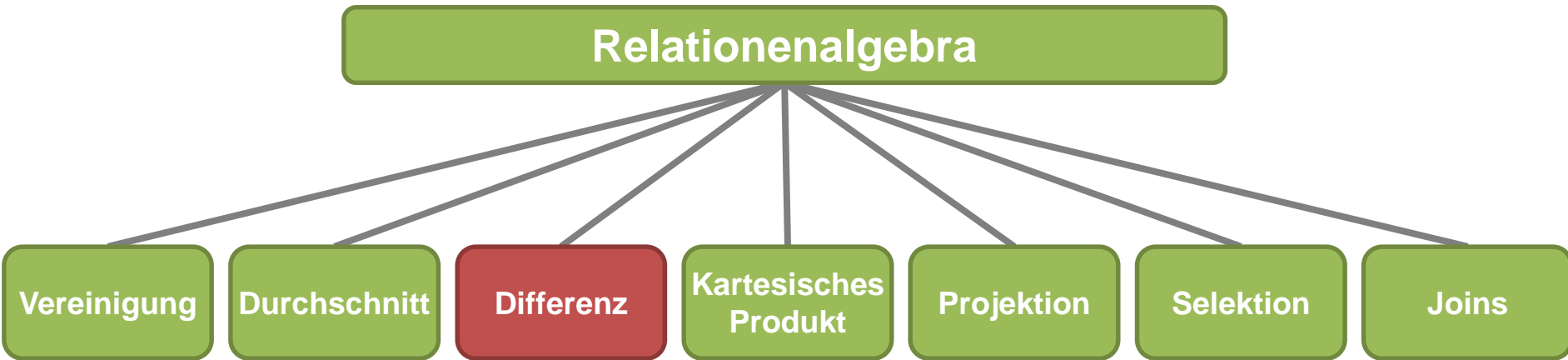



Tabelle 1

Vorname	Nachname
Homer	Simpson
Marge	Simpson
Bart	Simpson

Tabelle 2

Vorname	Nachname
Donald	Duck
Daisy	Duck
Bart	Simpson



Vorname	Nachname
Homer	Simpson
Marge	Simpson

# Datenmodellierung – Relationenalgebra

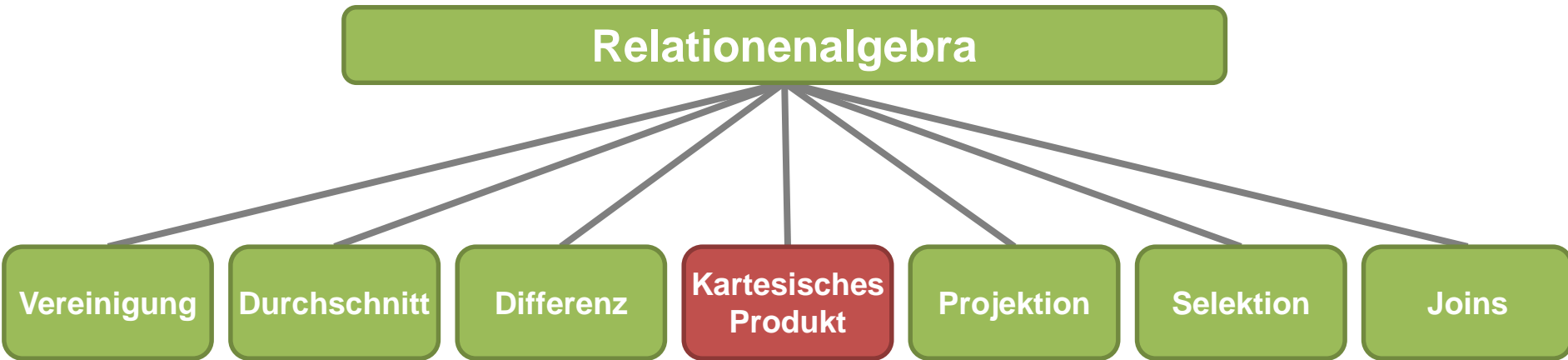


Tabelle 1

Vorname	Nachname
Homer	Simpson
Marge	Simpson
Bart	Simpson

Tabelle 2

Vorname	Nachname
Donald	Duck
Daisy	Duck
Bart	Simpson



Vorname	Nachname	Vorname	Nachname
Homer	Simpson	Donald	Duck
Homer	Simpson	Daisy	Duck
Homer	Simpson	Bart	Simpson
Marge	Simpson	Donald	Duck
Marge	Simpson	Daisy	Duck
Marge	Simpson	Bart	Simpson
Bart	Simpson	Donald	Duck
...	...	...	...

# Datenmodellierung – Relationenalgebra



## Relationenalgebra

Vereinigung

Durchschnitt

Differenz

Kartesisches  
Produkt

Projektion

Selektion

Joins

Tabelle 1

Vorname	Nachname
Homer	Simpson
Marge	Simpson
Bart	Simpson

Tabelle 2

Vorname	Nachname
Donald	Duck
Daisy	Duck
Bart	Simpson



Vorname

Homer

Marge

Bart

Wie lauten die Vornamen aller Familienmitglieder der Simpsons?



Bearbeiten Sie die Aufgabe „Tabellen“ im Übungsskript!



# Datenmodellierung – Relationenalgebra



## Relationenalgebra

Vereinigung

Durchschnitt

Differenz

Kartesisches  
Produkt

Projektion

**Selektion**

Joins

**Selektion von Daten → siehe nächste Woche!!!**

# Datenmodellierung – Relationenalgebra



## Relationenalgebra

Vereinigung

Durchschnitt

Differenz

Kartesisches  
Produkt

Projektion

Selektion

Joins

Innerer Join  
Outer Join  
Left Join  
Right Join

## Bei GIS wichtig

→ Verknüpfung von Geometriedaten (gespeichert in der Verktordatei (shapefile) und einer externen Tabelle

# Datenmodellierung – Relationenalgebra



## Joins

Tabelle 1

ID	Vorname	Name	Wohnort	...
0	Homer	Simpson	0	
1	Marge	Simpson	0	
2	Donald	Duck	1	

**Innerer Join**

Tabelle 2

ID	Straße	Hausnr.	Ort	...
0	Donat-Street	13	Springfield	
1	Daisy-Street	14	Entenhausen	
2	Edlmairstraße	6	Deggendorf	

## Datenmodellierung – Relationenalgebra



### Innerer-Join

Ergebnis

Tab1_ID	Tab1_Vorname	Tab1_Name	Tab1_Wohnort	Tab2_Straße	Tab2_Hausnr.	...
0	Homer	Simpson	0	Donut-Street	13	
1	Marge	Simpson	0	Donut-Street	13	
2	Donald	Duck	1	Daisy-Street	14	

Verarbeitet Beziehungen zwischen zwei Relationen mit mindestens einem gemeinsamen Attribut.

Das Ergebnis ist wiederum eine Tabelle die alle Attribute aus beiden Tabellen enthält. Die gemeinsamen Attribute werden nur einmal angeführt.

# Datenmodellierung – Relationenalgebra



## Joins

Tabelle 1

ID	Vorname	Name	Wohnort	...
0	Homer	Simpson	0	
1	Marge	Simpson	0	
2	Donald	Duck	1	
3	Daisy	Duck	Null	

Tabelle 2

ID	Straße	Hausnr.	Ort	...
0	Donat-Street	13	Springfield	
1	Daisy-Street	14	Entenhausen	
2	Edlmairstraße	6	Deggendorf	

**Outer Join**

→ Left-Outer-Join

→ Right-Outer-Join

## Datenmodellierung – Relationenalgebra



### Outer-Join (Left-Outer-Join)

Ergebnis

Tab1_ID	Tab1_Vorname	Tab1_Name	Tab1_Wohnort	Tab2_Straße	Tab2_Hausnr.	...
0	Homer	Simpson	0	Donut-Street	13	
1	Marge	Simpson	0	Donut-Street	13	
2	Donald	Duck	1	Daisy-Street	14	
3	Daisy	Duck	Null	Null	Null	

Auch der Datensatz aus Tabelle 1, der keinen Referenzdatensatz in Tabelle 2 hat wird in die Ergebnistabelle übernommen. Die fehlenden Attributwerte lauten NULL.



Bearbeiten Sie die Aufgabe „Demographie“ im Übungsskript!



## 4. Qualität von Geodaten



# Qualität von Geodaten

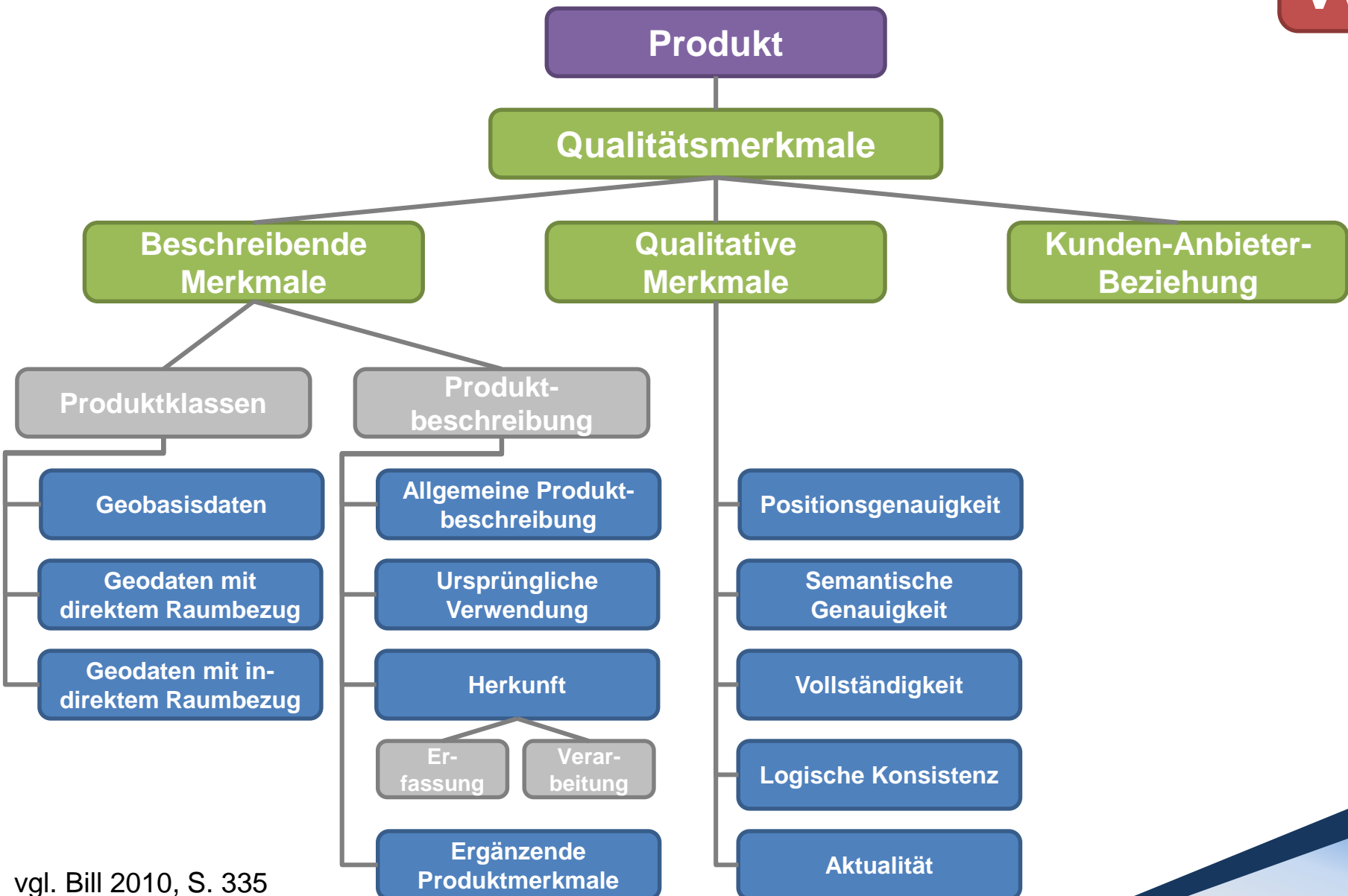
Beurteilung von Geodaten hinsichtlich ihrer Qualität fällt aufgrund des boomenden Marktes häufig schwer!

- Metainformationen sind vor allem für die Mehrfachnutzung und Weitergabe wichtig, um die Geodaten bewerten und zielgerichtet einsetzen zu können.
- Metainformationen: Informationen über die Informationen!



# Qualität von Geodaten DDGI-Qualitätsmodell

DDGI = Deutscher Dachverband für Geoinformation





## Qualität von Geodaten – Mögliche Fehlerursachen

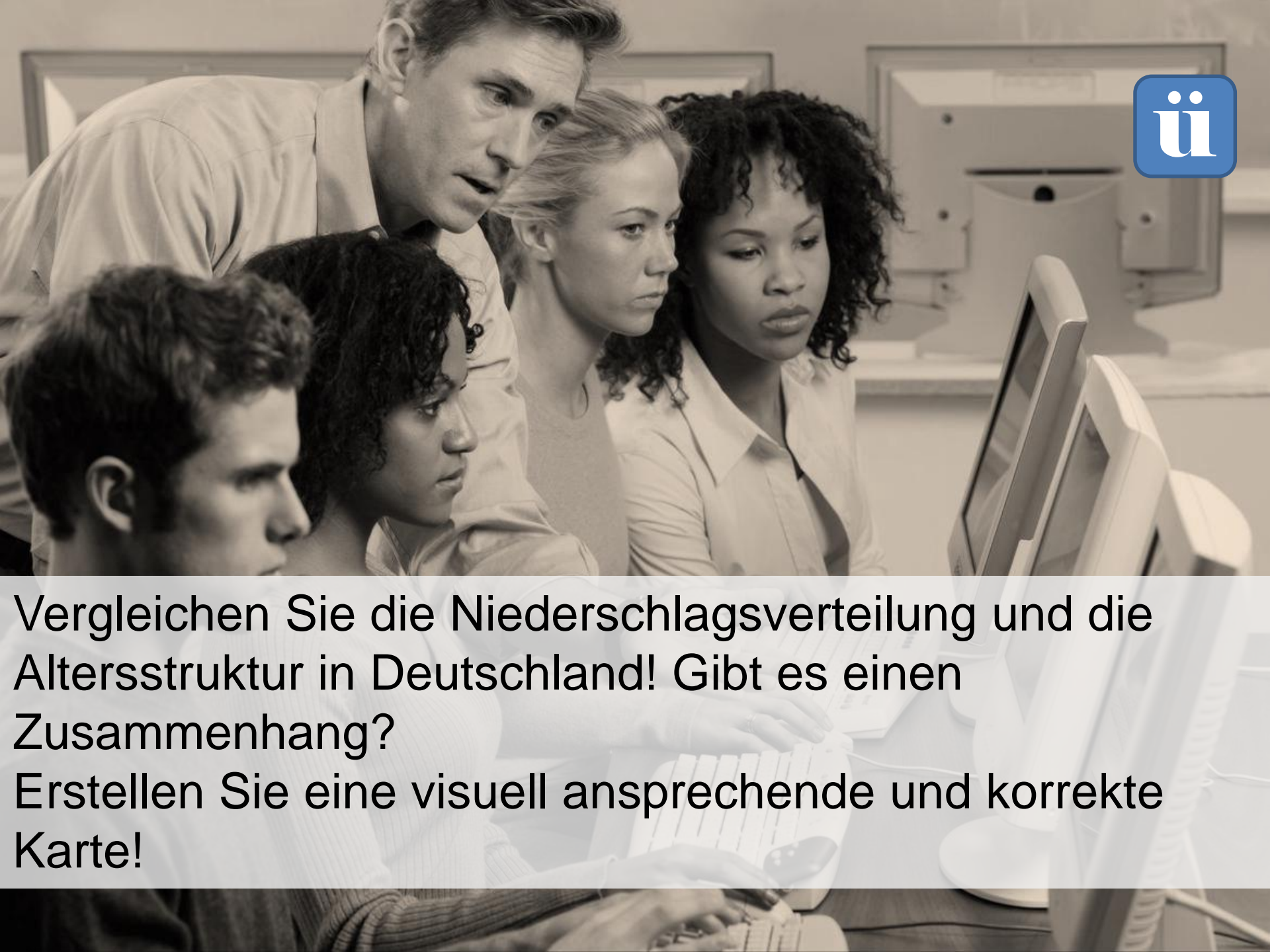
- Unterschiedliches Alter der Daten
- Verschiedene Dichte der Gebietsabdeckung (am Rand der Karte wurde ungenauer editiert oder war die Grundlage schlecht erfasst)
- Maßstab und Generalisierung
- Fehlende Gültigkeit und geringe Eignung der Daten
- EDV-technische Aufbereitung der Daten
- Verfügbarkeit Kosten und Urheberrechte

### Explizit für Geodaten

- Lage- und Positionsgenauigkeit (Geometrie)
- Inhaltsgenauigkeit (Attributinformation)
- Natürliche Variation der Daten (z.B. Messfehler, Interpretationsfehler und Auflösung)
- Fehlerhafte Verarbeitung (Rechenschärfe des PC)
- Fehlerhafte Annahmen (z.B. bei Berechnungen)
- Fehler in der Erfassung und Verarbeitung (z.B. Digitalisierung, Konvertierung und Klassifikation)



## 5. Übung



Vergleichen Sie die Niederschlagsverteilung und die Altersstruktur in Deutschland! Gibt es einen Zusammenhang?  
Erstellen Sie eine visuell ansprechende und korrekte Karte!



Prof. Dr. Roland Zink  
Fakultät Elektrotechnik und Medientechnik

Tel: +49 – 8551 – 91 764 – 28  
Email: roland.zink@th-deg.de

Edlmairstr. 6+8  
94469 Deggendorf

[www.th-deg.de/](http://www.th-deg.de/)

