# Project Documentation

## D-MIG

*17.02.2024*


Adian Dawuda

*adian.dawuda@plus.ac.at*

Felix Schachtschneider

*felix.schachtschneider@plus.ac.at*

# Table of Contents

# 1 Introduction/Abstract

This project provides a Spatial Data Infrastructure (SDI) that allows for the analyses and visualization of annual domestic migration flows within Germany. We utilize statistical migration data from the GENESIS-Online platform and federal-state boundary data from Eurostat to map these movements. The acquisition, integration, and storage of migration statistics and spatial information is conducted programmatically using the Python programming language and a PostGIS spatial database. This approach reduces manual effort and increases the overall repeatability and efficiency of the workflow. Our data is published using a Web Feature Service (WFS) created by GeoServer. An example dashboard visualization is created using ArcGIS Insights. This dashboard aims to provide an interactive platform for our stakeholders to explore an excerpt of the data and may serve as inspiration for further custom applications.

In this document, we first address our motivation and the potential stakeholders for our project. Next, the architecture of our spatial data infrastructure is showcased and details regarding the theory and implementation of each component are described in detail. Finally, we summarize the results and provide an outlook of potential future developments. Additionally, the appendix contains information about the structure and management of D-MIG.

# 2 Motivation

Internal migration plays a significant role in shaping the demographic landscape of Germany. This phenomenon, which influences population size, age, and sex distribution, has far-reaching implications for planning and policy-making at regional and national levels. Historically, the migration between and within East and West Germany post-reunification has led to marked demographic shifts, especially in the eastern regions, highlighting the significance of internal migration on population dynamics(Heiland, 2004; Kühntopf & Stedtfeld, 2012; Sander, 2014). Additionally, the role of international migration may indirectly influence demographics through changes in fertility rates (Sobotka, 2008).

In his foundational work, Ravenstein explores the interrelations between migration patterns and demographic changes (Corbett, 2003). Ravenstein identifies economic development and spatial inequalities as key drivers of migration. This remains relevant today and can help

to understand internal migration dynamics. In the current demographic landscape, characterized by declining fertility rates and increasing life expectancy, the impact of migration on demographic distribution becomes even more pronounced. This trend can be observed across Germany and Europe (Bujard, 2011; Bundeszentrale für politische Bildung, 2021; Wenau et al., 2019).

# 3 Target Audience/Stakeholders

The primary stakeholders for this project are German policymakers and government officials who are directly involved in managing and understanding internal migration within Germany. This group includes the fictitious stakeholders Martin Migration and Ms. Berta Bamf (these stakeholders simply serve as examples):

**Mr. Martin Migration**

Job: Assistant Migration Officer

Needs/Goals: To monitor migration movements, utilizing the latest data and analytical tools to inform migration policy and support decision-making processes.

**Ms. Berta Bamf**

Job: Official in Charge

Needs/Goals: To control migration movements and gain a deep understanding of migration patterns over time to implement effective migration management strategies and policy interventions.

This SDI specifically serves as a tool for our stakeholders to explore, understand, and act on migration trends across different federal states within Germany. Furthermore, our stakeholders benefit from the easily reproducible results and well-documented, automated workflow.

# 4 SDI Architecture

To meet the needs of our envisioned SDI, an intuitive, modular, and performant architecture is designed (Figure 1). This SDI makes use of a variety of technologies, standards, and software implementations. To provide a transparent and open infrastructure we focus on

utilizing open-source software. However, the use of proprietary software for visualization purposes is also demonstrated.

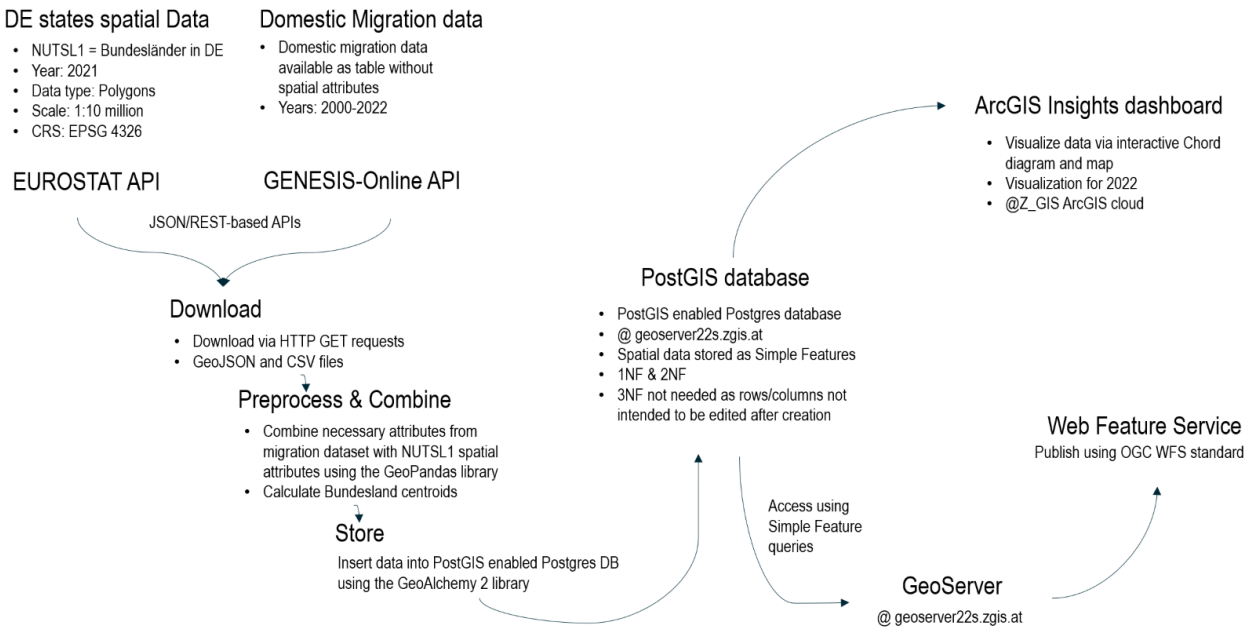



*Figure 1: D-MIG SDI architecture diagram*

German domestic migration data (section 5) and German federal state spatial polygon data (section 6) are sourced from the EUROSTAT and GENESIS-Online platforms. both the EUROSTAT and GENESIS-Online platforms offer JSON/REST-based APIs. The two scripts Eurostat.py and genesis.py download the data from the respective APIs (section 7). The script process.py then preprocesses and inserts the final data into a PostGIS-enabled Postgres database (section 8). Additionally, metadata of each final dataset is created (equivalent to one table in the database) (section 9) The results stored in the database are then published using GeoServer, making them available as an OGC Web Feature Service (section 10). Finally, the results are also interactively visualized as an ArcGIS Insights dashboard, which connects directly to the database (section 11).

# 5 German domestic migration data

In Germany, state-level migration data is collected by each federal state and summed up to generate a dataset of the entire domestic migration within Germany. To access the data, we make use of the GENESIS-Online platform. The Gemeinsames neues statistisches Informationssystem (GENESIS) is a metadata-based statistical information system

developed by the German federal and state statistical offices (Statistisches Bundesamt, 2023). GENESIS-Online is the official web platform that provides access to the GENESIS database. Additionally, registered users (registration is free of charge) gain access to additional services including a JSON/REST-based API (Statistisches Bundesamt, 2023). This allows users to define specific queries and download the data in a variety of formats. We make use of this API to access table 12711-0022 which contains yearly domestic migration statistics from 2000 until 2022 (Statistisches Bundesamt, 2024).

# 6 German federal state spatial data

As the GENESIS-Online German domestic migration data does not contain spatial attributes, these are separately sourced from the Nomenclature of Territorial Units for Statistics (NUTS). NUTS was introduced in 2003 and is a hierarchical system used by the European Union to divide the economic territory of its member states into regions at three different levels of detail: NUTS 1, NUTS 2, and NUTS 3 (Figure 2). The higher the level, the more fine-grained the regions are (with some exceptions like Luxembourg and Cyprus). In Germany, the NUTS 1 regions correspond directly to the 16 federal states (European Commission – Eurostat/GISCO, n.d.), which are also present in the domestic migration data. The EUROSTAT web platform provides access to the NUTS data as spatial files. In addition to the web interface, the platform also offers a JSON/REST-based API, through which the data can be queried and downloaded free of charge. We make use of this API for our automated workflow.

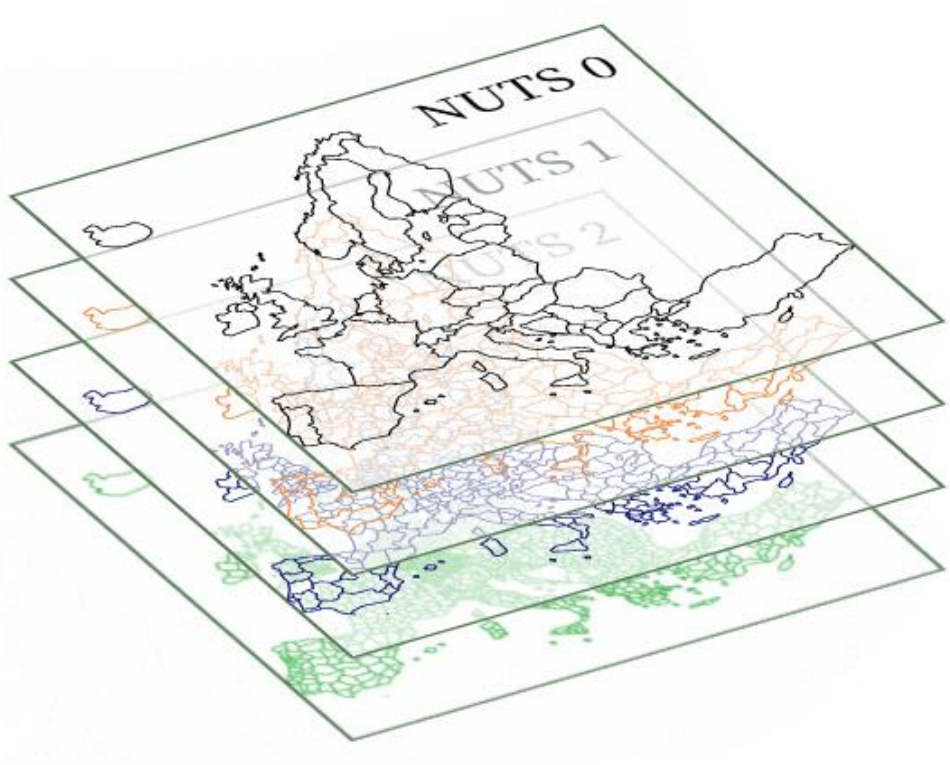



*Figure 2: NUTS levels (European Commission – Eurostat/GISCO, n.d.)*

# 7 Automated processing workflow

The process of downloading, preprocessing, and inserting the data into a PostGIS database is automated by developing a custom Python workflow consisting of multiple scripts. The following section describes the functionalities of each Python script. All code can be found on GitLab (git.sbg.ac.at/s1093093/dmig ) including detailed inline documentation.

## main.py

This script is responsible for the high-level automation of the entire workflow. It executes the three scripts eurostat.py, genesis.py, and process.py consecutively. If any execution errors occur, the exception output is printed.

Note: The script assumes the eurostat.py, genesis.py, and process.py scripts are located in the same directory.

## eurostat.py

This script makes use of the Eurostat GISCO (Geographical Information System of the Commission) REST-based API to download NUTS (Nomenclature of Territorial Units for

Statistics) Level 1 regions for 2021 (latest at the time of writing). In Germany, these regions are equivalent to the federal states. The regions are requested using an HTTP GET request and received as polygon features in the EPSG:3857 coordinate reference system. These features are then written to a GeoJSON file in the same directory. The default filename is set to NUTS_L1.geojson.

Third-party dependencies:

- requests

## genesis.py

This script makes use of the GENESIS-Online REST-based API to download German domestic migration data for a specified year. The data is represented at the state level. To access the API a GENESIS-Online account is needed. An account can be registered free of charge at https://www-genesis.destatis.de/genesis/online?Menu=RegistrierungForm. The data is requested by defining the dataset and the required features and sending an HTTP GET request. The data is received in tabular format and saved as a CSV file in the same directory as the script. The default filename is set to dmig[year].csv (e.g., dmig2022.csv).

Third-party dependencies:

- requests

## process.py

This script is responsible for processing the data. First, the migration data CSV file and NUTS GeoJson file are read and filtered to exclude unnecessary attributes (e.g., nationality, sex, other NUTS levels, geometries outside of Germany). In addition to the federal state polygon geometries, the centroids of each state are also calculated to enable more visualization possibilities. The filtered datasets are then merged together into a single dataset and exported to a PostGIS-enabled PostgreSQL database.

Third-party dependencies:

- geopandas
- pandas
- geoalchemy2

- shapely
- sqlalchemy

### parameters.py

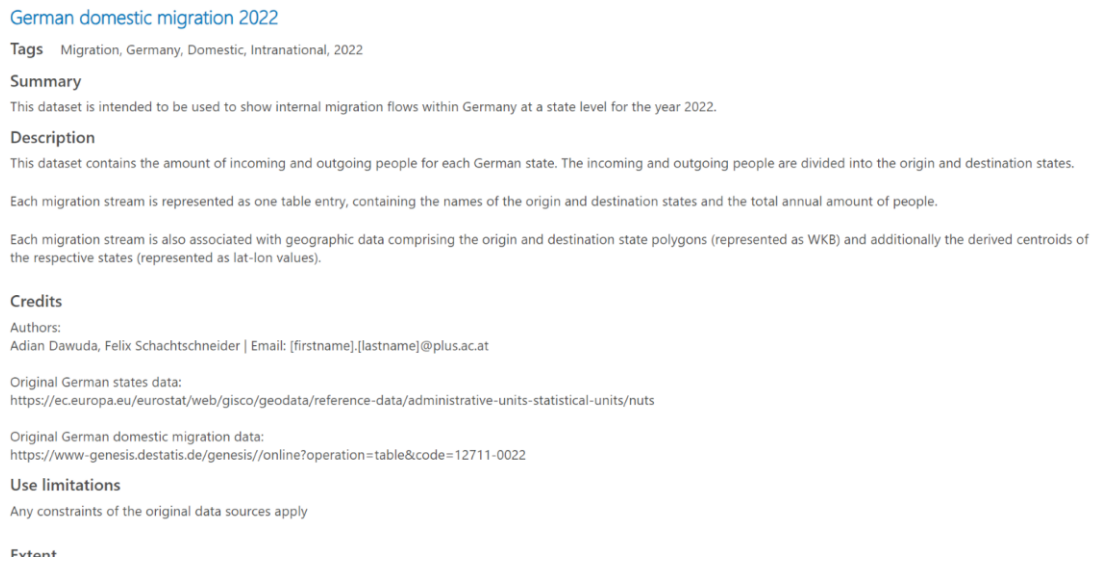
This Python file serves as a location to store all user-defined parameters needed to execute the workflow. This includes the year for which the domestic migration data is to be used and the connection parameters needed to insert the data into a PostGIS database.

## 8 PostGIS spatial database

PostGIS is an open-source extension that can run on a PostgreSQL database, turning it into a spatial database. This is done by adding geographic capabilities such as storing, querying, manipulating, and indexing spatial data. We make use of a PostGIS database running on Z_GIS servers to store the D-MIG datasets (one per year) with spatial attributes stored as Simple Features. Each D-MIG table adheres to 1NF and 2NF. Adherence to 3NF was not prioritized in the scope of this project, as the datasets are not originally intended to be edited after executing the Python workflow.

## 9 Metadata

For each year from 2000 until 2022 metadata adhering to the ISO 19115 and 19139 standards is documented. Figure 3 shows an excerpt of the D-MIG 2022 dataset's metadata.


German domestic migration 2022

**Tags** Migration, Germany, Domestic, Intranational, 2022

**Summary**

This dataset is intended to be used to show internal migration flows within Germany at a state level for the year 2022.

**Description**

This dataset contains the amount of incoming and outgoing people for each German state. The incoming and outgoing people are divided into the origin and destination states.

Each migration stream is represented as one table entry, containing the names of the origin and destination states and the total annual amount of people.

Each migration stream is also associated with geographic data comprising the origin and destination state polygons (represented as WKB) and additionally the derived centroids of the respective states (represented as lat-lon values).

**Credits**

Authors:
Adian Dawuda, Felix Schachtschneider | Email: [firstname].[lastname]@plus.ac.at

Original German states data:
https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts

Original German domestic migration data:
https://www-genesis.destatis.de/genesis//online?operation=table&code=12711-0022

**Use limitations**

Any constraints of the original data sources apply

**Extent**

*Figure 3: excerpt of the D-MIG 2022 dataset's metadata*

This metadata can be found in XML, HTML, and PDF form in the project's GitLab repository. Additionally, metadata of the original data sources are also uploaded here.

# 10 Geoserver WFS

GeoServer is a Java-based, open-source server used to publish spatial data on the web (GeoServer, 2024). Furthermore, it implements OGC protocols such as the Web Feature Service (WFS), Web Map Service (WMS), and Web Coverage Service (WCS).

We make use of GeoServer to publish our preprocessed data from the PostGIS database as a Web Feature Service (WFS). GeoServer accesses the spatial data from the database using Simple Feature queries and then publishes it as a globally available WFS. The WFS protocol is particularly useful in our context as it allows our stakeholders to interact with our data in a variety of software programs and even build custom applications around the data. However, it must be noted that some software platforms do not provide native support for OGC-compliant WFS, a prominent example being ArcGIS Insights.

# 11 Visualization

To visualize D-MIG data we utilized ArcGIS Insights to create an interactive dashboard (Figure 4). The main dashboard elements are the interactive chord diagram and map, both of which are linked together. This visualization is part of the D-MIG SDI and gives stakeholders a sample implementation using our data. The outer circle of the chord diagram contains the German federal states and the connections between the states, also called arcs, represent the migration movements. Our dashboard additionally includes background information about DMIG, a link for further information, and contact information. A current limitation of ArcGIS Insights is its lack of responsive design, a feature that is anticipated to be addressed in future updates.

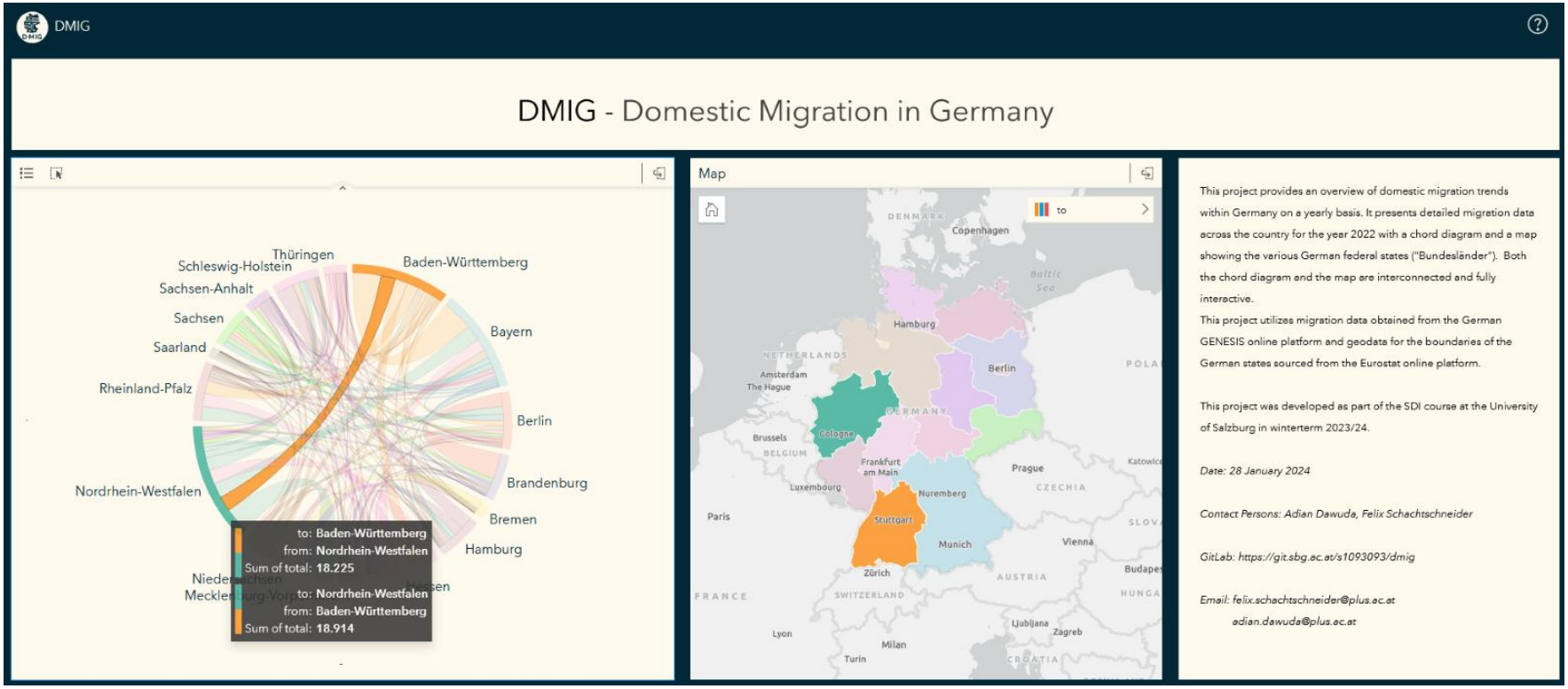



*Figure 4: ArcGIS Insight dashboard visualization*

# 12 Conclusion

Overall, this project successfully implements all of the planned features. The created SDI provides our stakeholders with the data to spatially analyze and visualize annual domestic migration flows within Germany. A sample visualization is also showcased as part of the SDI. The open-source data acquisition, storage, and publishing workflow makes our work widely accessible and transparent. Furthermore, our highly automated workflow from data acquisition to storage allows for easily reproducible results with very little change needed to alter the timeframe of interest.

In the case of positive evaluation from our stakeholders and funding for future work, further improvements to D-MIG could be made, such as increased modularity to incorporate other migration statistics and attributes into the workflow, further optimization of the database (possibly 3NF compliant), increased automation of the entire workflow, and increased accessibility (e.g. hosting Jupyter notebooks on a cloud platform).

## References


Bujard, M. (2011). *Family Policy and Demographic Effects: The Case of Germany*. https://www.researchgate.net/publication/259745491

Bundeszentrale für politische Bildung. (2021). *Daten-report 2021 Ein Sozial-bericht für die Bundesrepublik Deutschland WZB*.

Corbett, J. (2003). *CSISS Classics - Ernest George Ravenstein: The Laws of Migration*.

European Commission – Eurostat/GISCO. (n.d.). *Administrative Units / Statistical Units: NUTS*. Retrieved February 17, 2024, from https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts

GeoServer. (2024). *GeoServer User Manual*. https://docs.geoserver.org/latest/en/user/

Heiland, F. (2004). Trends in East-West German migration from 1989 to 2002. *Demographic Research*, *11*, 173–194. https://doi.org/10.4054/DemRes.2004.11.7

Kühntopf, S., & Stedtfeld, S. (2012). *Wenige junge Frauen im ländlichen Raum: Ursachen und Folgen der selektiven Abwanderung in Ostdeutschland*. www.bib-demografie.de

Sander, N. (2014). Internal migration in Germany, 1995-2010: New insights into East-West migration and re-urbanisation. *Comparative Population Studies*, *39*(2), 217–246. https://doi.org/10.12765/CPoS-2014-04en

Sobotka, T. (2008). Overview chapter 7: The rising importance of migrants for childbearing in Europe. In *Demographic Research* (Vol. 19, pp. 225–248). https://doi.org/10.4054/DemRes.2008.19.9

Statistisches Bundesamt. (2023). *Was bietet GENESIS-Online?* https://www-genesis.destatis.de/genesis/online?Menu=Hilfe

Statistisches Bundesamt. (2024). *Wanderungen zwischen den Bundesländern: Deutschland, Jahre, Nationalität, Geschlecht, Herkunfts-Bundesland, Ziel-Bundesland*. https://www-genesis.destatis.de/genesis//online?operation=table&code=12711-0022

Wenau, G., Grigoriev, P., & Shkolnikov, V. (2019). Socioeconomic disparities in life expectancy gains among retired German men, 1997-2016. *Journal of Epidemiology and Community Health*, *73*(7), 605–611. https://doi.org/10.1136/jech-2018-211742

# Appendix: Project management

This section provides an overview of the structure of the D-MIG project and the workflow we followed. Previous document versions and further details such as time sheets and other project management documents can be found on GitLab.

## Work Breakdown Structure

The Work Breakdown Structure (Figure 5) for this project provides an overview of D-MIG's following work packages: Project Management (WP1), Literature Review & Methodology (WP2), Data Acquisition & Pre-Processing (WP3), Data Processing & Analysis (WP4) and Dissemination (WP5). Each work package (WP) is further subdivided into separate work tasks (WT).

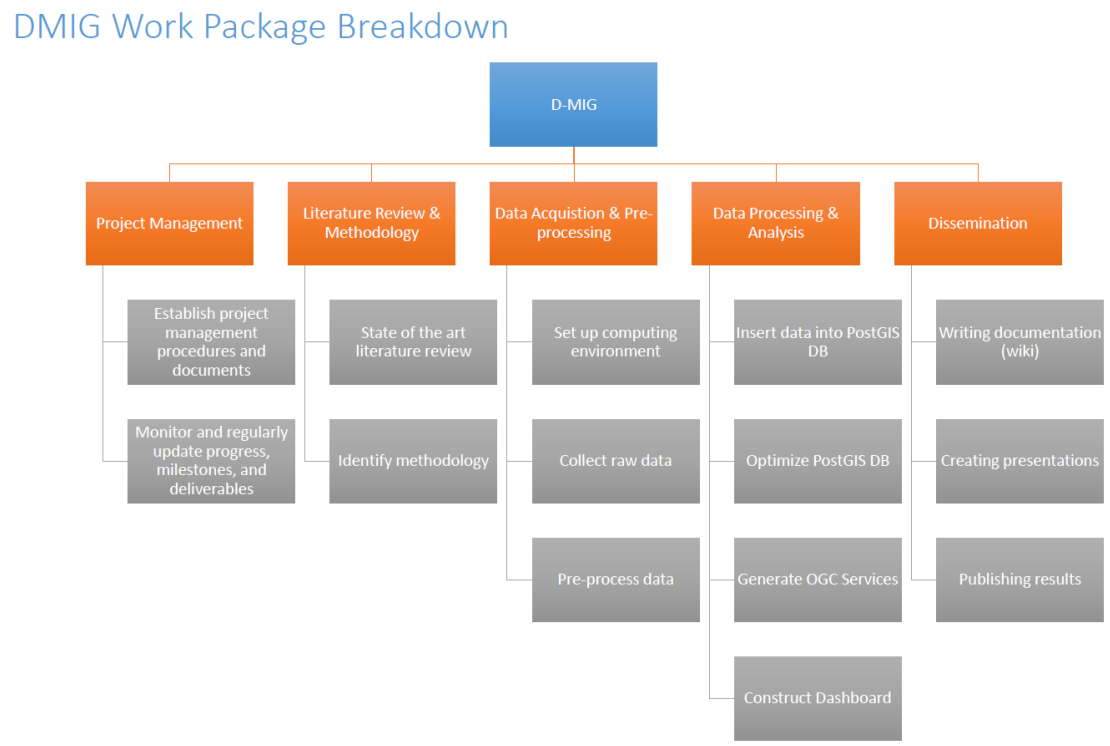



*Figure 5: D-MIG Work Breakdown Structure*

## Work Package 1: Project Management

This WP focuses on project management from October 3, 2023, to January 31, 2024. The main goal is to ensure that the project is completed according to schedule and all progress

is regularly documented. The tasks include setting up the project management environment (GitLab), creating all documents, and regularly monitoring and updating the project's progress, milestones, and deliverables.

### Work Package 2: Literature Review & Methodology

This WP focuses on reviewing the state-of-the-art literature and deciding on a methodology. It spans from October 3, 2023, to November 19, 2023. The main goal is to search for and organize important literature or information, to gain an overview of the state-of-the-art. Based on this we then decide on the specific methods used for this project.

### Work Package 3: Data Acquisition & Pre-Processing

This WP spans from November 20, 2023, to December 24, 2023. The main objective is to identify all necessary data and develop an automated download and preprocessing workflow using Python. This includes the tasks of setting up a computing environment, collecting raw data, and preprocessing the raw data.

### Work Package 4: Data Processing & Analysis

This WP spans from December 18, 2023, to January 28, 2024. The main objective is to construct a spatial data infrastructure for the collected and preprocessed data. This includes inserting preprocessed data into a PostGIS DB, optimizing the PostGIS DB, generating OGC web services, and constructing an ArcGIS Insights Dashboard.

### Work Package 5: Dissemination

This final WP spans from October 30, 2023, to January 31, 2024. Its goal is to publish the results achieved in this project. This includes describing the project in a documentation, creating two presentations, and publishing the code used for D-MIG.

## Gantt Chart

The Gantt Chart in Figure 6 provides an overview of all WPs and WTs with the estimated time, responsibility, start and end date, status, and weekly tasks.

| | |
|---|---|
| **Project Name** | Domestic Migration in Germany |
| **Project Acronym** | DMIG |
| **Project Manager** | Adian Dawuda, Felix Schachtschneider |
| **Project Deliverable** | Dashboard, Code, and Documentation |
| **Start Date** | 3.10.2023 |
| **End Date** | 31/01/2024 |
| **Overall Progress** | 100 % |

| WP | Time Estimate (hours) | Tasks | Responsible | Start | End | Status |
|---|---|---|---|---|---|---|
| **1** | **20 (10%)** | **Project Management** | | | | |
| 1,1 | 10 | Establish project management procedures and documents | Adian, Felix | KW40 | KW46 | Complete |
| 1,2 | 10 | Monitor and regularly update progress, milestones, and deliverables | Adian, Felix | KW40 | KW05 | Complete |
| **2** | **30 (15%)** | **Literature Review & Methodology** | | | | |
| 2,1 | 20 | State of the art literature review | Adian, Felix | KW40 | KW46 | Complete |
| 2,2 | 10 | Identify methodology | Adian, Felix | KW40 | KW46 | Complete |
| **3** | **50 (30%)** | **Data Acquisition & Pre-Processing** | | | | |
| 3,1 | 10 | Set up computing environment | Adian, Felix | KW47 | KW48 | Complete |
| 3,2 | 20 | Collect raw data | Adian, Felix | KW48 | KW50 | Complete |
| 3,3 | 20 | Pre-process data | Adian, Felix | KW50 | KW51 | Complete |
| **4** | **60 (25%)** | **Data Processing & Analysis** | | | | |
| 4,1 | 20 | Insert data into PostGIS DB | Adian, Felix | KW51 | KW01 | Complete |
| 4,2 | 10 | Optimize PostGIS DB | Adian, Felix | KW52 | KW01 | Complete |
| 4,3 | 10 | Generate OGC services | Adian, Felix | KW02 | KW02 | Complete |
| 4,4 | 20 | Construct Dashboard | Adian, Felix | KW03 | KW04 | Complete |
| **5** | **40 (20%)** | **Dissemination** | | | | |
| 5,1 | 15 | Writing documentation (wiki) | Adian, Felix | KW46 | KW05 | Complete |
| 5,2 | 15 | Creating presentations | Adian, Felix | KW44 | KW05 | Complete |
| 5,3 | 10 | Publishing results | Adian, Felix | KW02 | KW05 | Complete |
| | | **Completion** | | | KW05 | |

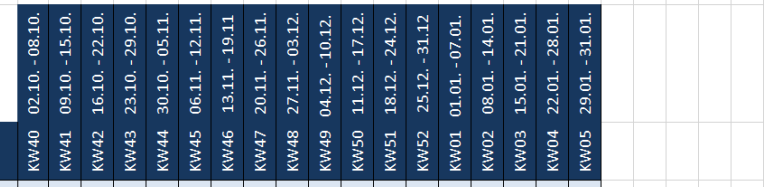



*Figure 6: D-MIG Gantt chart*