

---

# Analysis of the distribution of the population of Austria by altitude

*A Data Management Plan created using DMPonline*

**Creator:** David Wagner

**Affiliation:** Other

**Funder:** European Commission (Horizon 2020)

**Template:** Horizon 2020 DMP

**ORCID iD:** 0000-0002-5142-1054

**Project abstract:**

In order to get a better understanding of the altitude on which the Austrians live certain data had to be collected and transformed. At this point in time the analysis was only done at the "Bezirk" level of Austria. The final result is that every Bezirk in Austria was connected with the number of inhabitants in this Bezirk and furthermore the altitude of the main city in the Bezirk was added.

**Last modified:** 22-04-2019

**Copyright information:**

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

# Analysis of the distribution of the population of Austria by altitude - Initial DMP

---

## 1. Data summary

### Purpose

The purpose of the data generation is to enable analysis on which elevation Austrians typically live. As there is currently no data available that links the population of Austria on a granular level (for example Bezirk) to the altitude that the people live at.

### Data collected and reused:

- Population Data of Austria per Bezirk
  - <https://www.data.gv.at/katalog/dataset/3bfba412-7053-3a60-937a-8c3dd2c71294>
    - Unique ID of the data on data.gv.at: 3bfba412-7053-3a60-937a-8c3dd2c71294
  - Three CSV files:
    1. Bezirk Name mapping to internal code
      1. Name: OGD\_f0743\_VZ\_HIS\_GEM\_4\_C-GRGEMAKT-0
      2. Size: 7 Kb
      3. Location: ./data/raw
    2. Year mapping to internal Code
      1. Name: OGD\_f0743\_VZ\_HIS\_GEM\_4
      2. Size: 8 Kb
      3. Location: ./data/raw
    3. Actual Population data
      1. Name: OGD\_f0743\_VZ\_HIS\_GEM\_4
      2. Size: 40Kb
      3. Location: ./data/raw
- Altitude Data
  - from DBpedia
    - Accessed directly and stored in a pandas DataFrame
  - Missing values that were not found on DBpedia
    - Was looked up manually on Wikipedia
    - A csv file was created with the data
      - Name: MissingHM
      - Location: ./data/raw
      - Generated manually
      - Format: CSV
        - Non proprietary
      - Size: <2 Kb

### Data created:

- Comined table of Population of Austria in 2011 per Bezirk plus the altitude of the main city in the Bezirk.
  - Name: Population2011\_altitude
  - Location: ./data
  - Format: CSV
    - Non proprietary
  - Size: <9 Kb
  - DOI: <https://doi.org/10.5281/zenodo.2648511>

The data will be useful to anybody that wants to work with the elevation data from Austria.

## 2. FAIR data

**Metadata** that instructs on how to run the code is included, as well as metadata that defines the software requirements. Lastly, there is also metadata on the software architecture and a description of it.

The output data will be stored in the following repository:

- Zenodo
  - DOI: <https://doi.org/10.5281/zenodo.2648511>
  - License: [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)
  - Available since: 22nd of April 2019

The Jupyter Notebook, as well as all the documentation, input files and output were published on:

- Github with the integration to Zenodo
  - DOI: <https://doi.org/10.5281/zenodo.2648637>
  - License: [MIT License](#)
  - Available since: 22nd of April 2019

For the data files the **naming conventions** that were found at the data source were kept.

**Keywords** that relate to Austria, Population and Altitude were chosen.

**Versioning** will only occur every few years, as only then new population data will be available. Then changes to the input data will still not break the Jupyter notebooks, as the old data will still be included.

Versioning will be taken care of on Zenodo

**The metadata** that will be created is the following:

- README.md
  - explanation of the project and how to reproduce it/ run the code.
- requirements.txt
  - SW requirements that need to be installed in order to run the code
- metadata.xml
  - Metadata describing the tables and the origin of the project
- architecture.png
  - Picture displaying the architecture for preprocessing the data.
- description.txt
  - Short description of the architecture for preprocessing the data.

All of the data will be made openly available, as there are neither license issues nor is any personal data included.

The input data will be made available with the code on Github. However, it is also available on the [data.gv.at](#) homepage.

The output data will be made available on Zenodo.

The data will be uploaded to Github and Zenodo and DOI will be assigned.

The only software that is required is a browser, that allows to access Zenodo and Github. The data can even be read by a simple text editor, as it is just in csv format.

The output data will be stored in the following repository:

- Zenodo
  - DOI: <https://doi.org/10.5281/zenodo.2648511>
  - License: [Creative Commons Attribution 4.0 International](#)
  - Available since: 22nd of April 2019

The Jupyter Notebook, as well as all the documentation, input files and output were published on:

- Github with the integration to Zenodo
  - DOI: <https://doi.org/10.5281/zenodo.2648637>
  - License: [MIT License](#)
  - Available since: 22nd of April 2019

There are no restrictions on the data.

All of the data is stored as csv files, which are non proprietary.

Furthermore metadata that instructs on how to run the code is included, as well as metadata that defines the software requirements. Lastly, there is also metadata on the software architecture and a description of it.

No vocabulary is used.

The output data will have the following **license**:

- License: [Creative Commons Attribution 4.0 International](#)

The Jupyter Notebook, as well as all the documentation, input files and output are published with the following license:

- License: [MIT License](#)

There is no data embargo period.

Both the code and the data are **available since**: 22nd of April 2019

The data can be **reused** by third parties from the moment it is available, as defined by the licensing of it.

**Quality assurance** is possible, by checking against the input data and testing if any rows were lost. Furthermore, since we are dealing with the population of Austria it is also possible to test if the sums still add up to the total population of Austria.

The data will remain reusable of an indefinite timeframe.

### 3. Allocation of resources

- The main costs for making the data FAIR occur in the steps of making it interoperable, as here metadata,.. has to be created. This is a timeconsuming task that has to be covered. However, since I am a student nobody is paying for my time, therefore this part is also very affordable.
- The project is centered around data management from the beginning. The key contact for this is: David Wagner
- As long as the used services stay free there are very little direct costs for long term preservation. The main cost here also occurs due to the time that is required to keep documents and data up to date.
  - The potential value is that the future analysis is still possible.

### 4. Data security

#### Storage:

As the total size of all files is less than 1 Mb storage is not an issue.

#### Backup:

The data is stored locally for ease of use, but it is also saved on University Servers of TU Vienna.

- Backups are only made when the data changes, which does not happen often as we are dealing with altitude data of cities - both of which do not really change a lot over time. The only attribute that is prone to change is the number of inhabitants, but also this only happens every few years.

#### Recovery:

In case of fatal issues where the data is lost locally it will still be possible to access it from the original sources (data.gv.at and DBpedia).

#### Risks to data security:

- As we are not dealing with any personal data there is little risk of personal data being published.
- However, there is still the risk that data could be changed or deleted. However this is also of little concern since all the data that was used is publicly available so it can easily be checked if anything was changed.
  - Crosschecking with the original data sources is also a good option to continuously check the health of the data.

#### Access to the data:

- Access to the data (and the backup) will only be granted for the people that also conducted the experiment.

#### Safe transfer of data:

- Will be granted by using the VPN network of the Technical University of Vienna.

### 5. Ethical aspects

#### Ethical

No personal data is used for the experiment, so there are no real ethical issues that could arise.

- This also means that there is not sensible data that needs to be anonymised.

#### Legal

Data from data.gv.at:

- This is the main data source.
- Everything is under: Creative Commons Attribution License 3.0

Data from DBpedia:

- This is the second main data source where the altitude of cities is queried from.
- Everything is licensed under: [Creative Commons Attribution-ShareAlike 3.0 License](#) and the [GNU Free Documentation License](#)

Therefore there are not issues with reusing the raw data as well as the data that is produced by the experiment.

## 6. Other

Question not answered.