

Univ.-Prof. Dr.-Ing. Matthias Boehm
Graz University of Technology
Computer Science and Biomedical Engineering
Institute of Interactive Systems and Data Science
BMK endowed chair for Data Management

1 Data Management SS 2022: Exercise 01 – Data Modeling

Published: March 08, 2022 (last update: Mar 08)

Deadline: March 29, 2022, 11.59pm

This exercise on data modeling aims to provide practical experience in Entity-Relationship (ER) modeling, ER-relational mapping, and relational normalization. The expected result is a zip archive named `DBExercise01-<student_ID>.zip`, containing the partial results of the individual sub-tasks, submitted in TeachCenter.

1.1 ER Modeling (14/25 points)

Create an ER diagram in Modified Chen (MC) notation—including entity types, relationship types, attributes, cardinalities, and keys (create surrogate keys if natural unique identifiers are missing)—for managing the Graz Districts and Census Information¹. It is up to you if you use existing tools for data modeling, or draw this via presentation software or hand. There are multiple correct ways of modeling (alternatives, and level of details), however the diagram should capture the essential information of the following discourse:

- Graz has 17 *districts* (e.g., 6., Jakomini), each characterized by a unique `DistrictID`, one or multiple postal codes, a population count (as of 01/2022), an area (in km²), and a population density (population count / area).
- Every *street* (e.g., 1850, Inffeldgasse, 6) belongs to exactly one district, and has a unique Graz street code, and a street name.
- An *address* in Graz is characterized by its street, a street number, and an apartment number (where the apartment number might be N/A).
- *Schools* and *universities* have a unique name, an address, a phone number, and a type of educational institution.
- Every registered *person*—characterized by a first name, a last name, a gender (female, male, diverse), a day of birth (DoB), and age—has exactly one primary residence address, and an arbitrary number of secondary residence addresses. Furthermore, every person has a citizenship of at least one country; and every *country* is described by a unique ID, a unique three-letter country code, and a country name.

Partial Result: ERDiagram.pdf

¹Dataset link: https://github.com/tugraz-ids/datasets/tree/master/districts_graz/raw. A cleaned version of the dataset will be made available for Exercise 2 after additional data preparation and cleaning.

1.2 Mapping ER Diagrams into the Relational Model (11/25 points)

Create a relational schema—in third normal form—for the ER diagram designed in Task 1.1. This schema should include the relations and typed attributes, as well as all primary and foreign keys. Derived attributes can be excluded, and all unique attributes can be assumed to be always defined and functionally determine the remaining attributes of an entity type. It is up to you if you provide either a SQL script (`CreateSchema.sql`) with `CREATE TABLE` statements, or provide a text schema (`Schema.txt`) in the following text notation (one line per table):

```
<Table>( <Attribute 1>:<type>(PK), <Attribute 2>:<type>, ..., <Attribute n>(FK))
```

Here, PK and FK indicate primary and foreign keys, where multiple attributes with (PK) represent a composite primary key. If an attribute is both foreign key and (part of) a primary key, use `<Attribute m>(PK,FK)`. Please, adhere to this notation with unchanged parentheses and delimiters, and limit yourself to common data types (i.e., `int`, `numeric(p,s)`, `char(n)`, `varchar(n)`, or `date`) because this sub-task is subject to automated grading.

Partial Result: `Schema.txt`, or `CreateSchema.sql`