Databases Autumn 2020

Data Analysis Project
P2: Data Integration

Visualizing Traffic density data and comparing them to air pollution- and meteorological data in Basel, London and Los Angeles

December 6, 2020

Pascal Kunz, Etienne Mettaz

1 Updated ER-Diagram

This section present the current ER Diagramms of the integrated data.

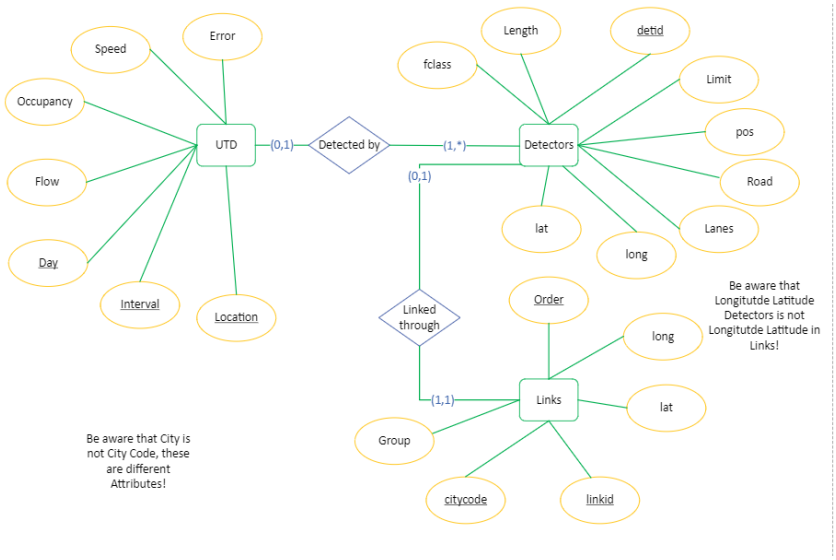

Figure 1: ER Diagramm of the utd19 traffic data

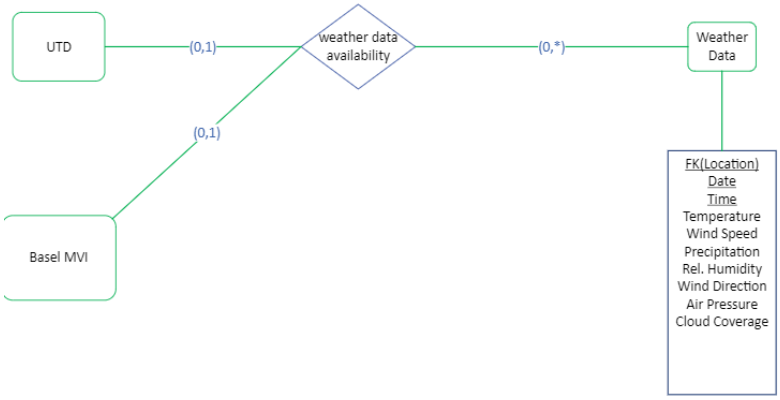

Figure 2: ER Diagramm of the weather datas. All attributes are written in a box for ease of read

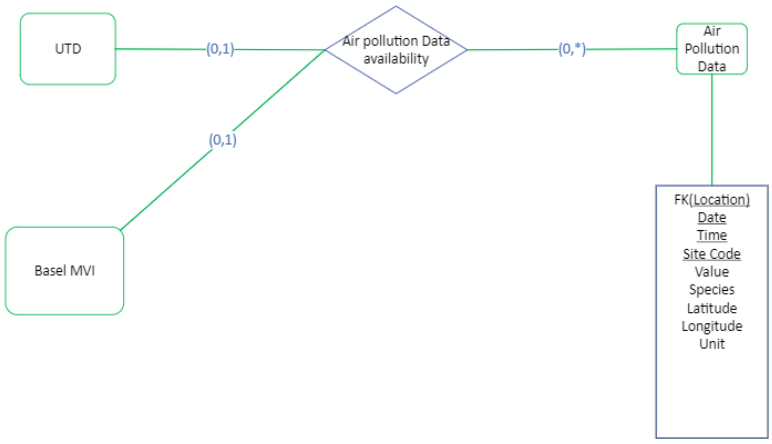

Figure 3: ER Diagramm of the air pollution datas. All attributes are written in a box for ease of read

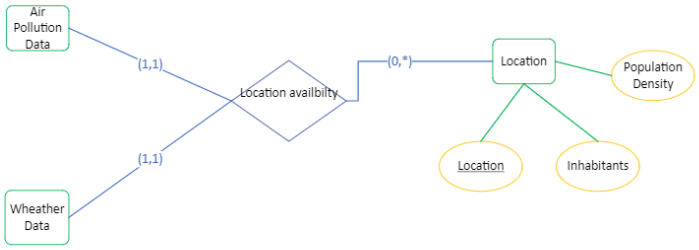

Figure 4: ER Diagramm of the location data

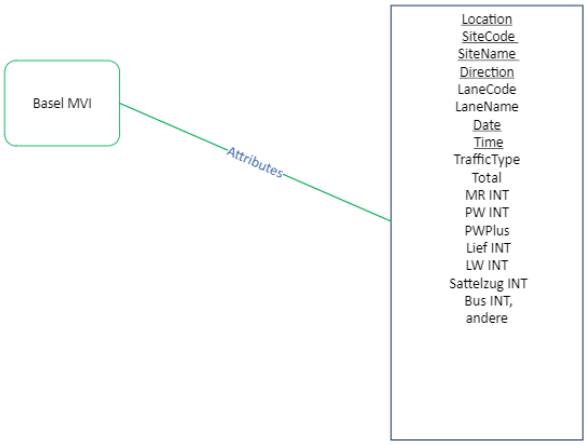

Figure 5: ER Diagramm of an additional traffic data source

2 Updated logical scheme

UTD19	(Date, Interval, DetID, Location, Flow, Occupancy, Speed)
Detectors	(DetID, City Code, Length, Road, Limit, Position, Lanes, Longitude, Latitude, fclass, Road)
Links	(LinkID, City Code, DetID, Order, Latitude, Longitude, Group)
Weather Data	(Date, Time, Location, Temperature, Wind Speed, Precipitation, Rel Humidity, Wind Direction, Air pressure, Cloud Coverage)
Air Pollution Data	(Date, Time, Site Code, Location, Species Latitude, Longitude, Value)
Basel MVI	(Location, Time, Day, SiteName, DirectionName, LaneCode, LaneName, Date, TimeFrom, Time To, DateTimeFrom, DateTimeTo, Year, ValuesApproved, ValuesEdited, Total, MR, PW, PW+, Lief, Lief+, Lief+Aufl., LW, LW+, Sattelzug, Bus, Andere)
Location	(Location, Inhabitants, Population Density)

3 Guide to replay the integration

We have decided to do the DataCleaning with Python using Pandas. Therefore we have created three folder with Code under p2// DataCleaning.

3.1 Weather Cleaning

The Weather Cleaning consists of four main files: 1 for each City and one file to unify them. The Datacleaning is done by running the four main functions of the respective files.

3.2 Air Pollution cleaning

The Air Pollution cleaning consists of five main files: 1 for each City, 1 for a Dictionary of Pollutant values and 1 file to unify them. The cleaned output files are being produced by running the four main functions of the respective files.

3.3 Traffic Cleaning

The Traffic Cleaning consists of three main files: the cleaning for the UTD Dataset, for the Basel MIV Dataset as well as for the links of the UTD Dataset. The cleaned output files are being produced by running the respective main functions.

3.4 Integration Code

Upon setting up the database, we can fill the schemes by running this code:

```
LOAD DATA INFILE 'Location.csv'
INTO TABLE location
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES ;

LOAD DATA  INFILE 'AQ_cleaned.csv'
IGNORE
INTO TABLE airpollution
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES ;

LOAD DATA  INFILE 'Weather_cleaned.csv'
IGNORE
INTO TABLE weather
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES ;

LOAD DATA  INFILE 'Detectors.csv'
```

```
IGNORE
INTO TABLE detectors
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES ;

LOAD DATA  INFILE 'LinksCleaned.csv'
IGNORE
INTO TABLE links
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES ;

LOAD DATA  INFILE 'utdcleaned.csv'
IGNORE
INTO TABLE UTD
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES ;

LOAD DATA  INFILE 'MVIcleaned.csv'
IGNORE
INTO TABLE mvi
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES ;
```