



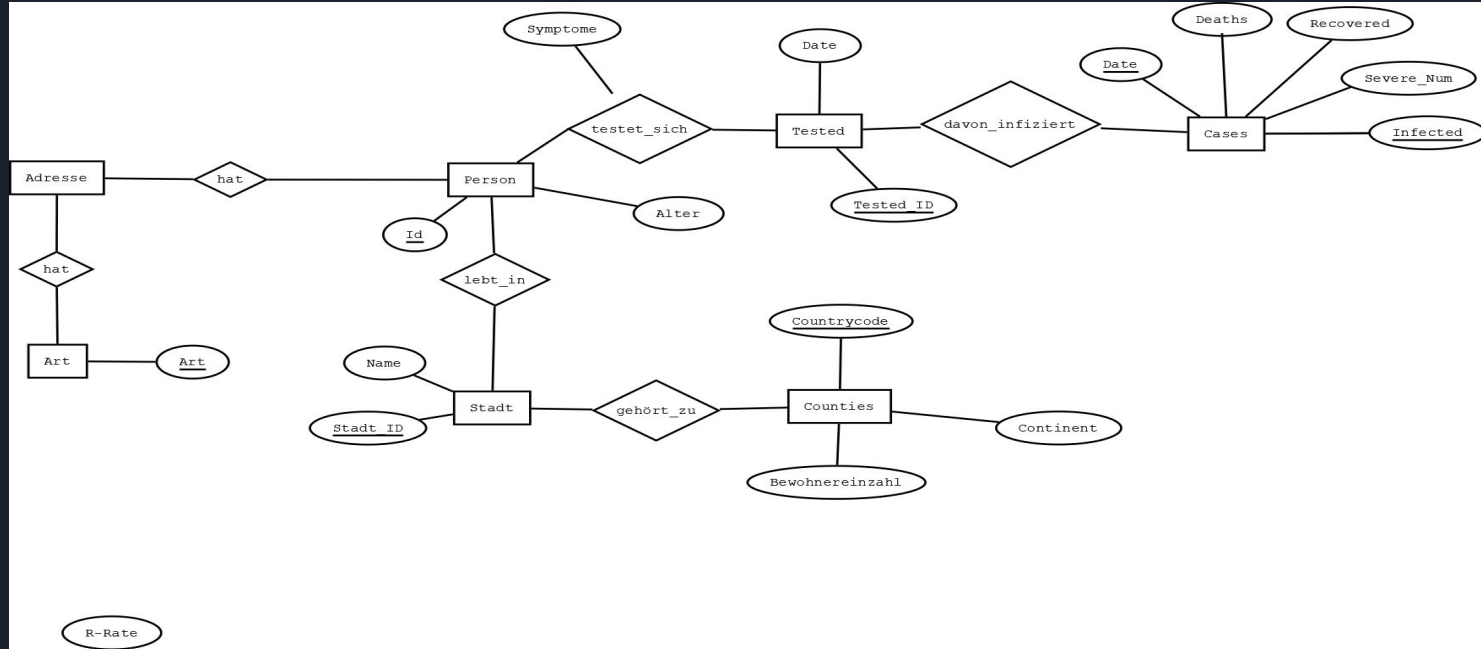
COVID-19 Datenbank

Phase 1: Modellierung

- Erstes Modell

Probleme:

- Nicht ausreichend Daten vorhanden
- Hoher Aufwand

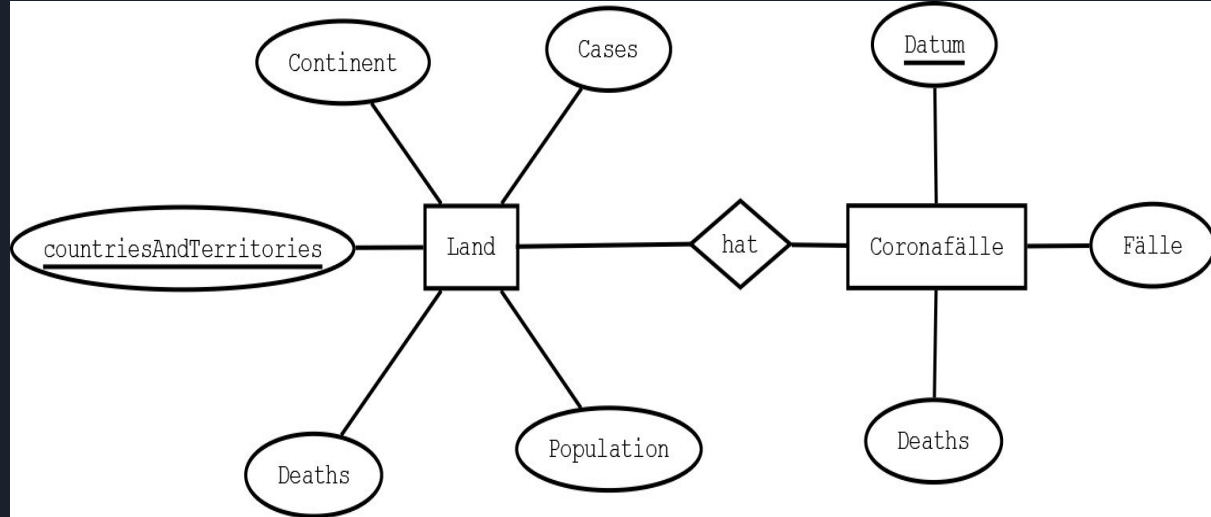


Phase 1: Modellierung

- Zweites Modell

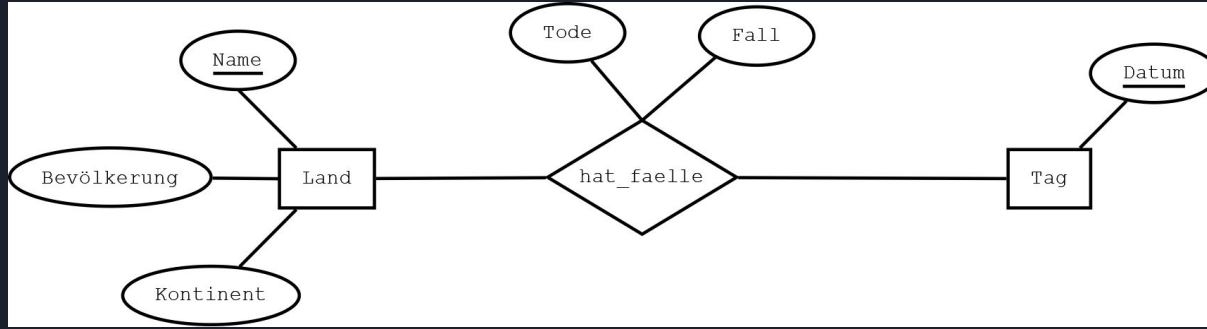
Probleme:

- Keine mögliche Ausgabe von Fällen in einzelnen Ländern



Phase 1: Modellierung

- Letztes Modell



- Simple ER-Modell zum gegebenen Datensatz
- Bevölkerung haben wir hinzugefügt



Phase 1: Modellierung

- Fazit

Was war an Ihrem Projekt herausfordernd?

- Was noch im Rahmen des Möglichen war

Welche Schwierigkeiten hat es gegeben und wie haben Sie diese gelöst?


- Wir haben uns Gedanken darüber gemacht welche Datenbank passen könnte und haben darauf unser ER-Modell aufgebaut

Was haben Sie gelernt?

- Dass man im ER-Modell schon mal Beispielabfragen machen könnte

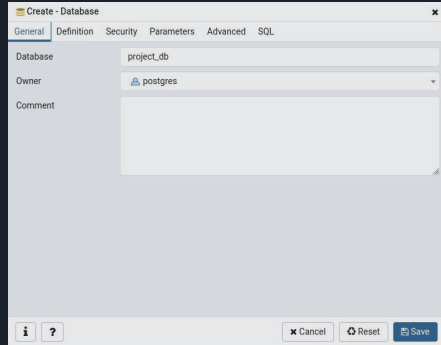
Welche Einsichten haben Sie gewonnen?

- Dass die vorgegebenen Datensätze ein Maßstab sind, wie wir die Datenbank gestalten sollen



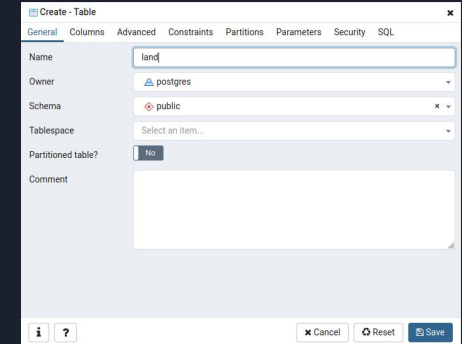
Phase 2: Datenverarbeitung

Die Datenbank

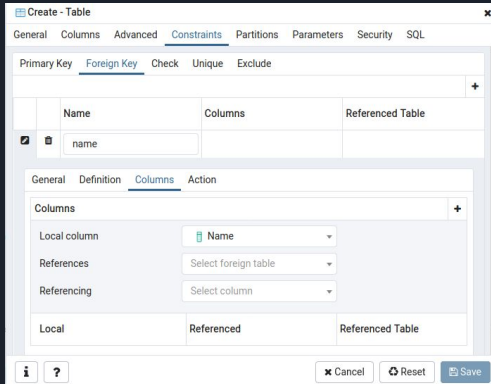


Erstellen der Datenbank

Erstellen von Tabellen nach relationalem Modell



Primärschlüssel und Fremdschlüssel einfügen



	Name	Data type	Length/Precision	Scale	Not NULL?	Primary key?
	Name	character varying			<input type="radio"/> No	<input checked="" type="radio"/> Yes

Datenimport

Problem: Datensatz fehlerhaft

```
145356 | | {  
145357 | |   "dateRep": "15/05/2020" }
```

Lösung:

- letzten zwei Zeilen löschen
- Klammern einfügen

```
145355 | | }  
145356 | | }
```

JSon kein geeignetes Format für die Datenbank



Konvertieren in .csv mit Python (Programm wurde mit hochgeladen)

Falsche Kodierung

Problem:

ERROR: extra data after last expected column
CONTEXT: COPY test, line 26: "Bonaire, Saint Eustatius and Saba"

Bonaire, Saint
Eustatius and Saba

ERROR: invalid byte sequence for encoding "UTF8": 0xe7 0x61 0x6f
CONTEXT: COPY test, line 51

Curaçao

Lösung:

In UTF-8 mithilfe von Python kodieren:

```
[record["countriesAndTerritories"].encode('utf-8')]
```

Negative Einträge / Doppelte Einträge

```
"dateRep": "03/06/2020",  
"day": "3",  
"month": "6",  
"year": "2020",  
"cases": "-766",  
"deaths": "107",  
"countriesAndTerritories": "France",  
"geoId": "FR",  
"countryterritoryCode": "FRA",  
"popData2018": "66987244",  
"continentExp": "Europe"
```

Lösung mit Python: negative Werte mit -1 multiplizieren

Gleiches Land und Datum aber zwei Einträge
dürfen keine abweichende Daten haben

Nur ein Datensatz übernehmen



Datum

ERROR: date/time field value out of range: "14/06/2020"



```
ALTER DATABASE "database" SET datestyle TO SQL,DMY;
```

```
ALTER USER "User" SET datestyle to SQL,DMY;
```




Fazit Datenverarbeitung

Auseinandersetzung mit PostgreSQL und
pgAdmin

Datenbereinigung

→ Alle mögl. Fälle betrachten

Geeignetes Programm für die
Datenbereinigung/Datenimport finden



Phase 3: Visualisierung & Datenanalyse

Visualisierung & Datenanalyse

Client

```
client on port:8080 connected
unformatted sql query: select%20*%20from%20hat_faelle
formatted sql query: select * from hat_faelle
```



Berlin

Browser
(Html)

```
//prepare query... remove the "%20" from query and swap %E3 and %C3 to
//get column metadata
ResultSetMetaData metaData = rs.getMetaData();
int columns = metaData.getColumnCount();

//get column names
String column_names = "";
for (int i = 1; i<=columns; i++)
    column_names = column_names + metaData.getColumnName(i) + ":";

//get column content
String column_content = "";
while (rs.next()) {
    for (int i = 1; i<=columns; i++)
        column_content = column_content + rs.getString(i) + ":";
}
```

Visualisierung & Datenanalyse

