



TWIN TRANSITION AND CHANGING PATTERNS OF SPATIAL MOBILITY: A REGIONAL APPROACH

MOBI-TWIN D1.2: COMPLETE MOBI-TWIN DATASET

Work package	WP1: Drivers of spatial mobility
Task	Task 1.3: Complete Mobi-Twin dataset
Authors	Tuomas Väisänen - UH (Main author), Milad Malekzadeh - UH, Maija Havusela - UH, Oula Inkeröinen - UH, Olle Järv - UH (Main author)
Lead Beneficiary	UH
Reviewer(s)	AUTH – Anastasia Panori UB – Vicente Royuela
Status	Final
Version	2.0
Due Date	31/05/2024
Issue Date	10/07/2024
Dissemination Level	PU

Abstract	The Mobi-Twin complete dataset provides data to deliver planned empirical analysis in the Mobi-Twin project. The dataset includes five datasets at NUTS 2 territorial division level in Europe: 1) regional characteristics of NUTS 2 regions; 2) mobility flow data; 3) microsimulation data for five pilot regions; 4) the Mobi-Twin Survey data; and 5) the NUTS 2 spatial layers. This documentation describes the creation and structure of the dataset.
Disclaimer	The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. COPYRIGHT@ MOBI-TWIN CONSORTIUM 2023



Funded by the
European Union

MOBI-TWIN has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement no. 101094402. Content reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.

Version	Date	Partner	Description
0.1	01/04/2024	UH	Draft outline
0.2	01/05/2024	UH	Draft in template
0.3	24/05/2024	UH	Draft in internal review
0.4	28/05/2024	AUTH, UB	Internal review feedback
0.5	29/05/2024	UH	Draft after comments
1.0	31/05/2024	UH	Submitted version
1.1	27/06/2024	UH	Revisions due to incorrect data sources
2.0	09/07/2024	WR	Revised submitted version

EXECUTIVE SUMMARY

The Complete Mobi-Twin dataset (D1.2) combines existing European-level survey and register datasets with publicly available open and big data sources to produce a data product containing information on mobility flows and regional characteristics from Europe. The dataset is at NUTS 2 (Nomenclature of territorial units for statistics; GISCO, 2024) regional level and covers data from 2005 to 2023. The dataset is the outcome of T1.3. and is an input later planned research in the Mobi-Twin project, ensuring that each member has access to and uses the same data.

The complete dataset includes five sections:

- 1) Regional characteristics,
- 2) Mobility data,
- 3) Microsimulation data for five pilot regions,
- 4) The Mobi-Twin Survey data,
- 5) The NUTS 2 spatial layers.

The regional characteristics dataset provides essential information on the NUTS 2 regions in Europe for understanding and redefining regional attractiveness in the twin transition. This data consists of seven themes ranging from variables describing digitalization and environmental characteristics of the regions to socio-economic, demographic, and typological information.

The mobility dataset provides information on mobility flows of the three identified mobility forms – long-term, short-term and circular mobility – previously defined in the report D1.1. State-of-the-art report on the drivers, forms and effects of spatial mobility on EU regions (Panori et al., 2024) ranging from permanent migration to short-term student mobility and seasonal work mobility within Europe. Additionally, cross-border mobility flow dataset was included as a separate, yet relevant mobility type for (cross-border) regional attractiveness.

Microsimulation dataset provides essential socio-economic and demographic input data for agent-based modelling in the five case study regions of the Mobi-Twin project. The Mobi-Twin Survey data is a final and clean dataset from the survey conducted in T1.2., including additional geographical and mobility type profiling variables derived from the initial survey questions. The NUTS 2 spatial layer dataset includes all official versions of the NUTS 2 territorial division (GISCO, 2024).

This documentation describes in detail the creation and structure of the dataset. The complete dataset can be updated with newer or corrected data during the project. The complete Mobi-Twin dataset will be made openly available after the project ends.

TABLE OF CONTENTS

1.	General Description.....	9
2.	Mobility data.....	11
2.1	Data sources.....	11
	Labour Force Survey.....	12
	Erasmus student exchange.....	13
	Twitter/X.....	14
	Processing Mobility types.....	14
2.2	Long-term mobility form.....	14
	Permanent migration.....	15
	Long-term student mobility.....	15
2.3	Short-term mobility form.....	15
	Seasonal work mobility.....	15
	Short-term student mobility.....	15
2.4	Circular mobility form.....	16
	Long-distance commuting.....	16
	Multilocal living.....	16
	Cross-border commuting.....	17
2.5	Missing mobility types.....	17
3.	Regional characteristics.....	18
3.1	Social fabric.....	18
3.2	Living conditions.....	19
3.3	Economy and labour market.....	20
3.4	Access and connectivity.....	22
3.5	Digitalization.....	23
3.6	Landscape and environment.....	24
3.7	Regional typology.....	25
	Methodology for regional typology.....	26
3.8	Regional characteristics at NUTS 3.....	27
4.	Microsimulation data.....	28
5.	Mobi-Twin survey.....	29

5.1	Methodology of the survey.....	29
5.2	Ethical and legal considerations.....	33
6.	Spatial layers.....	34
7.	References	36
	Mobility data.....	37
	Regional characteristics data	37

LIST OF FIGURES

Figure 1: The relation between the five sections of the complete Mobi-Twin dataset.....	9
Figure 2: The production and weighting of the mobility data for the complete Mobi-Twin dataset across the various sources of mobility data.....	12

LIST OF TABLES

Table 1: The list of regional characteristics at NUTS 3 included to the dataset.....	27
Table 2: Number of observations per country and age/gender group.....	32
Table 3: Number of observations per pilot region (NUTS2).....	32
Table 4: Calculated weights considering age/gender and region of residence.....	33
Table 5: The number of NUTS 2 units in Europe studied in the Mobi-Twin project (2003 – 2022). ...	34

LIST OF TERMS AND ABBREVIATIONS

LFS	Labour Force Survey
NUTS	Nomenclature of territorial units for statistics
LAU	Local Area Unit
SCI	Social Connectedness Index
ESPON	European Observation Network for Territorial Development and Cohesion
EDGAR	Emissions Database for Global Atmospheric Research
GHSL	Global Human Settlement Layer
ABM	Agent-based modelling
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
NACE	Statistical classification of economic activities in the European Community
WR	White Research
AUTH	Aristotle University of Thessaloniki
UB	University of Barcelona
UH	University of Helsinki
RUG	Rijksuniversiteit Groningen (University of Groningen)
RRI	Responsible Research and Innovation
ESF	European Science Foundation
GISCO	Geographical Information System of the Commission

1. GENERAL DESCRIPTION

The Mobi-Twin dataset is a curated collection of five datasets: regional characteristics, mobility flow data, the Mobi-Twin survey data, input data for agent-based modelling in the five pilot regions of the project, and spatial layers for official versions of NUTS 2 (Nomenclature of territorial units for statistics; GISCO, 2024) division over time.

The datasets are interconnected to each other based on the unique NUTS 2 identifier code (ID) and mappable via the spatial layer of NUTS 2 regions in Europe (Figure 1). These data have been collected by the Mobi-Twin partners from various sources and are presented in tabular and spatial formats. The dataset is the main output from data collection performed in T1.3 and will provide the input data for analyses done in WP2 and WP3.

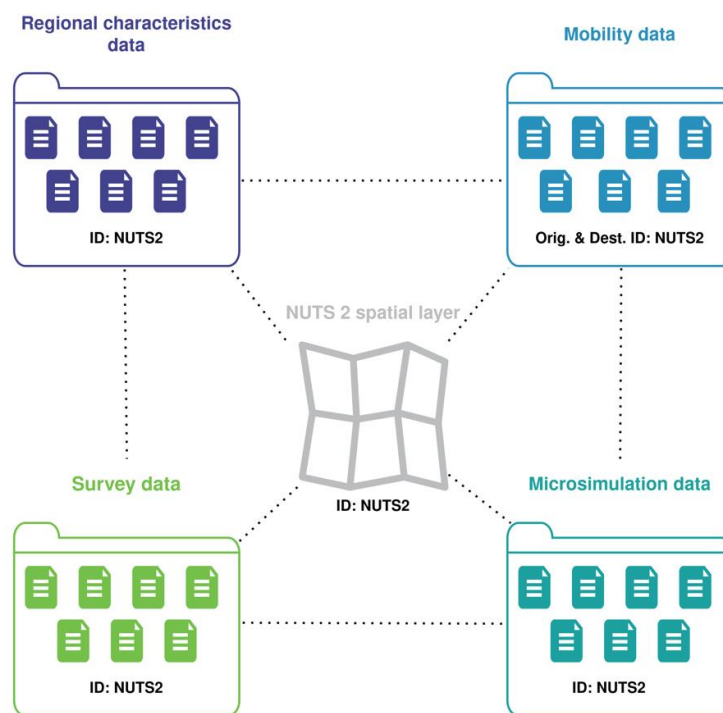


Figure 1. The relation between the five sections of the complete Mobi-Twin dataset. Author: Tuomas Väisänen. [Full-sized figure HERE](#).

The mobility data section covers seven types of mobility, each of which belongs to one of the three main mobility forms – long-term, short-term, and circular mobility (Section 2). The long-term mobility form covers permanent migration and long-term student mobility types. This mobility form refers to mobility where the individual is staying in the destination region for longer than 12 months. The short-term mobility form covers short-term student mobility and seasonal work mobility types. This mobility form refers to mobility where the individual is staying in the region for a duration between three and 11 months. Finally, the

circular mobility form covers mobility where the mobility between origin and destination region is habitual, frequent, and implies a return trip to the origin region. This form contains the following mobility types: long-distance commuting, cross-border commuting, and multilocal living.

The regional characteristics data section of the complete Mobi-Twin dataset provides information on the regions in Europe from 2005 until 2023 (Section 3). These characteristics have been further sectioned into seven themes, that capture different characteristics of these regions. These themes include social fabric, living conditions, economy and labour market, access and connectivity, digitalization, landscape and environment, and finally regional typologies. Each theme includes several variables describing each region throughout the years in the context of the theme. For instance, the social fabric files include variables describing gender balance, population at risk of poverty, income levels and median age of population. The information in the regional characteristics section of the dataset provides essential background information for understanding the mobilities through differences between the receiving and sending regions. To exemplify in the context of the twin transition, student mobility might be better explained by a large difference in the penetration rates of affordable high-speed broadband and mobile internet connections between the regions than climate differences.

The microsimulation data section provides a basic regional information on demographics and employment in the five pilot regions for agent-based modelling (Section 4). using the NUTS on its third level (NUTS 3). This data is used to model mobility patterns within the five focus NUTS 2 regions of the project:

- **Spain:** Castilla-La Mancha (ES42)
- **The Netherlands:** Groningen (NL11)
- **Italy:** Lombardy (ITC4)
- **Greece:** Central Macedonia (EL52)
- **Finland:** Northern and Eastern Finland (FI1D).

This information consists of tabular data on age/sex structure, marital status, education levels, employment status, and household sizes per NUTS 3 regions that make up the above-mentioned NUTS 2 regions. This data covers the years 2001, 2011, and 2021 where data is available, and some of the years in between depending on data availability per country. Any annual gaps in the data will be filled with interpolation in WP3.

The Mobi-Twin survey data section provides survey data on the populations in the five pilot countries regarding their mobility patterns relating to the twin transition (Section 5). The survey data was collected from the Netherlands, Spain, Italy, Greece, and Finland, but also on a more general level from all over Europe. It contains questions on past, current, and future regions of residence, but also on demographics, employment, and attitudes towards the twin transition. The survey data has been processed by the partners to extract mobilities between NUTS 2 regions, to classify respondents as digital nomads, return migrants and retirement migrants, and to provide weights for the respondents.

The NUTS 2 spatial layer dataset ensures the interoperability of the four Mobi-Twin datasets with each other (Figure 1), and potential outside sources of the data (Section 6). Here, each data record of the dataset is associated with a NUTS (Nomenclature of territorial units for statistics) unique code. Regional NUTS 2 regional codes enable enriching mobility flow information with regional characteristics of the origin and destination regions, which is essential for modelling the effects of the twin transition on inter-regional mobilities. Spatial NUTS divisions are available on three levels for statistical analyses of different scopes (GISCO, 2024). Mobi-Twin project focuses on the NUTS 2 level, as that is the level where regional policies are applied, and the availability of the data is better compared to the NUTS 3 or LAU (local area unit) levels.

The complete Mobi-Twin dataset will be disseminated as a complete dataset openly after the end of the publishing embargo granted to the project. The dataset will be hosted on the Zenodo repository to make the dataset citable with a DOI.

2. MOBILITY DATA

2.1 DATA SOURCES

The mobility data for the Mobi-Twin complete dataset was collected from three main sources: Labour Force Survey, Erasmus student exchange data, and Twitter (now known as X). These data sources required different pre-processing approaches to extract the relevant mobility types from them (Figure 2). We direct links to source data are in [Annex 1](#).

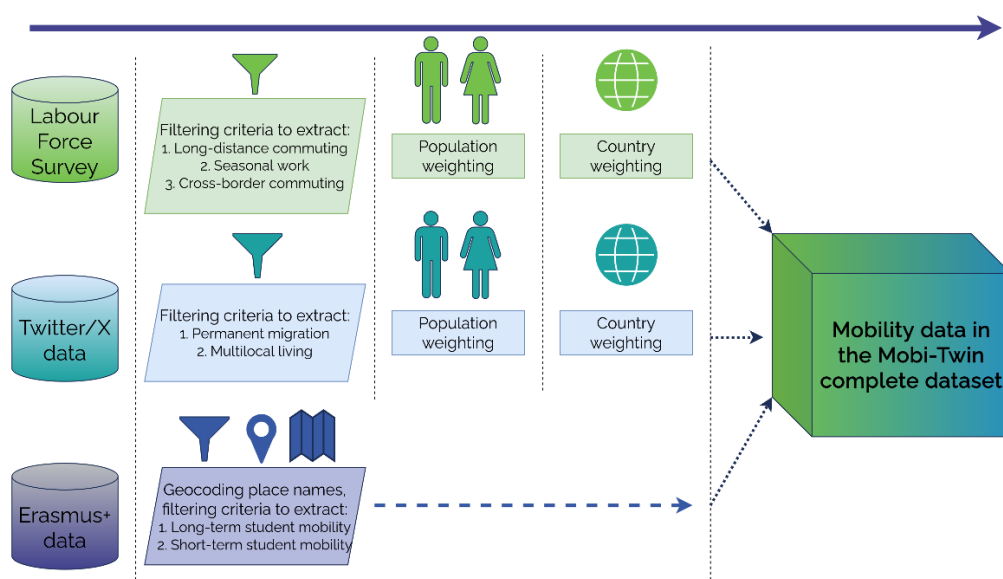


Figure 2: The production and weighting of the mobility data for the complete Mobi-Twin dataset across the various sources of mobility data. Author: Tuomas Väisänen. [Full-sized figure HERE](#).

LABOUR FORCE SURVEY

We used Labor Force Survey data to map three mobility types: **long-distance commuting, seasonal work mobility and cross-border commuting** (Figure 2). We used the annual data to extract seasonal work mobility, long-distance and cross-border commuting. We use the population weights provided in the LFS for aggregation on an annual level, to make the data comparable across Europe we apply a country weight using working age population per country per year to multiply the population weight (see Figure 2).

We trace the individual respondents in the data using the household identification and household member identification numbers (HHNUM, HHSEQNUM), and interview wave numbers (INTWAVE). However, this was feasible only on an annual and intra-country level as identical household and household member numbers can exist in different countries for the same year. The unique identification number for respondents (IDENT) is only available for LFS data since 2021, and thus it cannot be used.

Some responses in the LFS do not have proper NUTS 2 region codes and these vary per country per year. In these cases, the code either refer to NUTS 0 (country level, e.g., with the Netherlands and some cross-border movements), NUTS 1 (major region level, e.g., the UK) or they're not valid NUTS codes at all (e.g., the value 'EU28'). For some smaller countries the NUTS 2 codes are the same as the NUTS 0 codes (e.g., the Baltic countries), and we've treated them as such.

For responses that lacked a valid NUTS 2 code for the residence location but had a valid NUTS 1 code, we employed a stochastic approach to allocate these responses to sub-NUTS 2 regions. This allocation was performed using a random function, ensuring a fair distribution based on population weights. The population weights were calculated by dividing the population of each NUTS 2 region by the total population of the corresponding NUTS 1 region.

For responses that indicated a work region in another country, but on NUTS 0 instead of NUTS 2 level, we used a similar allocation technique, but this time based on Meta's Social Connectedness Index or SCI (Bailey et al., 2018) to allocate flows from the origin NUTS 2 region to NUTS 2 regions in the country of work indicated by the NUTS 0 code. The allocation was based on the distribution of SCI scores from the known origin NUTS 2 region to the NUTS 2 regions in the known destination country. The SCI represents the level of social connections between the regions derived from friendship links of Facebook users identified to be living in the NUTS 2 regions. As the SCI is not based on population distribution, but on social connections it likely represents long-distance commuting flows more realistically than mere population concentration.

ERASMUS STUDENT EXCHANGE

The Erasmus data consists of the Erasmus Mobility Report 2014–2020, which reports the student exchanges on individual level in the Key Action 1 section of the data (Directorate-General for Education, Youth, Sport and Culture, 2023). We filtered out students whose mobility duration was less than three months, the remaining data set was then divided into two, capturing **long-term** (longer than 11 months) and **short-term** (3 to 11 months) **student mobility** (see Figure 2) based on the duration of the mobility.

The locations in this data source are in written form for sending/receiving cities and institutions. We used the provided place names to geocode their locations using the Photon geocoding service implemented in GeoPy (2024). Geocoding is a process where the geocoder is given a place name, and it tries to resolve its location by comparing the given place name to place names it has in its gazetteer. If it finds a match, it returns an address and coordinate information in latitudes and longitudes. Due to various encoding schemes used in the data, some sending and receiving cities were not machine-readable, but the sending and receiving institutions were. In the case of unresolvable city names, we resorted to institution names. On top of the city and institution names, we used the sending and receiving country names to help the geocoder figure out ambiguous cases.

As Mobi-Twin concerns mobilities on the NUTS 2 level, we were interested in getting a rough understanding of the origins and destinations of these flows on the level of cities and towns, and not on a more granular level (e.g. accurate addresses) which is what geocoding is more commonly used for. After an initial run, we were able to geocode 84 % of the placenames in the data. For the remaining 2000 place names, we manually checked and corrected the placenames, as most of the remaining places had typographical errors or spellings in non-English languages in the name (e.g. Estocolmo instead of Stockholm).

After manual correction, we were able to successfully geocode 97 % of the place names in the data on the accuracy level of cities. The remaining placenames were either outside Europe or unresolvable. Using this location information, we then aggregated the origins and destinations to NUTS 2 level from year 2021 and mapped the interregional flows.

In total, the final full student mobility data consists of 1.65 million individual flows, out of which 1.61 million are short-term and 40 thousand are long-term mobilities.

TWITTER/X

We used a comprehensive dataset from Twitter (now known as X), encompassing geotagged tweets from across Europe between 2012 and 2022 (Poorthuis & Zook, 2017) to extract **permanent migration** and **multilocal living mobility**. Each tweet in this dataset is timestamped and geolocated at the H3 level (Uber's hexagonal hierarchical spatial index), providing precise spatial and temporal information. To ensure data quality, we filtered out the top 5% of users with the highest number of tweets, typically generated by automated bots. We excluded users with less than 180 days (about 6 months) of data to capture our defined mobility types. Each user in the dataset was assigned a unique identifier, enabling us to track individual movements and identify significant places (e.g., region of residence). We used Twitter data to extract long-distance commuting and digital nomadism flows, but for data scarcity reasons they are not included in the complete dataset.

To weight the outcomes derived from Twitter data, we calculated the penetration rate for each NUTS 2 region. Given that the number of Twitter users varies across different age groups, we accounted for this variation by incorporating population data segmented by age from Eurostat. Using the reported proportions of Twitter users within each age bracket, we applied weights to the population data accordingly. We then estimated the home locations of users to determine their respective NUTS 2 regions of origin. By calculating the ratio of the weighted population – considering the age brackets – to the number of Twitter users in each NUTS 2 region, we derived a penetration rate. Since the data is provided without a temporal variable due to data scarcity on the annual level, we averaged this penetration rate across 2012–2022. Finally, we multiplied this averaged rate by the outcomes to estimate realistic values for movement patterns.

PROCESSING MOBILITY TYPES

In the following sections, the details of how each mobility type is extracted from the corresponding datasets are elaborated. To ensure transparency and reproducibility of our methods, the code used for processing these datasets and extracting mobility types has been stored in a dedicated GitHub repository. The repository located at UH Digital Geography Lab – MobiTwin [codebase](#).

2.2 LONG-TERM MOBILITY FORM

PERMANENT MIGRATION

Permanent migration was identified using Twitter data by comparing users' residential locations. We first determined significant locations for each user at the NUTS3 level. These significant locations were defined by two criteria for each year: they had to be in the top 75th percentile for tweet frequency, and the user must have spent at least 30 days at that location within the year. The location with the highest dwell time was selected as the primary residence for each user. A migration event was recorded when a user's primary residence changed within two years, marking it as a permanent migration from the origin to the destination. To ensure consistency with other datasets, we then aggregated the data from the NUTS3 level to the NUTS2 level. Finally, to contend with data scarcity on the annual level we grouped the data temporally into three periods: 2012–2015; 2016–2019; 2020–2022.

LONG-TERM STUDENT MOBILITY

Long-term student mobility is extracted from the Erasmus data provided by the Directorate-General for Education, Youth, Sport and Culture (2024). The raw data provides student mobilities across the academic years starting from 2014 to 2020 on an individual level. That is, each individual mobility that is a part of Key Action 1 data represents an individual student. We filtered the mobilities that last longer than 11 months to have a clear separation between student mobilities that last longer than one academic year, and those that are shorter-term, such as one or two semesters. We used the place names provided in the data to map the geographical flows of students by geocoding them with the GeoPy (2024) Python library and the Photon geocoding service provided therein (for more detailed description see above).

2.3 SHORT-TERM MOBILITY FORM

SEASONAL WORK MOBILITY

Seasonal work mobility is extracted by using filtering criteria to extract respondents who explicitly state they are seasonal workers (JATTACH, SEEKWORK), but also workers who have a short-term contract (3-11 months in duration, TEMP, TEMPDUR) based on the delineation provided in D1.1 and work in a different region (REGION_W) than where they live (REGION).

SHORT-TERM STUDENT MOBILITY

Short-term student mobility is extracted from the Erasmus data provided by the Directorate-General for Education, Youth, Sport and Culture (2024). The raw data provides student mobilities across the academic years starting from 2014 to 2020 on an individual level. That is, each individual mobility that is a part of Key Action 1 of the Erasmus+

programme in the data represents an individual student. We filtered the mobilities that last longer than 3 months but less than 11 to have a clear separation between student mobilities that last for one or two semesters, and those that last longer than one academic year. Place names provided in the data was used to map the geographical flows of students by geocoding them with the GeoPy (2023) Python library and the Photon geocoding service provided therein (for a detailed description, see the section Processing Mobility types).

2.4 CIRCULAR MOBILITY FORM

LONG-DISTANCE COMMUTING

Long-distance commuting is extracted by filtering respondents who have a different region of residence (REGION) to the region of work (REGION_W) using the annual LFS data. Additionally, we only include respondents for whom the commute is 100 kilometres or longer in Euclidean distance. We do this by comparing the distance between the centroids of NUTS 2 regions. While this introduces some uncertainty of whether all included parties are long-distance commuters, the varying size of NUTS 2 regions themselves make estimating long-distance commuting challenging. Some NUTS 2 regions cover vast areas (e.g. Northern & Eastern Finland (F11D) covers nearly $\frac{3}{4}$ of the whole country), while some (e.g. Inner London) cover much smaller areas. Thus, long-distance commuting within large regions is not captured and countries with numerous smaller NUTS 2 regions are overrepresented in the data (e.g. Germany).

MULTILOCAL LIVING

Multilocal living was identified by first determining the significant locations for each user at the NUTS 3 level. Significant locations were defined by two criteria for each year: they had to be in the top 75th percentile in terms of tweet frequency, and the user must have spent at least 90 days at that location within a five-year window (two years before and two years after the considered year). This process yielded significant places for each user annually. Next, we calculated the number of movements between these significant places. Movements that constituted more than 50% of the total movements between significant places were likely due to long-distance commuting. We classified all less frequent pairs of significant places as multilocal living movement, indicating that individuals spent substantial time in these locations without showing a commuting pattern between them.

Between these two locations, we estimated primary and secondary locations based on dwell time, defining the primary location as where the user spent the most time and the secondary location as where they spent less time than the primary location. However, it is important to note that this distinction is approximate due to the limited temporal accuracy of Twitter data. Thus, these locations should be considered part of a multilocal living arrangement rather than a clear flow from origin to destination. Despite this, in the final dataset, we labelled the primary location as the origin and the secondary location as the

destination to maintain consistency with other datasets. Finally, we aggregated the data to the NUTS 3 level, and the final dataset was further aggregated to the NUTS 2 level.

CROSS-BORDER COMMUTING

Cross-border commuting is a subset of long-distance commuting (see above), characterized by individuals who travel between countries for work. The key criterion for identifying cross-border commuting is that the region of work for a respondent must be in a different country (COUNTRY_W) than the region of residence (COUNTRY). Unlike with long-distance commuting, here we do not apply the distance filter of 100 kilometres, and thus all commutes that are between two regions in different countries are included.

2.5 MISSING MOBILITY TYPES

Data sources available for this study (Twitter/X, LFS, Mobi-Twin survey) were either not able to capture four mobility types (digital nomadism, retirement migration, family reunification, and return migration) or had too few data records for valid datasets.

Digital nomadism involves individuals who frequently move between regions without returning to previous locations, that is they have a nomadic mobility pattern. While we attempted to extract this type of mobility from the Twitter data, we found that instances of digital nomadism were very few, and the movements were very sparse, only around ten individuals during a 11-year period. This suggests that digital nomadism may be underrepresented in the Twitter dataset or that this type of data is not well-suited to capture such movements. The Mobi-Twin survey had questions pertaining to a digital nomad lifestyle, but the number of respondents who matched the criteria were too low for mapping flows. We, however, included a typology in the survey for the few respondents who match these criteria.

Retirement migration describes migration where the respondent is nearing retirement age and moves to another region. Due to a lack of a unique identifier that would be consistent in the LFS data for the same country across quarters or years, extracting this mobility type became unfeasible.

Family reunification is another specific type of mobility, which is impossible to extract from available data sources on the NUTS 2 level. The LFS data has a question related to this (MIGREAS), but it refers only to the respondent's country of citizenship on the level of country groups or sub-continent (e.g., EU27, North Africa) and thus it is not usable for regional mobility patterns within Europe.

Return migration is recorded in the Mobi-Twin survey data, but the number of respondents who have returned to their country of birth or citizenship after living abroad is very small. Similarly, the number of respondents who indicate a high likelihood of migrating back to their country of birth, residence, or a region they previously lived in is also low. As

extracting return migration flows requires information on three locations per individual (original location, middle location, current location), the data sources could not capture it on a sufficient level for analysis. Like digital nomadism, return migrants are shown in the survey data through our typology, but their movements cannot be mapped meaningfully due to low numbers. To ensure solid and robust dataset for analytical purposes, these mobility types had to be excluded from the complete dataset.

3. REGIONAL CHARACTERISTICS

Regional characteristics data provide crucial background information on the NUTS 2 regions in Europe from 2005 onwards. These data are used to analyse differences between origin and destination regions for the different mobility types, to understand what the drivers of mobility are, but also what makes certain regions attractive. These characteristics are provided in the seven themes each of which describes the regions from the theme's perspective. The data structure follows these themes: there are seven folders corresponding to each theme, each contains a CSV file containing the data, and a readme file containing metadata about the variables.

The data were collected mostly from the openly available Eurostat and ESPON – European Spatial Planning Observation Network (ESPON, 2024) – m databases, but also some other sources ([Annex 1](#)) on the NUTS 2 regions in Europe from 2005 onwards. Most of the data has been collected from open data provided by Eurostat and ESPON, whereas some have been aggregated from more detailed data to the coarser NUTS 2 region level (e.g., average annual temperature data) or are results from analyses performed by other researcher previously (e.g., the Green Transition Vulnerability Index).

The files are in tabular form and associated with readme files that describe the variables provided, but also delineate the sources for these files. Additionally, some variables for regional characteristics are provided on the NUTS 3 regional level based on their availability from Eurostat.

3.1 SOCIAL FABRIC

The social fabric sub-topic under regional characteristics captures regional information pertaining to social systems from several perspectives. The social fabric data (a CSV file) is based on data downloaded from the Eurostat database and the ESPON database describing the regional characteristics related to different social phenomena. Time period used in data collection was 2005-2022, but some variables are not available for all years from all NUTS 2 regions. The data collected from Eurostat contains flags provided by Eurostat to identify provisional, estimated, and low-reliability values as well as for breaks in time series, differing definition and regions or years where the variable is not applicable.

- **NUTS 2 region code** (NUTS_2_CODE) – 4-letter code representing a NUTS 2 region
- **Year of data** (YEAR) – a year representing the data values
- **Women per 100 men** (WOMEN_PER_100_MEN) – proportion of women in total population in NUTS 2 regions.
- **Population at risk of poverty** (AT_RISK_OF_POVERTY) – share of population living in private households that are at risk of poverty in NUTS 2 regions
- **Students in tertiary education** (TERTIARY_STUDENTS) – number of students enrolled in tertiary education in NUTS 2 regions
- **Household income** (HOUSEHOLD_INCOME) – amount of disposable net income of households in NUTS 2 region in million euro
- **Severe material and social deprivation** (MA_SO_DEPRIVATION) – share of population living in private households that are experiencing severe material and social deprivation in NUTS 2 regions
- **Median age** (MEDIAN_AGE) – yearly median age of the population in NUTS 2 regions
- **Net migration** (NET_MIGR) – crude rate of net migration in NUTS 2 regions, with statistical adjustment
- **Share of immigrants** (SHARE_OF_IMMIGRANTS) – proportion of immigrants in total population in NUTS 2 regions
- **People balance index** (PEOPLE_BALANCE_IDX) – ratio between the difference in inflow-outflow and the maximum value of inflow or outflow. Value -1 indicates pure outflow, value 1 pure inflow and value 0 perfectly balanced flow.

3.2 LIVING CONDITIONS

The living conditions sub-topic under regional characteristics captures regional information pertaining to safety and population aspects from several perspectives. The living conditions data (a CSV file) is based on data downloaded from the Eurostat database and the ESPON database describing the regional characteristics related to different safety and population phenomena.

Time period used in data collection was 2005-2022, but some variables did not have all those years recorded at all or they had them just in certain NUTS 2 regions. For data collected from Eurostat, there are flags provided for provisional values and estimated values, for breaks in time series, differing definition and regions or years where the variable is not applicable.

- **NUTS 2 region code** (NUTS_2_CODE) – 4-letter code representing a NUTS 2 region
- **Year of data** (YEAR) – a year representing the data values
- **Police recorded homicides** (HOMICIDE_PER_100K) – number of police recorded intentional homicides per hundred thousand inhabitants in NUTS 2 regions
- **Police recorded assaults** (ASSAULT_PER_100K) – number of police recorded assaults per hundred thousand inhabitants in NUTS 2 regions
- **Police recorded robberies** (ROBBERY_PER_100K) – number of police recorded robberies per hundred thousand inhabitants in NUTS 2 regions
- **Police recorded burglaries** (BURGLARY_PER_100K) – number of police recorded burglaries per hundred thousand inhabitants in NUTS 2 regions
- **Police reported vehicle thefts** (VEHICLE_THEFT_PER_100K) – number of police recorded thefts of motorized land vehicles per hundred thousand inhabitants in NUTS 2 regions

- **Fertility rate** (FERTILITY_RATE) – number of births per woman in NUTS 2 regions
- **Mean age of women at childbirth** (MEAN_AGE_CHILDBIRTH) – mean age of women at childbirth in NUTS 2 regions
- **Mortality rate** (MORTALITY) – standardized death rate for all causes of death in NUTS 2 regions
- **Life expectancy** (LIFE_EXPECTANCY) – life expectancy for less than 1 year old inhabitants in NUTS 2 regions
- **Total population** (POPULATION_TOTAL) – total population number in NUTS 2 regions
- **Population of women** (POPULATION_WOMEN) – population number of women in NUTS 2 regions
- **Number of physicians** (PHYSICIANS_PER_100K) – number of physicians per hundred thousand inhabitants in NUTS 2 regions
- **Population density** (POPULATION_DENSITY) – number of people per square kilometre in NUTS 2 regions
- **Available hospital beds** (HOSP_BEDS_PER_100K) – number of available hospital beds per hundred thousand inhabitants in NUTS 2 regions
- **Index of human resources** (HUMAN_RESOURCES_IDX) – index of human resources, knowledge and skills (share of people ages 25-64 and 20-24 with upper secondary or tertiary education) in NUTS 2 regions
- **Index of good governance** (GOOD_GOVERNANCE_IDX) – index of good governance in NUTS 2 regions, derived from the European Quality of Government Index produced by the University of Gothenburg.

3.3 ECONOMY AND LABOUR MARKET

The economy and labour market sub-topic under regional characteristics captures regional information pertaining to economical and labour market aspects from several perspectives relevant to understanding changing mobilities. The economy and labour market data (a CSV file) is based on data downloaded from the Eurostat database describing the regions with different economic and labour market characteristics. Time period used in data collection was 2005-2022, but some variables are not available for all years or for all NUTS 2 regions. The data collected from Eurostat contains flags for the identification of provisional, estimated, low-reliability values, and confidential values as well as for breaks in time series and differing definitions.

- **NUTS 2 region code** (NUTS_2_CODE) – 4-letter code representing a NUTS 2 region
- **Year of data** (YEAR) – a year representing the data values
- **Gross domestic product** (GDP) – gross domestic product in current market prices in NUTS 2 regions in million euro
- **Gross value added** (GVA) – gross value added in volume in NUTS 2 regions, index (2015=100) is used as unit
- **Employed men aged 15 to 64** (EMPLOYED_M_AGE15-64) – share of employed men from ages 15 to 64 in NUTS 2 regions
- **Employed men aged 20 to 64** (EMPLOYED_M_AGE20-64) – share of employed men from ages 20 to 64 in NUTS 2 regions
- **Employed women aged 15 to 64** (EMPLOYED_F_AGE15-64) – share of employed women from ages 15 to 64 in NUTS 2 regions
- **Employed women aged 20 to 64** (EMPLOYED_F_AGE20-64) – share of employed women from ages 20 to 64 in NUTS 2 regions

- **Unemployed men aged 15 to 64** (UNEMPLOYED_M_AGE15-64) – share of unemployed men from ages 15 to 64 in NUTS 2 regions
- **Unemployed men aged 20 to 64** (UNEMPLOYED_M_AGE20-64) – share of unemployed men from ages 20 to 64 in NUTS 2 regions
- **Unemployed women aged 15 to 64** (UNEMPLOYED_F_AGE15-64) – share of unemployed women from ages 15 to 64 in NUTS 2 regions
- **Unemployed women aged 20 to 64** (UNEMPLOYED_F_AGE20-64) – share of unemployed women from ages 20 to 64 in NUTS 2 regions
- **Unemployed people aged 15 to 64** (UNEMPLOYED_T_AGE15-64) – share of total number of unemployed people from ages 15 to 64 in NUTS 2 regions
- **Unemployed people aged 20 to 64** (UNEMPLOYED_T_AGE20-64) – share of total number of unemployed people from ages 20 to 64 in NUTS 2 regions
- **Weekly hours of work of men aged 15 to 64** (WEEKLY_HRS_M_AGE15-64) – average number of work hours in main job per week of men ages 15 to 64 in NUTS 2 regions
- **Weekly hours of work of men aged 20 to 64** (WEEKLY_HRS_M_AGE20-64) – average number of work hours in main job per week of men ages 20 to 64 in NUTS 2 regions
- **Weekly hours of work of women aged 15 to 64** (WEEKLY_HRS_F_AGE15-64) – average number of work hours in main job per week of women ages 15 to 64 in NUTS 2 regions
- **Weekly hours of work of women aged 20 to 64** (WEEKLY_HRS_F_AGE20-64) – average number of work hours in main job per week of women ages 20 to 64 in NUTS 2 regions
- **Age dependency ratio** (AGE_DEP_RATIO) – share of non-working age population (ages 0-19 and 65+) compared to working age population (ages 20-64) in NUTS 2 regions
- **Old age dependency ratio** (OLD_AGE_DEP_RATIO) – share of non-working age population (ages 65+) compared to working age population (ages 20-64) in NUTS 2 regions
- **Young age dependency ratio** (YOUNG_AGE_DEP_RATIO) – share of non-working age population (ages 0-19) compared to working age population (ages 20-64) in NUTS 2 regions
- **Men in high-technology jobs** (HIGH_TECH_M) – share of men employed in high-technology sectors in NUTS 2 regions
- **Women in high-technology jobs** (HIGH_TECH_F) – share of women employed in high-technology sectors in NUTS 2 regions
- **Wages in construction** (WAGES_CONSTRUCTION) – wages in construction sectors (NACE level 1) in million euros in NUTS 2 regions
- **Wages in electricity, gas, steam and air-conditioning** (WAGES_EGSA) – wages in electricity, gas, steam and air-conditioning supply (NACE level 1) in million euros in NUTS 2 regions
- **Wages in manufacturing** (WAGES_MANUFACTURING) – wages in manufacturing sector (NACE level 1) in million euros in NUTS 2 regions
- **Wages in mining and quarrying** (WAGES_MINING_QUARRYING) – wages in mining and quarrying (NACE level 1) in million euros in NUTS 2 regions
- **Wages in water supply** (WAGES_WATER_SUPPLY) – wages in sewerage, waste management and remediation activities (NACE level 1) in million euros in NUTS 2 regions
- **Wages in wholesale, retail trade and repair of motor vehicles** (WAGES_WRTR) - wages in wholesale, retail trade and repair of motor vehicles and motorcycles (NACE level 1) in million euros in NUTS 2 regions
- **Wages in accommodation and foodservice** (WAGES_ACCOM_FOODSERVICE) – wages in accommodation and food service sectors (NACE level 1) in million euros in NUTS 2 regions
- **Wages in information and communication** (WAGES_INFO_COMM) – wages in information and communication sectors (NACE level 1) in million euros in NUTS 2 regions
- **Wages in transportation and storage** (WAGES_TRANSPORT_STORAGE) – wages in transportation and storage sectors (NACE level 1) in million euros in NUTS 2 regions
- **Wages in real estate** (WAGES_REAL_ESTATE) – wages in real estate sectors (NACE level 1) in million euros in NUTS 2 regions

- **Wages in admin** (WAGES_ADMIN_SUPPORT) – wages in administrative and support service activities (NACE level 1) in million euros in NUTS 2 regions
- **Wages in professional, scientific and technical activities** (WAGES_PROF_SCIEN_TECH) – wages in professional, scientific and technical activities (NACE level 1) in million euros in NUTS 2 regions
- **Left-behindedness index** (LEFT_BEHINDEDNESS_IDX) – Index of left-behindedness, which is characterized by different combinations of issues left-behind regions usually have. Originating from the work of Velthuis et al. (2023).

3.4 ACCESS AND CONNECTIVITY

The access and connectivity sub-topic under regional characteristics capture regional information pertaining to connectivity of networks, systems and resources from several perspectives. The access and connectivity data (a CSV file) is based on data downloaded from the Eurostat database describing the regional characteristics related to different connectivity phenomena.

Time period used in data collection was 2005-2022, but some variables did not have all those years recorded at all or they had them just in certain NUTS 2 regions. For data collected from Eurostat, there are flags provided for provisional values, low reliability values, confidential values and for breaks in time series and regions or years where the variable is not applicable.

- **NUTS 2 region code** (NUTS_2_CODE) – 4-letter code representing a NUTS 2 region
- **Year of data** (YEAR) – a year representing the data values
- **Kilometres of motorways** (MOTORWAYS) – kilometres of motorways per thousand square kilometres in NUTS 2 regions
- **Kilometres of railroads** (RAILROADS) – kilometres of railroads per thousand square kilometres in NUTS 2 regions
- **Number of airports** (AIRPORTS) – number of airports in NUTS 2 regions
- **Number of air passengers** (AIR_PASSENGERS) – how many thousand passengers transported by air in NUTS 2 regions
- **Freight loaded onto aircrafts** (AIR_FREIGHT_LOADED) – how many thousand tonnes of freight loaded onto aircrafts in NUTS 2 regions
- **Freight unloaded from aircrafts** (AIR_FREIGHT_UNLOADED) – how many thousand tonnes of freight unloaded from aircrafts in NUTS 2 regions
- **Number of harbours** (HARBORS) – number of harbours in NUTS 2 regions
- **Number of passengers embarking on maritime transportation** (MARITIME_EMBARKED) – how many thousand passengers embark on maritime transportation in NUTS 2 regions
- **Number of passengers disembarking from maritime transportation** (MARITIME_DISEMBARKED) – how many thousand passengers disembark from maritime transportation in NUTS 2 regions
- **Freight loaded onto maritime transportation** (MARITIME_FREIGHT_LOADED) – how many thousand tonnes of freight loaded onto maritime transportation in NUTS 2 regions
- **Freight unloaded from maritime transportation** (MARITIME_FREIGHT_UNLOADED) – how many thousand tonnes of freight unloaded from maritime transportation in NUTS 2 regions
- **Tourists arriving at accommodation** (ARR_TOURIST_ACCOM) – number of tourists that arrive at accommodation establishments in NUTS 2 regions.

3.5 DIGITALIZATION

The digitalization sub-topic under regional characteristics captures regional information pertaining to the digital transition from several perspectives. The digitalization data (a CSV file) is based on data downloaded from the Eurostat database and Ookla Open Data (Ookla, 2024) describing the regional characteristics related to digitalization phenomena.

Time period used in data collection was 2005-2022, but some variables did not have all those years recorded at all or they had them just in certain NUTS 2 regions. The cellular data from Ookla is from the years 2019-2023 and has been aggregated to the NUTS 2 regional level. For data collected from Eurostat, there are flags provided for low reliability values and breaks in time series.

- **NUTS 2 region code** (NUTS_2_CODE) – 4-letter code representing a NUTS 2 region
- **Year of data** (YEAR) – a year representing the data values
- **Households with broadband access** (BROADBAND_ACCESS) – share of households in NUTS 2 regions that have access to broadband
- **Households with internet access** (INTERNET_ACCESS) – share of households in NUTS 2 regions with internet access at home
- **Individuals with internet access away from home** (INTERNET_AWAY) – share of individuals who have access to the internet away from home or work
- **Individuals who use the internet to contact public authorities** (INTERNET_PUBL_AUTH) – share of individuals who access the internet to interact with public authorities
- **Individuals who use the internet for selling** (INTERNET_SELLING) – share of individuals who access the internet to sell goods or services
- **Last online purchase in the last three months** (LAST_ONLINE_PUR_3MO) – share of individuals whose last online purchase was within the last three months
- **Last online purchase between three and twelve months ago** (LAST_ONLINE_PUR_3-12MO) – share of individuals whose last online purchase was between three and twelve months ago
- **Last online purchase in the last twelve months** (LAST_ONLINE_PUR_12MO) – share of individuals whose last online purchase was within the last twelve months
- **Last online purchase more than a year ago** (LAST_ONLINE_PUR_YEAR+) – share of individuals whose last online purchase was more than one year ago
- **Last online purchase more than a year ago or never** (LAST_ONLINE_PUR_YEAR+_NEVER) – share of individuals whose last online purchase was more than one year ago or never
- **Individuals who have never used a computer** (NEVER_USED_COMPUTER) – share of individuals who have never used a computer
- **Individuals who use the internet for online banking** (ONLINE_BANKING) – share of individuals who use the internet for online banking
- **Average download speed with a non-cellular connection** (FIX_AVG_D_KBPS) – a median value in kilobits of the average download speed with a non-cellular connection (e.g. WiFi, ethernet)
- **Average upload speed with a non-cellular connection** (FIX_AVG_U_KBPS) – a median value in kilobits of the average upload speed with a non-cellular connection (e.g. WiFi, ethernet)
- **Average latency with a non-cellular connection** (FIX_AVG_LAT_MS) – a median value in milliseconds of the average latency with a non-cellular connection (e.g. WiFi, ethernet)
- **Average download speed with a cellular connection** (MOB_AVG_D_KBPS) – a median value in kilobits of the average download speed with a cellular connection (e.g. 4G LTE, 5G NR)
- **Average upload speed with a cellular connection** (MOB_AVG_U_KBPS) – a median value in kilobits of the average upload speed with a cellular connection (e.g. 4G LTE, 5G NR)

- **Average latency with a cellular connection** (MOB_AVG_LAT_MS) – a median value in milliseconds of the average latency with a cellular connection (e.g. 4G LTE, 5G NR).

3.6 LANDSCAPE AND ENVIRONMENT

The landscape and environment data covers the NUTS 2 regions with variables describing environmental and landscape-related characteristics. The data is CSV file collected from Eurostat, ESPON, European Environment Agency, EDGAR – Emissions Database for Global Atmospheric Research (Crippa et al., 2023), and the WorldClim (Harris et al., 2020) databases. Time period used in data collection was 2005-2022, but some variables did not have all those years recorded at all or they had them just in certain NUTS 2 regions.

- **NUTS_2_CODE** – 4-letter code representing a NUTS 2 region
- **YEAR** – a year representing the data values
- **AGRICULTURE** – share of land used for agriculture
- **COOLING_DEGREE_DAYS** – The number of cooling degree days, describing the demand for air conditioning a building.
- **HEATING_DEGREE_DAYS** – The number of heating degree days, describing the demand for energy needed to heat a building.
- **SOIL_EROSION_WATER** – Tonnes per hectare of soil erosion in agricultural areas, forests, and seminatural areas.
- **FISHING_AQUACULTURE** – The share of land used for fishing and aquaculture
- **FORESTRY** – The share of land used for forestry.
- **HEAVY_ENV_IMPACT** – The share of land used in activities that have a heavy environmental impact.
- **NATURE_RESERVERS** – The share of land used for natural reserves.
- **SERVICES_RESIDENTIAL** – The share of land used for services and residential areas.
- **OTHER_PRIM_SECTOR_ACTS** – The share of land used for other primary sector activities
- **UNUSED_ABANDONED** – The share of land that is unused or abandoned.
- **GREEN_INFRA** – An indicator showing multifunctionality of green infrastructure in the region. Available only for 2012.
- **PV_POTENTIAL** – The potential for photovoltaic energy (GWh/year) at a fixed price of 12 c/kWh. Available only for 2009.
- **MEAN_MIN_TEMP** – Average minimum annual temperature in Celsius. Calculated from data with spatial resolution of 2.5 minutes.
- **MEAN_PRECIPITATION** – Average annual precipitation in millimetres. Aggregated from data with spatial resolution of 2.5 minutes.
- **CIRC_ECON_EMPLOYMENT** – The number of people employed in the circular economy material providers' sector. Available for 2010 and 2015.
- **CIRC_ECON_TURNOVER** – Total turnover generated by circular economy material providers' activities. Unit is in million euros. Available for 2010 and 2015.
- **NATURAL_HAZARD_SUSCEP** – Territorial susceptibility to natural hazards. The scale is between 1 and 2. Available for 2018.
- **TAPC_IDX** – Index of territorial assets and physical conditions. Combines the potential of onshore wind, photovoltaic, and biomass energy potentials, with share of Natura 2000 areas in NUTS 2 regions. Available for 2013
- **RGTV_IDX** – Index of regional green transition vulnerability. Based on the work of Rodríguez-Pose & Bartalucci (2023).

- **BIO_GEO TYPOLOGY** – Biogeographical region typology. Based on the work of Roekaerts (2002) and the EEA (2016).
- **AIR_QUALITY_AVG** – Air pollution as a population-weighted average. The unit is ug/m³, and the pollutant is PM_{2.5}.
- **GHG_AGRICULTURE** – Emissions of greenhouse gases from agriculture. Unit is kilotonnes of CO₂eq.
- **GHG_BUILDINGS** – Emissions of greenhouse gases from energy-use of buildings. Unit is kilotonnes of CO₂eq.
- **GHG_ENERGY** – Emissions of greenhouse gases from power generation. Unit is kilotonnes of CO₂eq.
- **GHG_INDUSTRY** – Emissions of greenhouse gases from combustion in industry. Unit is kilotonnes of CO₂eq.
- **GHG_TRANSPORT** – Emissions of greenhouse gases from road, rail, pipeline and off-road transportation. Unit is kilotonnes of CO₂eq.
- **GHG_WASTE** – Emissions of greenhouse gases from solid waste incineration, landfills, and wastewater handling. Unit is kilotonnes of CO₂eq.

3.7 REGIONAL TYPOLOGY

The regional typology data (a CSV file) is based on data downloaded from the Eurostat database and the GHSL – Global Human Settlement Layer (Schiavina et al., 2023; European Commission, Eurostat, 2021) database describing the regional characteristics related to geographical analyses. The datasets used are from the years 2005, 2010, 2015, 2016, 2020 and 2021. The GHSL data was calculated to given typologies on the NUTS 2 level based on the same methodology as is used by Eurostat to calculate them for the NUTS 3 level (European Commission and Eurostat, 2019).

The variables in this data are:

- **MOUNT_TYPE_1** – Share of population who live in areas where more than half live in mountain regions.
- **MOUNT_TYPE_2** – Share of population who live in regions where at least 50 % of surface area is in mountain regions.
- **MOUNT_TYPE_3** – Share of population who live in regions where more than half of the population and surface area are in mountain regions
- **MOUNT_TYPE_4** – Share of population who live in other than mountain regions
- **COAST_TYPE_1** – Share of population who live in coastal regions
- **COAST_TYPE_2** – Share of population who live in coastal regions where at least half of the population are within 50 kilometres from the coastline
- **COAST_TYPE_3** – Share of population who live in non-coastal regions.
- **ISLAND_PERCENTAGE** – Share of population that live on islands.
- **URB_RUR TYPOLOGY** – Classification of the region to
 - Predominantly rural (50 % or more of the population live in rural regions)
 - Intermediate (More than 20% but less than 50 % live in rural regions OR 50 % or more live in rural regions, but the urban centre population is 200 000 or more and is at least 25 % of the population)
 - Predominantly urban (20 % or less live in rural regions OR more than 20 % but less than 50 % live in rural regions, but the urban centre population is at least 500 000 and has at least 25 % of the total population.
- **URBAN_SHARE** – Share of population that live in urban regions

- **RURAL_SHARE** – Share of population that live in rural regions

METHODOLOGY FOR REGIONAL TYPOLOGY

We used the GHSL data to calculate typologies on the NUTS 2 level using the same methodology as Eurostat uses to calculate them for the NUTS 3 level. The methodology is, as follows:

The MOUNT_TYPE - variables were calculated from mountain area types per NUTS 3 region:

1. Mountain area types per NUTS 3 region
2. Population data per NUTS 3 region
3. Calculating the total population per NUTS 2 region
4. Calculating the percentage of population living in each mountain area type per NUTS 2 region

The COAST_TYPE - variables were calculated from coastal area types per NUTS 3 region:

1. Coastal area types per NUTS 3 region
2. Population data per NUTS 3 region
3. Calculating the total population per NUTS 2 region
4. Calculating the percentage of population living in each coastal area type per NUTS 2 region

The ISLAND_PERCENTAGE - variable was calculated from island areas per NUTS 3 region:

1. Island areas per NUTS 3 region
2. Population data per NUTS 3 region
3. Calculating the total population per NUTS 2 region
4. Calculating the percentage of population living in island areas per NUTS 2 region

The URB_RUR TYPOLOGY - variable was calculated from the GHSL population data containing the total population, urban centre population, urban cluster population and rural population on municipality level (GID_2 of GADM). To authors' knowledge, this is the first time given variable is calculated to NUTS 2 level. The calculation was, as follows:

1. Joining the population data to NUTS 2 region codes
2. Calculating the sum of populations living in urban centres and urban clusters per NUTS 2 region
3. Calculating the share of both urban regions (URBAN_SHARE) and rural regions (RURAL_SHARE) per NUTS 2 region
4. Defining a classification function:
 - a. if RURAL_SHARE is 50 % or more → region is predominantly rural
 - If the urban centre population is 200 000 or more and it is at least 25 % of the total population → then region is intermediate
 - b. If RURAL_SHARE is more than 20 % but less than 50 % → region is intermediate
 - If the urban centre population is 500 000 or more and it is at least 25 % of the total population → then region is predominantly urban
 - c. if RURAL_SHARE is 20 % or less → region is predominantly urban

The URBAN_SHARE - variable was calculated by dividing the sum of urban centres' population and urban cluster's population by the total population per each NUTS 2 region.

The RURAL_SHARE variable was calculated by dividing the rural regions' population by the total population per NUTS 2 region.

3.8 REGIONAL CHARACTERISTICS AT NUTS 3

Similarly to NUTS 2 level, the same regional characteristics at NUTS 3 level are provided as a single file. Not all characteristics are available for NUTS 3 level, nor for all years the NUTS 2 data is available for. The list of characteristics included at NUTS 3 level are presented in Table 1 (see variable descriptions above).

Table 1: The list of regional characteristics at NUTS 3 included to the dataset.

Theme	characteristic	Theme	characteristic
Social fabric	MEDIAN_AGE	Digitalization	FIX_AVG_D_KBPS
	NET_MIGR		FIX_AVG_U_KBPS
	GENDER_BALANCE		FIX_AVG_LAT_MS
Living conditions	HOMICIDE_100K		MOB_AVG_D_KBPS
	ASSAULT_100K		MOB_AVG_U_KBPS
	ROBBERY_100K	MOB_AVG_LAT_MS	
	BURGLARY_100K	Landscape and environment	COOLING_DEGREE_DAYS
	VEHICLE_THEFT_100K		HEATING_DEGREE_DAYS
	FERTILITY_RATE		SOIL_EROSION
	MEAN_AGE_CHILDBIRTH		NAT_HAZ_SUSC
	POPULATION_TOTAL	Regional typology	MOUNT_TYPE
POPULATION_WOMEN	COAST_TYPE		
POP_DENSITY	ISLAND_AREA		
Economy and labour market	GDP		
	EMPLOYMENT		
	AGE_DEP_3 RD		
	AGE_DEP_YOUNG		
	AGE_DEP_OLD		

4. MICROSIMULATION DATA

Microsimulation data providing input for an agent-based modelling is collected from the five pilot regions from the national statistical authorities of the pilot countries (Spain, Netherlands, Italy, Greece & Finland). These pilot regions are the following:

- Spain: Castilla-La Mancha (ES42)
- The Netherlands: Groningen (NL11)
- Italy: Lombardy (ITC4)
- Greece: Central Macedonia (EL52)
- Finland: North & East Finland (FI1D)

The regional level of analysis for ABM is the NUTS 3 level, so the models can account for intra NUTS 2 region mobility as well. That is, the microsimulation data consists of data from NUTS 3 regions that make up one NUTS 2 region. The data is presented in separate tabular files and provides simple counts of individuals/households in each category across the demographic and employment variables per year per region. The years the data is generally available for are 2001, 2011, and 2021 for the pilot regions. However, some countries have data only for census years, and others have annual or semi-annual data available, in-between years are also partly available. Furthermore, the 2021 census round has been shown to be a challenge to finalize data in several European countries, thus some data is not available as the statistical authorities of the country have not been able to produce the data at all. Here we provide the most recent available data.

Technically, the microsimulation data consists of CSV files describing the NUTS 3 regions using common demographic and socio-economic variables annually. These files are provided for each country separately per theme. Below is a general description of the data:

- **Age/sex data**
 - Count of individuals belonging to age/sex brackets that represent female/male residents within 10-year age brackets
 - E.g., male_20-29, female_20-29, male_30-39, femal_30-39.
- **Marital status data**
 - Count of individuals who are married, not married, divorced, or widowed
- **Education data**
 - Count of individuals belonging to the ISCED education categories
- **Household size data**
 - Count of households based on how many members they consist of
 - For some countries, data also includes how the number of children/adults in households
- **Employment data**
 - Count of individuals who are employed, unemployed, students, retired, or outside the labour force for some other reason
 - Count of individuals belonging to different occupational categories (ISCO)
 - Values before 2008 use ISCO-88, after use ISCO-08
 - Count of individuals employed in different NACE sectors
 - Values before 2008 use NACE rev 1, after use NACE rev 2.

5. MOBI-TWIN SURVEY

5.1 METHODOLOGY OF THE SURVEY

The methodology comprises a quantitative research design for analysing spatial mobility that combines elements of survey design and stakeholder engagement for data collection and integration into a larger dataset. Based on the input from D1.1 and the literature review of previous related studies in green and digital transition, and spatial mobility, WR developed a standardised questionnaire to assess citizens' spatial mobility decisions in relation to different life events, forms of mobility, regional attractiveness factors and determinants of immobility. All contributing partners supported WR in the design of the questionnaire by providing feedback. The pilot partners translated the questionnaire items in the languages of the pilot countries, namely Greek, Finnish, Italian, Spanish and Dutch.

Questionnaire development: The survey consisted of closed-ended questions, with response options consisting of multiple choice and Likert scale questions. The survey has been divided into the following seven (7) sections (plus an introductory “Welcome note” section) covering the main topics to be addressed:

1. Demographic data - to obtain information on the different socio-economic backgrounds of the survey participants. This included questions on income level, household composition, employment status and urban or rural residence. In addition, questions on cultural and regional differences and technological knowledge were integrated to obtain a holistic view of the respondent's context. By exploring these demographic aspects, the survey seeks to decipher nuanced patterns and correlations between individual characteristics and spatial mobility decisions, enriching the dataset with multiple perspectives.
2. Place of residence (current and previous) – to obtain information about insights into patterns of migration, preferences on place of residence, and the impact of factors such as working arrangements and the COVID-19 pandemic on mobility and lifestyle choices.
3. Intention to change residence – to evaluate various factors such as career advancement opportunities, quality of life indicators, and proximity to social support networks.
4. Traditional factors of (im)mobility – relevant information according to the literature include the following: employment opportunities, educational institutions, social circles and amenities, with the objective to capture how they still influence mobility and compare them with non-traditional factors, such as the green and digital transitions, the so-called Twin Transition.
5. Factors related to digitalization – to capture how emerging factors, and digitalization in particular, influence the (im)mobility of individuals. These factors, based on the literature review, include access to high-speed internet, active participation in social

networks, readily available e-services from the public sector, a diverse range of e-commerce options from private companies, and support for flexible and remote working arrangements within the community.

6. Factors related to the green transition – to capture how emerging factors, and particularly the green transition, influence the (im)mobility of individuals. The literature suggests that these factors include air quality, promotion of renewable energy, affordable energy prices, clean water supply, access to green and blue spaces, green infrastructure, circular economy principles, and community support for green behaviours.
7. Attitudes towards life satisfaction and corruption – to capture societal values and behaviours, promoting ethical governance, and fostering transparency.

Data collection: The survey was designed and structured using the SurveyMonkey platform. A non-random quota sampling approach was employed to ensure that certain quotas were reached (i) at the EU level and (ii) at the corresponding pilot regions' country level. The survey was then outsourced to collect responses via two channels:

(i) outsourcing the survey to Dynata company for performing external data collection focusing on gathering responses from pilot regions (Central Macedonia, Groningen, Castilla-La Mancha, Northern and Eastern Finland, Lombardy) and expanding the data collection in the corresponding pilot countries (Greece, Netherlands, Spain, Finland, Italy).

and

(ii) disseminating the survey through the Prolific platform, which is a paid crowdsourcing platform. The dissemination of the survey through Prolific was implemented and monitored by White Research as a supplementary channel that was used after the Dynata outsourcing procedure had finished. In this case, respondents were paid 1.14 EUR for a 10-minute questionnaire. We opted for this platform to reach a wider audience and collect many responses within a short time. This allowed us to obtain a robust sample size for data analysis. To ensure that we targeted the correct participants, we added filters such as country of residence in Prolific's screening process. Specifically, we screened for participants from all 27 EU Member States. Additionally, we pre-screened participants to ensure that we received quality and serious responses. To do this, we only included participants who had previously participated in at least 20 studies and had an approval rate of their contributions of at least 95%. Along this process, the survey response rates were closely monitored by WR.

Sample size: The survey targeted approximately 7,000 respondents at the EU level and over 1,000 respondents per pilot country: Greece, the Netherlands, Spain, Finland, and Italy. By M7 of the project (December 2023), WR collected a total of 11,622 complete responses. The data was then cleaned and geocoded with the critical support of AUTH, UB, and UH, which assigned regional codes to respondents. After the cleansing process

and geocoding, the final dataset includes the following characteristics at the country level, with pilot countries highlighted in green.

The total number of valid responses for our analysis, and therefore, the sample size for the EU-level data is 9,146 surveys (Table 2) with a confidence level of 95%, and for the worst case ($p = 0.50$) implied a maximum sampling error of $\pm 1\%$. In the case of the pilot areas, both countries and regions, the sample sizes presented in Table 2 relate to the following sampling errors with a confidence level of 95%:

- ES42 - Castilla-la Mancha: 610 observations - $\pm 4\%$
- FI1D - Northern and Eastern Finland: 582 observations - $\pm 4\%$
- EL52 - Central Macedonia: 775 observations - $\pm 3.5\%$
- ITC4 - Lombardy: 915 observations - $\pm 3.2\%$
- NL11 - Groningen: 230 observations - $\pm 6.5\%$

Table 2: Number of observations per country and age/gender group.

NUTSO	male_18-34	male_35-54	male_55+	female_18-34	female_35-54	female_55+	Total
AT	20	0	1	15	10	1	47
BE	13	6	1	14	6	1	41
BG	0	0	0	1	0	0	1
CZ	16	5	0	16	5	0	42
DE	143	57	6	128	27	8	369
DK	4	3	1	1	2	0	11
EE	7	4	0	14	6	0	31
EL	300	310	72	319	225	44	1271
ES	305	308	94	389	341	73	1511
FI	134	169	159	153	214	147	976
FR	59	17	6	78	17	4	181
HU	52	18	1	35	11	3	120
IE	8	13	4	23	25	5	78
IT	453	310	195	515	296	167	1937
LU	0	0	0	2	1	0	3
LV	10	1	1	10	1	0	23
NL	252	142	154	314	224	146	1232
PL	258	48	3	224	37	2	573
PT	228	58	7	275	42	8	618
SE	19	11	2	15	4	1	52
SI	11	1	0	14	2	0	28
SK	1	0	0	0	0	0	1
Total	2293	1481	707	2555	1496	610	9146

Given that the focus of the project lies at the regional NUTS2 level, the responses for our pilot-specific regions are given in Table 3. In the case of Groningen, it was not possible to collect a high number of responses from this specific region, so we have used data from nearby regions belonging in the same NUTS1 region to supplement our analysis and ensure representativeness (NL12 - Friesland 21 responses and NL13 - Drenthe 207 responses).

Table 3: Number of observations per pilot region (NUTS2).

NUTS2 Pilot region	Observations
ES42 – Castilla-La Mancha	610
FI1D – Northern and Eastern Finland	582
EL52 – Central Macedonia	775
ITC4 – Lombardy	915
NL11 – Groningen	230

To balance our sample across gender, age, and regional levels for pan-European analysis, we have calculated the corresponding weights. Due to a very low number of responses in specific regions, we were unable to weight at the NUTS2 or NUTS0 levels. Instead, we calculated regional weights based on EU macro regions (Northern, Western, Southern, and Eastern Europe). Additionally, individual weights were calculated based on gender (male/female) and age groups (18-34, 35-54, 55+), using data available at the EU level for 2023. The derived weights for our survey, based on EU macro regions and individual characteristics, are provided in Table 4 below. Each weight has been applied to individuals according to their personal characteristics and place of residence.

Table 4: Calculated weights considering age/gender and region of residence.

REGIONS	m_18-34	m_35-54	m_55+	f_18-34	f_35-54	f_55+	region
Northern	0.875	0.961	1.331	0.701	0.754	1.666	0.573
Western	0.488	1.325	2.242	0.413	1.041	2.750	2.202
Southern	0.455	0.931	2.923	0.370	1.022	4.385	0.553
Eastern	0.274	1.972	32.977	0.303	2.511	34.485	1.827

The final dataset was provided to UH, which integrated the survey dataset into the full MOBI-TWIN dataset under T1.3.

5.2 ETHICAL AND LEGAL CONSIDERATIONS

The survey was designed in line with Responsible Research and Innovation (RRI) principles and reviewed by the European Science Foundation (ESF), which is responsible for aligning the scope and objectives of MOBI-TWIN with the values, needs and expectations of society based on RRI (T4.1). Particular attention was paid to the inclusion of underrepresented population groups such as ethnic minorities, migrants, women, the LGBTIQ+ community, people with physical disabilities and people at risk of poverty. Given the large number of respondents (around 12,000 people took part in the survey) the survey was also reviewed by the project's Ethics Advisor.

To ensure the ethical and legal conduct of the survey, the following measures were taken:

1. General Data Protection Regulation compliance: All data processing activities comply with the guidelines of the General Data Protection Regulation.
2. A welcome note was added to the survey informing participants about the content of the project, the location of the project activities, the method of data collection, the description of the data, the description of the methods of data use, the retention and possible deletion, the procedure for withdrawing consent and the risks for individuals.

3. The participation was based on freely given informed consent and the participants were reminded that they are free to withdraw from the research at any time of the survey completion, if they feel uncomfortable.
4. Anonymisation: All data collected were anonymised to protect the identity of participants.
5. Compliance with RRI principles.
6. Co-operation with the ethics advisor.

6. SPATIAL LAYERS

The spatial layers used in the Mobi-Twin project are the official versions of the NUTS 2 level territorial division provided by GISCO (GISCO, 2024) as the regional level for this project. The dataset includes the six official NUTS versions from years 2003, 2006, 2010, 2013, 2016 and 2021. Most of the regional data is available for the three most recent NUTS division versions. The dataset includes NUTS 2 divisions, where all oversee regions (the Caribbean, Indian Ocean nor the Svalbard region) are excluded as the Mobi-Twin project focuses on units located in Europe (Table 5).

Table 5. The number of NUTS 2 units in Europe studied in the Mobi-Twin project (2003 – 2022).

	2003	2006	2010	2013	2016	2021
NUTS 2 unit count (n)	309	314	312	316	327	328

Over time, some of the NUTS 2 territorial units have changed logically without changing the overall territorial division of a country radically (e.g. existing region split into several smaller regions or merged existing regions into one larger region) and in some cases the existing territorial division is spatially redefined that is geographically incompatible with the previous division. For the first case, tracking and linking NUTS 2 units over years is feasible (e.g. linking unit code changes; merging unit data). In the latter cases, longitudinal analysis of given territorial units is not possible.

For instance, by redefining the territorial division of NUTS 2 in Ireland during the 2010s or structural changes in the Southern Norway does not allow to link NUTS 2 region data for longitudinal trend analysis from these countries/regions.

To overcome the systematic challenge of changing NUTS units and improve the usability of NUTS (e.g. NUTS 2) level datasets in future, the authors propose converting the historical

NUTS level data into newer NUTS territorial division for data consistency and enabling the longitudinal perspective to the data.

7. REFERENCES

- Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives*, 32(3), 259–280.
- Crippa, M., Guizzardi, D., Pagani, F., & Pisoni, E. (2023). *GHG Emissions at sub-national level (v2.0) (2.0)* [dataset]. European Commission, JRC. <https://doi.org/10.2905/D67EEDA8-C03E-4421-95D0-0ADC460B9658>
- Directorate-General for Education, Youth, Sport and Culture. (2023). *Erasmus mobility statistics 2014–2020—Data Europa EU*. <https://data.europa.eu/data/datasets/erasmus-mobility-statistics-2014-2020?locale=en>
- European Commission and Eurostat. (2019). *Methodological manual on territorial typologies: 2018 edition* (2018 edition). Publications Office of the European Union. <https://data.europa.eu/doi/10.2785/930137>
- European Commission and Eurostat (2021). *Applying the degree of urbanisation: A methodological manual to define cities, towns and rural areas for international comparisons: 2021 edition*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2785/706535>
- ESPON (2024). *ESPON Database Portal*. <https://database.espon.eu/>
- GeoPy (2024). *Geopy* [Python]. geopy. <https://github.com/geopy/geopy> (Original work published 2010)
- GISCO. (2024). *Territorial units for statistics (NUTS) — Eurostat*. <https://ec.europa.eu/eurostat/web/gisco/geodata/statistical-units/territorial-units-statistics>
- Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1), 109.
- Ookla (2024). *Speedtest® by Ookla® Global Fixed and Mobile Network Performance Maps* [Python]. Ookla. <https://github.com/teamookla/ookla-open-data> (Original work published 2020).
- Panori, A., Kakderi, C., Komninos, N., Mariotti, I., Rossi, F. M., and Alfieri, L. (2024). State-of-the-art report on the drivers, forms and effects of spatial mobility on EU regions. Deliverable D1.1 of the MOBI-TWIN Project.
- Poorthuis, A., & Zook, M. (2017). Making Big Data Small: Strategies to Expand Urban and Geographical Research Using Social Media. *Journal of Urban Technology*, 24(4), 115–135.
- Rodríguez-Pose, A., & Bartalucci, F. (2023). The green transition and its potential territorial discontents. *Cambridge Journal of Regions, Economy and Society*.
- Schiavina, M., Melchiorri, M., & Freire, S. (2023). *GHS-DUC R2023A - GHS Degree of Urbanisation Classification, application of the Degree of Urbanisation methodology (stage II) to GADM 4.1 layer, multitemporal (1975-2030)* [dataset]. <https://doi.org/10.0.11.89/DC0EB21D-472C-4F5A-8846-823C50836305>
- Velthuis, S., Royer, J., Le Petit-Guerin, M., Cauchi-Duval, N., Franklin, R., Leibert, T., MacKinnon, D., and Pike, A. (2023) Locating 'left-behindness' in the EU15: a regional typology. Beyond Left Behind Places Project Working Paper 03/23. Centre for Urban and Regional Development Studies (CURDS), Newcastle University, UK

ANNEX 1: DATA SOURCES

MOBILITY DATA

Annex Table 1. Links to data source and time period of available data for the mobility type.

Name	Available years	Source	Link
Permanent migration	20012-2022	Twitter/X	https://developer.x.com/en/products/twitter-api
Long-term student mobility	2014-2020	Erasmus	https://data.europa.eu/data/datasets/erasmus-mobility-statistics-2014-2020?locale=en
Short-term student mobility	2014-2022	Erasmus	https://data.europa.eu/data/datasets/erasmus-mobility-statistics-2014-2020?locale=en
Seasonal work	2005-2022	LFS	https://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey
Long-distance commuting	2005-2022	LFS	https://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey
Multilocal living	2012-2022	Twitter/X	https://developer.x.com/en/products/twitter-api
Cross-border commuting	2005-2022	LFS	https://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey

REGIONAL CHARACTERISTICS DATA

SOCIAL FABRIC

Annex Table 2. Links to data source for the social fabric variables.

Variable	Link to data
WOMEN_PER_100_MEN	https://ec.europa.eu/eurostat/databrowser/view/demo_r_pjanind2/default/table?lang=en
AT_RISK_OF_POVERTY	https://ec.europa.eu/eurostat/databrowser/view/ilc_li41__custom_9428261/default/table?lang=en
TERTIARY_STUDENTS	https://ec.europa.eu/eurostat/databrowser/view/educ_uoe_enrt06__custom_9424513/default/table?lang=en
HOUSEHOLD_INCOME	https://ec.europa.eu/eurostat/databrowser/view/nama_10r_2hhinc/default/table?lang=en
MA_SO_DEPRIVATION	https://ec.europa.eu/eurostat/databrowser/view/ilc_mdtd18__custom_9428400/default/table?lang=en
MEDIAN_AGE	https://ec.europa.eu/eurostat/databrowser/view/demo_r_pjanind2/default/table?lang=en
NET_MIGR	https://ec.europa.eu/eurostat/databrowser/view/demo_r_gind3__custom_9429268/default/table?lang=en
SHARE_OF_IMMIGRANTS	https://ec.europa.eu/eurostat/databrowser/view/cens_11ctzo_r2__custom_9460698/default/table?lang=en
PEOPLE_BALANCE_IDX	https://database.espon.eu/indicator/2587/#metadata-download

ECONOMY AND LABOUR MARKET

Annex Table 3. Links to data source for the economy and labour market variables.

Variable	Link to data
----------	--------------

GDP	https://ec.europa.eu/eurostat/databrowser/view/nama_10r_2gdp__custom_9823161/default/table?lang=en
GVA	https://ec.europa.eu/eurostat/databrowser/view/nama_10r_2qvagr__custom_9823352/default/table?lang=en
EMPLOYED_M_AGE15-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfe2emprtn__custom_9826724/default/table?lang=en
EMPLOYED_M_AGE20-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfe2emprtn__custom_9826748/default/table?lang=en
EMPLOYED_F_AGE15-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfe2emprtn__custom_9826792/default/table?lang=en
EMPLOYED_F_AGE20-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfe2emprtn__custom_9826826/default/table?lang=en
UNEMPLOYED_M_AGE15-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfur2gac__custom_9823679/default/table?lang=en
UNEMPLOYED_M_AGE20-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfur2gac__custom_9823774/default/table?lang=en
UNEMPLOYED_F_AGE15-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfur2gac__custom_9823470/default/table?lang=en
UNEMPLOYED_F_AGE20-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfur2gac__custom_9823584/default/table?lang=en
UNEMPLOYED_T_AGE15-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfur2gac__custom_9823877/default/table?lang=en
UNEMPLOYED_T_AGE20-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfur2gac__custom_9823929/default/table?lang=en
WEEKLY_HRS_M_AGE15-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfe2ehour__custom_9824011/default/table?lang=en
WEEKLY_HRS_M_AGE20-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfe2ehour__custom_9824138/default/table?lang=en
WEEKLY_HRS_F_AGE15-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfe2ehour__custom_9824283/default/table?lang=en
WEEKLY_HRS_F_AGE20-64	https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfe2ehour__custom_9824338/default/table?lang=en
AGE_DEP_RATIO	https://ec.europa.eu/eurostat/databrowser/view/demo_r_pjanind2__custom_9824495/default/table?lang=en
OLD_AGE_DEP_RATIO	https://ec.europa.eu/eurostat/databrowser/view/demo_r_pjanind2__custom_9824564/default/table?lang=en
YOUNG_AGE_DEP_RATIO	https://ec.europa.eu/eurostat/databrowser/view/demo_r_pjanind2__custom_9824629/default/table?lang=en
HIGH_TECH_M	https://ec.europa.eu/eurostat/databrowser/view/htec_emp_reg2__custom_9824715/default/table?lang=en
HIGH_TECH_F	https://ec.europa.eu/eurostat/databrowser/view/htec_emp_reg2__custom_9824777/default/table?lang=en
WAGES_CONSTRUCTION	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9824840/default/table?lang=en
WAGES_EGSA	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9824909/default/table?lang=en
WAGES_MANUFACTURING	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9824935/default/table?lang=en
WAGES_MINING_QUARRYING	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9824970/default/table?lang=en

WAGES_WATER_SUPPLY	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9825060/default/table?lang=en
WAGES_WRTR	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9825126/default/table?lang=en
WAGES_ACCOM_FOODSERVICE	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9825183/default/table?lang=en
WAGES_INFO_COMM	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9825216/default/table?lang=en
WAGES_TRANSPORT_STORAGE	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9825252/default/table?lang=en
WAGES_REAL_ESTATE	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9825309/default/table?lang=en
WAGES_ADMIN_SUPPORT	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9825377/default/table?lang=en
WAGES_PROF_SCIEN_TECH	https://ec.europa.eu/eurostat/databrowser/view/sbs_r_NUTS_06_r2__custom_9825421/default/table?lang=en
LEFT_BEHINDEDNESS_IDX	https://research.ncl.ac.uk/beyondleftbehindplaces/publicationsanddownloads/Locating%20'left-behindness'_working%20paper%200123.pdf

LIVING CONDITIONS

Annex Table 4. Links to data source for the living conditions variables.

Variable	Link to data
HOMICIDE_PER_100K	https://ec.europa.eu/eurostat/databrowser/view/crim_gen_reg__custom_9864936/default/table?lang=en
ASSAULT_PER_100K	https://ec.europa.eu/eurostat/databrowser/view/crim_gen_reg__custom_9865026/default/table?lang=en
ROBBERY_PER_100K	https://ec.europa.eu/eurostat/databrowser/view/crim_gen_reg__custom_9865092/default/table?lang=en
BURGLARY_PER_100	https://ec.europa.eu/eurostat/databrowser/view/crim_gen_reg__custom_9865164/default/table?lang=en
VEHICLE_THEFT_PER_100K	https://ec.europa.eu/eurostat/databrowser/view/crim_gen_reg__custom_9865203/default/table?lang=en
FERTILITY_RATE	https://ec.europa.eu/eurostat/databrowser/view/demo_r_find2__custom_9865340/default/table?lang=en
MEAN_AGE_CHILDBIRTH	https://ec.europa.eu/eurostat/databrowser/view/demo_r_find2__custom_9865592/default/table?lang=en
MORTALITY	https://ec.europa.eu/eurostat/databrowser/view/hlth_cd_asdr2__custom_9865649/default/table?lang=en
LIFE_EXPECTANCY	https://ec.europa.eu/eurostat/databrowser/view/demo_r_mlifexp__custom_9865782/default/table?lang=en
POPULATION_TOTAL	https://ec.europa.eu/eurostat/databrowser/view/demo_r_dzjan__custom_9865853/default/table?lang=en
POPULATION_WOMEN	https://ec.europa.eu/eurostat/databrowser/view/demo_r_dzjan__custom_9866015/default/table?lang=en
PHYSICIANS_PER_100K	https://ec.europa.eu/eurostat/databrowser/view/hlth_rs_physreg__custom_9866080/default/table?lang=en

POPULATION_DENSITY	https://ec.europa.eu/eurostat/databrowser/view/demo_r_d2jan__custom_9866297/default/table?lang=en https://ec.europa.eu/eurostat/databrowser/view/reg_area3__custom_9866823/default/table?lang=en
HOSP_BEDS_PER_100K	https://ec.europa.eu/eurostat/databrowser/view/hlth_rs_bdsrg2__custom_9866413/default/table?lang=en
HUMAN_RESOURCES_IDX	https://database.espon.eu/indicator/544/#map
GOOD_GOVERNANCE_IDX	https://database.espon.eu/indicator/539/#map

ACCESS AND CONNECTIVITY

Annex Table 5. Links to data source for the access and connectivity variables.

Variable	Link to data
MOTORWAYS	https://ec.europa.eu/eurostat/databrowser/view/tran_r_net__custom_9885998/default/table?lang=en
RAILROADS	https://ec.europa.eu/eurostat/databrowser/view/tran_r_net__custom_9886195/default/table?lang=en
AIRPORTS	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/transport-networks
AIR_PASSENGERS	https://ec.europa.eu/eurostat/databrowser/view/tran_r_avpa_nm__custom_9886383/default/table?lang=en
AIR_FREIGHT_LOADED	https://ec.europa.eu/eurostat/databrowser/view/tran_r_avgo_nm__custom_9886542/default/table?lang=en
AIR_FREIGHT_UNLOADED	https://ec.europa.eu/eurostat/databrowser/view/tran_r_avgo_nm__custom_9886642/default/table?lang=en
HARBORS	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/transport-networks
MARITIME_EMBARKED	https://ec.europa.eu/eurostat/databrowser/view/tran_r_mapa_nm__custom_9886731/default/table?lang=en
MARITIME_DISEMBARKED	https://ec.europa.eu/eurostat/databrowser/view/tran_r_mapa_nm__custom_9886799/default/table?lang=en
MARITIME_FREIGHT_LOADED	https://ec.europa.eu/eurostat/databrowser/view/tran_r_mago_nm__custom_9886867/default/table?lang=en
MARITIME_FREIGHT_UNLOADED	https://ec.europa.eu/eurostat/databrowser/view/tran_r_mago_nm__custom_9886987/default/table?lang=en
ARR_TOURIST_ACCOM	https://ec.europa.eu/eurostat/databrowser/view/tour_occ_arn2__custom_9887070/default/table?lang=en

LANDSCAPE AND ENVIRONMENT

Annex Table 6. Links to data source for the landscape and environment variables.

Variable	Link to data
AGRICULTURE	https://ec.europa.eu/eurostat/databrowser/view/lan_use_ovw__custom_10147245/default/table?lang=en
COOLING_DEGREE_DAYS	https://ec.europa.eu/eurostat/databrowser/view/nrg_chddr2_a__custom_10147288/default/table?lang=en
HEATING_DEGREE_DAYS	https://ec.europa.eu/eurostat/databrowser/view/nrg_chddr2_a__custom_10147321/default/table?lang=en
SOIL_EROSION_WATER	https://ec.europa.eu/eurostat/databrowser/view/aei_pr_soiler__custom_10147355/default/table?lang=en

FISHING_AQUACULTURE	https://ec.europa.eu/eurostat/databrowser/view/lan_use_ovw__custom_10147406/default/table?lang=en
FORESTRY	https://ec.europa.eu/eurostat/databrowser/view/lan_use_ovw__custom_10147454/default/table?lang=en
HEAVY_ENV_IMPACT	https://ec.europa.eu/eurostat/databrowser/view/lan_use_ovw__custom_10147519/default/table?lang=en
NATURE_RESERVES	https://ec.europa.eu/eurostat/databrowser/view/lan_use_ovw__custom_10147568/default/table?lang=en
SERVICES_RESIDENTIAL	https://ec.europa.eu/eurostat/databrowser/view/lan_use_ovw__custom_10147641/default/table?lang=en
OTHER_PRIM_SECTOR_A CTS	https://ec.europa.eu/eurostat/databrowser/view/lan_use_ovw__custom_10147676/default/table?lang=en
UNUSED_ABANDONED	https://ec.europa.eu/eurostat/databrowser/view/lan_use_ovw__custom_10147703/default/table?lang=en
GREEN_INFRA	https://database.espon.eu/indicator/1034/#metadata-download
PV_POTENTIAL	https://database.espon.eu/indicator/582/#metadata-download
MEAN_MAX_TEMP	https://www.worldclim.org/data/monthlywth.html
MEAN_MIN_TEMP	https://www.worldclim.org/data/monthlywth.html
MEAN_PRECIPITATION	https://www.worldclim.org/data/monthlywth.html
CIRC_ECON_EMPLOYME NT	https://database.espon.eu/indicator/1060/#metadata-download
CIRC_ECON_TURNOVER	https://database.espon.eu/indicator/1061/#metadata-download
NATURAL_HAZARD_SUS CEP	https://database.espon.eu/indicator/2225/#metadata-download
TAPC_IDX	https://database.espon.eu/indicator/540/#metadata-download
RGTV_IDX	https://single-market-economy.ec.europa.eu/publications/regional-vulnerability-green-transition_en
BIO_GEO_TYOLOGY	https://www.researchgate.net/figure/Biogeographical-regions-Roekaerts-2002-of-the-study-area-in-Europe-overlying-NUTS_2_fig1_297758320 https://www.eea.europa.eu/data-and-maps/figures/biogeographical-regions-in-europe-2/map_2-1_biogeographical-regions https://bmcecolevol.biomedcentral.com/articles/10.1186/s12862-019-1576-z https://www.researchgate.net/figure/Map-of-the-Neotropical-region-depicting-the-transition-zones-sub-regions-according-to-the_fig1_348753830 https://www.oneearth.org/bioregions-2020/
AIR_QUALITY_AVG	https://discomap.eea.europa.eu/App/AQViewer/index.html?fqn=Airquality_Dissemination.NUTS_3_sel
GHG_AGRICULTURE	https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/subnational_NUTS_2/v80_FT2022_GHG_NUTS_2_v2/EDGARv8.0_total_GHG_GWP100_AR5_NUTS_2_1990_2022.zip
GHG_BUILDINGS	https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/subnational_NUTS_2/v80_FT2022_GHG_NUTS_2_v2/EDGARv8.0_total_GHG_GWP100_AR5_NUTS_2_1990_2022.zip
GHG_ENERGY	https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/subnational_NUTS_2/v80_FT2022_GHG_NUTS_2_v2/EDGARv8.0_total_GHG_GWP100_AR5_NUTS_2_1990_2022.zip
GHG_INDUSTRY	https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/subnational_NUTS_2/v80_FT2022_GHG_NUTS_2_v2/EDGARv8.0_total_GHG_GWP100_AR5_NUTS_2_1990_2022.zip
GHG_TRANSPORT	https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/subnational_NUTS_2/v80_FT2022_GHG_NUTS_2_v2/EDGARv8.0_total_GHG_GWP100_AR5_NUTS_2_1990_2022.zip
GHG_WASTE	https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/subnational_NUTS_2/v80_FT2022_GHG_NUTS_2_v2/EDGARv8.0_total_GHG_GWP100_AR5_NUTS_2_1990_2022.zip

DIGITALIZATION

Annex Table 7. Links to data source for the digitalization variables.

Variable	Link to data
BROADBAND_ACCESS	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_broad_h__custom_10201032/default/table?lang=en
INTERNET_ACCESS	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_iacc_h__custom_10201095/default/table?lang=en
INTERNET_AWAY	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_iumd_i__custom_10201125/default/table?lang=en
INTERNET_PUBL_AUTH	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_gov_i__custom_10201158/default/table?lang=en
INTERNET_SELLING	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_iuse_i__custom_10201227/default/table?lang=en
LAST_ONLINE_PUR_3MO	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_blt12_i__custom_10201279/default/table?lang=en
LAST_ONLINE_PUR_3-12MO	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_blt12_i__custom_10201345/default/table?lang=en
LAST_ONLINE_PUR_12MO	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_blt12_i__custom_10201419/default/table?lang=en
LAST_ONLINE_PUR_YEAR+	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_blt12_i__custom_10201471/default/table?lang=en
LAST_ONLINE_PUR_YEAR+_NEVER	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_blt12_i__custom_10201508/default/table?lang=en
NEVER_USED_COMPUTER	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_cux_i__custom_10201574/default/table?lang=en
ONLINE_BANKING	https://ec.europa.eu/eurostat/databrowser/view/isoc_r_iuse_i__custom_10201648/default/table?lang=en
FIX_AVG_D_KBPS	https://ookla-open-data.s3.amazonaws.com/
FIX_AVG_U_KBPS	https://ookla-open-data.s3.amazonaws.com/
FIX_AVG_LAT_MS	https://ookla-open-data.s3.amazonaws.com/
MOB_AVG_D_KBPS	https://ookla-open-data.s3.amazonaws.com/
MOB_AVG_U_KBPS	https://ookla-open-data.s3.amazonaws.com/
MOB_AVG_LAT_MS	https://ookla-open-data.s3.amazonaws.com/

REGIONAL PROFILE

Annex Table 8. Links to data source for the regional profile variables.

Variable	Link to data
MOUNT_TYPE_1	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts
MOUNT_TYPE_2	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts
MOUNT_TYPE_3	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts

MOUNT_TYPE_4	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts
COAST_TYPE_1	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts
COAST_TYPE_2	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts
COAST_TYPE_3	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts
ISLAND_PERCENTAGE	https://ec.europa.eu/eurostat/web/nuts/overview
URB_RUR TYPOLOGY	https://human-settlement.emergency.copernicus.eu/download.php?ds=DUC
URBAN_SHARE	https://human-settlement.emergency.copernicus.eu/download.php?ds=DUC
RURAL_SHARE	https://human-settlement.emergency.copernicus.eu/download.php?ds=DUC



TWIN TRANSITION AND CHANGING PATTERNS OF SPATIAL MOBILITY: A REGIONAL APPROACH

The Partners of the MOBI-TWIN Consortium:

Name	Country
White Research SRL	Belgium
Aristotelio Panepistimio Thessalonikis	Greece
Rijksuniversiteit Groningen	The Netherlands
Universitat de Barcelona	Spain
Helsingin Yliopisto	Finland
Fondation Europeene de la Science	France
Politecnico di Milano	Italy
Erevnitiko Panepistimiako Institutou Periferiakis Anaptixis	Greece
Accelignce Ltd.	Cyprus

[CONNECT WITH MOBI-TWIN FOR MORE UPDATES:](#)

[LinkedIn](#), [Twitter](#), [Website](#)