

**Geographia Polonica**

2026, Volume 99, Issue 3, pp. 333-357

<https://doi.org/10.7163/GPol.0330>

INSTITUTE OF GEOGRAPHY AND SPATIAL ORGANIZATION
POLISH ACADEMY OF SCIENCES
www.igipz.pan.pl

www.geographiapolonica.pl

OLD WINE IN NEW BOTTLES: A PAN-EUROPEAN REGIONAL DATASET OF POPULATION AND GDP SINCE 1960

Piotr Rosik  • Marcin Mazur  • Rafał Wiśniewski  • Patryk Duma 

Institute of Geography and Spatial Organization
Polish Academy of Sciences

Twarda 51/55, 00-818 Warsaw: Poland

e-mails: rosik@twarda.pan.pl (corresponding author) • m.mazur@twarda.pan.pl • rafwis@twarda.pan.pl • p.duma@twarda.pan.pl

Abstract

This article presents a harmonised procedural framework for constructing a new open-access pan-European regional database covering population, GDP, and derived GDP per capita for 1546 regions over the period 1960-2021; the empirical illustrations presented in the article cover the period 1960-2020. The resulting resource is distributed in three supplementary Excel workbooks, one for each variable. Drawing on national and regional sources, including the Maddison Project Database as the main GDP anchor, ARDECO, the Rosés-Wolf regional GDP database, database.earth, and country-specific sources, it provides continuous annual time-series coverage and broad cross-country comparability. Its main contribution lies in combining pan-European spatial coverage, a six-decade temporal span, and the systematic harmonisation of heterogeneous national and regional data sources. The resource enables analyses of regional income convergence, spatial dynamics, and long-term demographic and economic polarisation across Europe. By extending beyond the European Union, it fills major spatial and temporal gaps in existing resources and creates new opportunities for spatial economic and regional research.

Keywords

Pan-European spatial coverage • regional population • GDP per capita • NUTS3 • GDP retropolation • Maddison Project Database • ARDECO

Introduction

Regional and local datasets, including demographic and socio-economic data, often with high spatial resolution, are by no means a novelty (Bernal et al., 2021). The richest such datasets can be found in selected OECD countries, particularly those with a long tradition

of statistical and cartographic research (Kotavaara, 2012). However, what is attainable, sometimes even over extended time periods, at the local, regional, or national level, often becomes much more difficult to achieve at the scale of an entire continent. This stems primarily from the fact that each continent is highly heterogeneous, both in terms of the level

of development of its constituent regions and the availability of regional data, especially when analyses span long time series. Even Europe, the most advanced continent in terms of integration, including “statistical integration,” remains strongly differentiated spatially with regard to the dissemination of basic data over longer periods of time (Laaboudi et al., 2024).

Thus, although for individual European countries (through their national statistical offices) or groups of countries (primarily those covered by Eurostat data collection; see Walaszek, 2025) it is possible to use ready-made databases prepared by professional teams of statisticians at the NUTS or LAU level, this “traditional” group of datasets also has its limitations, including, among others, the absence of longer time series, extending over several decades, on GDP, particularly for countries that joined the EU in this century, as well as the lack of comparability between data for countries within the Eurostat system and those outside it (Antal, 2019, see also the broader discussion).

On the other hand, in recent years an increasing number of demographic and economic databases have emerged, including those covering the entire world and utilising so-called grid-based datasets (Schiavina et al., 2019; Kummu et al., 2018). These data can be freely aggregated to the chosen boundaries of any regions for selected continents. However, such datasets, for example those based on nighttime lights, also have significant methodological limitations, both in terms of the length of time series and the actual correspondence with official statistical data on GDP or population collected by individual statistical offices (Phan, 2023).

These limitations explain why a consistent database on fundamental demographic (population) and economic (GDP) variables at the regional level has so far been lacking for the European continent. At the national scale, however, several global initiatives have attempted to compile historical data, such as the Maddison Project Database (MPD) (Bolt & van Zanden, 2024), which

collects information for very long time series of aggregates such as GDP (PPP), GDP per capita, and population. Although this database does not include official subnational (regional) data, it may serve as a valuable reference point for estimating historical data at the regional level.

The aim of this article is to present a harmonised procedural framework and a new open-access database on population, GDP and derived GDP per capita at the regional level (1546 regions) for the entire European continent, covering all years across six decades, that is, the period 1960-2021. The presented database is the first of its kind for the entire continent, integrating existing data sources at both the national and regional levels. Maddison is used primarily as the national GDP anchor, while population totals and regional population structures are constructed using ARDECO, database.earth, City Population, national statistical sources, and census-based estimates. Regional GDP structures are complemented by ARDECO, Rosés-Wolf, SASI, national statistical sources, and other country-specific data. Against this background, its main contribution lies primarily in its pan-European spatial coverage, long temporal scope, and systematic harmonisation of heterogeneous national and regional sources, rather than in proposing a fully new statistical methodology. The resulting dataset is distributed as three separate supplementary Excel workbooks (Supplementary files S1-S3), covering population, GDP and GDP per capita, respectively.

The potential scientific applications of this newly developed regional database on GDP and population in long time series for the whole of Europe, the titular “old wine” spanning from the Portuguese to the Crimean “vineyards”, are manifold. They include, among others, analyses of regional income convergence (as in Rosés & Wolf, 2018, or Smętkowski, 2015), long-term consequences of cohesion policy (Žuber et al., 2025), spatial econometrics analyses (Özyurt & Dées, 2015), the delineation of population- or GDP-weighted centroids of the European

continent, as well as studies of changes in accessibility (see Rosik et al., 2025a,b) and of relationships between, for instance, accessibility dynamics and regional development (see Rosik & Wójcik, 2023), among others.

The article begins with a literature review focusing on datasets on population and GDP at the regional level, including grid-based datasets, regional databases, and, as background, an overview of national-level databases. The following section presents the study area, with particular attention to the adopted definition of the European continent and its division into regions. The subsequent section describes the applied methodology in a step-by-step manner, outlining the research procedure. The results section presents selected outcomes, both in terms of continuous regional monitoring (chart) and changes observed throughout the entire analysed period (cartographic representation). The article concludes with key findings and a discussion, which also highlights the limitations associated with the scientific use of the new database and outlines desirable directions for further development.

Literature review

International datasets on population and GDP differ, among other aspects, in their level of data aggregation, ranging from the lowest level, that is, grid-based data, to the national level. Within this brief literature review, we first present the potential and scope of grid-based datasets, then move to the national level, and finally focus on existing European databases that collect population and GDP data at the regional level. In each case, we point out the advantages and limitations of the respective types of datasets in the context of the objectives of our article.

Grid-based datasets. The rapid development in recent years of grid-based data on population and GDP, typically at the global scale, has been the result of the work of large research teams (see Table 1, for a broader review of studies, see Archila Bustos et al., 2020, and Leyk et al., 2019), for example those associated with the European Commission's Global Human Settlement Layer (GHSL-POP; Schiavina et al., 2019), the University of Southampton (WorldPop), research conducted

Table 1. Pan-European Gridded Datasets for Population and GDP: Sources, Resolution and Temporal Coverage

Variable	Data source (author/year)	Grid resolution	Years in dataset	Link to the source of data
Population	GHSL-POP (Schiavina et al., 2019); EC-JRC	250 m	1975, 1990, 2000, 2015	https://human-settlement.emergency.copernicus.eu/ghs_pop.php
	WorldPop – WorldPop Team (University of Southampton)	100m and 1 km	Global1: 2000-2020; Global2 2015-2030 (new projections)	https://hub.worldpop.org/Glob-al1_2000-2020 ; https://hub.worldpop.org/project/categories?id=3
	GPWv4	30 arc-seconds (~1 km)	2000, 2005, 2010, 2015, 2020	https://www.earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-gpwv4-popdens-4.0#overview
GDP	G-Econ (Nordhaus, 2006)	1-degree latitude by 1-degree longitude	1990, 1995, 2000, 2005	https://gecon.yale.edu/data-and-documentation-g-econ-project
	Kummu et al. (2018)	5 arc-minutes	1990-2015	https://www.nature.com/articles/sdata20184
	Murakami et al. (2021)	5 arc-minutes	1850-2010 and scenarios for 2010-2100	https://www.frontiersin.org/journals/built-environment/articles/10.3389/fbuil.2021.760306/full

by NASA (Gridded Population of the World, GPWv4), or Yale University (G-Econ; Nordhaus, 2006) (see Table 1). In general, the grid level has the advantage over national and regional levels in that it allows aggregation to any chosen level of spatial hierarchy. However, from the perspective of econometric studies, a drawback is that GDP is often estimated using auxiliary variables, such as population, urban area or nighttime light intensity, rather than being calculated independently of other variables. Another limitation is the lack, to date, of longer time series extending more than a few decades into the past (with the exception of Murakami et al., 2021). Therefore, although the grid-based approach undoubtedly holds great potential for future development, especially in the case of population data linked to national censuses, as in the case of the WorldPop project, the review of grid data in this article serves primarily as background context.

National databases. The national level, from the perspective of this article, serves primarily as a reference point for estimating

population and GDP at the regional level. Unlike grid-based data, at this level it is possible to obtain long or even very long time series for most countries, based on various data sources, including the Maddison Project Database, the World Bank, the IMF, the OECD, and the United Nations (see Tab. 2). Most of the databases listed in Table 2 have global coverage, with certain exceptions resulting from the lack of long-term data for smaller countries (e.g. the Vatican) or autonomous regions (e.g. Transnistria). Supplementary population data not included in Table 2 can be found at <https://database.earth/population> (for example, for Andorra, the Faroe Islands, Guernsey, Jersey, Monaco, San Marino, the Isle of Man, or the Vatican). In the case of GDP, additional data can be obtained, for instance, from <https://database.earth/economy> (for example, for Liechtenstein). For a small number of European territories not fully covered by standard international datasets, supplementary national GDP data were incorporated from alternative official sources, including the *Abstracts*

Table 2. Country-Level Long-Run Datasets for Population and GDP: Sources and Temporal Coverage

Variable	Data source	Years available (typical coverage)
Population	Maddison Project Database (MPD) – population series	Long-run historical to 2022 (annual, common post-1950)
	World Bank – World Development Indicators (SP.POP.TOTL)	1960 – present (annual)
	IMF – World Economic Outlook (WEO) Database – Population	1980 – present (annual, with forecasts)
	OECD Data – Population	1960 – present (annual)
	UN DESA – World Population Prospects (WPP 2022/2024)	1950-2100 (annual)
GDP	Maddison Project Database (MPD) – GDP (real, PPP)	Long-run historical to 2022 (annual, common post-1950)
	World Bank – World Development Indicators (NY.GDP.MKTP.KD)	1960 – present (annual)
	IMF – World Economic Outlook (WEO) Database – GDP	1980 – present (annual, with forecasts)
	OECD Data – National accounts / GDP (SNA)	1960 – present (annual)
	UN National Accounts – Main Aggregates Database	1970 – present (annual)

Source: authors’ own elaboration.

of Statistics Reports published by the Government of Gibraltar (available since 1987) and publicly accessible compilations for the Vatican City providing annual data since 1960 (e.g. GeoRank). Of course, national-level databases covering very long time series also have their own limitations, such as the measurement of market production rather than non-market output, or assumptions concerning the relationship between data and the underlying economic activities represented (Bolt & van Zanden, 2024).

Regional databases in Europe. The primary source of statistical information on population and GDP in Europe is Eurostat, where within the NUTS system the lowest level at which both population and GDP data can be obtained simultaneously is NUTS3. Historically, population and GDP data at the NUTS3 level are available through the so-called ARDECO database (Annual Regional Database of the European Commission)

(<https://territorial.ec.europa.eu/ardeco/explore?lng=en>).

The population data from 1960 onward in the ARDECO database are available not only for the 27 EU member states but also for 10 other countries: Albania, Switzerland, Iceland, Liechtenstein, Montenegro, North Macedonia, Norway, Serbia, Turkey, and the United Kingdom, making a total of 37 countries. As for GDP at the NUTS3 level, the ARDECO database provides data starting from 1980 for most Western European countries that are members of the European Union, as well as for Norway, Switzerland, the United Kingdom, and Turkey, covering a total of 17 countries (although, for example, in Germany the data from 1980 refer only to the former West Germany). Unfortunately, a detailed examination of the ARDECO data revealed that some of these data are linear in nature, meaning that the shares of regional GDP remain constant across regions even for

Table 3. Main European regional and supplementary datasets used or discussed in the construction of the database

Data source	Spatial level	Variables	Temporal coverage	Role in this study
Eurostat	NUTS regions	Population, GDP and regional accounts indicators	Mainly recent decades	Reference source for harmonised European regional statistics
ARDECO	NUTS2/NUTS3	Population, GDP and other regional indicators	From 1960 for population; from 1980 for GDP in selected countries	Main regional source for recent years and for countries covered by the European Commission database
Rosés-Wolf database	Mainly NUTS2, with some lower-level historical units	Regional GDP	1900-2022 in the current online database; benchmark observations in historical periods	Key source for extending regional GDP backwards before ARDECO coverage in selected Western European countries
SASI model database	NUTS3 / model regions	Regional GDP and accessibility-related variables	Selected benchmark years, e.g. 1996, 2001, 2006	Supplementary source for selected countries/periods not fully covered by ARDECO or Rosés-Wolf
National statistical offices	Nationally defined regions	Population, GDP/GVA	Country-specific	Source for countries outside ARDECO coverage, especially former USSR countries and selected territories
City Population / database.earth	National/subnational units	Population	Country-specific	Supplementary source for population data in countries and territories not covered by ARDECO

longer time series, as is the case, for instance, in Switzerland.

An important complement to the ARDECO database in Europe consists of other statistical databases. For population data for regions in countries not covered by the ARDECO database, specifically in Bosnia and Herzegovina and in particular in the countries and areas of the former USSR (including Ukraine, Belarus, Moldova, and the European parts of Russia and Kazakhstan), information can be drawn from two sources: <https://www.city-population.de/> and <https://database.earth/population> (see Methods for details on data construction).

In the case of regional GDP, the supplementation of the ARDECO database concerns three main issues: (1) backward extensions, that is, for example, for the years 1960 to 1980, (2) supplements for European countries not included in the ARDECO database, and (3) the replacement of data in cases where the shares of regional GDP remain identical over longer time series, such as in Switzerland. For the first case, that is, data for 1960 to 1980, recent years have brought the possibility of filling this gap for selected European countries thanks to the large-scale Rosés-Wolf Database on Regional GDP (<https://cepr.org/node/424487>), which provides data on the economic development of European regions at the NUTS2 level for the years 1900 to 2022. The accompanying book by Rosés & Wolf (2018) contains regional GDP data for individual European countries for the years 1900 to 2010 (from 1950 onward in ten-year intervals) for Austria, Belgium, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom, including, in several cases, data for smaller geographical units such as NUTS3 regions or various former administrative divisions beyond the NUTS system.

As for the supplementation of regional GDP data for countries and territories not covered by the ARDECO database, this is most often possible only through data obtained directly from national statistical offices (for

example, for Russia, Ukraine, and Belarus) or through the use of alternative databases, such as the one employed for the SASI model (Wegener, 2008), which provides information, among others, on GDP for NUTS3 regions of Switzerland, Slovakia, and Iceland, for example for the years 1996, 2001, and 2006.

Extending a pan-European regional database back to 1960 requires addressing the consequences of major political and territorial changes that affected the statistical reporting of economic data. These include countries that previously formed parts of larger political entities (such as the former USSR, Yugoslavia, Czechoslovakia, and Serbia and Montenegro), territories whose status changed over time (including Ukraine following the annexation of Crimea by Russia and Moldova including Transnistria), as well as historically divided states (such as the German Democratic Republic and the Federal Republic of Germany). Similar challenges arise for the European parts of transcontinental countries, including Russia, Kazakhstan, and Turkey. From a technical perspective, these situations generate data gaps analogous to those observed at the regional level and require the combination of international datasets and domestic statistical sources, as described in the Methods section.

Ultimately, consistent principles for the retropolation of regional GDP have not yet been established. For the period preceding ARDECO (that is, before 1980 for most Western European countries and before the 1990s for Central and Eastern Europe), it is possible, where data availability allows, to rely on share-based backcasting of GVA or GDP using allocation indicators such as employment (similarly to Rosés & Wolf, 2018), wages, energy consumption, indirect taxes, and others, reconciled with national accounts through temporal benchmarking methods (Chow & Lin, 1971; Fernández, 1981), in line with ESA/SNA guidelines and the Eurostat Manual on Regional Accounts Methods (Eurostat, 2013). The key limitations include changes in NUTS boundaries and sectoral classifications over time, the absence

of regional deflators, definitional inconsistencies over the long term, and the risk of model error if a single proxy variable dominates (Eurostat, 2013; Dagum & Cholette, 2006). A good practice is to apply a sectoral (bottom-up) approach combining the triangulation of multiple indicators and a final reconciliation with national aggregates, with full documentation of chain-linking and revision procedures (Eurostat, 2013).

These approaches define an important methodological benchmark for historical regional GDP reconstruction. However, applying a full sectoral bottom-up procedure consistently to all European countries and regions over the period 1960-2021 would require harmonised sectoral accounts, regional deflators, stable territorial correspondences and extensive archival work in national statistical yearbooks, which are not available in a comparable form across the continent. The present study therefore adopts a pragmatic harmonisation and retropolation strategy, designed to maximise spatial and temporal coverage while explicitly documenting the resulting limitations.

Study Area

The aim of the database is to collect population and GDP data for the years 1960 to 2021 across 1546 transport regions. The database is designed to spatially cover: (1) the entire European continent, and (2) areas that are or have been part of the European Community or the European Union, including those located outside continental Europe, that is, Cyprus, the Spanish territories of Ceuta, Melilla, and the Canary Islands, the Portuguese regions of Madeira and the Azores, the overseas territories of France, and Greenland, which was a member of the EEC between 1973 and 1985. For the purposes of this article, we adopted the traditional, encyclopaedic, physico-geographical definition of the continent. However, since the analysis operates with entire regions, we applied a general rule that NUTS1 or sNUTS2 regions (Kolosov, 2012, that is, for example, oblasts in Russia) are not divided between

continents. Therefore, if a given region lies at least partly within Europe, it is included in the database in its entirety. Detailed boundary decisions, including the treatment of transcontinental regions and overseas territories, are described in the Appendix.

Based on the above definition of Europe, in the next step we divided the study area into regions, grouping them into four categories: (1) regions within the area covered by the ARDECO database, (2) regions located within the territory of the former Yugoslavia, (3) regions within the area of the former USSR, and (4) other territories. The starting point was the ARDECO database (excluding Extra-Regio), that is, the number of NUTS3 regions according to the 2021 classification, comprising 1514 regions in 37 countries. For historical reasons, we added one additional region within the ARDECO area by dividing Berlin into East Berlin and West Berlin, and in Turkey, from the 81 regions listed in the ARDECO database, we included only the four belonging to the European continent. As a result, the ARDECO coverage amounted to 1438 regions.

In the area of the former Yugoslavia, we added 14 regions, namely Kosovo and 13 regions in Bosnia and Herzegovina (based on the division into political entities and cantons, with Republika Srpska further divided into two regions due to its specific shape).

In the area of the former USSR, the criterion for regional delineation was the sNUTS2 level, that is, oblasts. In this way, we added 84 regions: (1) 46 oblasts within the previously defined European part of Russia, (2) 3 oblasts located in the European part of Kazakhstan, (3) 7 oblasts in Belarus, (4) 25 oblasts in Ukraine (Donbas, that is, the Donetsk and Luhansk oblasts, remains part of Ukraine in the database regardless of the ongoing conflict since 2014 and the full-scale Russian invasion in 2022), (5) due to the illegal annexation of Crimea by Russia in 2014 and the de facto removal of Crimea and Sevastopol from Ukrainian control since that year, the database treats Crimea and Sevastopol as a single, separate transport region, outside

both Ukraine and Russia, and (6) Moldova was divided into two regions, namely Moldova proper and Transnistria. These territorial decisions were adopted for methodological rather than political reasons. Their purpose is to preserve the comparability of time series and to reflect the availability and continuity of statistical data under changing *de facto* administrative conditions. Crimea and Sevastopol were treated as a separate transport region because the statistical source changes after 2014, from UKRSTAT to ROSSTAT, which would otherwise create a discontinuity within either the Ukrainian or Russian regional system. Donetsk and Luhansk were retained within Ukraine because they remain part of the Ukrainian regional statistical framework, although their post-2014 GDP estimates should be interpreted with caution due to the conflict-related disruption of economic activity and data reporting. Transnistria was separated from Moldova proper because of its distinct *de facto* administrative and statistical status. These choices may affect analyses focused specifically on Ukraine, Moldova, or conflict-affected territories, but they have only limited influence on pan-European aggregate patterns because of the relatively small number of affected regions.

In addition, we included 10 territories located outside the areas of ARDECO, the former Yugoslavia, and the former USSR, namely: (1) Andorra, (2) the Faroe Islands, (3) Guernsey, (4) Jersey, (5) Gibraltar, (6) Greenland, (7) the Isle of Man, (8) Monaco, (9) San Marino, and (10) Vatican City. Altogether, the database covers a total of **1546** regions.

Methods

National level. For the purposes of this article, national-level anchors were constructed separately for population and GDP. For GDP, the Maddison Project Database was selected as the main national-level source. This choice was motivated by several reasons. First, the Maddison Project Database is a project developed by renowned economic historians at the University of Groningen, continuing the

legacy of Angus Maddison, who was a world-leading scholar in quantitative macroeconomic history. Although at the national level the Maddison estimates differ slightly from those reported by Eurostat, the database is widely used by economic historians engaged in the reconstruction of regional GDP across Europe (for example, Rosés & Wolf, 2021). Second, the 2023 version of the Maddison database allows for long-term comparisons of changes in GDP (PPP), including for historically significant but now non-existent political entities relevant to Europe, namely: (1) the former Yugoslavia and (2) the former USSR.

For population, national-level totals were not uniformly anchored to the Maddison Project Database. Instead, they were constructed using the source hierarchy documented in the supplementary Excel metadata, combining ARDECO, database.earth, country-specific sources and, where necessary, estimates based on database.earth and census data. This approach was adopted in order to maintain consistency between national totals, regional structures and the available census-based information, particularly for countries and territories not fully covered by ARDECO.

For territories and areas not covered by the principal GDP sources, missing national-level GDP data were supplemented from alternative sources. The World Bank database was used as the main complementary source for GDP, as it provides longer and more consistent time series than other international providers. The decision not to use GDP data from the IMF, OECD, or the United Nations was made in order to reduce the number of data providers and thereby increase the internal consistency and comparability of national GDP series within the constructed database. Additional national-level GDP data for territories absent from all of the above-mentioned databases were obtained from individual sources identified in the literature review section.

Regional level. For recent years, the ARDECO database constituted the primary source of regional-level data. In countries for which population data at the regional level are not available in ARDECO—specifically

Bosnia and Herzegovina and the countries and areas of the former USSR, including Ukraine, Belarus, Moldova, and the European parts of Russia and Kazakhstan—population figures for selected census years were obtained from *CityPopulation* (<https://www.citypopulation.de/>). Regional GDP data from ARDECO were complemented for earlier periods mainly with information derived from Rosés and Wolf and from the SASI database. For the remaining European territories, as well as for areas with missing national-level GDP data, regional data were partly addressed through the use of domestic supplementary data sources. This approach was applied in particular to countries of the former Soviet Union, for which regional GDP data were obtained directly from official national statistical offices, including ROSSTAT (Russia), BELSTAT (Belarus), UKRSTAT (Ukraine), the Bureau of National Statistics of Kazakhstan, and the National Bureau of Statistics of the Republic of Moldova. All of these datasets are publicly available without access restrictions. The temporal coverage of regional GDP series varies across countries, including 1998–2021 for Russia, 2000–2021 for Kazakhstan, 2009–2021 for Belarus, 2004–2021 for Ukraine, and 2013–2021 for Moldova. For the Autonomous Republic of Crimea and the city of Sevastopol, data were sourced from UKRSTAT up to and including 2013, and from ROSSTAT from 2014 onwards, following the annexation of Crimea by Russia. Additional regional GDP data were also incorporated for Guernsey (available since 2004 from the *Guernsey Annual GVA and GDP Bulletins*) and for Jersey (available since 2012 from official national accounts statistics).

Potential inconsistencies between national- and regional-level data were resolved by applying a single overarching principle throughout the entire study period: values established at the national level were treated as primary, and all raw regional values within a given country were subsequently adjusted to ensure full consistency with the corresponding national totals. Therefore, the final value of population or GDP for each region

i in year y was calculated as the region's share in the sum of raw regional values, multiplied by the value established at the national level for country c . This procedure can be written as:

$$V_{iy}^{final} = \frac{V_{iy}}{\sum_{r=1}^n V_{ry}} \cdot V_{cy} \quad (1)$$

where:

V_{iy}^{final} – is the final value assigned to region i in year y ,

V_{iy} – is the raw value for region i ,

$\sum_{r=1}^n V_{ry}$ – is the sum of raw values for all n regions within country c , and

V_{cy} – is the national-level value for country c in year y .

This procedure ensured that regional values were fully additive to the national total in each country-year cell.

Whenever data availability permitted, this procedure was applied hierarchically across multiple territorial tiers, using intermediate 'national' aggregation levels. This included, for example, the case of the former Yugoslavia: (1) allocating total values for Yugoslavia to its constituent 'national' regions, namely Slovenia, Croatia, Bosnia and Herzegovina, Serbia and Montenegro, and North Macedonia; (2) distributing values for Serbia and Montenegro between Serbia and Montenegro; (3) allocating values for Serbia with Kosovo between Serbia and Kosovo; and (4) subsequently disaggregating values to individual regions within Serbia excluding Kosovo.

An analogous multi-tier approach was applied to countries with a continental division. This included, for example: (1) distributing total values for the USSR among its constituent 'national' regions (Estonia, Latvia, Lithuania, Belarus, Ukraine, Moldova, Russia, Kazakhstan, Georgia, Armenia, and Azerbaijan); (2) allocating national totals for Russia between its European and Asian parts; and (3) further disaggregating values to individual regions within the European part of Russia.

Moreover, in cases where regional GDP data from the Rosés–Wolf or SASI databases were available only for higher-level, outdated, or otherwise inconsistent territorial divisions, a stepwise disaggregation procedure was applied. Regional shares were gradually redistributed across successive levels using proportional structures derived from the ARDECO database or, where necessary, based on population shares, until full correspondence with the regional delineation adopted in this study was achieved.

In order to avoid a purely technical extrapolation of national-level GDP series in cases of single-region countries with gaps in historical data, such territories were treated analogously to regions embedded within larger, economically integrated national systems. Specifically, Andorra was linked to Spain; the Faroe Islands and Greenland to Denmark; the Isle of Man, the Channel Islands, and Gibraltar to the United Kingdom; Monaco to France; Kosovo to Serbia; San Marino and the Vatican City to Italy; and Liechtenstein to Switzerland. In these cases, the extension of GDP time series was based on the observed dynamics of the small economy, benchmarked against the economic cycles of the strongly connected larger economy, rather than on extrapolations of absolute GDP levels. This approach allowed the dependence of small economies on larger economic systems and their cyclical dynamics to be taken into account.

A specific case of a single region composed of parts that previously belonged to different countries was identified for Berlin, which was divided into West Berlin and East Berlin. The estimation of the time-varying population shares of these two subregions was conducted using standard time-series extension procedures. However, because the two parts historically belonged to countries that differed substantially in terms of economic development, the total GDP of Berlin was divided between its eastern and western parts based on both population size and GDP per capita observed in the remaining territories of the former German Democratic

Republic and the Federal Republic of Germany, respectively.

Given the combination of multiple data sources, a harmonisation procedure was required to ensure comparability of GDP figures across countries and over time. Differences in base years for constant-price series and inconsistencies in temporal coverage across databases necessitated additional adjustments. While supplementary regional GDP data were used exclusively to establish domestic proportional structures and could therefore be expressed directly as shares, national-level GDP data obtained from alternative sources – specifically for the Channel Islands, Gibraltar, and former USSR countries currently outside the European Union – required conversion into a common currency framework. For this purpose, official annual average exchange rates published by the Federal Reserve Bank were used.

After collecting and harmonising data from all considered sources, remaining gaps in the time series provided by individual sources were filled using a consistent time-series extension framework. At the national level, missing observations were estimated in terms of absolute values, while at the regional level the corresponding regional shares were estimated accordingly. Two main cases were distinguished: interpolation and extrapolation/retrapolation.

First, when both an earlier and a later observation were available, for example for intercensal years, missing values for year y were interpolated using the last preceding known value and the first subsequent known value. Interpolation was performed assuming a geometric progression, which allows cumulative dynamics to be taken into account, analogously to compound growth processes. National-level values were estimated as:

$$V_{cy}^{est} = V_{cy_0} \cdot \left(\frac{V_{cy_1}}{V_{cy_0}} \right)^{\left(\frac{y-y_0}{y_1-y_0} \right)} \quad (2)$$

An analogous procedure was applied at the regional level, where interpolated regional shares were calculated as:

$$Share_{iy}^{est} = Share_{iy_0} \cdot \left(\frac{Share_{iy_1}}{Share_{iy_0}} \right)^{\frac{(y-y_0)}{(y_1-y_0)}} \quad (3)$$

where:

V_{cy}^{est} denotes the interpolated national-level value for country c in year y ;

V_{cy_0} and V_{cy_1} are the known values at the last preceding and first subsequent observation years y_0 and y_1 , respectively.

$Share_{iy}^{est}$ denotes the interpolated share of region i in the corresponding country total in year y , while

$Share_{iy_0}$ and $Share_{iy_1}$ are the corresponding known regional shares at y_0 and y_1 .

Second, in many cases, only one endpoint of the estimated period was available, while the period to be reconstructed extended either back to 1960 or forward to 2021. This was the case, for example, when the first available census observation occurred after 1960. In such situations, either the earlier national value or the earlier regional share was missing, and values for the remaining years were estimated by extrapolating the observed trend beyond the range of empirical data. Since this procedure was applied predominantly backward in time, it can also be referred to as retroprojection.

Although the extrapolation procedure is formally similar to interpolation, it differs from it in two essential respects. First, because only one endpoint of the estimated period is known, the underlying dynamics must be inferred from the available observations. For this purpose, a regression function was estimated using all available observations for the given country or region. Second, the choice of the functional form depended on the direction of the observed dynamics and on which endpoint of the extrapolated period was constrained by the available data.

The geometric-growth formulation was retained for extrapolation only in cases where values decreased with increasing temporal distance from the known endpoint of the estimated period, corresponding to positive dynamics in retroprojection or negative dynamics

in forward extrapolation. In such cases, the regression function was specified as:

$$V_{cy} = a \cdot \left(\beta_c \right)^y + \varepsilon_y \quad (4)$$

and, analogously, for regional shares:

$$Share_{iy} = a \cdot \left(\beta_i \right)^y + \varepsilon_y \quad (5)$$

In each case, the annual growth dynamics were determined by the estimated value of the parameter. Final estimated values were then calculated as:

$$V_{cy}^{est} = V_{cy_0} \cdot \left(\beta_c \right)^{(y-y_0)} \quad (6)$$

and, for regional shares:

$$Share_{iy}^{est} = Share_{iy_0} \cdot \left(\beta_i \right)^{(y-y_0)} \quad (7)$$

where:

a – is the regression constant,

β_c and β_i are the estimated annual multiplicative growth factors for the national-level value and regional share, respectively,

ε_y – is the error term, and

y_0 – denotes the known endpoint year from which extrapolation or retroprojection is performed.

In cases where the application of compound geometric growth over a long estimated period, combined with the absence of an upper bound, could result in unrealistically high values, a linear regression framework was applied instead. Under this specification, national-level values were modelled as:

$$V_{cy} = \left(\beta_c \cdot y \right) + a + \varepsilon_y \quad (8)$$

and, analogously, regional shares were specified as:

$$Share_{iy} = \left(\beta_i \cdot y \right) + a + \varepsilon_y \quad (9)$$

Estimated values for each missing year were then calculated using a linear projection based on the value observed at the known

endpoint of the estimated period (y_0), according to:

$$V_{cy}^{est} = V_{cy_0} + \beta_c \cdot (y - y_0) \quad (10)$$

and, for regional shares:

$$Share_{iy}^{est} = Share_{iy_0} + \beta_i \cdot (y - y_0) \quad (11)$$

In Equations (8)–(11),

β_c and β_i denote the estimated linear trend coefficients for national-level values and regional shares, respectively;

a – is the intercept,

ε_y – is the error term, and

y_0 – is the known endpoint year.

After extending the time series of regional shares, a normalisation coefficient was applied for each year and each multi-regional country in order to ensure that regional shares summed to unity. Final estimated regional values were therefore calculated as:

$$\begin{aligned} V_{iy}^{est} &= Share_{iy}^{est} \cdot Coef_{cy} \cdot V_{cy}^{est} = \\ &= Share_{iy}^{est} \cdot \frac{1}{\sum_{r=1}^n Share_{ry}^{est}} \cdot V_{cy}^{est} \end{aligned} \quad (12)$$

where:

$Coef_{cy}$ – is the normalisation coefficient for country c in year y ,

n – is the number of regions within the country,

r – indexes these regions, and

V_{cy}^{est} – is the estimated national-level value.

This normalisation ensured that, after interpolation or retropolation of regional shares, the sum of regional estimates within each country remained consistent with the corresponding national-level estimate.

To further harmonise regional GDP data, cases in which regional shares for a given country were available from more than one source – namely, in addition to the ARDECO database, the Rosés-Wolf database and/or the SASI database – were addressed

by constructing a single weighted average of regional shares. This situation applied to regions in Austria, Belgium, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, and the United Kingdom in the case of Rosés-Wolf data; Iceland and Slovakia in the case of SASI data; and Switzerland in the case of both Rosés-Wolf and SASI data.

The weighted average of regional GDP shares across three or two data sources was calculated separately for each year over the entire period 1960–2021. For smoothing purposes, the procedure was applied uniformly to all years, including those for which data were directly available from a given source. A single overarching rule was adopted: regardless of the source, the weight assigned to each available observation for region i decreases with the temporal distance between the year of observation y_known and the year being estimated y . Weights were specified using an exponential decay function:

$$Share_{iy}^{est_ult} = \frac{\sum_{y_known=1}^n \left(e^{-\frac{\sqrt{(y_known - y)^2}}{2}} \right) \cdot Share_{iyknown}^{est}}{\sum_{y_known=1}^n \left(e^{-\frac{\sqrt{(y_known - y)^2}}{2}} \right)} \quad (13)$$

where:

$Share_{iy}^{est_ult}$ denotes the final harmonised share of region i in the corresponding country total in year y ,

y_known denotes the year of an available source observation,

n – is the number of observations included in the weighted average, and

$Share_{iyknown}^{est}$ is the corresponding regional share. The exponential term represents the temporal weight assigned to each observation according to its distance from year y .

Under this specification, the weight assigned to an observation is halved with each increase of approximately 1.386 years in temporal distance.

The approximately 1.4-year bandwidth was selected on the basis of methodological judgement as a conservative harmonisation rule;

Table 4. Dominant provenance of regional population and GDP series by country group and broad period

Variable	Country/provenance group	No. of regions	1960-1979	1980-1999	2000-2021	Principal basis
Population	ARDECO-covered countries and directly sourced single-zone territories	1449	D	D	D	Annual regional or territorial source series; no regional allocation is required for single-zone territories.
	Census-based extensions: Belarus, Bosnia and Herzegovina, European Kazakhstan, Moldova, European Russia, Ukraine and Crimea	97	B	B	B	Regional structures derived from census benchmarks, with interpolation between observations and backward extension before the first usable benchmark.
GDP	ARDECO direct/back-cast group	342	B	B→D	D	Direct annual ARDECO regional series begin between 1980 and 2004, depending on the country; preceding years are back-cast from regional-share trends.
	ARDECO and Rosés-Wolf benchmark group	1056	H	H	H	Time-weighted harmonisation of annual ARDECO regional shares and historical Rosés-Wolf benchmark shares; selected subgroups rely on ARDECO only.
	SASI-supported group: Iceland, Slovakia and Switzerland	36	M/H	M/H	M/H	SASI model shares combined, where available, with ARDECO and Rosés-Wolf inputs; extrapolation is used outside the source years.
	Official post-Soviet regional-source group: Belarus, European Kazakhstan, European Russia, Ukraine and Crimea	82	B	B	B→D	Official annual regional GDP series begin between 1998 and 2008, depending on the country or territory; preceding years are back-cast.
	Proxy-based and special regional allocations: Bosnia and Herzegovina and Moldova/Transnistria	15	M	M	M	SHDI and population proxies with backward extrapolation for Bosnia and Herzegovina; a fixed literature-based share for Transnistria.
	Single-zone states and territories	15	N	N	N	No regional allocation is required: the national series is also the final regional series. Detailed national-series provenance is reported in the supplementary metadata.

Notes: D – directly sourced annual regional data; B – regional values interpolated or back-cast from observed source series or benchmark structures; H – hybrid harmonisation of annual and benchmark regional sources; M – model- or proxy-based estimation, including extrapolation outside model benchmark years; B→D – transition from estimated to directly sourced annual data at country-specific dates; N – no regional disaggregation required. The categories are intentionally broad because several country series combine annual observations, historical benchmarks and model-based inputs. Because all regional population and GDP values are reconciled with the corresponding national totals, the table primarily classifies the provenance of the regional allocation structure.

it was not empirically calibrated, and no sensitivity analysis of alternative bandwidths was conducted. Because overlapping sources differ in spatial classification, accounting conventions, temporal coverage, and territorial aggregation, the procedure gives priority to observations closest to the year being estimated. Its purpose is not long-range interpolation but the local smoothing of transitions between sources, while limiting the influence of more distant observations. The weighted average should therefore be interpreted as a pragmatic local harmonisation device rather than as a statistically optimised smoothing model.

Prior to applying this procedure, regional shares from the Rosés-Wolf database were harmonised in cases where data aggregated under two inconsistent territorial divisions were provided simultaneously. In such cases, specifically for Finland, Norway, Portugal, and Sweden, an arithmetic mean of the values derived from both territorial divisions was used when regional shares for a given year differed depending on the aggregation applied in the Rosés-Wolf database.

To provide an at-a-glance overview of data provenance, Table 4 summarises the dominant source and estimation logic used for the regional component of the population and GDP series across broad country groups and periods.

Detailed country- and year-specific information on data sources and methodological pathways is provided in the supplementary Excel workbook for population (Supplementary file S1) and the supplementary Excel workbook for GDP (Supplementary file S2). A third workbook contains regional GDP per capita (Supplementary file S3), calculated directly from the corresponding GDP and population series. All three workbooks cover the same 1546 regions and the period 1960-2021.

Regional GDP per capita was subsequently calculated for each region i and year y by dividing regional GDP by the corresponding regional population:

$$GDPpc_{iy} = \frac{GDP_{iy}}{Population_{iy}} \quad (14)$$

where:

$GDPpc_{iy}$ – denotes GDP per capita,

GDP_{iy} – regional GDP, and

$Population_{iy}$ – regional population for region i in year y .

Regional GDP per capita is therefore a directly derived output variable, calculated after the construction of the underlying GDP and population series.

Results

The full database covers the period 1960-2021; however, the illustrative results presented in this section refer to 1960-2020. The aim of the Results section is illustrative rather than explanatory. It first presents changes in population density and GDP per capita between 1960 and 2020, followed by two additional illustrations: year-by-year trajectories for the regions of Belarus and a comparison of reconstructed Danish regional GDP per capita series with Rosés-Wolf benchmark values.

Changes in population density. During the analysed period 1960-2020, population numbers declined in as many as 447 out of the 1546 regions included in the database. These are, of course, mainly rural and border regions, including those located along the external borders of the European Union (e.g. both in Russia's Pskov Oblast and in neighbouring Latgale in Latvia). A particularly strong polarisation between depopulating peripheries and regions of population growth, such as metropolitan areas and those situated along the Mediterranean coast, is visible in Spain and France. Intensive depopulation processes affecting entire countries can be observed in several Central and Eastern European states, namely Bulgaria, Lithuania, and Croatia. In the latter case, the most severe decline occurred in the region of Lika-Senj County (*Ličko-senjska županija*), where by 2020 the population had fallen to only 40% of its 1960 level. On the other hand, apart from the aforementioned metropolitan areas of most EU countries,

an unprecedented population growth has been recorded in certain French overseas territories, the Canary Islands, as well as in Albania's Tirana and Turkey's Istanbul (with population increases exceeding threefold). Remarkable growth in population was also observed in the regions located in the European part of Kazakhstan. Conversely, the country showing the least regional variation in population change over the analysed period turned out to be the Czech Republic (see Fig. 1).

Changes in GDP per capita. The database reveals a pronounced spatial differentiation in changes in GDP per capita over the six-decade period. There are regions where GDP per capita has declined since

1960. These include, above all, two oblasts in eastern Ukraine, Donetsk and Luhansk, where the outbreak of hostilities in 2014 led to a several-fold drop in GDP, resulting in a sharp decrease in the observed GDP per capita in the database. At the same time, according to the adopted research methodology, these regions were the industrial heartland of the USSR in 1960. Over the analysed period, GDP per capita has also decreased in Moldova, Transnistria, the Ukrainian Zakarpattia oblast, and, interestingly, in certain regions of the United Kingdom, such as the Scottish Borders and some districts of Greater Manchester. In the context of the UK, the lack of development in these regions reflects a continuous decline in their share of national

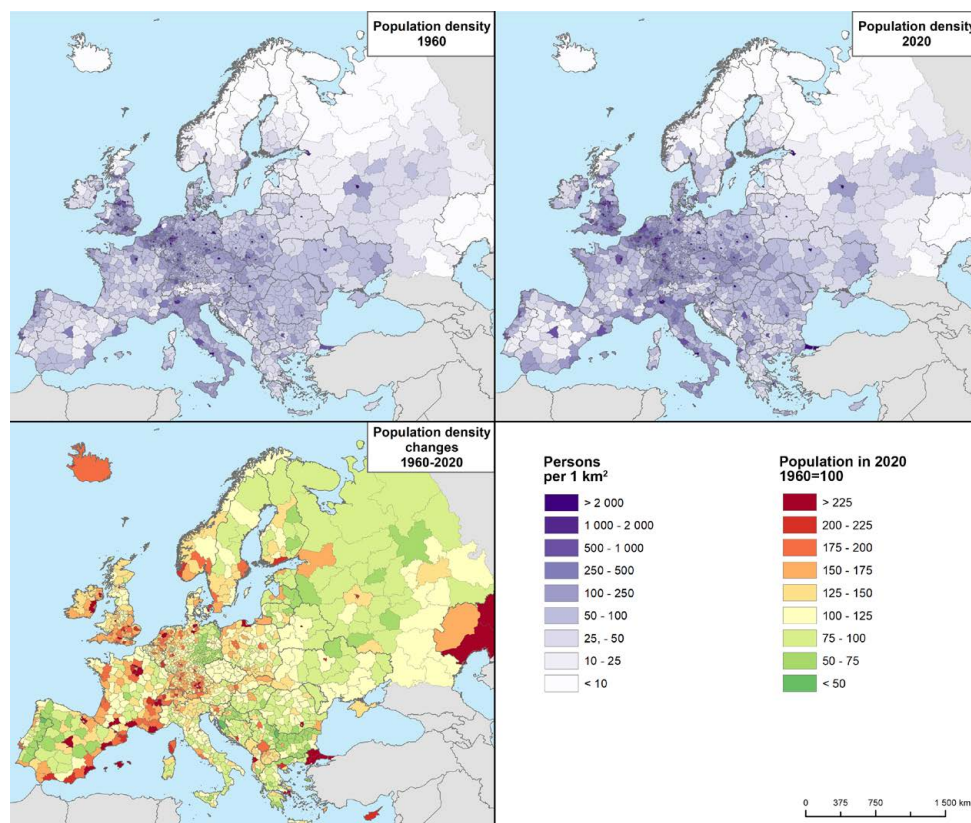


Figure 1. Estimated changes in population density in European regions, 1960-2020*

* The spatial coverage of the database excludes French overseas territories, the Canary Islands, Madeira, the Azores, Greenland, and Svalbard.

GDP as reported in the ARDECO database after 1980.

Conversely, many regions have experienced an unprecedented real increase in GDP per capita. According to the adopted methodology, in addition to Norway's Svalbard and Jan Mayen, Gibraltar recorded a fiftyfold increase in GDP per capita. Among regions with more than a tenfold increase are several capital regions of Central and Eastern Europe (Moscow, Sofia, Bucharest) as well as of Western Europe (Munich and Liverpool). Such high levels of growth were also recorded in industrial regions such as Wolfsburg (headquarters of Volkswagen) and Ingolstadt, Kreisfreie Stadt (headquarters of Audi) and in tourist destinations (Malta, the Spanish island

NUTS3 regions of El Hierro, La Gomera, and La Palma, as well as Portugal's Madeira and the Azores). This group also includes regions surrounding major urban agglomerations in countries that started from relatively low levels of development in the 1960s and experienced rapid growth in recent decades, such as Ilfov surrounding Bucharest in Romania, as well as Wroclawski or Krakowski sub-regions in Poland.

As already mentioned, one of the main advantages of the presented database is that, for the first time in a pan-European context, it enables comparisons of population and GDP changes in regions located also outside the European Union. A good example is the presentation of results for Belarusian regions,

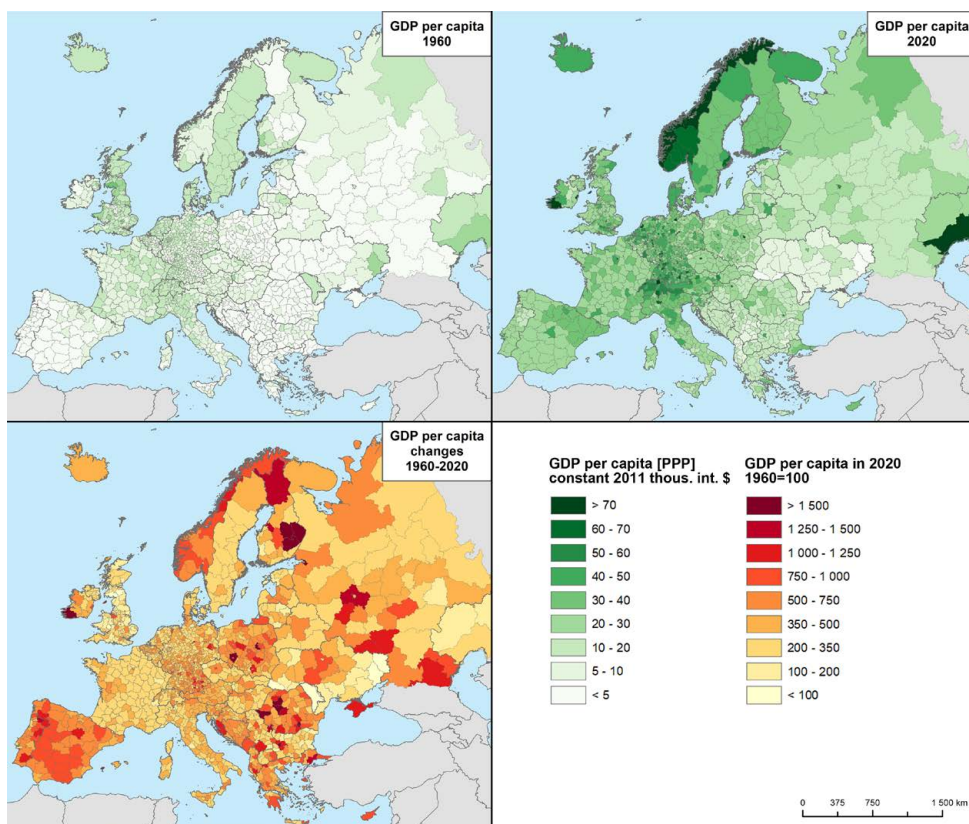


Figure 2. Estimated changes in GDP per capita in European regions, 1960-2020*

* The spatial coverage of the database excludes French overseas territories, the Canary Islands, Madeira, the Azores, Greenland, and Svalbard.

including the six oblasts and Minsk City, both in terms of the evolution of population size and GDP *per capita*. Figure 3 clearly illustrates the divergent trajectories of individual regions over time, as well as the smoothness of the observed trends, without abrupt year-to-year fluctuations, despite significant variations in both population numbers and GDP *per capita* in Belarus throughout the analysed period (Fig. 3).

To provide an internal consistency check for a country where historical benchmark data are available, we additionally compare the reconstructed annual GDP per capita series for Danish regions with the Rosés-Wolf benchmark

values. Both datasets were expressed as regional GDP per capita indices relative to the national average (Denmark = 100), which makes the two sources directly comparable. The comparison shows that the reconstructed series broadly preserve the regional hierarchy observed in the Rosés-Wolf data in several benchmark years, particularly around 1970. At the same time, deviations in later years are visible, especially for Copenhagen City and Greater Copenhagen, reflecting the fact that the final database combines Rosés-Wolf with ARDECO and applies a harmonisation/smoothing procedure rather than reproducing any single input source mechanically.

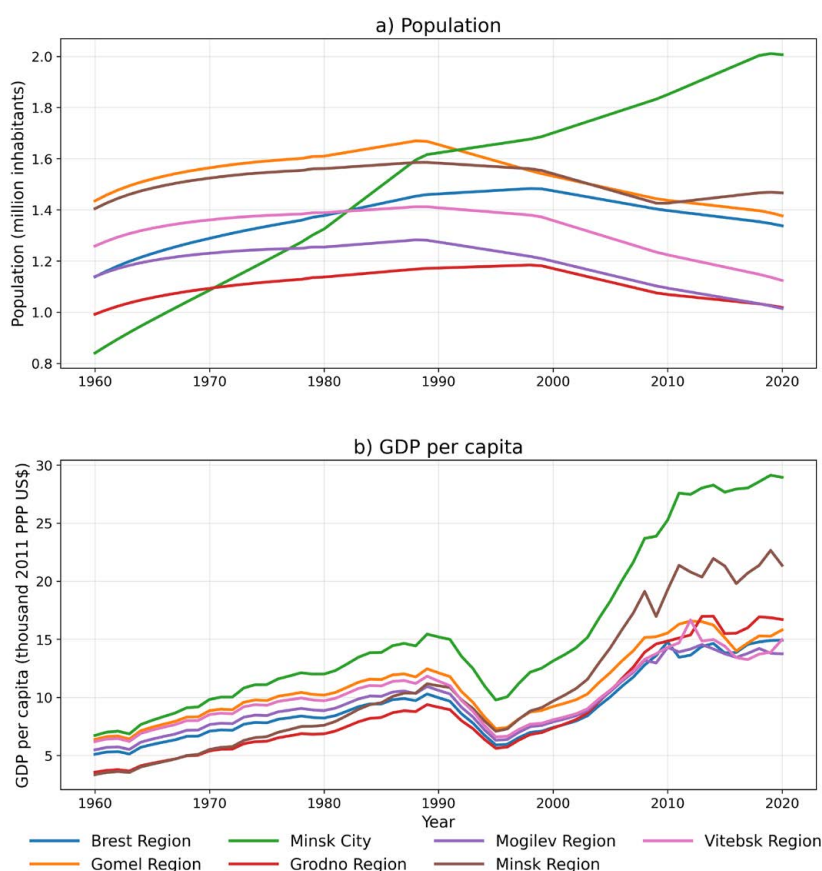


Figure 3. Evolution of population and GDP per capita in the six oblasts of Belarus and Minsk City, 1960-2020*

* For GDP, estimates are based on BELSTAT data for the years 2009-2020

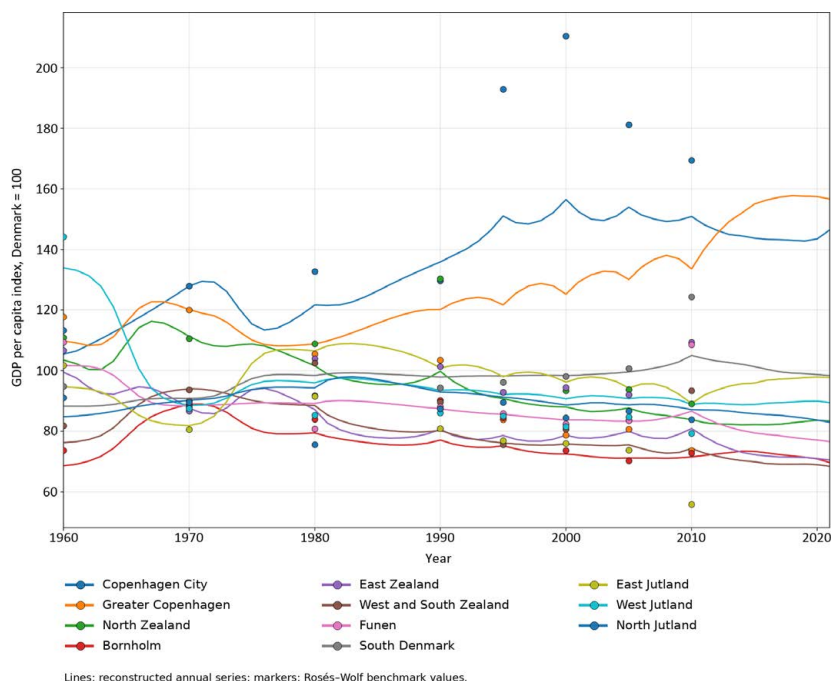


Figure 4. Comparison of reconstructed regional GDP per capita series with Rosés-Wolf benchmark values for Danish regions, 1960-2020

Note: Both datasets are expressed as regional GDP per capita indices relative to the national average (Denmark = 100). Annual lines show the reconstructed database; markers show Rosés-Wolf benchmark observations for selected years.

Source: authors' own elaboration based on the constructed database and Rosés-Wolf data.

Conclusions

The empirical results derived from the newly developed database reveal strong spatial polarisation in demographic and economic dynamics across Europe between 1960 and 2020. Population decline affected nearly one-third of regions, particularly rural and border areas, especially though not exclusively in Central and Eastern Europe, while metropolitan and coastal regions, particularly in Spain and France, experienced substantial population growth. Changes in GDP per capita also show pronounced spatial differentiation: for example, sharp declines occurred in industrial regions of eastern Ukraine and in Moldova, whereas extraordinary growth was observed in capital regions, industrial hubs

(e.g. Wolfsburg, Ingolstadt), and major tourist destinations.

The presented pan-European regional database, covering the period 1960-2021, is available as three downloadable supplementary Excel workbooks (Supplementary files S1-S3): one containing population data, one containing GDP data and one containing the corresponding GDP per capita series. The population and GDP workbooks also contain detailed metadata documenting the principal data sources and methodological pathways applied to different country groups and periods, while the GDP per capita workbook provides a transparent calculation based on the two underlying variables. As mentioned in the introduction, it may therefore be applied, for instance, to analyses of regional

income sigma, beta, and club convergence. Until now, studies on regional convergence in European countries have primarily focused on EU regions at the NUTS3 level using the ARDECO database (e.g. Postiglione et al., 2020, for the period 1981-2014, or López-Villuendas et al., 2023, for time-series clustering covering 2000-2017). In contrast, the database described here allows:

1. extending the spatial coverage to include regions from the entire European continent, including those outside the EU, and
2. expanding the temporal scope to the 1960s and 1970s—periods not covered by the ARDECO dataset.

Furthermore, population was not used as the general basis for estimating regional GDP, although it entered a limited number of explicitly documented auxiliary allocation procedures. GDP per capita was subsequently calculated from the final GDP and population series. This makes it possible to explore the relationships between income convergence and population change, for example in the context of depopulation processes in peripheral, rural, or border regions. In the near future, the authors intend to use the database to identify the trajectories of population- and GDP-weighted centroids of Europe and the European Union over the period 1960-2020.

Discussion

The database presented in this article represents the first version, which the authors intend to further refine in the coming years, as opportunities arise to explore alternative methodological solutions. This particularly concerns Central and Eastern European countries, where approaches such as the Geary & Stark (2002) method could be applied, along with the potential use of historical data on employment and wages by economic sector across regions (see, for example, the application of this method by Rosés & Wolf, 2018). Another potential avenue for addressing data gaps involves the use of machine learning techniques, as demonstrated by Koch et al. (2024).

The authors anticipated at an early stage that the database would combine regional data on GDP and population and therefore deliberately avoided using population figures to estimate GDP in order to prevent circular inference (*idem per idem*). Nevertheless, the compiled database is subject to several methodological limitations that should be taken into account by users and that were unavoidable in order to ensure completeness and consistency in spatial coverage and temporal continuity.

First, as summarised in Table 4, the provenance of the regional series varies across country groups and periods, reflecting heterogeneous sources, methods, and assumptions. In some cases, regional values are based directly on raw statistical data, whereas in others raw data serve primarily as inputs for model-based extensions of established temporal trends. Consequently, the relative proportions of empirically observed and modelled data differ across regions and time spans, and certain periods may rely entirely on model-based estimates. In this context, detailed knowledge of region-specific demographic dynamics, historical events, and economic cycles can facilitate interpretation of the results.

Second, another limitation concerns the interpretation of retropolated regional GDP values. In cases where historical regional GDP shares had to be extrapolated backwards from later observations, the procedure necessarily relies on the continuation of observed regional shares or their estimated trends. This may introduce a fixed-weight or Laspeyres-type bias, especially where regions experienced major structural change, such as deindustrialisation in Western Europe or the post-1989 transformation of centrally planned economies in Central and Eastern Europe. In such cases, the retropolated GDP values should not be interpreted as fully welfare-consistent historical real GDP levels. Rather, they should be treated as harmonised indicators of relative regional economic dynamics, constrained to national totals and designed primarily for comparative spatial

analysis. This limitation is inherent to large-scale long-run regional retropolation exercises, particularly where consistent sectoral regional accounts, regional deflators and historical administrative correspondences are unavailable. Importantly, the magnitude of this potential fixed-weight or Laspeyres-type bias has not been empirically bounded in the present study. An alternative-anchor sensitivity analysis is therefore left for future work.

Third, a related limitation concerns the combination of PPP-adjusted national GDP aggregates with regional shares derived from domestic-price national accounts or statistical office data. In the database, Maddison national GDP series provide the top-down national anchor, while regional allocation is based on domestic regional structures whenever available. This approach ensures national-level consistency and allows cross-country comparison at the aggregate level, but it does not eliminate all index-number issues. Consequently, cross-country aggregation of regional GDP values should be interpreted with caution, as sums of regions drawn from different countries do not necessarily correspond to a fully consistent PPP-based aggregate in an index-number sense. In particular, regional shares are not independently PPP-adjusted and therefore the resulting regional GDP estimates should not be interpreted as fully price-consistent regional welfare measures. Rather, they represent national-anchor-consistent regional allocations intended for comparative analysis of relative spatial patterns and long-term regional dynamics.

Fourth, a gradual weighting scheme was used to smooth transitions between sources and limit cumulative divergence over long time horizons. Because weighted shares may combine observations that differ in spatial classification, price base, territorial aggregation, or GVA/GDP concept, the resulting series may differ from any single source and should not be interpreted as fully coherent regional price-index paths. Normalisation ensures consistency with national totals within each country-year, but not intertemporal index-number consistency for individual

regions. The procedure should therefore be understood as a pragmatic harmonisation device that gives greater weight to temporally proximate observations, without applying additional source-specific quality assessments. The approximately 1.4-year bandwidth was selected on the basis of methodological judgement and was neither empirically calibrated nor subjected to sensitivity analysis.

When considering datasets that are formally and statistically comparable—in terms of definitions, spatial aggregation, temporal resolution, and available time spans—even countries included within the Eurostat system present certain challenges. For example, despite efforts made in recent decades to harmonise the definition of a person considered to be a resident of a given territorial unit within Eurostat (Regulation (EU) No 1260/2013), significant differences persist between countries, for instance due to varying approaches to residence registration obligations. On the other hand, spatial aggregation of data within territorial units also lacks comparability across countries. For example, there are notable differences in the average area of NUTS3 units between Spain and Germany, and even greater disparities at the level of local administrative units (e.g., LAU2). This issue is related to the so-called Modifiable Areal Unit Problem (MAUP) (see Basile, 2008; Unwin, 1996). In the context of regional GDP estimates, MAUP also interacts with the index-number limitations discussed above. Because NUTS3 units differ substantially in size, population, and economic structure across countries, aggregation to higher spatial levels or to selected macro-regional groupings may be sensitive to the chosen spatial aggregation path. Users should therefore interpret convergence analyses or cross-regional aggregates with caution, especially when comparing subsets of regions defined at different spatial scales or with strongly heterogeneous territorial structures.

The lack of comparability among available data also extends to the temporal dimension. Historically, population censuses have been conducted with varying regularity and

at different times. The methodologies used for estimating population figures in intercensal periods likewise differ across countries. Moreover, there is a scarcity of continent-wide studies based exclusively on GDP data published by national statistical offices (for example, in the countries of the former USSR; see Timiryanova et al., 2021), which often differ in terms of spatial aggregation levels and temporal coverage. Another challenge lies in the proper calculation of regional GDP for economies that once operated under central planning, such as Poland, where a major administrative reform in the 1990s abolished 49 voivodeships, and existing regional GDP estimates do not extend significantly back in time (see GUS, 1997).

These limitations do not undermine the value of the database, but they define the interpretative framework within which the data should be used, particularly in comparative analyses of regional economic dynamics.

Acknowledgments

This research was funded by the Polish National Science Centre under decisions no. 2020/37/B/HS4/00544 and 2024/53/B/HS4/00304. Within the framework of the first project (2020/37/B/HS4/00544), the regional delimitation of transport regions was developed and population data were assigned to these regions. The second project (2024/53/B/HS4/00304) supported the development of an advanced methodological framework for integrating and harmonising multiple data sources on regional GDP.

Editors' note:

Unless otherwise stated, the sources of tables and figures are the authors', on the basis of their own research.

References

- Antal, D. (2019). Working with Regional Statistics. RPubS by RStudio. https://rpubs.com/antaldaniel/eurostat_regional_problems
- Archila Bustos, M. F., Hall, O., Niedomysl, T., & Ernstson, U. (2020). A pixel level evaluation of five multitemporal global gridded population datasets: a case study in Sweden, 1990-2015. *Population and Environment*, 42(2), 255-277. <https://doi.org/10.1007/s11111-020-00360-8>
- Basile, R. (2008). Regional economic growth in Europe: A semiparametric spatial dependence approach. *Papers in Regional Science*, 87(4), 527-545. <https://doi.org/10.1111/j.1435-5957.2008.00175.x>
- Bernal, I., Van Halderen, G., & Sejersen, T. (2021). Big data for population and social statistics.
- Bolt, J., & Zanden, J. L. van. (2024). Maddison style estimates of the evolution of the world economy: A new 2023 update. *Journal of Economic Surveys*, 1-41. <https://doi.org/10.1111/joes.12618> (The data for the MPD 2023 release are available from April 1, 2024: <https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/>)
- Center For International Earth Science Information Network-CIESIN-Columbia University. (2016). *Gridded Population of the World, Version 4 (GPWv4): Population Density* [Data set]. Palisades, NY: NASA Socio-economic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4NP22DQ> Date Accessed: 2025-10-14; <https://www.earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-gpwv4-popdens-4.0>
- Chow, G. C., & Lin, A.-L. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53(4), 372-375. <https://doi.org/10.2307/1928739>
- City Population. *Population statistics for countries, administrative areas, cities and agglomerations*. <https://www.citypopulation.de/>

- Dagum, E. B., & Cholette, P. (2006). *Benchmarking, temporal distribution, and reconciliation methods for time series*. New York, NY: Springer. <https://doi.org/10.1007/0-387-35439-5>
- database.earth. *Population by country and territory*. <https://database.earth/population>
- European Commission. (2025). Annual Regional Database of the European Commission (ARDECO). Directorate-General for Regional and Urban Policy. <https://territorial.ec.europa.eu/ardeco/explorer?lng=en>
- Eurostat. (2025). Regional statistics by NUTS classification. European Commission. <https://ec.europa.eu/eurostat/web/regions/database>
- Eurostat. (2013). *Manual on regional accounts methods*. European Commission Publications Office.
- Federal Reserve Bank of St. Louis. FRED Economic Data. <https://fred.stlouisfed.org/>
- Fernández, R. B. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics*, 63(3), 471-476. <https://doi.org/10.2307/1924371>
- Geary, F., & Stark, T. (2002). Examining Ireland's post-famine economic growth performance. *The Economic Journal*, 112(482), 919-935. <https://doi.org/10.1111/1468-0297.00064>
- GeoRank https://georank.org/economy/united-states/vatican#google_vignette
- GHS data package 2019 – Public release GHS P2019. Publications Office, 2019. <https://data.europa.eu/doi/10.2760/290498>
- Gibraltar. Abstracts of Statistics Reports <https://www.gibraltar.gov.gi/statistics>
- Guernsey Annual GVA and GDP Bulletins. <https://gov.gg/article/120766/Search-archive-of-statistical-publications#selectednavItem162852>
- GUS. (1997). Produkt krajowy brutto według województw za 1995 rok. Zakład Badań Statystyczno-Ekonomicznych GUS i PAN, Urząd Statystyczny w Katowicach, 1997
- International Monetary Fund. (2024). *World Economic Outlook (WEO) database*. <https://www.imf.org/en/Publications/SPROLLs/world-economic-outlook-databases#sort=%40imfdate%20descending>
- Jersey national accounts statistics. <https://stats.je/statistic/national-accounts/#anchor-0>
- Koch, P., Stojkoski, V., & A. Hidalgo, C. (2024). Augmenting the availability of historical GDP per capita estimates through machine learning. *Proceedings of the National Academy of Sciences*, 121(39), e2402060121. <https://doi.org/10.1073/pnas.2402060121>
- Kolosov, V. (2012). *Territorial units in Russia, Ukraine, Belarus and Moldova and NUTS classification*. (Doctoral dissertation, ESPON – Inspire Policy Making with Territorial Evidence).
- Kotavaara, O. (2012). Accessibility, population change and scale dependency–exploring geospatial patterns in Finland, 1880-2009. *Nordia Geographical Publications*, 41(4), 48.
- Kummu, M., Taka, M., & Guillaume, J. H. A. (2018). Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015. *Scientific Data*, 5, 180004. <https://doi.org/10.1038/sdata.2018.4>
- Laaboudi, C., Karlberg, M., & Islam, M. (2024). Open data dissemination at Eurostat: State of the art. *Statistical Journal of the IAOS*, 40(1), 81-90. <https://doi.org/10.3233/SJI-230115>
- Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., ... & Pesaresi, M. (2019). The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, 11(3), 1385-1409. <https://doi.org/10.5194/essd-11-1385-2019>
- López-Villuendas, A. M., & del Campo, C. (2023). Regional economic disparities in Europe: Time-series clustering of NUTS3 regions. *International Regional Science Review*, 46(3), 265-298. <https://doi.org/10.1177/01600176221125703>
- Murakami, D., Yoshida, T., & Yamagata, Y. (2021). Gridded GDP projections compatible with the five SSPs (shared socioeconomic pathways). *Frontiers in Built Environment*, 7, 760306. <https://doi.org/10.3389/fbuil.2021.760306>
- Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences*, 103(10), 3510-3517. <https://doi.org/10.1073/pnas.0509842103>

- OECD. (2025). *Population (indicator)*. <https://www.oecd.org/en/data/indicators/population.html>;
Real Gross Domestic Product (GDP)
<https://www.oecd.org/en/data/indicators/real-gross-domestic-product-gdp.html>
- Özyurt, S., & Dees, S. (2015). *Regional dynamics of economic performance in the EU: To what extent spatial spillovers matter?* ECB Working Paper Series, no. 1870.
- Phan, D. H. (2023). Lights and GDP relationship: What does the computer tell us?. *Empirical Economics*, 65(3), 1215-1252. <https://doi.org/10.1007/s00181-023-02377-y>
- Postiglione, P., Cartone, A., & Panzera, D. (2020). Economic convergence in EU NUTS3 regions: A spatial econometric perspective. *Sustainability*, 12(17), 6717. <https://doi.org/10.3390/su12176717>
- Regulation (EU) No 1260/2013 of the European Parliament and of the Council of 20 November 2013 on European demographic statistics Text with EEA relevance.
<https://eur-lex.europa.eu/eli/reg/2013/1260/oj/eng>
- Rosés, J. R., & Wolf, N. (2021). Regional growth and inequality in the long-run: Europe, 1900-2015. *Oxford Review of Economic Policy*, 37(1), 17-48. <https://doi.org/10.1093/oxrep/graa062>
- Rosés, J. R., & Wolf, N. (Eds.). (2018). *The Economic Development of Europe's Regions: A Quantitative History Since 1900*. Routledge. <https://doi.org/10.4324/9780429449789>
- Rosés, J. R., & Wolf, N. (2025). *Rosés-Wolf database on regional GDP, version 7*. CEPR.
<https://cepr.org/node/424487>
- Rosik, P., & Wójcik, J. (2023). *Transport infrastructure and regional development: A survey of literature on wider economic and spatial impacts*. *Sustainability*, 15 (1), 548. <https://doi.org/10.3390/su15010548>
- Rosik, P., Mazur, M., Komornicki, T., Goliszek, S., Stępnik, M., Duma, P., ... & Pomianowski, W. (2025a). National and regional potential accessibility convergence by decay and decades in Germany, France, Spain, Poland and Romania in the years of 1960-2020. *Journal of Transport Geography*, 123, 104106. <https://doi.org/10.1016/j.jtrangeo.2024.104106>
- Rosik, P., Mazur, M., Stępnik, M., Komornicki, T., Goliszek, S., Duma, P., & Churski, P. (2025b). What drives potential accessibility decomposition? Temporal and spatial variability of the impact of infrastructure and population components in France, Spain, and Poland in the years 1960-2020. *Journal of Transport and Land Use*, 18(1), 1-27. <https://doi.org/10.5198/jtlu.2025.2508>
- Schiavina, M., MacManus, K., J., & Freire, S. (2019). *GHS-POP R2019A – GHS population grid multitemporal (1975-1990-2000-2015)*. R2019A. European Commission, Joint Research Centre (JRC) [Dataset].
- Smętkowski, M. (2015). Spatial patterns of regional economic development in Central and Eastern European countries. *Geographia Polonica*, 88(4), 539-556. <https://doi.org/10.7163/GPol.0033>
- Timiryanova, V., Krasnoselskaya, D., Lakman, I., & Popov, D. (2021). Inter-and Intra-Regional Disparities in Russia: Factors of Uneven Economic Growth. *Sustainability*, 13(24), 13754. <https://doi.org/10.3390/su132413754>
- United Nations Statistics Division. (2025). *National Accounts – Main Aggregates Database*.
<https://unstats.un.org/unsd/snaama/Downloads>
- United Nations, Department of Economic and Social Affairs, Population Division. (2025). *World Population Prospects 2024*. <https://population.un.org/wpp/>
- Unwin, D. J. (1996). GIS, spatial analysis and spatial statistics. *Progress in Human Geography*, 20(4), 540-551. <https://doi.org/10.1177/030913259602000408>
- Walaszek, M. (2025). Changes in actual population growth across European subregions (NUTS 3) in the years 2018-2021. *Geographia Polonica*, 98(3), 357-377. <https://doi.org/10.7163/GPol.0307>
- Wegener, M. (2008). SASI model description. *Spiekermann & Wegener Urban and Regional Research, Dortmund*, 49.
- World Bank. (2025). *World Development Indicators: Population, total (SP.POP.TOTL)*.
<https://data.worldbank.org/indicator/SP.POP.TOTL>. *GDP (constant 2015 US\$ – NY.GDP.MKTP.KD; current US\$ – NY.GDP.MKTP.CD)*. <https://data.worldbank.org/indicator/NY.GDP.MKTP.KD>

WorldPop Team. *WorldPop 'Global1' 2000-2020 datasets*. University of Southampton.
https://hub.worldpop.org/Global1_2000-2020

WorldPop Team. (n.d.). *World Pop 2015-2030 datasets* <https://hub.worldpop.org/project/categories?id=3>

Żuber, P., Bachtler, J., Churski, P., Komornicki, T., Matczak, R., Mikołajczyk, A., ... & Zaucha, J. (2025). Will territorial cohesion survive after 2028? Assessing the implications of the reform of cohesion policy for future structural funds. *Geographia Polonica*, 98(4), 411-435. <https://doi.org/10.7163/GPol.0309>

Appendix – Boundary decisions for the study area

In the north and west, we adopted the Arctic and the Mid-Atlantic Ridge as the boundary of the continent. This approach includes Jan Mayen and Svalbard (Norway), Iceland, and the Faroe Islands. To enable analyses covering areas located outside continental Europe but formerly or currently belonging to the European Community or the European Union, additional EU and Community territories situated beyond the European continent were included, namely the French overseas territories in Latin America (Guadeloupe, Martinique, and Guyane), as well as Madeira and the Azores (Portugal) and Greenland. In the east, we adopted the Ural Mountains and the Ural River, including the northwestern part of Kazakhstan, the Caspian Sea, and the Kuma–Manych Depression between the Don River and the Caucasus as the continental boundary. This differs from the more traditional schoolbook boundary running along the watershed of the Greater Caucasus, the so-called “Caucasus Ridge”, and was intended to exclude the entire Caucasus region from Europe, including all Caucasian countries and the Trans-Caucasian part of Russia. The boundary then continues along the Black Sea, the Bosphorus, and the Dardanelles. We assumed that the European part of Turkey, that is, Thrace, as well as the entire NUTS1 region encompassing Istanbul, located on both sides of the Bosphorus Strait, belong to Europe. In the south, we defined the Mediterranean Sea, including its islands, Sicily, Sardinia, Corsica, the Balearic Islands, Crete, and Malta, as the southern boundary of Europe. Following the rule of avoiding the division of NUTS1 regions, all Greek islands were also included. However, in the extended version of the database, designed for analyses covering the European Community or the European Union, we additionally included the Canary Islands, Ceuta, and Melilla (Spain), Cyprus, as well as the French overseas territories located in Africa, namely La Réunion and Mayotte.