GRANULAR

EXPLORING RURAL DATA LANDSCAPES

A BENCHMARK OF PERFORMANCE AND COSTS IN THE EU AND BEYOND

D 3.2

DECEMBER 2023

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. UK participants in the GRANULAR project are supported by UKRI- Grant numbers 10039965 (James Hutton Institute) and 10041831 (University of Southampton).

D3.2 Exploring rural data landscapes – A benchmark of performance and costs in the EU and beyond

Project name	GRANULAR: Giving Rural Actors Novel data and re-Useable tools to Lead public Action in Rural areas
Project No	Horizon Europe Grant Number (101061068); UKRI Grant Numbers James Hutton Institute (10039965) and University of Southampton (10041831).
Type of funding scheme	Horizon Europe Research and Innovation Action (RIA)- UK Research & Innovation Grant
Call ID & topic	HORIZON-CL6-2021-COMMUNITIES-01-01
Website	www.ruralgranular.eu
Document type	Deliverable
Status	[ ] Draft [X] Submitted to the European Commission [ ] Approved by the European Commission
Dissemination level	Public
Authors	Kull M., Weckroth M., Vihinen H., Hiltunen A., Niskanen L. (Luke); Ysebaert R., Guérois M. (CNRS); Balagué J. (Départ. des Pyrénées-Orientales); Hopkins J., Miller D. (Hutton); Stjernberg M., Tapia C. (Nordregio); Georgieva I., McCallum I., Hofer M. (IIASA); Kurdys-Kujawska A. (TU Koszalin), Chasset L., Depontailler L. (Pays Pyrénées Méditerranée), Voepel H. (UoS), Martins B., Doval Ruiz M. (UVigo), Jain P., Berchoux T. (IAMM)
Work Package Leader	International Institute for Applied Systems Analysis (IIASA)
Project coordinator	Mediterranean Agronomic Institute of Montpellier (IAMM)

Citation: Kull, M., Weckroth, M., Vihinen, H., Hiltunen, A., Niskanen, L., Ysebaert, R., Marianne, G., Balagué, J.-C., Hopkins, J., Miller, D., Stjernberg, M., Tapia, C., Georgieva, I., McCallum, I., Hofer, M., Kujawska, A., Chasset, L., Depontailler, L., Voepel, H., … Berchoux, T. (2023). Exploring rural data landscapes – A benchmark of performance and costs in the EU and beyond. GRANULAR. https://doi.org/10.5281/zenodo.13744714

This license allows users to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. UK participants in the GRANULAR project are supported by UKRI- Grant numbers 10039965 (James Hutton Institute) and 10041831 (University of Southampton).

Table of contents

1. Contents

Table of contents **3**

2. Executive summary **5**

3. Introduction **6**

3.1. Introduction to Task 3.2. and purpose of this deliverable 6

3.2. Data cost dimensions and considerations 6

3.3. Structure of the Document 8

4. Methodology and approach **8**

4.1. Survey: national and regional statistical offices, authorities and data providers in the EU and beyond 8

4.2. Inspirational examples from Granular partner countries – Data, tools and approaches 9

4.3. Data Fiches: From initial datasets to data fiches 9

5. The costs of and for rural data – a journey across regional and national statistical offices and authorities in the EU and beyond **12**

5.1. Main databases openly available, containing data about rural development issues 13

5.1.1. Bulgaria 13

5.1.2. Croatia 13

5.1.3. Cyprus 13

5.1.4. Finland 13

5.1.5. France 14

5.1.6. Greece 14

5.1.7. Hungary 14

5.1.8. Ireland 15

5.1.9. Italy 15

5.1.10. Moldova 15

5.1.11. Poland 15

5.1.12. Portugal 16

5.1.13. Scotland 16

5.1.14. Serbia 16

5.1.15. Slovakia 16

5.1.16. Slovenia 17

5.1.17. Spain / Galicia 17

5.1.18. Sweden 17

5.2. Data types and domains available at no costs 18

5.3. Customers, costs and revenues 20

6. Inspirational examples – Data, tools and approaches from Granular partner countries. 23

6.1. Finland: Indicator to follow Subjective Wellbeing (SWB) development of the rural population during the CAP programming period – by Mikko Weckroth (Luke) 23

6.2. Finland: The Rural Barometer – by Hilkka Vihinen & Michael Kull (LUKE) 24

6.3. France: Monitoring mobility and road traffic at local scale – Case *Conseil Départemental des Pyrénées-Orientales (CD66)* – by Louise Chasset, Lenaïc Depontailler (Pays Pyrénées Méditerranée) & Jean-Claude Balagué (Département des Pyrénées-Orientales) 26

6.4. Galicia – Spain: Telecare for the elderly at home in the rural areas of Ourense – by María Isabel Doval Ruiz & Breixo Martins (University of Vigo) 28

6.5. Poland: Functional and spatial diagnosis for social revitalization at the local (municipal) level – by Agnieszka Kurdys-Kujawska (TU Koszalin) 29

6.6. Scotland: Scottish National Islands Plan Survey (2020): results explorer – by Jonathan Hopkins and David Miller (Hutton) 31

6.7. Nordic Countries: The Nordic Service Mapper - by Mats Stjernberg (Nordregio) 33

6.8. EU: Integrating text-to-image and image-to-text techniques to enhance accessibility and understanding of rural land-use data - Cross Modality Framework – by Pallavi Jain (IAMM) 35

6.9. Global: Geo-Wiki Earth Observation & Citizen Science – by Ivelina Georgieva (IIASA) 36

6.10. Key lessons learned, cost considerations, policy implications and ways forward 38

7. Discussion, conclusions and going forward 42

8. Conclusion 44

9. Acknowledgements 45

10. References 45

11. Appendix 1 – Overview of national databases openly available, containing data about rural development issues 48

12. Appendix 2 - Data Fiches 53

12.1. Accessibility 53

12.2. Agriculture 54

12.3. Climate 57

12.4. Demography 59

12.5. Digitalisation 66

12.6. Economic Development 67

12.7. Energy 70

12.8. Health 71

12.9. Infrastructure 72

12.10. Mobility 75

12.11. Recreation 76

12.12. Transversal 77

2. Executive summary

Through **collaboration with and input from regional, national, and international data providers, Living Labs, and statistical offices**, we evaluated the **performance and costs associated with acquiring data relevant for rural territories at various geographical scales**. The resulting document presents a **compilation of 27 data fiches,** a comprehensive exploration of data costs and dimensions through **a survey, and nine inspirational examples of data, tools, and approaches**. The document seeks to inform GRANULAR's Living and Replication Labs, as well as data users and providers from various levels, about practical insights into data collection types and costs for rural territories, their challenges, and opportunities. The outlined **cost categories, spanning data infrastructure, governance, model training, security, and more,** provide a robust framework for understanding and managing **the complexities associated with enriching knowledge and advancing data-driven initiatives for policy making in rural contexts**.

The Survey engaged national statistical offices and data providers, yielding 17 responses from 16 different countries, and focused on data availability, costs, and user information. The survey delves into the landscape of data accessibility applied to rural development across several European countries. The report highlights key open databases, data domains, and types of data available at no cost, revealing variations in resolution and thematic coverage. The majority of surveyed offices make over 500 data files freely accessible, with primary domains including demography, economy, agriculture, and tourism. While most respondents adhere to Eurostat guidelines, data types vary, encompassing tabular, grid-level, vector, and raster data. Notably, the report touches upon the diverse bases for user charges, with some offices levying fees for data compilation or structuring. The analysis underscores the crucial role of individuals, research organizations, and the private sector as the main users of these datasets, emphasizing the multifaceted demand for demographic, economic, and development-focused information.

Data fiches were compiled for 27 datasets that capture a wide range of rural data types, and include a description of the data, including indicator class, data class, spatial and temporal information, a short description and how to cite it. Cost information include data infrastructure costs, e.g. software or hardware needed or costs for data repositories / storage. Data Governance & Management costs, i.e. costs needed to work with the data e.g. comprises information about staff costs needed to work with / access the data, for data analysis, quality assurance and "cleaning" data. Authors also reflect on data documentation costs. Data covered the 4 rural functions identified in the Rural Compass (productive, residential, environmental, recreational), with a variety of domains, such as accessibility, agriculture, climate, demography, digitalization, economic development, energy, health, infrastructure, mobility, recreation, and transversal.

The inspirational examples showcase data collection and provision methodologies that have been implemented at local and international scales. Each example was documented with a structured overview including costs, photos, and a matrix of key insights, policy implications, and future considerations. The aim is to inform actors on the diversity of methods and data that can be collected and the costs that are associated with such initiatives. Examples include:

- Indicator to monitor the subjective well-being of the rural population during the CAP programming period in Finland (National)
- Rural Barometer Finland (National)
- Monitoring mobility and road traffic at local scale in France (Local)
- Telecare for the elderly at home – Galicia / Spain (Local)
- Functional & spatial diagnosis for social revitalization – Poland (National)
- Scottish National Islands Plan Survey (Regional)
- Web-mapping tool to visualize proximity to different services - Nordic countries (International)
- Enhancing accessibility & understanding of rural land use data – EU (International)
- Earth Observation & Citizen Science – Geo-Wiki (Locally informed, international coverage)

The **holistic approach** chosen in this document with three focus areas enables a **comprehensive understanding of rural data, costs, and innovative examples for evidence-based policy-making**.

Effectively utilizing open data in rural territories within the EU demands **strategic investments in capacity building**, with an emphasis on training local personnel in Geographic Information System (GIS) and data management. The estimated annual cost of employing entry-level GIS technicians ranges from €30,000 to €40,000, reflecting the necessity to bridge the skills gap for meaningful spatial data analysis. Additionally, ensuring the reliability of open data for decision-making requires rigorous quality assessment processes, including validation with local databases and complementary data collection, with associated annual expenses ranging from €25,000 to €35,000.

3. Introduction

3.1. Introduction to Task 3.2. and purpose of this deliverable

Task 3.2. builds on D3.1 Screening Rural Data Sources, where an initial screening of data availability was conducted for the generation of new and novel datasets to support indicators of rural sustainability for Europe. More than 90 existing datasets relevant for rural territories, along with accompanying meta-data have been recorded. The screened datasets address the majority of GRANULAR's Rural Compass[1] indicators, with the majority of datasets representing demography, infrastructure and environment.

T3.2. has prioritised data sources and methods identified in T3.1, considering the performance and costs of acquiring data at an appropriate geographical scale. Options identified in T3.1 were considered to make them relevant for users both in the context of the Living Labs and Replication Labs (https://www.ruralgranular.eu/living-labs/).

This document showcases the performance and costs of different data types and tools and discuss viable options going forward. Different regional, national, and international data providers were consulted, as were the Living Labs as such. Consultation and co-writing took place during various stages of producing this deliverable and always included the identification of costs for users and those occurred during the development of a tool or compiling data. We also consulted national and regional statistical offices and authorities to get an overview of available data at different levels of granularity, the costs of and for data and data users. Finally, project partners, including Living Labs were invited to share and write about inspirational examples from their region or country, including on how they have used and produced data and indicators or tools to work with data.

This document was written to inform and inspire both GRANULAR Living and Replication Labs as well as data users and providers in rural areas and beyond. It thus responds to one of key aims of GRANULAR, i.e. to enrich knowledge for rural actors on the diversity of rural areas, their functional characteristics, challenges and opportunities. Consequently, this document contains:

- *A journey across national and regional statistical offices and authorities in the EU and beyond to better understand costs of and for rural data,*
- *9 inspirational examples of data, tools, and approaches and*
- *Data fiches summarizing key information for 27 different datasets (in the appendix).*

These elements all include different cost dimensions as far as they were available.

3.2. Data cost dimensions and considerations

Most of the datasets identified in D3.1 Screening Rural Data Sources contain a free and open license or allow partial access, often complying with the FAIR Principles and the INSPIRE Directive. Just a few datasets require

[1] The Rural Compass is a tool aimed at orientating and providing policy direction, going from policy design in rural areas to policy monitoring and evaluation. The Rural Compass is conceived as a multi-dimensional set of indicators and trends. The assessment of rural areas against those indicators and trends can be used to map and assess rural communities and their functional characteristics. It can be found here https://www.ruralgranular.eu/tools/.

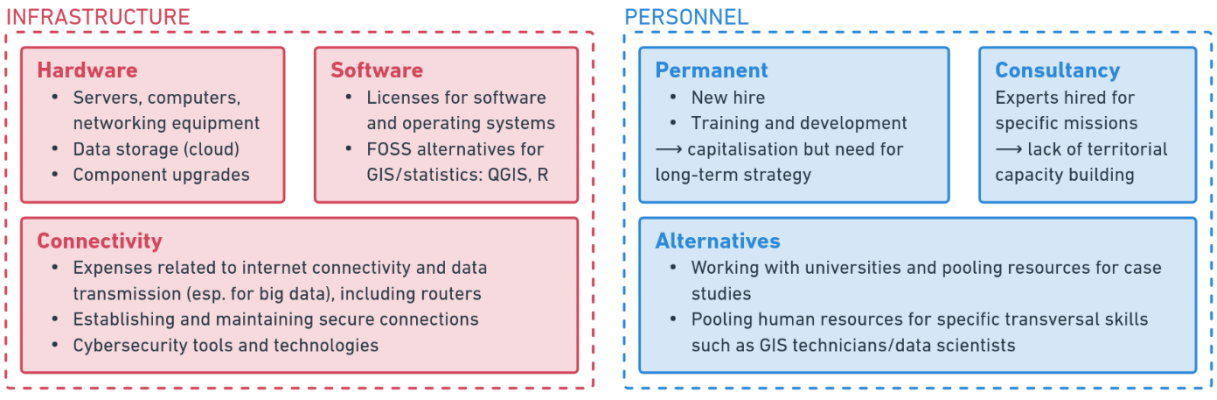
purchase. Likewise, statistical offices at different levels of governance and throughout the EU provide numerous data, datasets and maps being free of charge for the users (chapter 5). Whilst the inspirational examples in chapter 6 are also free of charge for different user groups, there are different development costs to consider, particularly if they are to be replicated elsewhere.

Overall, there are different cost and performance dimensions that need to be thought about when working with or processing data. Whilst we try to specify the particular costs for each dataset and inspirational example as best as possible and as far as applicable, there are a number of general data costs categories defined in the literature (e.g. Becker 2017, Colas et al 2014, Martinez et al. 2021, Saltz et al. 2017b, Sivarajah et al. 2016), which will also be considered where applicable as far as information is available. These include (Figure 1):

- ***Data infrastructure:*** including software to work with the data, hardware to process data, data storage, or data repositories. This also includes processes for storing, retrieving, and sharing data. This also includes "the big data perspective" with data volume increasing and velocity intensifying computation requirements and dependence on IT resources (Martinez et al. 2021, Saltz et al. 2017b, Sivarajah et al. 2016). Finally,
- ***Data governance & management:*** the needs and capabilities of people to conduct data analysis, training needs of staff, or costs for staff effort for quality assurance. An intermediary (e.g. Martinez et al. 2021) that understands both the language of data analytics and the domain of application, creating an understanding between data scientists and stakeholders, customers, businesses etc. This dimension also includes model training and retraining costs, e.g. for machine learning models, which may be resource intensive and costly. (Martinez et al. 2021).
- ***Data quality and scale:*** there might be a need to clean "dirty" data and reflection on the potential to be suitable. Coordinated data cleaning and quality assurance are needed to assure a robust validation (Martinez et al. 2021). Data scale implies reflecting complexity, adjust necessary architecture and infrastructure and the corresponding costs (Becker 2017, Colas et al 2014 & Martinez et al. 2021).
- ***Data security & privacy*** including compliance with data protection regulations, encryption methods, anonymization, user education and training (Martinez et al. 2021, Saltz & Shamshurin 2016, Colas et al 2014, Sivarajah et al. 2016). In this connection, there should be awareness of dependency on legacy systems and data integration *(Colas et al 2014, Becker 2017).*
- ***Creating institutional memory & retaining institutional knowledge:*** personnel costs associated with knowledge documentation and sharing etc. (Byrne 2017, Martinez et al. 2021, Colas et al 2014, Becker 2017).

Martinez et al. 2021 conclude that in addition to an understanding to what data might be available (see also Saltz et al. 2017a), its representativeness for the problem at hand (Saltz & Shamshurin 2016) and its limitations (e.g. Byrne 2017) is critical for the success of data projects.

Figure 1. Data costs categories



3.3. Structure of the Document

This document is organised as follows. After this introduction, section 4 explains the methodology used to gather different insights and lessons learned for this report – for the data fiches, the survey and the inspirational examples. Chapter 5 is a journey throughout Europe and its national and regional statistical offices, all providing insights on accessibility and availability of data, different costs dimensions and additional costs users must be aware of. Chapter 6 provides inspirational examples of data, tools and applications from around the EU. Chapter 7 is a discussion and conclusion of our findings and discusses possible ways forward. The data fiches – 27 different datasets of relevance for rural areas and actors – are to be found in the appendix.

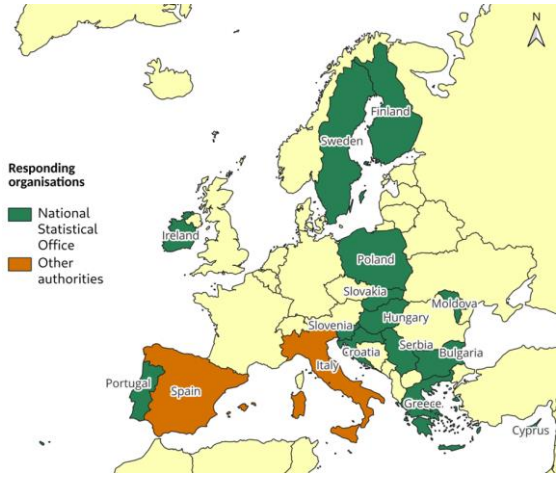
4. Methodology and approach

Our three inspirational examples were all built on their own particular methodology and approach. Hence, we outline methodologies separately for each, starting with the survey, shortly describing how the inspirational examples were selected, composed and structured and finally explain how and why the data fiches have been chosen and composed.

4.1. Survey: national and regional statistical offices, authorities and data providers in the EU and beyond

GRANULAR's aim is to identify, develop and provide novel data and reusable tools to understand the characteristics, dynamics and drivers of rural areas and hence support place and evidence-based policy making. In connection to this, we want to better understand and map the performance and costs of acquiring data at different geographical scales, including from the national level and the statistical offices and authorities from around the EU and beyond. We thus invited national statistical offices and authorities across the EU and beyond, including all countries where the GRANULAR Living and Replication Labs are located, to take part in a survey. We inquired about data availability, pricing policies and costs of data as well about customers and user groups and their demands. The survey contains 3 sections: 1) data availability, 2) data costs and 3) users. More than 20 questions are both open-ended, selection and multi-selection, ranking and open-ended questions[2]. We received 17 answers from both EU and non-EU countries. Most respondents were national statistical authorities. In two cases respondents answered on behalf of other authorities[3] (Figure 2).

Figure 2. Location of responding organizations.



[2] All questions are available from the lead author.
[3] One of the cases was a regional authority from Spain, which was contacted due to the relevance and connection to the Living Lab Ourense. The other respondent was from Italy.

In some cases, respondents sent their replies and additional answers directly to us (e.g. Scotland, Sweden, Moldova). In some cases, such as Spain, also regional offices were included, since they are working closely with Living Lab actors.

We used the Webropol platform to collect the answers. The survey was open between September and November 2023. Survey results are presented and discussed in chapter 5. We also include links to the main databases relevant for rural development issues, and present the data domains contained in open databases and the finest resolution of data free of charge in each of the country that participated in the survey.

4.2. Inspirational examples from Granular partner countries – Data, tools and approaches

To better understand and map the performance and costs of acquiring data at different geographical scales, we invited project partners including Living Labs, whether they have good examples of data collection or tools that are already being implemented in their territory and invited them to contribute to a collection of inspirational examples from different scales. This means, examples ranging from an indicator to an international or transboundary mapping tool and from the local level to national level examples and beyond. The idea and motivation were that the examples will not only be included in this deliverable, but also published on the GRANULAR website.

All examples follow a similar structure and are composed of a description of the example and costs for developing it. Authors explain the motivation and objective, their experience thus far, future considerations, and who has developed it.

The cost dimensions covered – and as far as the authors had them available – include:

Budget of the whole initiative, such as:

- *Data Governance & Management costs*
- *Staff costs or PM*
- *Costs for Data analysis*
- *Quality assurance, "cleaning data"*

Data Infrastructure costs, such as:

- *infrastructure needed (software, hardware)*
- *Data repositories / storage needs*

Costs for data processing & visualization

Nice photos from the area and the tool, including maps, add some additional flavor. A matrix will summarize key learnings, cost dimensions, policy implications and future considerations from the inspirational examples.

4.3. Data Fiches: From initial datasets to data fiches

D3.1 Screening Rural Data Sources developed a Rural Data Table with more than 90 datasets was developed.[4] Those datasets were initially ranked according to their degree of relevance for GRANULAR, in terms of supporting GRANULAR's Rural Compass and on a scale from 4 (very relevant) to 1 (lower relevance). For the purpose of this document, those data ranked 4 (very relevant) and 3 (relevant) were selected. The motivation is twofold. First, we wish to provide LL and RL and other rural stakeholders with cost and accessibility information about these relevant datasets for their work. Second, we want to summarise the current knowledge about these datasets to support the continuous work of GRANULAR, its Repository and the Rural Compass.

Datasets that had certain shortcomings, for instance regarding temporality, old data, or quality availability of metadata, were replaced. The selected datasets for the data fiches are summarized in table 1.

[4] Data collected by GRANULAR will be made available in the project's Repository, available at https://platform.ruralgranular.eu/collection/All/1.

Table 1: Datasets selected for analysis by Indicator Class & Rural Compass.

Title	Description	Author / Producer	Free & open Access[5]	Rural Compass
Accessibility				
Open-source routing machine	Accessibility to services (by car and by bike)	OSRM	Yes	Residential
Agriculture				
Carbon budget in the EU agricultural soils	C budget in the EU agricultural soils including lateral C fluxes	JRC	Yes	Productive
Cover Crops across Europe	Disaggregated map of cover crops occurrence for Europe and the UK	JRC	Yes	Productive
EuroCrops - Land Parcel Identification Systems	Combines all publicly available self-declared crop reporting datasets	EU	Yes	Productive
Climate				
Temperature, Precipitation	Climate model data	ECMWF	Yes	Environmental
WorldClim	Satellite derived climate data (temperature, rainfall)	Worldclim	Yes	Environmental
Demography				
GHSL-POP	Distribution of population, expressed as the number of people per cell	JRC	Yes	Residential
WorldPop	High resolution world population weighted-density	UoS	Yes	Residential
WorldPop	High resolution world population density	UoS	Yes	Residential
WorldPop	High resolution world population (age and sex structures by 5-year classes)	UoS	Yes	Residential
WorldPop	High resolution world population (Population Counts)	UoS	Yes	Residential
Data4good	Population	Meta	Partial	Residential

[5] Relates to the free and open access and availability of the derived datasets, as well as raw datasets (for the analysis) and the model or code.

Total Population	Population	ARDECO	Partial	Residential
Digitalisation				
Rural household internet access in 2021	Share of rural household with internet access in the selected European countries 2021	Statista	n.a.	n.a.
Economic Development				
Accommodation	Overnight stays	Trip Advisor	No	Productive
Accommodation	Overnight stays	AirBnB	No	Productive
EU SILC	Longitudinal Household Survey	EuroStat	Partial	Productive
Energy				
Suitability map for solar energy (PV)s	Suitability for installation of large-scale PV systems in Europe	JRC	Yes	Productive
Health				
Healthcare services locations & number of beds in Europe	Dataset making centrally, geo-localized healthcare information available	EuroStat	Yes	Residential
Infrastructure				
OpenStreetMap	Specific point locations for potential destinations for accessibility indicators calculation. Topographic mapping of features across the globe; good coverage of Europe; high quality of mapping.	OSM	Yes	Residential
Mapillary	Access street-level imagery & map data from all over the world	Mapillary	Partial	Residential
Tourism	Tourism capacity and density based on booking.com, TripAdvisor and Eurostat data	JRC	Yes	Recreational
Mobility				
Mobility heat maps, raw data (jogging, biking)	The heatmap shows 'heat' made by aggregated, public activities.	Strava	No	Residential

Recreation				
Potential quiet rural areas	These spacious areas allow for example extensive walks without crossing noisy areas.	Data.Europa.EU	n.a.	n.a.
Transversal				
Local Administrative Units (LAU)	Smaller official territorial division for Europe	GISCO	Yes	Other
GEOSTAT 1km population grid	EU reference grid for statistics. Only population currently, but data diversity will certainly grow in the future. Only populated cells (not a regular grid)	GISCO	Yes	Other
NUTS geometries	NUTS division	GISCO	Yes	Other

The data fiches (in the appendix) contain general information about the dataset and cost information. The first page informs about the data type, spatial and temporal extent, a short description of the data, how to cite it and the methodology used to produce the data. The second page is about the costs. Whilst most datasets are free and open access, data analysis, quality assurance, and infrastructure costs need to be considered.

Data domains covered by the fiches are accessibility, agriculture, climate, demography, digitalization, economic development, energy, health, infrastructure, mobility, recreation, and transversal.

Rural Compass categories included are environmental, productive, recreational, residential, and other.

The data fiches are in the appendix and will be made available online at https://www.ruralgranular.eu/tools/.

5. The costs of and for rural data – a journey across regional and national statistical offices and authorities in the EU and beyond

This section presents and discusses the survey results[6] by focusing on:

- *Main databases openly available in the participating countries, containing data about rural development issues.*
- *Data types and domains available at no costs.*
- *Customers, costs, and revenues.*

Most of the information compiled below was derived from the survey. In some cases, additional research to complement missing information was undertaken. In cases no answers were provided to particular questions, we leave the specific category out in the section 5.1. descriptions. This concerns, for instance, and in several cases, data types and domains.

[6] Appendix 1 contains a summary table providing a quick overview over the results.

5.1. Main databases openly available, containing data about rural development issues

5.1.1. Bulgaria

Main database openly available, containing data about rural development issues: The IS Infostat platform (https://infostat.nsi.bg) publishes data relevant for rural development. It includes business, demographic social, macroeconomic, environment energy and multi-domain statistics. Data on population and housing census is available free of charge. Paid databases provide users access to more detailed data at lower levels after disaggregation. Grid data for population (total, age groups and sex) for 2011, 2021 is free of charge.

Data domains contained in open databases: Demography, energy and health.

Finest resolution of data free of charge: Grid data for population: Total population, Age groups 0-14; 15-64 ;65+ and sex, for 2011 and 2021.

5.1.2. Croatia

Main database openly available, containing data about rural development issues: There are several databases available but not strictly connected with rural development issues. The so-called PC-Axis databases are available at https://web.dzs.hr/PX-Web_e.asp?url=%22Eng/Archive/stat_databases.htm%22. Some data available is at municipality level. Grid 1000 is accessible, free of charge, at the GeoSTAT - Web GIS portal of the Croatian Bureau of Statistics (https://geostat.dzs.hr). The available grid-level data is on population – number of populations, number of population by large age groups, population by educational attainment, population by activity, business register (active business entities). Tourism data on accommodation capacities and tourist arrivals and nights. The PC-Axis databases https://web.dzs.hr) has some data by municipalities. Some census data by settlements is available at https://podaci.dzs.hr/en/statistics-in-line/. All data published online is free of charge. If special data processing is needed for grid-level data, it is charged according to the subject's hourly rate.

Data domains contained in open databases: agriculture, demography, economy, energy, environment, tourism / recreation, transport.

Finest resolution of data free of charge: 1km grid. Available data free of charge at the GeoSTAT - Web GIS portal for the Croatian Bureau of Statistics https://geostat.dzs.hr/?lang=en.

5.1.3. Cyprus

Main database openly available, containing data about rural development issues: The main database by Statistics Cyprus is CYSTAT-DB, available at https://cystatdb.cystat.gov.cy/pxweb/en/8.CYSTAT-DB/. It contains data on agriculture, livestock, fishing, business register, construction, education, energy, environment, external trade, health, industry, information society, innovation, labor market, living conditions, social protection, national accounts, population, price indices, public finance, research and development, services, tourism and trade.

Data domains contained in open databases: Demography, Health, Education, ICT Usage.

Finest resolution of data free of charge: 1000m for grid-level data.

5.1.4. Finland

Main database openly available, containing data about rural development issues: The main database published by Statistics Finland is called StatFin (https://www.stat.fi/tup/statfin/index_en.html). StatFin is freely accessible and includes data on population, economy, housing, transport, tourism, consumption, prices, wages and salaries, energy, enterprises etc. The Paavo database (https://www.stat.fi/tup/paavo/index_en.html) contains data by postal code area on the population structure, education, income, housing, workplaces, households' life stage etc. There is also a grid-level database with grid sizes of 250 m x 250 m, 1 km x 1 km and 5 km x 5 km (https://www.stat.fi/tup/ruututietokanta). The grids cover the whole of Finland, but this is not for free and charged based on number of licenses and grid size. Data is available on the areas' population structure, level of education, income of inhabitants and households, size and stage in life of households, buildings and dwellings, workplaces, and main activities of inhabitants. Population structure grid data 1km x 1km is free of charge.

Data domains contained in open databases: agriculture, demography, economy, energy, environment, housing, infrastructure, mobility, tourism / recreation, transport.

Finest resolution of data free of charge: Population structure grid data 1km x 1km.

5.1.5. France

The geoservices.ign.fr site (https://geoservices.ign.fr/) and its *Géoservices* catalogue (https://geoservices.ign.fr/catalogue) by the *Institut National de l'Information Géographique et Forestière (IGN)* are of relevance for anyone interested in geodata and web services related to rural development. Data on the site is free of charge and available under an open license and accessible without registration. It contains, vector databases, maps, ortho-images, cadastral parcels, 3D models as well as other applications and services. Information includes the Common Agricultural Policy, forest geographical reference system, "Land-Sea Boundary" data, a Renewable Energy Map Portal and Good Agricultural and Environmental Condition.
A second data source is by the Centre for Studies on Risks, the Environment, Mobility and Urban Planning (CEREMA). Its catalogue with over 300 datasets, maps and series is open and available at https://catalogue.cdata.cerema.fr/geonetwork/srv/eng/catalog.search#/home. Data covers such fields like land cover, natural risks, energy use, habitats and biotopes and hydrography.
The *Office français de la biodiversité* with its partners feeds an information system on biodiversity, with 14 datasets, services and maps, incl. characterization of Natura 2000 sites, protected natural areas, inventory of Natural Areas of Ecological, Faunal and Floristic Interest etc. *(https://www.ofb.gouv.fr/)*
Thematic data and datasets on agriculture, culture and heritage, sustainable development and energy, economics and statistics, eductaion and reserach, international and EU issues, health and social issues, tourism and recreation, territories and transport can be found at https://www.geoportail.gouv.fr/. This is the national portal for territorial knowledge providing open and interoperable data to facilitate the exchange and sharing of data in support of public policies.

Finest resolution of data free of charge: grid data

5.1.6. Greece

Main database openly available, containing data about rural development issues: Currently, there is no dissemination database openly available to the public. However, data files available for the public are uploaded at the website of the Hellenic Statistics Authority (ELSTAT) in the form of time series, tables and Public Use Files (PUFs). ELSTAT publishes statistical data on its website at a level of analysis, where statistical confidentiality is not violated, and all users can access them. Most of the data on ELSTAT's website refer to NUTS 2 level. Depending on the limitations set due to statistical confidentiality, some data refer also to NUTS 3 level and few data (mainly population data) to LAU level. Statistical data that allow the indirect identification of statistical units are provided to users under certain conditions. Tailor-made data based on user requirements are generally priced. The cost of providing these data depends on the number of man-days required for their compilation by ELSTAT staff. The pricing policy of ELSTAT is available here: https://www.statistics.gr/en/microdata_pricing

Finest resolution of data free of charge: R&D, innovation, population data at NUTS-2 level.

5.1.7. Hungary

Main database openly available, containing data about rural development issues: There is no dedicated database for rural development data, but the main database contains data on agriculture, environment, and many aspects of territorial data. The data can be downloaded in csv and xlsx format and are free of charge. The database can be accessed here: https://statinfo.ksh.hu/Statinfo/themeSelector.jsp?&lang=en. Predefined data tables in the STADAT system also contain data on the abovementioned topics, however, this is not a database, but ready-made tables are available, which can also be downloaded in csv and xlsx, also free of charge. The STADAT-system is available here: https://www.ksh.hu/stadat_eng.
The third main resource where users can find territorial data is the Interactive Mapping Application, accessible here: https://map.ksh.hu/timea/?locale=en. Data can be downloaded from this interface, from the attribute table in csv. Unlike the aforementioned two sources, this application also contains grid-level data.

Data domains contained in open databases: agriculture, demography, economy, energy, environment, health, housing, infrastructure, tourism / recreation, transport.

Finest resolution of data free of charge: Grid-level data. Charging for the data is not determined by the level of resolution or the theme but by the capacity it requires from the statisticians to prepare the data.

5.1.8. Ireland

Main database openly available, containing data about rural development issues: The Central Statistics Office (CSO) of Ireland produces a wide range of statistics. This includes business sectors including agriculture and fisheries, census data, data on the economy, environment, labour market, people and society as well as other themed publications. CSO's *PxStat Open Data Platform* is available at https://data.cso.ie/#. Data published on the platform is also provided by other public sector databases, such as by the Department of Agriculture, Food and the Marine, the Department of Housing, Local Government and Heritage, the Sustainable Energy authority etc.

Data domains contained in open databases: see above.

Data types available: grid-level, tabular and vector data.

Finest resolution of data free of charge: Small-area data, for which are units with an average of 50-100 households in each. Grid-level data is free, and if it is not available, €80-200 are charged for customers from the private sector.

5.1.9. Italy

Main database openly available, containing data about rural development issues: IstatData is the latest aggregate data dissemination platform of the Italian National Institute of Statistics (Istat) and available at https://esploradati.istat.it/databrowser/#/en. It makes use of the open-source tools “Data Browser” and “Meta & Data Manager” developed by Istat following the international SDMX (Statistical Data and Metadata eXchange) standard for exchanging and sharing statistical data and metadata. Currently, six themes are covered: National Accounts, Population and Households, Household Economic Conditions, Agriculture, Enterprises, Welfare and Pension. Time Series (https://seriestoriche.istat.it/index.php?id=18&L=1) contains over 1,500 time series organized into 22 thematic areas made available to inform about the environmental, social and economic changes in Italy

Data domains contained in open databases: see above.

Data types available: tabular and vector data.

Finest resolution of data free of charge: Agricultural plot level.

5.1.10. Moldova

Main database openly available, containing data about rural development issues: The statistical databank of Statistics Moldova is available at https://statbank.statistica.md/. It contains data on environment, population and demographic processes, social statistics, economic statistics, gender statistics, and regional statistics.

Data domains contained in open databases: Accessibility, agriculture, demography, economy, energy, environment, health, housing, infrastructure, mobility, tourism / recreation, transport.

Data types available: raster and tabular data.

Finest resolution of data free of charge: Currently the office disseminates at the level of communes (with a commune being formed of one or several villages) – both in the Statistical databank and as static maps in its publications. In 2024, the office plans to carry out a Population and Housing Census and will then have available spatial data both at vector and grid-level. At the moment, the office does not provide spatial data against a fee.

5.1.11. Poland

Main database openly available, containing data about rural development issues, incl. data domains: The Knowledge Database (https://dbw.stat.gov.pl/en) by Statistics Poland contains 31 domain areas including Demography, Education, Energy, Social economy, Municipal and housing infrastructure, Agriculture, Labor Market,

Transport, Tourism, Living conditions, Health and healthcare. For some indicators data is available by voivodships, in the case of demographic data and of local government unit budgets also for lower levels of territorial division. The Knowledge Database is a publicly available and free of charge.
Furthermore, there is also a Centre for Rural Statistics (https://olsztyn.stat.gov.pl/en/) and a Small Areas Statistics Centre (https://poznan.stat.gov.pl/en/) in addition to regional statistics centers throughout the country. (https://stat.gov.pl/en/regional-statistics/). EUROSTAT Economic accounts for agriculture - values at current prices and Economic accounts for agriculture - values at n-1 prices are also available.

Finest resolution of data free of charge: NUTS 2.

5.1.12. Portugal

Main database openly available, containing data about rural development issues of Statistics Portugal is available here https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_bdc_tree&contexto=bd&selTab=tab2. Themes also include Agriculture, forest and fisheries.

Data domains contained in open databases: agriculture, demography, economy, energy, environment, health, housing, mobility, tourism / recreation, transport in addition to culture, Prices, living conditions.

Data types available: grid-level and tabular data.

Finest resolution of data free of charge: For the Population and housing Census grid of 1km2 is free of charge. For some of the other domains, data is at parish-level.

5.1.13. Scotland

Main database openly available, containing data about rural development issues: The main database openly available, containing data about rural areas are the 1) National Performance Framework (https://nationalperformance.gov.scot/measuring-progress/national-indicator-performance) with 26 out of 81 indicators providing data for Rural Scotland and the 2) Rural Scotland Key Facts 2021 (https://www.gov.scot/publications/rural-scotland-key-facts-2021/documents/) and a considerable number of sources. There are several public sector open data portals, most allow for publication of data at the 3* level of openness (csv or equivalent).[7] Many of Scottish Government statistics is available online, for free and without restrictions. It contains around 300 open datasets and reference material, mainly at the 5* level – the highest level of openness, with associated metadata.

Finest resolution of data free of charge: Grid-level data.

5.1.14. Serbia

Main database openly available, containing data about rural development issues of the Statistical Office of the Republic of Serbia is available here: https://data.stat.gov.rs/?caller=SDDB&languageCode=en-US.

Data domains contained in open databases: accessibility, agriculture, demography, economy, energy, environment, health, housing, infrastructure, tourism / recreation, transport.
Finest resolution of data free of charge: Spatial resolution depends on the coverage of statistical survey, from which the dataset is produced. In the annual plan of statistical surveys, the spatial resolution of available datasets which are free of charge is defined.

5.1.15. Slovakia

Main database openly available, containing data about rural development issues of Slovak Statistics is called DataCube and is available at https://datacube.statistics.sk/. It contains multidimensional tables for indicators of economic and socio-economic development, for the following areas: demographic and social statistics, macroeconomic statistics, business statistics, sector statistics, environment, multi-domain statistics and selected tables of the Eurostat database. 1 km x 1 km grid-level data is available from the 2011 and 2021 census at https://slovak.statistics.sk/wps/portal/ext/themes/demography/census/indicators/.

[7] According to Tim Berners-Lee, open data can be published at various levels of openness (see 5-star Open Data (5stardata.info).

Data domains contained in open databases: accessibility, agriculture, demography, economy, energy, environment, health, housing, infrastructure, tourism / recreation, transport.

Finest resolution of data free of charge: Municipalities - especially demographic data. All data in the database are free of charge.

5.1.16. Slovenia

Main database openly available, containing data about rural development issues of Statistics Slovenia is available at *https://pxweb.stat.si/SiStat/en/Podrocja/Index/583/regionalni-pregled*.

Data domains contained in open databases: agriculture, demography, economy, energy, environment, housing, infrastructure, mobility, tourism / recreation, transport.

Finest resolution of data free of charge: Grid-level data available at the STAGE pages at https://gis.stat.si/#lang=en. All data at Statistics Slovenia is free of charge.

5.1.17. Spain / Galicia[8]

Main database openly available, containing data about rural development issues of the Galician Institute of Statistics is available at https://www.ige.gal/web/index.jsp?idioma=gl.

Data domains contained in open databases: demography, economy, tourism / recreation.

Finest resolution of data free of charge: Spatial distribution of the characteristics of the population of Galicia by grid of 1km2.

5.1.18. Sweden

Main database openly available, containing data about rural development issues
The majority of Statistics Sweden's statistics are openly available at *"Statistikdatabasen"* (Statistical database) (https://www.statistikdatabasen.scb.se/pxweb/en/ssd/). In addition, other data openly available from Statistics Sweden includes geospatial data (https://www.scb.se/en/services/open-data-api/open-geodata/). Furthermore, there are 29 different public agencies, which publish official statistics, and partly cover other topics than Statistics Sweden: https://www.scb.se/en/About-us/official-statistics-of-sweden/government-agencies-responsible-for-official-statistics/.

Data domains contained in open databases: agriculture, demography, economy, energy, environment, housing, infrastructure, mobility, tourism / recreation, transport.

Data types available: grid-level, point, tabular and vector data.

Finest resolution of data free of charge: Grid-level data in Sweden is free of charge. In the database mentioned previously, some statistics are made available according to DeSO, "Demografiska statistikområden" (in English: "Demographic Statistical Areas"). There are also corresponding GIS layers available. For more information, please see https://scb.se/hitta-statistik/regional-statistik-och-kartor/regionala-indelningar/deso---demografiska-statistikomraden/ (Swedish only) or https://www.scb.se/en/services/open-data-api/open-geodata/deso--demographic-statistical-areas/.

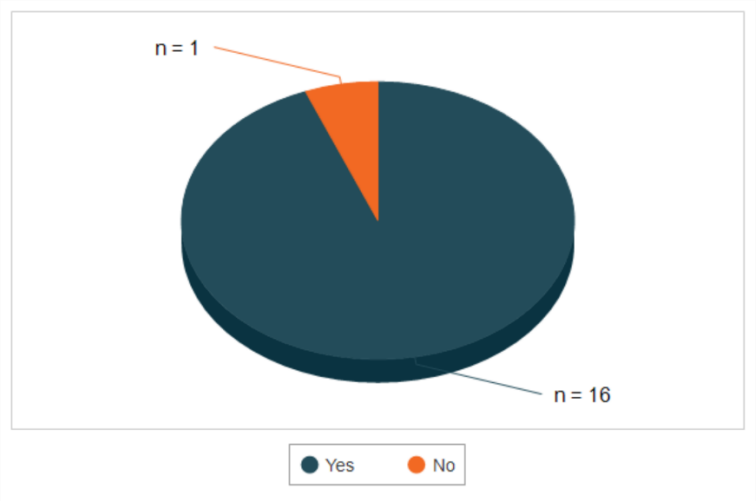
In addition to the database mentioned above, there are also data at the following spatial resolution, which fully or partly go below municipal level: Grid statistics, preschools and agency and municipal offices, localities and small localities, holiday-home areas, retail trade areas, activities zones, and RegSO (Regional Statistical Areas). For accessing the GIS layers, please see: https://www.scb.se/en/services/open-data-api/open-geodata/. Furthermore, in addition to Statistics Sweden, other agencies also publish similar data (including geospatial data/GIS layers).

[8] The answers were provided by a regional authority related to the Living Lab Ourense.

5.2. Data types and domains available at no costs

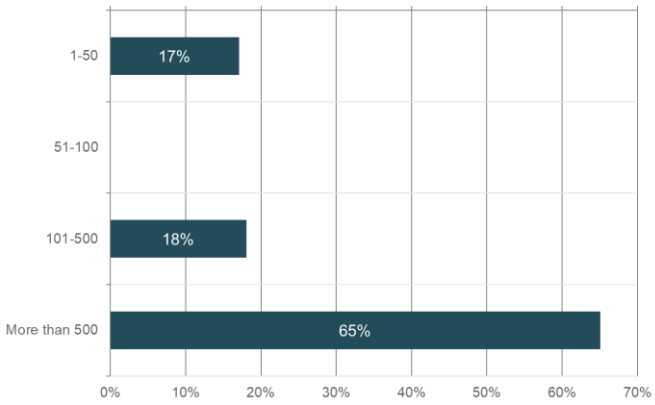
All but one[9] of the authorities and offices that took part in the survey make data openly available (Figure 3).

Figure 3. Is there any data openly available?



The majority of responding authorities and offices make more than 500 data files available free of charge (Figure 4).

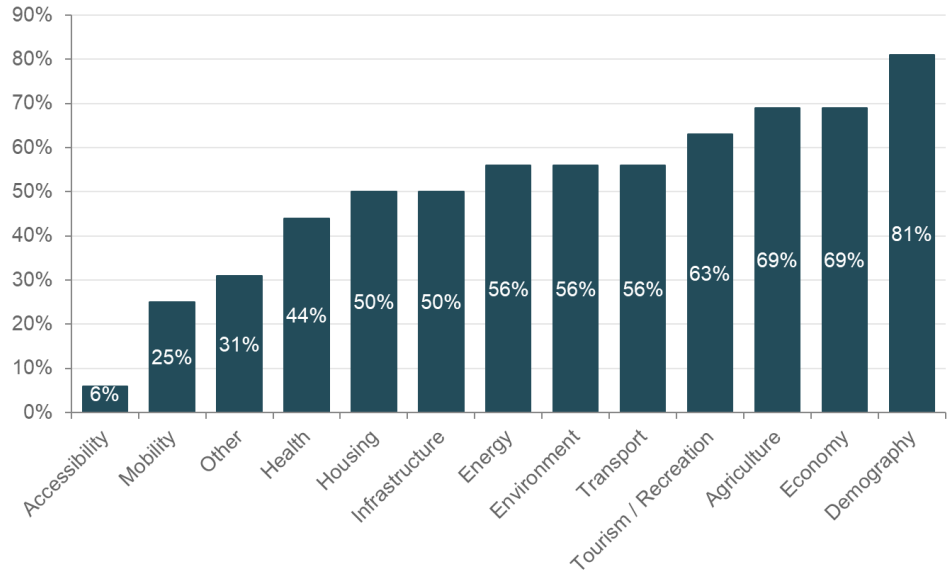
Figure 4. Data files made available by each organization.



The main domains most frequently available and with relevance for rural development and governance are demography, economy, agriculture, and tourism / recreation (Figure 5). More than half of the respondents also provide data on energy environment and transport. Mobility data is provided by only one-fourth of the offices. Accessibility data, according to the survey, is provided by one only country (non-EU).

[9] The respondent is from an Italian authority.Yet, as shown above, the national statistical authority publishes various datasets.

Figure 5. Data domains contained in open database(s).

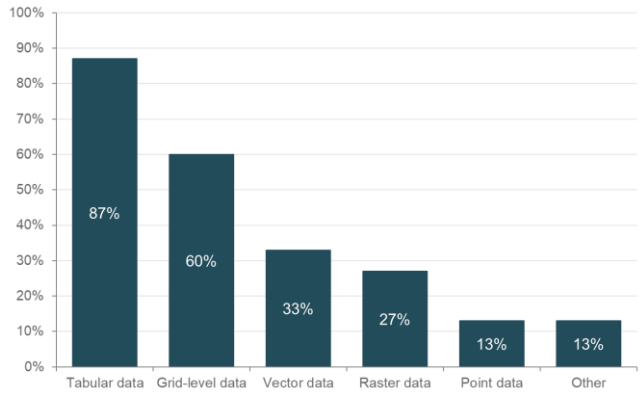


There are differences between the countries. Statistics Finland, for instance, has open databases on all but health data, whilst the Greek national authority has an open database only for R&D and innovation (see also above and according to the survey).

More than 60% of the responding offices use published guidelines/frameworks. In case they do, this is Eurostat (all of them) and UN Statistics Division (two respondents) guidelines.

Most of the offices make tabular data available (87%) and more than half of them grid-level data. More than one-third provide vector data and slightly less also raster data. Only two offices said they make point data available (Figure 6).[10] In addition, one office stressed that they publish time series and anonymized microdata of statistical surveys.

Figure 6. Data types available.

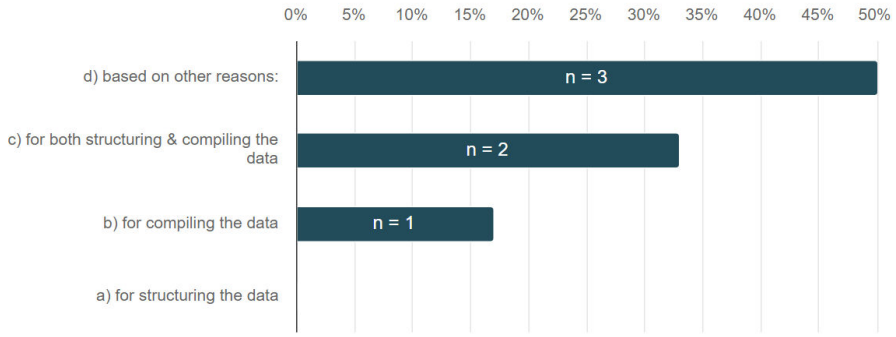


[10] For a distinction of these data types, see for instance https://gisgeography.com/spatial-data-types-vector-raster/.

5.3. Customers, costs and revenues

In addition to whether and what kind of data is accessible for free (see above), we asked the authorities and offices about the reasons for charging customers in case costs apply for users. Unfortunately, only less than half of the respondents provided information on this question. Of those who answered, 1 authority (Greece) charges for compiling data, 2 for both compiling and structuring data (Finland and Italy) and 3 base their costs on other reasons (Croatia, Hungary and Sweden) (Figure 7).

Figure 7. Reasons for charging customers

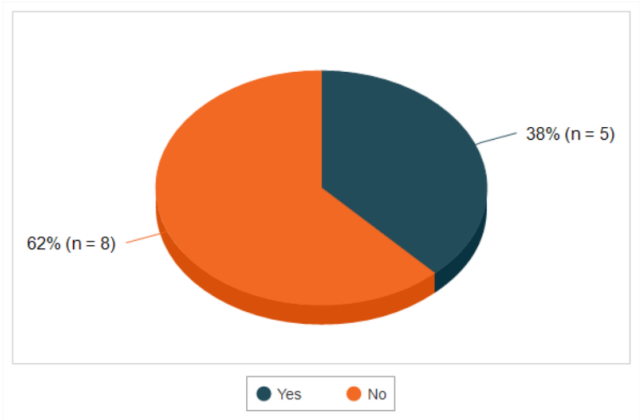


Other reasons range from subscriptions to specific databases, retrievals from internal databases, individual calculations based on the number of man-days required for compiling the data or any other individual tasks on behalf of the office.

We also asked whether public officials must pay for accessing data or whether specific discounts for students or academia are offered.

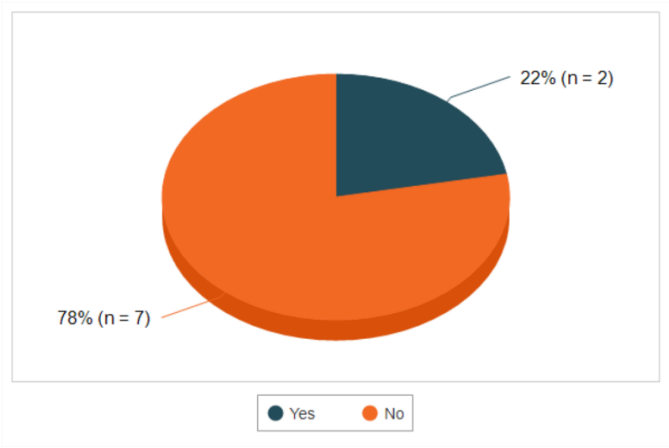
Most of the offices, who answered this question, stated that public officials do not have to pay for accessing data (Figure 8).

Figure 8. Do public officials have to pay for accessing data?



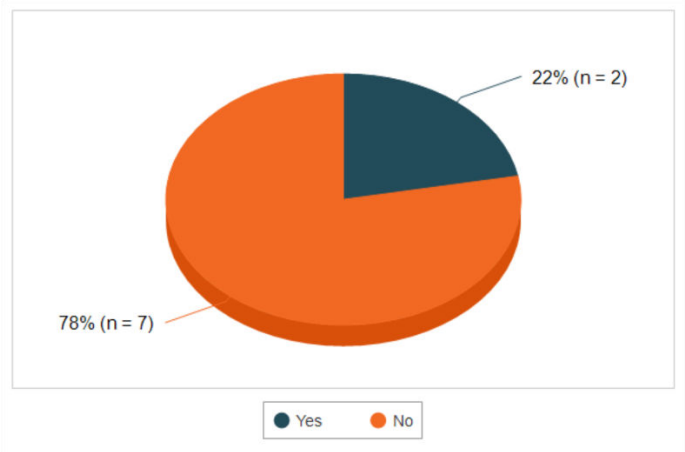
Most of the offices do not offer specific discounts for students (Figure 9).

Figure 9. Specific discounts offered for students?



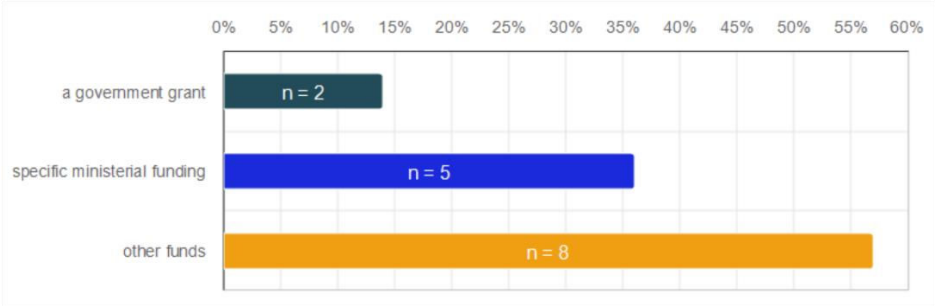
Discounts for academia are also rather an exception (Figure 10).

Figure 10. Specific discounts offered for academia (e.g. teachers, researchers)?



Most of the offices are neither supported by a government grant nor specific ministerial funding (Figure 11). This is in most of the cases coming from the state budget.

Figure 11. Our office is supported by…



Some offices, like Sweden, do commission projects upon request in addition to state funding. Finland also mentioned chargeable services and other central government authorities and financing from the EU.

We asked the respondents, who and which groups are mainly using their data and in a ranking exercise. Main users are individuals, followed by research organizations and private sector (Figure 12).

Figure 12. Main user groups

One respondent added that public sector, research, and private sector are all using their data and are important, as are private individuals. Furthermore, according to one respondent, "*actual usage between the groups might also differ in relation to the size of the groups"*.

Regarding the type of data customers mainly ask for and for what purpose, most organizations referred to demography and population census data. There is also demand for labor market and employment data, R&D, (rural) business development and funding data, as well as price development, tourism and immigration data. As one respondent put it, whilst *"the majority of users probably access data on their own, without asking questions"* and need mainly tabular data, there is a demand for more complex and granular data, as well as indicators to follow up. Respondents also referred to the interest to understand rural-urban interrelations, the need for good data for decision-making purposes and to capture development and change.

6. Inspirational examples – Data, tools and approaches from Granular partner countries.

Section 6 presents 9 inspirational examples, selected to inspire different types of actors to get familiar with interesting work carried out throughout the EU and beyond, and possibly also to pursue a similar exercise. We provide cases from different scales – reaching from the local and Living Lab level to the EU and international levels. We also highlight diverse chains of development. Importantly, we invited developers to share their experience and considerations about motivation, enablers and potential challenges. The examples include reflections about various cost dimensions, and contact information for follow-up conversation. The examples are presented in alphabetical order and following the country / region they are stemming from and going up to the Nordic Region, the EU and the global level. At the end of this section, a matrix will summarize key learnings, cost dimensions, policy implications and future considerations from the inspirational examples.

6.1. Finland: Indicator to follow Subjective Wellbeing (SWB) development of the rural population during the CAP programming period – by Mikko Weckroth (Luke)

The Natural Resources Institute Finland (LUKE) has been actively involved in the planning of Finnish CAP plans as a policy support provided to the Ministry of Agriculture and Forestry (MMM). In earlier work, LUKE has utilized datasets on subjective wellbeing (SWB), which include exact location indicators, enabling detailed spatial analysis. Hence, LUKE was commissioned to develop a measure for evaluating one specific aim in Finnish CAP plan. For that end, LUKE researchers developed a novel indicator to follow development of SWB of the rural population during the CAP programming period 2023-27.

Photo: Edit Kul

Technically, this indicator was created by merging a large (N=38 000) annual FinSote Survey (currently Healthy Finland Survey) collected by the Finnish Institute for Health and Welfare (THL), to 7 classes urban–rural classification system constructed by the Finnish Environment Institute, based on 250 X 250 m statistical squares. SWB of the rural population is measured by the standardized Mental Health Index MHI5 and a single item question on perceived loneliness. More specifically, the index on SWB of the rural population is adjusted by age and gender (by estimated marginal means), in order to control for changes in socio-demographic composition in rural areas. Additionally, the index is adjusted to national mean of SWB to control for national level changes in SWB and thus to reflect only the relative changes of SWB in the rural population. The indices generated from the data are processed and stored in LUKE's database and published in LUKE's indicator portal (https://www.luke.fi/en/statistics/indicators/cap-indicators).

In summary, researchers at THL and LUKE, as well as civil servants from MMM involved in the project, see this as inspiring case and example of cooperation between sectoral research institutes and affiliated ministries. However, the institutional structure that allows data availability is rather fragile, since it is based on double affiliation of single researcher. The project represents an inspiring pilot on using a large SWB dataset in CAP policy evaluation.

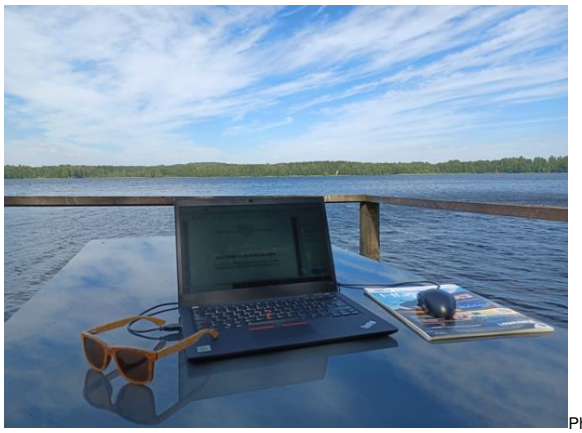


Photo: Janne Poikolainen

Costs

Development Costs: The responsible investigator for this project at LUKE (Mikko Weckroth) has an affiliation with THL. The survey data itself is collected by THL and budgeted on the basis of the agreement with the responsible Ministry of Social Affairs and Health. However, in a case there would not be a researcher with shared affiliation, the whole procedure would require involvement of the Finnish Social and Health Data Permit Authority (Findata) that grants permits for the secondary use of social and health care data (https://findata.fi/en/). The pricing for creating this dataset and giving permits would need to be estimated by Findata but would most likely be 8.000 – 12.000 €. In sum, the survey data is collected by THL - not for the purpose of developing this indicator but was utilized for its development.

Data accessibility for users: Indexes are free at Agrigaattori portal for users.

Data infrastructure costs: None for the user.

Data governance & management costs: None for the user. However, for the development of the indicator, expertise in statistics and data management, particularly in the field of (subjective) wellbeing is needed. Data analysis expertise is needed, too, e.g. calculating indexes and linking responses to urban-rural classification.

Further information: mikko.weckroth@luke.fi

6.2. Finland: The Rural Barometer – by Hilkka Vihinen & Michael Kull (LUKE)

The aim of the Rural Barometer is to shed light on how Finnish citizens, public decision-makers, business decision-makers, the media and rural experts perceive the countryside. It includes themes such as: elements of the good life, images of rural areas, regional identity, multi-local living, entrepreneurship and livelihoods, rural development, opinions on policy measures and the future of rural areas. The Rural Barometer provides a statistically representative sample of the Finnish people's views on the state and future of rural areas. The 4th Rural Barometer is currently conducted and continues the series of Rural Barometers done in 2009, 2013 (published 2014) and 2020. The Barometer is commissioned by The Rural Policy Council (MANE) and implemented by LUKE and subcontractors, e.g., the survey is designed and interpreted by LUKE.


Photo: Janne Poikolainen

The Rural Barometer 2020 – Approach

Responses were collected through an online survey (Finnish-speaking population) and a telephone-informed survey (Swedish-speaking population). Overall, there were 1788 respondents, that is Finnish citizens aged 15-79.[11]

The Rural Barometer 2020 – Selected Findings

"A place for the good life" - The image of a good life is much or very much associated with the countryside by 61% of Finns. More positive images are associated with rural than urban areas.

"I am both rural and urban" - 37% of Finns have a double-identity: they consider themselves both rural and urban.

"The countryside is the land of dreams" - 20% of young urban dwellers consider the countryside to be the place of their dreams.

"Rural development should be based primarily on the needs of rural people." - 39% of Finns strongly agree with this statement.

Photo: Michael Kull

Costs

Development costs: In the beginning, when the whole survey is drafted, about 6 PM (senior experts) are needed, depending on how wide-ranging the survey is supposed to be as to different dimensions, and how many respondent groups there will be. After the first run about 3 PM should be enough for updating / improving questions and lists of specific respondents.

Data Governance & Management costs: Senior experts on rural studies for the substance, and if possible, professional survey companies to run the citizen sample with the help of their permanent panels. In Finland, the survey (covering two languages and the above-mentioned different respondent groups) costs about € 42-45,000. For one language only it would be €30-35,000.

Data analysis, processing & visualization: Planning work and interpretation done by a rural studies experts; in our case about 5 PM. Much depends on how many "mobile" parts the survey has, and on the quality of reporting. If PPT presentations and newsletters suffice, it is possible to do it with 1-1,5 PM. A proper scientific report needs 3-4 PM.

Data Infrastructure costs: It would be possible to conduct the survey using, for instance, Webropol, which could be (almost) free of charge. However, to guarantee statistically representative samples and quality, professional companies with their existing software and hardware might be used.

Data repositories / storage: We use the Finnish social science data archive (FSD) (https://www.tuni.fi/en/research/finnish-social-science-data-archive), a national infrastructure with open access and a part of CESSDA (PAN-EUROPEAN RESEARCH INFRASTRUCTURE). FSD is funded by the Ministry of Education and Culture, and it does not cost us to preserve the barometer data there.

Costs for data documentation and assistance in data use: Data is stored in FSD. In most cases, policy-makers or the media use the interpretations and presentations we have made available in the Barometer report or on the website of MANE. Depending on quality and quantity, costs vary between 1 to 6 PM.

[11] More specifically, 294 young urban citizens, aged 15-29) took part. 579 respondents were from the Swedish-speaking population (aged 15-79). Respondents were also business decision-makers, e.g. CEOs, development managers, CFOs and others in similar positions (industry, construction, trade and services (280 respondents) and public decision-makers, e.g. MPs, mayors and presidents of city/commune councils, provincial governors and councillors, MEPs (237 respondents). Sixty-eight media representatives, incl. editors-in-chief, news managers, regional editors, managing editors, political journalists, journalists (newspapers, TV, radio) took part. 121 "Rural experts", i.e. experts, researchers, civil servants, third sector actors working on and following rural issues took part as well.

Data security costs: Included in the costs of the survey company and in our own budget. Data security costs have increased a lot during the last 10-15 years, so it is wise to include them in the budget and PM.

The Barometer requires, by definition, repetition. 2–3-year intervals would be optimal. It is also important to design the survey really well in the beginning, since the value of the Barometer decreases each time you change something. in Finland, we have decided to have approximately one "mobile" section among the 6 themes of the barometer – a set of questions that changes depending on what is regarded of current, urgent interest each time.

In addition, it is warmly recommended to use experts on qualitative attitude research at least in the beginning, when designing the survey. It would also be great if we could, someday, have several European Rural Barometers, where at least some of the sections would be formulated exactly the same way, to render comparisons possible.

Further information: hilkka.vihinen@luke.fi & michael.kull@luke.fi

6.3. France: Monitoring mobility and road traffic at local scale – Case *Conseil Départemental des Pyrénées-Orientales (CD66)* – by Louise Chasset, Lenaïc Depontailler (Pays Pyrénées Méditerranée) & Jean-Claude Balagué (Département des Pyrénées-Orientales)

The *Département des Pyrénées Orientales (CD66)* has set up a system for observing and monitoring traffic levels on the roads it manages. As a local authority in the South of France, the CD66 is responsible for a network of 2,154 km of roads, serving an area of over 4,000 km² with a population of 480,000 (source: INSEE 2020).

Photo: Pays Pyrénées Méditerranée

To help manage its road network, CD66 produces this data with several objectives in mind:

- *To understand traffic trends and help decision-making for long term general studies,*
- *Design road projects, in particular for sizing pavement structures,*
- *Adapt measures taken in terms of worksite operations, for example by choosing the time slots with the least impact on traffic during roadside mowing operations.*

CD66 is responsible for all the stages involved in producing this data: from the installation and maintenance of the meters to the collection, consolidation, and analysis of the data, right through to making the information available on open data platforms. Two operators and one data administrator are working on this system, which relies on a network of permanent, rotating and on-demand meters, as well as the ROUTE+ software and a virtual server for data storage.

Photo: Pays Pyrénées Méditerranée

Costs

BUDGET OF THE WHOLE INITIATIVE: A total of around €165,000 per year in operating costs for the whole the department. This include human resources (€100,000/year), costs of replacing and maintaining meters (€ 62,000 / year) including ROUTE+ software maintenance (€3,300 / year) and costs for acquisition and training (€ 30,000).

DATA GOVERNANCE & MANAGEMENT COSTS

Staff costs or PM: around €100,000 per year.

Costs for Data analysis: The processing and display of the data is carried out in house by the department's civil servants, whose salary costs are aggregated.

Consolidation of collected data: the software draws attention to aberrant results, on the basis of which the operators and the data administrator check it. The director of the roads department finally validates the data. Data cleansing is carried out internally by the department's civil servants, whose salary costs are globalised.

DATA INFRASTRUCTURE COSTS:

SOFTWARE. The system is based on the ROUTE+ which costed around €30,000 to acquire (+ agent training).

Annual maintenance costs €3,300 excluding VAT/year.

HARDWARE. The system is based on meters, the hardware replacement and maintenance cost is approximately €62,000/year.

Data repositories / storage: The data is stored on one of the 700 virtual servers of the Conseil Départemental des Pyrénées-Orientales.

COSTS FOR DATA PROCESSING & VISUALIZATION: The processing and display of the data is carried out in house by the department's civil servants, whose salary costs are aggregated. Processing the data and then making it available on the open data platforms takes around 12 working days.

DATA SECURITY COSTS: The data is stored on one of the CD66's virtual servers, protected by a firewall. Data access is allocated manually. Costs are therefore considered negligible today.

Further information: Jean-Claude Balagué, Département des Pyrénées-Orientales, jeanclaude.balague@cd66.fr.

6.4. Galicia – Spain: Telecare for the elderly at home in the rural areas of Ourense – by María Isabel Doval Ruiz & Breixo Martins (University of Vigo)

Photo: GRANULAR team University of Vigo

Home telecare is an uninterrupted telephone service with specific communications and computer equipment, especially designed for elderly people, who live alone, and in order to pay immediate attention in the event of an emergency. The system consists of an alarm unit carried by the person, a telephone terminal and a computerized switchboard that receives the calls and is located in the Care Centre. The service allows users to communicate with the center, which is staffed by specialized personnel, in the event of any emergency situation by simply pressing the button on the device elderly people carry with them. This system is specific to rural local authorities (under 20,000 inhabitants). It is noteworthy that only one local entity, out of a total of 92, exceeds this number of inhabitants.

The service aims to provide an immediate response that allows communication between the user and the care center 24 hours a day, 365 days a year. This offer is complemented by a diary-reminder service for certain tasks, such as the control of chronic medication or medical and social consultations. Other systems related to home automation and care for the elderly are also being implemented. An example of this is the system for detecting the opening of the doors of these people. Through a centralized system, it is possible to know the number of times a person opens their door per day and if the opening range falls below a certain number, specific protocols are activated.

The extracted data

The background that allows for a strong territorial analysis of this health care or welfare system is the continuous internal system of data collection. The data collection has both a statistical and a territorial level. On the one hand, a large amount of information is collected for each procedure performed by each user. Depending on the year, the number of users varies between 2000 and 3000 people, and the amount of information is large. Every time the user presses the emergency button, the duration of the call and even, through a qualitative system, information on the reasons for each of the emergencies can be collected. This analysis has led to preliminary internal conclusions that one of the main causes of emergencies is motivated by unwanted loneliness in the last stage of life.

On the other hand, perhaps the most innovative information collection is implemented from a territorial point of view. This is due to the geolocation of each of their users. In other words, there is the possibility of generating maps of user points which, together with the above statistics and the geolocation of the service centers, can generate very useful cartographic information systems. For example, by combining the number of emergency vehicles and users, we can create maps of the flow of emergency vehicles throughout the province with wide-ranging densities that allow us to draw relevant conclusions. On the other hand, it would allow us to obtain information on the distortion between the location of services and users, areas of service reinforcement, areas of social exclusion or even distortions of services in relation to the degree of urbanization or rural-urban areas.

Visualization and opening of data

As far as the visualization system is concerned, it is difficult to implement due to the high degree of protection of the data processed. While it would be possible to display purely statistical data in anonymized form, this would be very complex in terms of the territorial representation of the data. This is due to the fact that all mobility and service-user distortion flows are based on the exact location of each user and therefore, the very location of the households

eliminates the anonymity of the users. In any case, our Living Lab is starting to collect this data in order to analyze it and then, if possible, develop appropriate reports or visualization systems.

Costs

The estimated budget of the whole initiative is complex as, in practice, it depends on different administrations. Potential travel is carried out by regional institutions and the telecare program is funded by the provincial authority. Similarly, data analysis is carried out by the university in collaboration with the provincial administration. In any case, on the basis of the call for tenders in 2021, the contract is estimated to be at around €1,300,000.

Further information: granular@uvigo.es & mdoval@uvigo.gal

6.5. Poland: Functional and spatial diagnosis for social revitalization at the local (municipal) level – by Agnieszka Kurdys-Kujawska (TU Koszalin)

Employees of the Faculty of Economics at KUT (Małgorzata Czerwińska-Jaśkiewicz, Ph.D.; Patrycjusz Zarębski, Ph.D.), collaborated with the Science for the Environment Foundation in their research on the creation of a revitalization program in the communes of the West Pomeranian Voivodeship located in the Special Exclusion Zone. The primary objective of this investigation was to identify a degraded area through an objective and comprehensive assessment of social, economic, technical, environmental, and spatial issues. Based on the research, a model (including diagnostic instruments) for socio-economic diagnosis was created, and a model for implementing social revitalization was proposed.

To assess degraded areas in the commune, 22 indicators were used to describe the intensity of the phenomenon within a given commune. These indicators were grouped according to their nature and information content into social, economic, structural, spatial-functional, and environmental indicators. Data were obtained from public statistical portals, including the Local Data Bank of the Central Statistical Office, the Commune Office, the District Labor Office, and the Commune Social Welfare Center. The indicators adopted for evaluation were also optional, which was helpful in considering the specific features of a given area, and the internal diversity of the analyzed analytical units. The nature of these indicators was also determined. The indicators describing the phenomena were assessed according to the principle that they indicate a crisis when their value exceeds the median for the entire set of analyzed units. On this basis, a critical indicator was built. It is the basis for defining a degraded area, i.e. in a state of crisis due to the concentration of negative social phenomena and negative phenomena of a different nature, i.e. economic or environmental, spatial-functional or technical.

Problem phenomena and the causes of their occurrence were identified through an in-depth diagnosis based on direct interviews, questionnaires, and focus group interviews. Revitalization areas were designated based on an in-depth diagnosis using four research methods: individual direct survey - with village heads in the commune; IDI – Individual In-Depth Interview, i.e. an individual in-depth interview – with people in power in the commune (mayor); FGI (focus group Interview) – with local leaders in the commune; interviews, animations, and research walks with residents of villages. Field research was carried out by appropriately selected and instructed interviewers - mostly representatives of commune residents. These were people selected in terms of competencies, skills, and personal predispositions, with good knowledge of the local environment.

The high usefulness of statistical data was ensured by the inclusion of local animators and independent mediators in the research process, who, on the one hand, were responsible for social activation in a specific region on an ongoing basis (and knew it well), and, on the other hand, were independent experts looking at a given area objectively. The designation of revitalization areas took into account the participation of residents and local leaders. Statistical diagnosis was made based on a set of 14 independent indicators. The revitalization area was designated in places where, as a result of statistical analysis, an accumulation of problem phenomena was observed. In a further stage of research, the development potential and the most urgent areas of intervention were considered. For this purpose, various tools and methods of collecting information were used: interviews (based on a survey questionnaire); direct animation meetings with various stakeholders; preparation of maps of resources and needs directly by residents with the support of the Local Animator, as well as tutors and external specialists; focus meetings.

Source: https://www.facebook.com/ndsfund/?locale=pl_PL.

The model for diagnosing socio-economic structures and the original concept of implementing social revitalization in rural areas were used to develop 18 Local Revitalization Programs (e.g. https://bip.dobragmina.pl/strony/menu/139.dhtml)

108 investment projects were implemented in municipalities based on Local Revitalization Programs with the total amount of PLN 6,000,955 (appr. 1,390,000 €). 720 residents (direct participants of revitalization projects) from 18 communes of the Special Inclusion Zone (including 490 people at risk of social exclusion) were covered by substantive and animation support in the field of revitalization. Three projects used the diagnostic research methodology to delimit degraded areas and implemented the social revitalization model.[12]

Revitalization projects were implemented in 2016-2021 by the Koszalin Regional Development Agency S.A., the Koszalin Science for the Environment Foundation, Aktywa Plus, and 4C from Szczecin. These projects were supported by funds from the European Social Fund and the state budget under the Regional Operational Program of the West Pomeranian Voivodeship for 2014-2020 and in communes located in the Special Exclusion Zone (based on the developed Local Revitalization Programs).[13]

Source: https://www.facebook.com/ndsfund/?locale=pl_PL.

Costs:

Data management and management costs: None for the user.

[12] Project No. 1. "Revitalization in the communes of the West Pomeranian Voivodeship located in the Special Inclusion Zone", project co-financed by the EU from the European Social Fund and the state budget under the Regional Operational Program of the West Pomeranian Voivodeship for 2014-2020, contract number: UDARPZP.07.01. 0032K601/1600 of September 21, 2016, implementation date: October 1, 2016 - December 31, 2016; Project No. 2. "Social Revitalization", a project financed by the EU from the European Social Fund and the state budget under the Regional Operational Program of the West Pomeranian Voivodeship for 2014-2020, Co-financing agreement No.: RPZP.07.01.0032K102/1800 of September 20, 2018. implementation date: 01/01/2019 -01/01/2021.

[13] See for instance https://nenp.facebook.com/spolecznarewitalizacja/videos/gmina-brojce-rewitalizacja-spo%C5%82eczna/250991843084847/; https://ne-np.facebook.com/spolecznarewitalizacja/videos/gmina-radowo-ma%C5%82e-spo%C5%82eczna-rewitalizacja/4262501857190377.

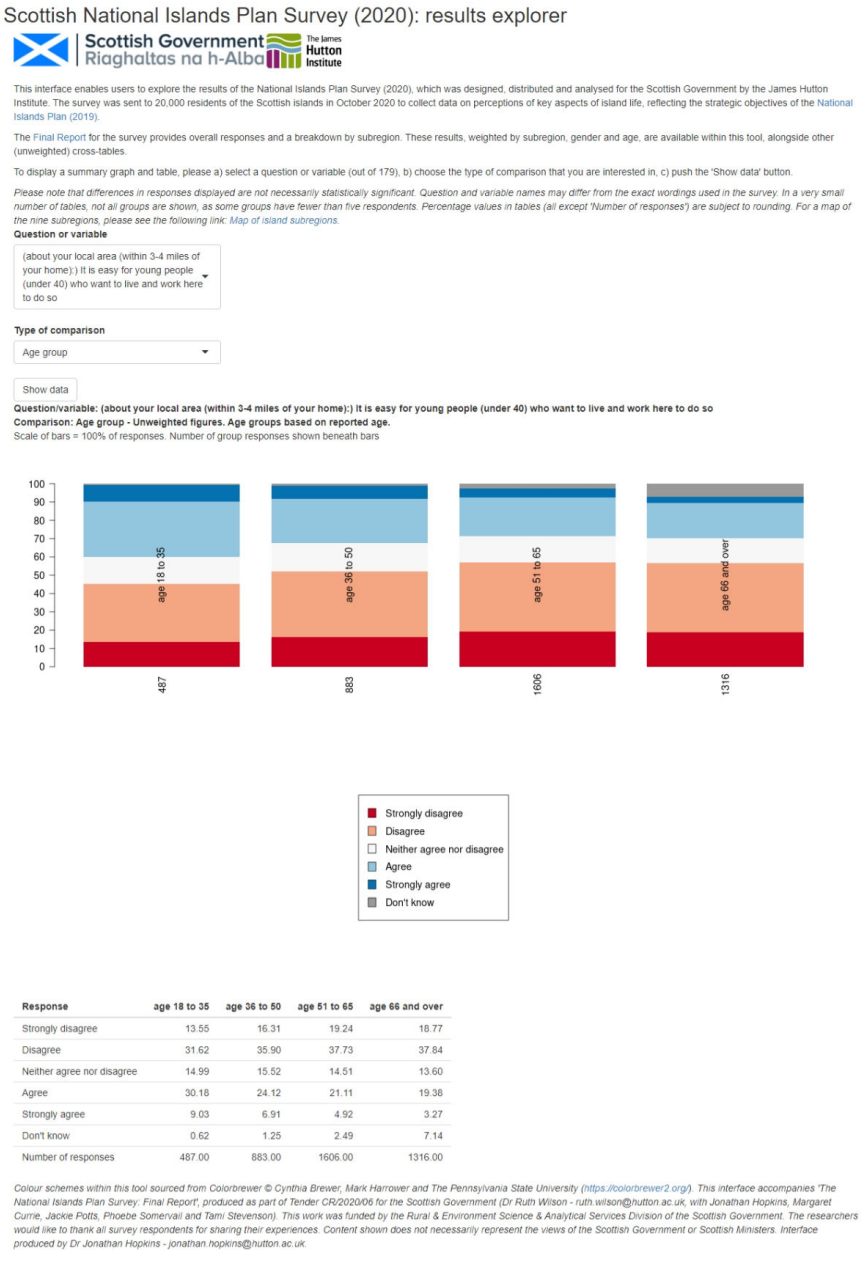
Indicator development: specialist knowledge in statistics and data management was needed, especially in the field of social, economic, spatial-functional, technical, and environmental phenomena in rural areas. Knowledge of data analysis is also necessary.

Further information: agnieszka.kurdys-kujawska@tu.koszalin.pl.

6.6. Scotland: Scottish National Islands Plan Survey (2020): results explorer – by Jonathan Hopkins and David Miller (Hutton)

In 2020, researchers at the James Hutton Institute were contacted by the Scottish Government to design, implement and analyze the National Islands Plan Survey, which collected data from residents of Scotland's islands on perceptions and opinions on island life, which were aligned with strategic objectives of the National Islands Plan. The survey implemented a customized subregional geography to inform both survey sampling and the reporting of results, in order to identify diversity in lived experiences beneath the level of local authority regions; these regions have subsequently been developed into an official geography for Scotland through further work by National Records of Scotland and the Scottish Government. The survey achieved 4,347 responses from 59 islands. However, in order to expand the volume of results reported, a simple interactive tool was published, enabling end users to generate graphs and data summaries for 179 variables and tabulations with geographical, demographic and economic characteristics of respondents. The tool demonstrates holistic experiences of island life and illustrates variations in these between people and regions, and highlights the extended evidence base for a type of geography for which there was limited data available.

The tool was published in 2021 and was developed as part of a short research project, using the Shiny package within R. Cross-tables and weighted results summaries were pre-generated for the tool, with reproducible code produced for their calculation. Some results showing figures for island regions, and the islands overall, were weighted by island region, age group and gender cohorts in order to account for differences in response rates. Data provided to the tool were screened, with variables and values redacted in places, to avoid the disclosure of data for low numbers of people.



Screenshot of the tool, showing a comparison of perceptions by age group.

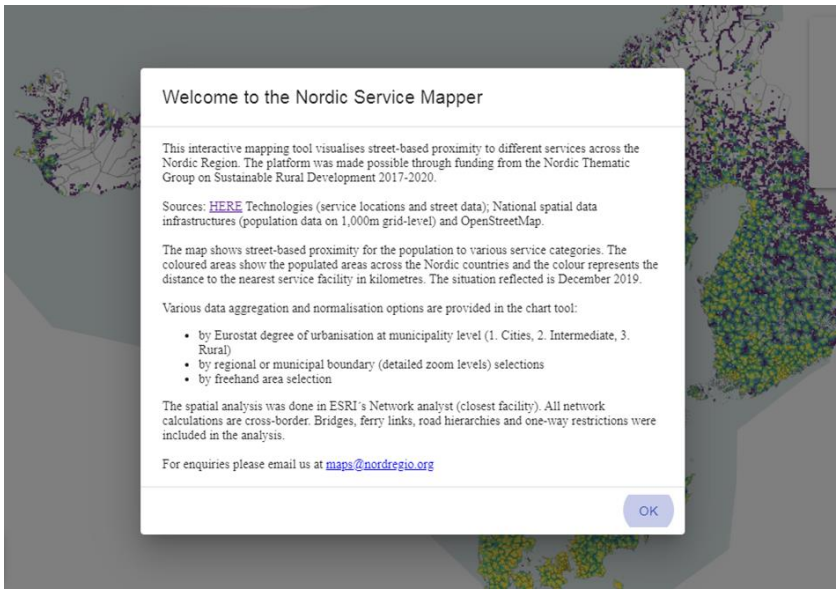
Costs

The project involved a team of six researchers with one leading on the tool development and publication. The tool utilised existing skills in R/Shiny programming. Data hosting costs for the tool are currently zero, through a free account at shinyapps.io, although a subscription was taken out for that account previously. The value of the whole contract for the Survey was approximately €70,000.

Further information: jonathan.hopkins@hutton.ac.uk

6.7. Nordic Countries: The Nordic Service Mapper - by Mats Stjernberg (Nordregio)

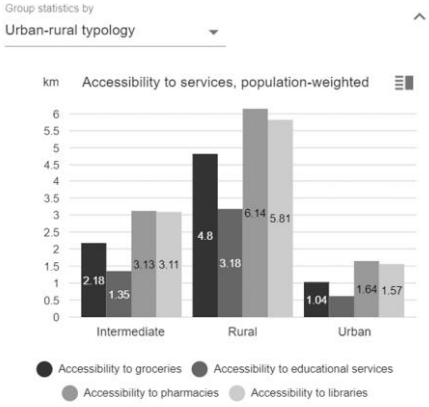
Nordic Service Mapper is an interactive online web-mapping tool that visualizes the proximity to different services across the Nordic Region. It covers the territories of Denmark, Finland, Iceland, Norway, and Sweden as well as the Faroe Islands and Åland. The tool includes four different types of services, namely grocery stores, pharmacies, libraries, and schools. The tool was published in 2021 and it reflects the situation in December 2019.



Landing page of Nordic Service Mapper which can be access at http://nordicservicemapper.org/

The tool shows street-based proximity for the population to various service categories at different geographical levels. Various data aggregation and normalisation options are provided in the tool, including access to services at regional and municipality level, according to the Eurostat Degree of Urbanisation at municipality level (cities, intermediate, rural), and at a more spatially detailed grid-level and based on a freehand selection of areas.

Nordic Service Mapper also includes an infographic tool which allows for visualising service accessibility according to selected territories and service types in the form of charts.



Visualization of accessibility to different service types in Finnish municipalities based on the Eurostat Degree of Urbanization classification in Nordic Service Mapper's infographic tool.

The calculations are based on spatial analysis carried out in ESRI´s Network Analyst (closest facility). Bridges, ferry links, road hierarchies and one-way restrictions were considered and included in the calculations. All network calculations have been done without considering national borders, which also enables for cross-border analysis. The data sources used include data on service locations which were purchased from HERE Technologies, population data on 1,000m × 1,000m grid-level and road network data from OpenStreetMap.

Zooming in on the accessibility to groceries across a cross-border area in northern Finland and Sweden

Costs

The Nordic Service Mapper tool was developed by Nordregio in collaboration with Ubigu. The tool was created by commission of the Nordic Thematic Group on Sustainable Rural Development (2017–2020) as part of the project Regional disparities and the geography of service within the Nordic countries. Mats Stjernberg and Oskar Penje at Nordregio led the project and the accessibility calculations were carried out in-house at Nordregio. The work to create the web-mapping platform was led by Ubigu.

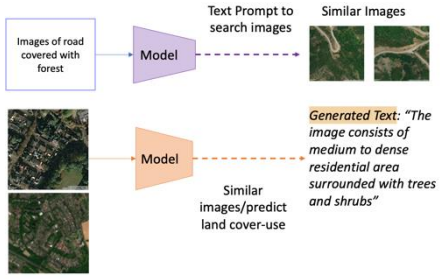
Data costs: The main costs were the costs for purchasing data on service points from Here technologies as well as for working hours for carrying out the accessibility calculations and developing the web-mapping platform. The costs for purchasing the service point data was around € 12.000.

Data analysis, processing & visualization: The number of PM for the carrying out the calculations and data harmonisation including different methodological considerations was around four months. The development of the web mapping platform also required around four months of work.

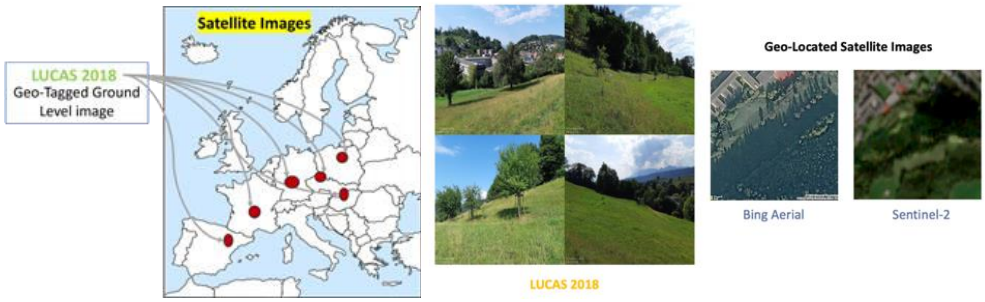
Further information: mats.stjernberg@nordregio.org

6.8. EU: Integrating text-to-image and image-to-text techniques to enhance accessibility and understanding of rural land-use data - Cross Modality Framework – by Pallavi Jain (IAMM)

The tool integrates text-to-image and image-to-text techniques to enhance accessibility and understanding of rural land-use data. It enables efficient retrieval, association, and analysis of specific land use categories or features within satellite imagery using textual queries or visual analysis.



The tool currently utilizes sophisticated vision-language models to link ground-level images with satellite imagery, providing crucial context for understanding rural land use dynamics. It utilizes geolocations from LUCAS survey data to gather satellite imagery from Bing and Sentinel-2 sources.



Further, the textual modelling aspect enables the implementation of Text-to-Image and Image-to-Text search methods, establishing essential connections between textual descriptions and visual representations. These methodologies enable seamless interaction between textual and image data domains, enhancing the project's comprehensive approach.

Through the comprehensive insights gained from the visual-language approach, the tool aims to empower policymakers, researchers, and land managers. This support facilitates well-informed decisions in land use planning, conservation strategies, and resource management. The ultimate goal is to leverage these advanced models to effectively understand, monitor, and manage rural land use.

Costs

Data Collection: The project initially uses LUCAS 2018 survey data to acquire ground-level visual context across Europe. Bing Aerial (15cm resolution) and Sentinel-2 (10m & 20m resolution) data are collected from Microsoft Bing and Planetary Computer API services, amounting to 230,000 location images obtained so far.

Hardware Infrastructure: For model training, four GPUs are utilised to support the computational requirements. The estimated cost for the infrastructure is 8,000€

Data Costs: As of now, Bing and Planetary Computer services from Microsoft offer free APIs for data collection. However, downloading 230K images takes nearly a week.

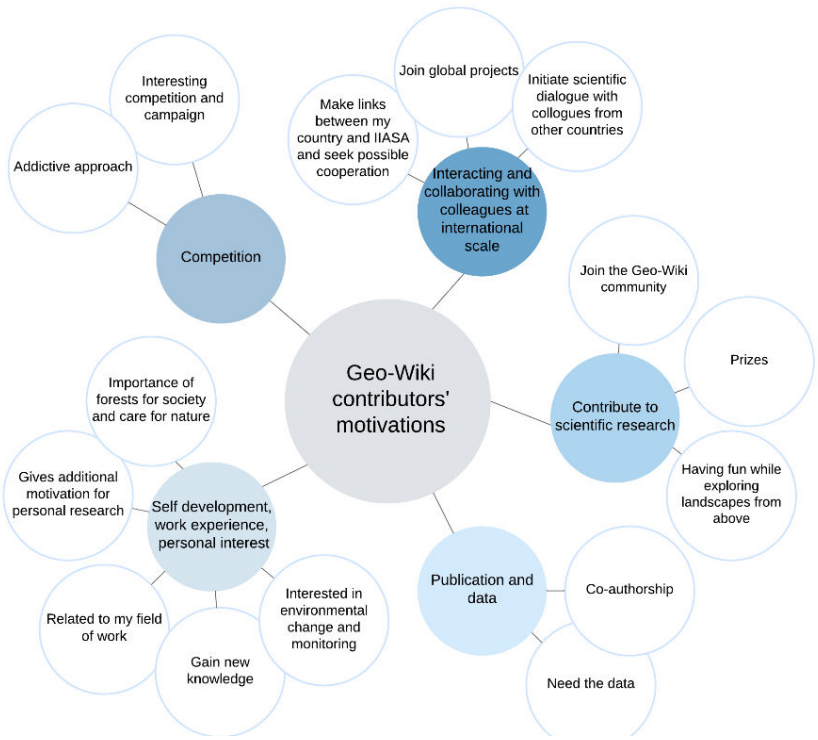
Tool Cost: The development of the algorithm is estimated at 24PM. The foundational expenses for the tool would encompass creating a user interface, an API, and backend infrastructure, along with utilising online cloud servers like AWS, Google Cloud, or Microsoft Azure for model deployment. Additionally, there might be additional API fees associated with utilizing satellite image retrieval services.

Further information: jain@iamm.fr.

6.9. Global: Geo-Wiki Earth Observation & Citizen Science – by Ivelina Georgieva (IIASA)

The Geo-Wiki platform provides anyone with the means to engage in monitoring of the Earth's surface by classifying satellite, drone or ground-level imagery. Data can be input via desktop or mobile devices, with campaigns and games used to incentivize input. These innovative techniques have been used to successfully integrate citizen-derived data sources with expert and authoritative data to address pressing policy-related questions (e.g. European environmental policy, SDG indicators and more).

Geo-Wiki was established in 2010 in the Novel Data Ecosystems for Sustainability research group, part of the Advancing Systems Analysis Program at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria. Since its inception, Geo-Wiki has grown rapidly, with currently over 22,000 registered users having contributed more than 18 million image classifications from around the world. Furthermore, the Geo-Wiki toolbox has expanded to include numerous applications which help to address a variety of global challenges (e.g., land use change, food security, pollution and more).



Motivations of contributors joining our citizen scientists campaigns.

Since its creation, multiple citizen science campaigns in the form of competitions have been carried out, asking volunteers to perform visual interpretation of VHR satellite imagery in the Geo-Wiki platform on topics related to land use and land cover changes. Some of the recent campaigns include defining the drivers of tropical forest loss, validating the global human settlement layer, and defining the human impact on forests. Within these campaigns we have involved hundreds of volunteers from more than 20 countries worldwide who had interest to contribute to science and become part of the growing Geo-Wiki community. The figure above illustrates their motivations.

To facilitate the process of visual interpretation, volunteers have access to specifically defined (for the validation task at hand) Geo-Wiki functionalities. The principle set of

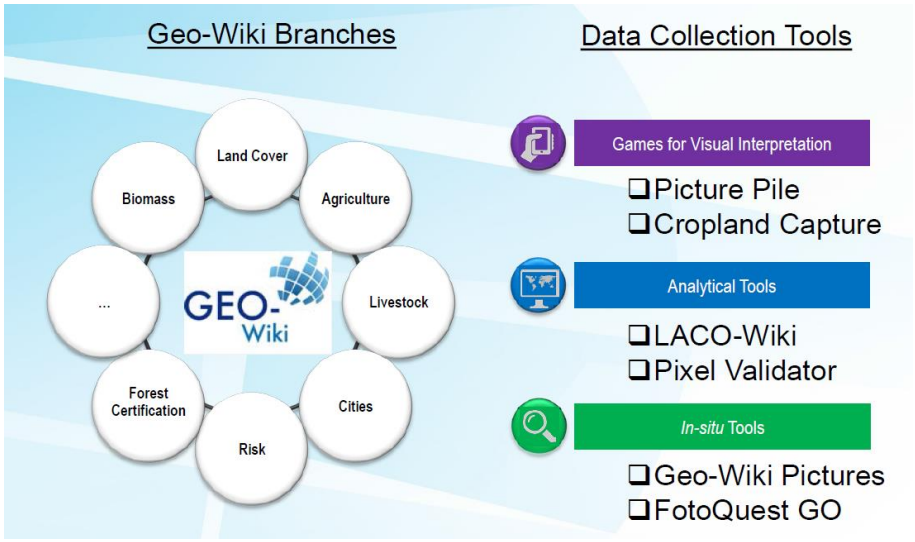
tools which the platform includes are implemented features like Sentinel Hub time series imagery, an NDVI tool for measuring the Normalized Difference Vegetation Index and Google Earth history imagery. The quality of contributions has been controlled from a group of experts during and after the campaigns and scientific publications are used to share the data, which are later uploaded in public repositories to ensure transperancy of the entire process.

Main development costs:

Initially, an early beta-version of Geo-Wiki was developed in partnership with the University of Freiburg, Germany and the University of Wiener Neustadt, Austria, as part of the Geobene project. The development of the first, beta-version lasted 6 months and the further development of the Geo-Wiki v.1. required around 1,5 person years of work. There have been ongoing feature developments and improvements since 2011. In 2013 the first big refactoring of the Geo-Wiki platform happened, which lasted approximately 10 months. This included database refactoring, as well as moving to newer technologies and frameworks. Furthermore, we moved to using a content management system (CMS) for e.g., the user management.

Software costs:

No software costs (we are using only open source/free software)



Overview of the Geo-Wiki branches and data collection tools

Hosting costs: Costs for servers and services needed for hosting the Geo-Wiki system from 2011 until now:
total: 35.000€ (Nov 2011 - Oct 2023)
Initial costs in Nov. 2011: 70€/month.
Current costs in Oct. 2023: 500€/month (the monthly hosting costs increased because of more collection tools & platforms & more servers and services were needed).

A screenshot of the Geo-Wiki application branch for measuring human impact on tropical forests.

Further information: georgieva@iiasa.ac.at & mccallum@iiasa.ac.at

6.10. Key lessons learned, cost considerations, policy implications and ways forward

This section summarizes the inspirational examples presented above. In the matrix below (table 2), you can find each case and key lessons learned. We highlight also the expertise needed to conduct such an exercise and what types of costs to consider. In addition to policy implications of each case we also try to consider what could be the next steps, both in relation to the case and as outlined by its authors but also in terms of potential replication elsewhere.

Table 2. Inspirational examples: Lessons learned, cost considerations, policy implications and ways forward

Case	*Key lessons learned*	*Expertise needed*	*Cost considerations*	*Policy Implications*	*Future considerations*
Indicator to monitor the subjective wellbeing of the rural population during the CAP programming period - Finland	Effective collaboration, innovative indicator development, and a commitment to transparency.	Expertise in statistics, data management, and data analysis, especially in the field of SWB, for successful indicator development.	Development costs linked to the shared affiliation of the key researcher, highlighting the importance of such affiliations in avoiding additional expenses. Finnish Institute for Health and Welfare (THL) collects data for other purposes. Potential additional costs ranging from 8,000 to 12,000 € if a researcher with shared affiliation was unavailable, requiring permits from the Finnish Social and Health Data Permit Authority (Findata).	The case will contribute valuable insights for public policy by using large SWB datasets for policy evaluation in the context of the CAP programming period.	Work on SWB in Granular will continue. Exchange with other national authorities in charge of the CAP welcome. (Fragility) of the institutional structure and the need for specific expertise are identified as important considerations for future initiatives.
Rural Barometer Finland	Careful planning, expertise, and consideration of various disciplines contribute to the success and sustainability of such surveys.	(Senior) experts in the field of rural studies providing subject matter expertise to ensure the survey content aligns with the objectives. Experts in qualitative & quantitative research needed, too.	Development expenses of approximately 9 PM for initial drafting and updates, data governance and management costs ranging from €30,000 to €45,000 for survey execution, along with additional expenses for data analysis, infrastructure, documentation, and security.	Barometers may guide policymakers in allocating resources to address specific challenges identified in rural areas. The emphasis on data infrastructure, storage, and documentation underscores the importance of a data-driven approach to governance. The Barometer's use of citizen samples and the involvement of different respondent groups can encourage public participation in policymaking.	Continuation of the Barometer in a 2–3 year interval. Eventual development of a European Rural Barometer / a Barometer elsewhere with standardized / harmonized sections for comparability might be worth to exchange about.

Monitoring mobility and road traffic at local scale in France	Well-planned and integrated approach to traffic monitoring, combining technology, human resources, and governance structures to effectively manage road networks and contribute to informed decision-making in planning.	Combination of various expertise incl. traffic engineering, data management, geospatial analysis and planning.	A total of around €165,000 per year in operating costs. Key cost dimensions are categorized into several areas. The largest portion of the budget is for human resources (€100,000 p.a.), covering salaries for operators, data administrators, and civil servants involved in data analysis, consolidation, validation, processing, and visualization. Software costed appr. €30,000 (+agent training). Annual maintenance costs €3,300 excl. VAT.	Wide-ranging implications for public policy, impacting areas such as transportation, regional planning, resource allocation, and environmental sustainability, through better understanding traffic trends and through flexible and adaptive approach to monitoring traffic levels.	Data generated by the system can guide future road design, adapting worksite operations and infrastructure investments by identifying areas with high traffic volumes or congestion, influencing decisions on road expansions or alternative transportation solutions.
Telecare for the elderly at home – Galicia / Spain	Implementing telecare programs for elderly individuals living alone, especially in rural areas is multifaceted. The importance of a holistic and data-driven approach to improve the well-being of the target population is emphasized.	A diverse set of expertise incl. telecommunications, healthcare, data science and analytics able to address the technological infrastructure, data analysis, user interface design, community acceptance. Collaboration with academic institutions for research and continuous evaluation.	Costs are distributed across different administrations, and the collaboration between regional institutions, the provincial authority, and the university plays a role in the overall initiative. The estimated budget is around €1,300,000.	Helps policymakers to address issues of unwanted loneliness in the elderly population, suggesting potential interventions to enhance social support. Geolocation data analysis underscores the importance of aligning service distribution with demographic realities, guiding policymakers in optimizing resource allocation in rural areas	System is still in the process of collecting and analyzing data, and efforts are being made to develop appropriate reports or visualization systems while considering data protection concerns. As society ages, policymakers may need to consider scaling and adapting such systems to meet the increasing demand for elderly care, potentially reshaping healthcare infrastructure and service delivery models.
Functional & spatial diagnosis for social revitalization - Poland	Research & subsequent programs address complex development issues in rural areas. Integration of community perspectives & utilization of various research methods underscore the comprehensive nature of the project	Multidisciplinary approach and collaboration between academic, foundation, and regional development entities. Indicator development required specialist knowledge in statistics, qualitative & quantitative dada collection and data management, particularly in the context of social, economic, spatial-functional, technical, and environmental phenomena in rural areas.	Data costs were mentioned as none for the user. 18 Local Revitalization Programs and 108 investment projects were implemented with a total cost of PLN 6,000,955 (appr. 1,390,000 €), supported by funds from the ESF and the state budget.	Integrated approach to rural development; evidence-based decision-making; community engagement & participation; tailored interventions for local contexts; funding mechanisms for rural revitalization; capacity building in rural areas; social inclusion; environmental sustainability.	Outcomes of the revitalization projects and the development of Local Revitalization Programs may present a model that could be replicated in other regions facing similar challenges. Policymakers may consider the potential for scaling up successful models to benefit a broader range of communities.
Scottish National Islands Plan Survey	Comprehensive approach to data collection, analysis, and visualization provided	Expertise needed for areas such as survey design and implementation, geospatial	Value of the whole contract for the survey, including the tool	National Islands Plan Survey and the associated tool provide a rich source of information for policymakers,	The survey, if repeated in the future, could facilitate longitudinal studies, allowing for the tracking

	valuable insights into the experiences of residents on Scotland's islands, contributing to the development and implementation of the National Islands Plan.	analysis, data analysis and statistics, tool development, communication, and policy integration.	development, was approximately €70,000. The cost-effective use of existing skills and free hosting options for the tool also adds to the efficiency of the project.	enabling them to design and implement targeted policies that address the specific needs and aspirations of residents on Scotland's islands.	of changes in perceptions and opinions over time. Similar surveys and tools could be developed for other regions.
Web-mapping tool to visualise proximity to different services - Nordic countries	Valuable tool for visualizing and analyzing the proximity of various services in the Nordic Region, considering factors like degree of rurality, road infrastructure, and cross-border dynamics.	Spatial analysis, web mapping platform development, knowledge of data harmonization techniques, Geographic information systems (GIS), accessibility calculations, OSM road network data.	Main costs included the purchase of service point data, amounting to around €12,000. Development of web-mapping platform and the accessibility calculations took around four months each.	Insights into service accessibility at different geographical levels may inform public policies aimed at reducing regional disparities by identifying areas with limited access to essential services – also across national borders.	Ongoing updates and incorporation of more recent data could enhance its adaptability. The tool's cross-border analysis may facilitate collaboration and coordinated efforts in addressing common challenges across Nordic territories.
Enhancing accessibility & understanding of rural land use data - EU	The tool utilizes sophisticated vision-language models to link ground-level images with satellite imagery, providing crucial context for understanding rural land use dynamics.	Integrating vision-language models, utilizing geolocation data, tool development, analysis of rural land use data from ground-level images and satellite imagery.	Hardware: for model training, four GPUs are utilized to support the computational requirements = estimated cost for the infrastructure is 8,000€. Different services offer free APIs for data collection, but downloading images take nearly a week. Tool cost / development of the algorithm is estimated at 24PM. Potential additional fees for satellite image retrieval services and cloud server deployment.	Successful integration of vision-language models in analyzing rural land use data offers policymakers a tool to make informed decisions in land use planning, conservation, and resource management.	The ultimate goal is to leverage these advanced models to effectively understand, monitor, and manage rural land use.

Earth Observation & Citizen Science – Geo-Wiki	Geo-Wiki is a remarkable example of harnessing the collective capacity of individuals worldwide to contribute meaningfully to scientific research and address pressing global challenges. Monitoring the Earth's surface through citizen engagement and data classification, platform integration of citizen-derived data with expert and authoritative sources helps to address many policy-relevant questions.	Combination of citizens and experts including citizen science platforms, earth observation, remote sensing, environmental policy, data management, technology development, geospatial tools.	Development costs in months / years: Early Beta-Version (6PM), version 1 (1.5 years), big refactoring (10 PM) + ongoing feature development & improvement. No software costs / open source/free software used. Total Hosting Costs (Nov 2011 - Oct 2023): 35,000€ Initial Hosting Costs 70€/month / current Hosting Costs (Oct 2023): 500€/month. Increase in hosting costs due to the need for more collection tools & platforms, additional servers, and services.	Integration of citizen-derived data with expert sources has substantial public policy implications by providing robust and diverse datasets for evidence-based decision-making in areas like environmental policy and SDGs. The case highlights the potential for collaborative citizen science initiatives to inform and shape policies addressing pressing global challenges.	As the platform continues to evolve, its impact on policymaking, environmental monitoring, and public engagement holds the potential to shift further towards more inclusive and data-driven approaches to address complex global issues.

7. Discussion, conclusions and going forward

In this deliverable, we approached the question of data and tools availability, as well as how different cost dimensions unfold, including staff costs, capacity to work with data and infrastructure costs. We highlighted both costs for users as well as development and maintenance costs. Many of the tools and methods presented in this deliverable serve to get some initial information and knowledge for different types of user groups, ranging from the individual to policymaker or planner at different governance levels. Our strategy was to approach this rather complex field by focusing on three different areas that concern the GRANULAR project and paired with the hope that they serve to inform and inspire people in the Living Labs and beyond:

1) *Datasets/ tools identified in previous WP3 work and considered relevant for the work of GRANULAR. WP3 partners will continue to work with some of those. We presented them in this deliverable in the form of data fiches.*
2) *A survey of national statistical offices and authorities – also to show what type of data is available from these sources*
3) *Inspirational examples of data and tools that both Living Labs and GRANULAR partners work with, including a summary matrix of key learnings, cost dimensions, policy implications and future considerations.*

The Data Fiches

While the European Union's (EU) commitment to open data has significantly enhanced transparency and accessibility of both official and research data, the effective use of such open data in rural territories comes with its own set of challenges and associated costs. This discussion will focus on two critical aspects that emerged from the 27 data fiches of this deliverable: the necessity for capacity building in local territories in order to train personnel in Geographic Information System (GIS) and data management, and the necessity for quality assessment through validation with available local databases, complementary data collection and/or ground-truthing.

As shown in this deliverable, while numerous datasets are open and freely accessible, the effective use of this data necessitates basic GIS and data management skills. Hiring an entry-level GIS technician is essential to navigate and meaningfully analyse spatial data, ensuring its relevance to local contexts. In the EU, the estimated cost of employing such personnel ranges from €30,000 to €40,000 annually.

Another significant challenge associated with open data is the variability in the quality of datasets. To ensure the reliability of information for decision-making, data must undergo quality assessment before its use, through the following processes: (i) leveraging locally available datasets or proxies; and/or (ii) engaging in complementary data collection efforts and groundtruthing activities. While locally available datasets or proxies might serve as a cost-effective initial validation step, it is often necessary to acquire additional data to supplement and enhance existing datasets. Hiring personnel for such activities, including surveying and data verification, might lead to annual expenses ranging from €25,000 to €35,000. Such comprehensive quality assessments can improve the reliability of open data to ensure its usability for decision-making locally.

Survey

Survey results provide the reader with information – on a country-by-country level – about the main databases openly available, containing data about rural development issues, the data domains in the open databases and about the data types and finest resolutions openly available. All but one of the offices that took part in the survey make data openly available. The main domains most frequently available and with relevance for rural development and governance are demography, economy, agriculture and tourism / recreation. Since mobility data was provided by only one-fourth of the offices and accessibility data, according to the survey, by only one non-EU country, GRANULAR will make a very valuable contribution to fill this gap by providing new and novel datasets. Most of the offices make tabular data available and more than half of those who responded to the survey, also grid-level data.

Statistical offices also provide insights into how data is accessed, the associated costs, funding sources, and the primary user groups and their preferences. They stressed that their customers have a demand for complex and granular data, as well as indicators to follow up. There is also a need to understand rural-urban interrelations and differences. Sound data is required for decision-making purposes and to understand and follow development and change.

Future plans with repercussions for Living and Replication Labs, and beyond, include those bilateral efforts by Statistics Sweden, Statistics Finland and the Finnish Environment Institute, possibly constructing a new local area definition, inspired by the Swedish DeSO (Demographic Statistical Areas) and RegSO (Regional Statistical Areas) to form coherent areas based on population clusters/ neighborhoods.

Inspirational Examples

As a general conclusion, the identification and tracking of costs, especially for the inspirational examples, was by far not a straightforward exercise and probably the most challenging dimension of case description. Whilst some of the examples are free for the user, development costs had different dimensions. Developers faced different challenges but also managed to solve them through different means.

The inspirational examples were mostly free for us as a potential user or provided data at no costs. Yet, much also depends on what the user wants to do with the data and tool. For a larger project, and more in-depth analyses, staff with specific training, such as GIS, geography or statistics is needed. Likewise, for the replication of tools, naturally costs occur, which we tried to describe as best as possible.

As a general observation, all data and tools providers faced different development costs, such as in relation to software or hardware, data management systems or repositories, purchasing or generating data and last but not least personnel costs. Tools development and calculations etc. require knowledge of data management, GIS, statistics, geography or computing / software development.

Data visualization is an important cost factor for many initiatives (GeoWiki, Rural Barometer, Service Mapper). In the case of telecare for the elderly in Ourense, and regarding visualization, difficulties are faced due to the high degree of protection of the data processed.

Some of the examples also used external expertise, such as for web-mapping platforms (Nordic Service Mapper), survey design and implementation (Rural Barometer) or data collected for a different purpose (SWB indicator).

The Geo-Wiki is a citizen science initiative par excellence. Anyone can engage via desktop or mobile devices in monitoring the Earth's surface. Further, volunteers facilitate the process of visual interpretation and data integration. Costs occurred when the platform was developed and when refactoring / changing technologies and frameworks as well as for hosting (services & servers). In addition, costs are incurred to establish and manage data collection campaigns, along with data curation.

In the case of developing the indicator to follow SWB in the CAP, data accessibility and keeping costs fairly low was based on key people's affiliation with free access to data. Yet, this means dependency on individuals and fragility if no more stable access to data at institutional level is build.

Geo-Wiki uses only open source/free software, thus, there are no software costs. Furthermore, the combination of citizens and experts is remarkable. Likewise, in the cross-modality framework example, different services offered free APIs for data collection but other costs, such as for tool and algorithm development, occurred.

Developers also stressed that whilst and since initiatives have started and investments have been made, it made much sense to continue with and invest in newer data sets, also to identify trends and changes (e.g. Nordic Service Mapper, Rural Barometer, Geo-Wiki). In the case of monitoring mobility and road traffic at local scale in France, initial software investment in addition to annual maintenance and operational costs, enable the operators to better understand traffic trends, help decision-making, design road projects, and adapt measures for worksite operations and thus help to save costs elsewhere. The model for diagnosing socio-economic structures and the concept of implementing social revitalization in rural areas developed in the communes of the West Pomeranian Voivodeship were used to develop 18 Local Revitalization Programs, co-funded by national and EU ESF funds.

8. Conclusion

As a conclusion, this deliverable showed that local actors have to face a diverse array of costs in order to grasp the diversity within their territory. Traditional surveys demand investments in hiring and training surveyors, distributing materials, and managing data entry processes, while more technologically-driven approaches (e.g. remote sensing) incur significant expenses, such as the acquisition and interpretation of imagery coupled with the need for robust technology infrastructure. Moreover, deploying sensor/meters networks necessitates induces costs for installation, maintenance, and ensuring data quality. Even data collection methods based on citizen science have infrastructure costs for running servers and personnel costs for data quality control.

On the analytical front, despite the free availability of many datasets that are relevant and useful for rural areas at the appropriate granularity, their use is not without costs. Basic GIS and data management skills are essential for effectively using and analyzing such data, and the necessity for complementary quality assessments for some datasets further adds to the financial requirements. It thus appears necessary to improve local capacities in GIS, data management and data collection skills through capacity building programs to ensure that rural territories can effectively leverage open data for informed policymaking.

The key take-away from this analysis is that research teams together with rural territories (including Living and Replication Labs) could leverage funding to build comprehensive capacity-building programs. Such programs could be designed to deliver tailored online training that equip rural communities with the skills necessary for the effective use of geospatial open data and data collection methods, with thematic applications relevant to current transitions that rural areas face. Fostering collaboration and knowledge sharing about the use of open data for local policy making should also be actively encouraged, promoting peer-to-peer learning and exchange of best practices among rural territories. Moreover, supporting the development and deployment of innovative human resources management by pooling personnel with specific skills between territories could ensure the sustainability of such approaches. Lastly, streamlining data access and management processes is recommended to simplify the procedures for rural territories, thereby reducing the logistical burden associated with accessing, managing, and analyzing open data. By embracing these recommendations, the EU can significantly enhance the capacity of rural regions to leverage open data for informed decision-making.

The **EU has been promoting the concept of open data** as part of its **broader digital agenda** and that certain data should be freely available for anyone to use, reuse, and redistribute. Public sector information should be available for reuse, promoting transparency and innovation. Implications of this study for the EU's open data policy include:

- ***Capacity Building and Training:*** *The need for capacity building in local territories was highlighted, adhering to the Open Data Directive, may involve training personnel in public sector bodies to effectively manage and release data in accordance with the directive.*
- ***Quality Assessment and Validation:*** *The release of high-quality public sector information is encouraged in the Directive. D3.2. emphasis on quality assessment and validation aligns with the Directive's objective of ensuring the reliability and usability of data for decision-making.*
- ***Cost Considerations:*** *D3.2 discusses the costs associated with utilizing open data, including the need for hiring personnel, conducting quality assessments, and data management. This aligns with the Directive's goal of fostering fair competition and innovation by minimizing the costs associated with reusing public sector information.*
- ***User Preferences and Complex Data:*** *The Directive encourages public sector bodies to take into account user needs and preferences. D3.2 findings on user demands for complex and granular data resonate with the Directive's focus on providing valuable information to users.*
- ***Collaboration and Knowledge Sharing:*** *D3.2 suggests collaboration, knowledge sharing, and peer-to-peer learning, which aligns with the Directive's spirit of promoting cooperation among public sector bodies to improve the availability and reuse of data.*
- ***Sustainability and Accessibility:*** *The recommendations in the document, such as comprehensive capacity-building programs and streamlining data access processes, align with the Directive's goal of ensuring the sustainability and accessibility of public sector information.*

- ***Innovation and Use of Open Data:*** *D3.2 discussion of inspirational examples and the need for ongoing investments in newer datasets align with the Directive's goal of fostering innovation and maximizing the use of open data for societal and economic benefits.*

Concerning the next steps in the project it is suggested that:

- *Work with data presented here continues, both as regards what was presented in the fiches, via different tasks under WP3 as well as regards the inspirational examples.*
- *Living and Replication Labs may take a close look at the different data types and survey results presented here and discuss whether some of the examples are of further interest or might potentially even been replicated in their areas.*
- *Some of data, such as the Strava dataset, as the largest collection of human-powered transport information in the world, would be worth exploring further by GRANULAR or the Living Labs, e.g. through applying for free access.*
- *In the discussions – supported by WP 3 and other experts - Living and later Replication Labs should of course reflect on their own needs, specific circumstances and eventual development challenges.*
- *Readers are warmly welcome to reach out to the authors and contacts of each inspirational example and data fiche to engage in discussing the options going forward.*
- ***Continued Work with Data*** *aligns with the Directive's encouragement for continuous efforts to improve the availability and quality of public sector information.*
- ***Encouraging Living and Replication Labs*** *to examine survey results and inspirational examples aligns with the Directive's emphasis on collaborative efforts to enhance data reuse. Reflecting on their needs and circumstances resonates with the Directive's encouragement for public sector bodies to tailor their practices to local conditions.*
- ***Further exploration of specific datasets****, such as the Strava dataset, aligns with the Directive's goal of encouraging the release of diverse and valuable datasets.*

9. Acknowledgements

The editors of this deliverable would like to express – once again – their thanks and gratitude to all colleagues, and especially to the members of the Living Labs, for all their contributions to this work and for their inspiration. The co-creation spirit, also during the final editing phase, and contributions of LP and WP lead was much appreciated as were comments and feedback from all partners. We are very grateful to all colleagues in the national and regional statistical authorities for taking the time to fill in the survey and for follow-up discussions.

10. References

ARDECO - Annual Regional Database of the European Commission. Available at https://economy-finance.ec.europa.eu/system/files/2022-10/Reference%20Metadata%20AMECO_September%202022.pdf.

Barranco, R. (2022). UDP - Tourism capacity. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/659a45dd-5bc2-4aaf-8bcd-bb337ba03f92.

Batista e Silva, F., Herrera, M. M., Rosina, K., Barranco, R. R., Freire, S., & Schiavina, M. (2018). Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management*, 68, 101-115. https://doi.org/10.1016/j.tourman.2018.02.020.

Becker D. (2017). Predicting outcomes for big data projects: big data project dynamics (bdpd): research in progress, in: 2017 IEEE International Conference on Big Data (Big Data), 2017, pp.2320–2330.

Bondarenko M., Kerr D., Sorichetta A., and Tatem, A.J. 2020. Census/projection-disaggregated gridded population datasets for 51 countries across sub-Saharan Africa in 2020 using building footprints. WorldPop, University of Southampton, UK. doi:10.5258/SOTON/WP00682.

Byrne C., 2017). Development Workflows for Data Scientists, O'Reilly Media.

Castillo, C. P., e Silva, F. B., & Lavalle, C. (2016). An assessment of the regional potential for solar power generation in EU-28. *Energy policy*, 88, 86-99. https://doi.org/10.1016/j.enpol.2015.10.004.

Colas M, Finck I, Buvat J, Nambiar R, Singh R. (2014). Cracking the data conundrum: How successful companies make big data operational, Capgemini Consulting.

Edwards R., Bondarenko M., Tatem A. and Sorichetta A. Unconstrained subnational Population Weighted Density in 2000, 2005, 2010, 2015 and 2020 (100m resolution). WorldPop, University of Southampton, UK. doi:10.5258/SOTON/WP00703.

European Environment Agency (2016). Potential quiet areas in Europe, based upon Quietness Suitability Index (QSI). Retrieved from EEA's website: http://data.europa.eu/88u/dataset/e9151c34-da65-48b9-a2ca-b9b835480812.

European Environment Agency (2016). Quiet areas in Europe. Technical report No 14/2016. Available at https://www.eea.europa.eu/publications/quiet-areas-in-europe.

European Commission, Joint Research Centre (JRC) (2019). ENSPRESO - SOLAR - PV and CSP. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/18eb348b-1420-46b6-978a-fe0b79e30ad3.

European Commission – Eurostat/GISCO. NUTS geometries available at https://ec.europa.eu/eurostat/fr/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts.

European Union (2019). Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). ELI: http://data.europa.eu/eli/dir/2019/1024/oj.

Eurostat. (2020). Healthcare services in Europe. Retrieved from Eurostat website: https://gisco-services.ec.europa.eu/pub/healthcare/metadata.pdf.

Eurostat. (2023). Healthcare services locations. Retrieved from Eurostat website: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/healthcare-services.

Eurostat (2023). ICT usage in households and by individuals (isoc_i): reference metadata in Euro SDMX Metadata Structure (ESMS). EU Statistics Rural household internet access in [reference year(s)]. Available at https://ec.europa.eu/eurostat/cache/metadata/en/isoc_i_esms.htm.

EU-SILC. Scientific use files available at https://cros-legacy.ec.europa.eu/EU-SILC-SUF/forum_en.

Fendrich, A. N., Matthews, F., Van Eynde, E., Carozzi, M., Li, Z., d'Andrimont, R., Lugato, E., Martin, P., Ciais, P., Panagos, P., 2023. From regional to parcel scale: A high-resolution map of cover crops across Europe combining satellite data with statistical surveys. *Science of the Total Environment*, pp.162-300. https://doi.org/10.1016/j.scitotenv.2023.162300.

Fick, S.E. and R.J. Hijmans (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302-4315.

Freire S., MacManus K., Pesaresi M., Doxsey-Whitfield E., Mills J. (2016). Development of new open and free multi-temporal global population grids at 250 m resolution. Geospatial Data in a Changing World; Association of Geographic Information Laboratories in Europe (AGILE).

Gupta Uma & Cannon S (2020). A Practitioner's Guide to Data Governance: A Case-Based Approach, Emerald Publishing Limited. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/nrifi-ebooks/detail.action?docID=6238638.

Hersbach, H., Bell, B., Berrisford, P., et al. (2017). Complete ERA5 from 1940: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service (C3S) Data Store (CDS). DOI: 10.24381/cds.143582cf.

López-Antequera, M., Gargallo, P., Hofinger, M., Bulò, S.R., Kuang, Y., & Kontschieder, P. (2020). Mapillary Planet-Scale Depth Dataset. *European Conference on Computer Vision*.

Lugato, E., Smith, P., Borrelli, P., Panagos, P., Ballabio, C., Orgiazzi, A., Fernandez-Ugalde, O., Montanarella, L., Jones, A. (2018). Soil erosion is unlikely to drive a future carbon sink in Europe. *Science Advances*. 4, eaau3523.

Martinez I, Viles E, Olaizola I. (2021). Data Science Methodologies: Current Challenges and Future Approaches. Big Data Research, Volume 24. https://doi.org/10.1016/j.bdr.2020.100183.

OpenStreetMap. Available at https://www.openstreetmap.org/.

Pezzulo, C., Hornby, G., Sorichetta, A. et al. (2017). Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. Sci Data 4, 170089. https://doi.org/10.1038/sdata.2017.89.

Saltz J & Shamshurin I (2016). Big data team process methodologies: a literature re-view and the identification of key factors for a project's success, in: 2016 IEEE International Conference on Big Data (Big Data), 2016, pp.2872–2879.

Saltz J, Shamshurin I, Connors C (2017a). A framework for describing big data projects, in: W. Abramowicz, R. Alt, B. Franczyk (Eds.), Business Information Systems Workshops, Springer International Publishing, Cham, 2017, pp.183–195.

Saltz J, Shamshurin I, Connors C (2017b). Predicting data science sociotechnical execution challenges by categorizing data science projects, J. Assoc. Inf. Sci. Technol. 68(12) (2017) 2720–2728, https://doi .org /10 .1002 /asi .23873.

Sarsfield, S (2009). The Data Governance Imperative. IT Governance Ltd. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/nrifi-ebooks/detail.action?docID=480410.

Schiavina M., Freire S., Carioli A., MacManus K. (2023). GHS-POP R2023A - GHS population grid multitemporal (1975-2030). European Commission, Joint Research Centre (JRC). PID: http://data.europa.eu/89h/2ff68a52-5b5b-4a22-8f40-c41da8332cfe. DOI: https://doi.org/10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE.

Schneider, M., Chan, A., & Körner, M. (2023). EuroCrops [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10118572.

Sivarajah U, Kamal M, Irani, Weerakkody V. (2016). Critical analysis of big data challenges and analytical methods. J. Bus. Res. 70 (2017) 263–286, https://doi.org/10.1016/j.jbusres.2016.08.001.

Sorichetta, A., Hornby, G., Stevens, F. et al. (2015). High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. Sci Data 2, 150045. https://doi.org/10.1038/sdata.2015.45.

Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLoS ONE* 10(2): e0107042. https://doi.org/10.1371/journal.pone.0107042.

Strava Metro. Human powered mobility. Available at https://metro.strava.com.

Tatem, A.J., Garcia, A.J., Snow, R.W. et al. (2013). Millennium development health metrics: where do Africa's children and women of childbearing age live?. Popul Health Metrics 11, 11. https://doi.org/10.1186/1478-7954-11-11.

Tiecke, T.G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E.B., & Dang, H.H. (2017). Mapping the world population one building at a time. *https://arxiv.org/abs/1712.05839v1*.

WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076).

11. Appendix 1 – Overview of national databases openly available, containing data about rural development issues

Country	Main database with rural-development relevant data	Data domains in open databases	Data types available	Finest resolution of data free of charge
Bulgaria	The IS Infostat platform (https://infostat.nsi.bg) publishes data relevant for rural development. It includes business, demographic social, macroeconomic, environment energy and multi-domain statistics. Data on population and housing census is available free of charge. Paid databases provide users access to more detailed data at lower levels after disaggregation. Grid data for population (total, age groups and sex) for 2011, 2021 is free of charge.	Demography, energy and health.	N.a.	Grid data for population: Total population, Age groups 0-14; 15-64 ;65+ and sex, for 2011 and 2021.
Croatia	Several databases are available but not strictly connected with rural development issues. The so-called PC-Axis databases are available at https://web.dzs.hr/PX-Web_e.asp?url=%22Eng/Archive/stat_databases.htm%22. Some data available is at municipality level. Grid 1000 is accessible, free of charge, at the GeoSTAT - Web GIS portal of the Croatian Bureau of Statistics (https://geostat.dzs.hr). The available grid-level data is on population – number of populations, number of population by large age groups, population by educational attainment, population by activity, business register (active business entities). Tourism data on accommodation capacities and tourist arrivals and nights. The PC-Axis databases https://web.dzs.hr) has some data by municipalities. Some census data by settlements is available at https://podaci.dzs.hr/en/statistics-in-line/. All data published online is free of charge. If special data processing is needed for grid-level data, it is charged according to the subject's hourly rate.	Agriculture, demography, economy, energy, environment, tourism / recreation, transport.	N.a.	1km grid. Available data free of charge at the GeoSTAT - Web GIS portal for the Croatian Bureau of Statistics https://geostat.dzs.hr/?lang=en.
Cyprus	The main database by Statistics Cyprus is CYSTAT-DB, available at https://cystatdb.cystat.gov.cy/pxweb/en/8.CYSTAT-DB/. It contains data on agriculture, livestock, fishing, business register, construction, education, energy, environment, external trade, health, industry, information society, innovation, labor market, living conditions, social protection, national accounts, population, price indices, public finance, research and development, services, tourism and trade.	Demography, Health, Education, ICT Usage.	N.a.	1000m for grid-level data.
Finland	The main database published by Statistics Finland is called StatFin (https://www.stat.fi/tup/statfin/index_en.html). StatFin is freely accessible and includes data on population, economy, housing, transport, tourism, consumption, prices, wages and salaries, energy, enterprises etc. The Paavo database (https://www.stat.fi/tup/paavo/index_en.html) contains data by postal code area on the population structure, education, income, housing, workplaces, households' life stage etc. There is also a grid-level database with grid sizes of 250 m x 250 m, 1 km x 1 km and 5 km x 5 km (https://www.stat.fi/tup/ruututietokanta). The grids cover the whole of Finland, but this is not for free and charged based on number of licenses and grid size. Data is available on the areas' population structure, level of education, income of inhabitants	Agriculture, demography, economy, energy, environment, housing, infrastructure, mobility, tourism / recreation, transport.	N.a.	Population structure grid data 1km x 1km.

	and households, size and stage in life of households, buildings and dwellings, workplaces, and main activities of inhabitants. Population structure grid data 1km x 1km is free of charge.			
France	The geoservices.ign.fr site (https://geoservices.ign.fr/) and its *Géoservices* catalogue (https://geoservices.ign.fr/catalogue) by the *Institut National de l'Information Géographique et Forestière (IGN)* are of relevance for anyone interested in geodata and web services related to rural development. Data on the site is free of charge and available under an open license and accessible without registration. It contains, vector databases, maps, ortho-images, cadastral parcels, 3D models as well as other applications and services. Information includes the Common Agricultural Policy, forest geographical reference system, "Land-Sea Boundary" data, a Renewable Energy Map Portal and Good Agricultural and Environmental Condition. A second data source is by the Centre for Studies on Risks, the Environment, Mobility and Urban Planning (CEREMA). Its catalogue with over 300 datasets, maps and series is open and available at https://catalogue.cdata.cerema.fr/geonetwork/srv/eng/catalog.search#/home. Data covers such fields like land cover, natural risks, energy use, habitats and biotopes and hydrography. The *Office français de la biodiversité* with its partners feeds an information system on biodiversity, with 14 datasets, services and maps, incl. characterization of Natura 2000 sites, protected natural areas, inventory of Natural Areas of Ecological, Faunal and Floristic Interest etc. *(https://www.ofb.gouv.fr/)* Thematic data and datasets on agriculture, culture and heritage, sustainable development and energy, economics and statistics, eductaion and reserach, international and EU issues, health and social issues, tourism and recreation, territories and transport can be found at https://www.geoportail.gouv.fr/. This is the national portal for territorial knowledge providing open and interoperable data to facilitate the exchange and sharing of data in support of public policies.	N.a.	N.a.	Grid-level data.
Galicia / Spain	The main database openly available, containing data about rural development issuesof the Galician Institute of Statistics is available at https://www.ige.gal/web/index.jsp?idioma=gl.	Demography, economy, tourism / recreation.	N.a.	Spatial distribution of the characteristics of the population of Galicia by grid of 1km2.
Greece	Currently, there is no dissemination database openly available to the public. However, data files available for the public are uploaded at the website of the Hellenic Statistics Authority (ELSTAT) in the form of time series, tables and Public Use Files (PUFs). ELSTAT publishes statistical data on its website at a level of analysis, where statistical confidentiality is not violated, and all users can access them. Most of the data on ELSTAT's website refer to NUTS 2 level. Depending on the limitations set due to statistical confidentiality, some data refer also to NUTS 3 level and few data (mainly population data) to LAU level. Statistical data that allow the indirect identification of statistical units are provided to users under certain conditions. Tailor-made data based on user requirements are generally priced. The cost of providing these data depends on the number of man-days required for their compilation by ELSTAT staff. The pricing policy of ELSTAT is available here: https://www.statistics.gr/en/microdata_pricing.	N.a.	N.a.	R&D, innovation, population data at NUTS-2 level.
Hungary	There is no dedicated database for rural development data, but the main database contains data on agriculture, environment, and many aspects of territorial data. The data can be downloaded in csv and xlsx format and are free of charge. The database can be accessed here:	Agriculture, demography, economy, energy,	N.a.	Grid-level data. Charging for the data is not determined by the level of resolution or the

	https://statinfo.ksh.hu/Statinfo/themeSelector.jsp?&lang=en. Predefined data tables in the STADAT system also contain data on the abovementioned topics, however, this is not a database, but ready-made tables are available, which can also be downloaded in csv and xlsx, also free of charge. The STADAT-system is available here: https://www.ksh.hu/stadat_eng. The third main resource where users can find territorial data is the Interactive Mapping Application, accessible here: https://map.ksh.hu/timea/?locale=en. Data can be downloaded from this interface, from the attribute table in csv. Unlike the aforementioned two sources, this application also contains grid-level data.	environment, health, housing, infrastructure, tourism / recreation, transport.		theme but by the capacity it requires from the statisticians to prepare the data.
Ireland	The Central Statistics Office (CSO) of Ireland produces a wide range of statistics. This includes business sectors including agriculture and fisheries, census data, data on the economy, environment, labour market, people and society as well as other themed publications. CSO's *PxStat Open Data Platform* is available at https://data.cso.ie/#. Data published on the platform is also provided by other public sector databases, such as by the Department of Agriculture, Food and the Marine, the Department of Housing, Local Government and Heritage, the Sustainable Energy authority etc.	See left.	Grid-level, tabular & vector data.	Small-area data, for which are units with an average of 50-100 households in each. Grid-level data is free, and if it is not available, €80-200 are charged for customers from the private sector.
Italy	*:* IstatData is the latest aggregate data dissemination platform of the Italian National Institute of Statistics (Istat) and available at https://esploradati.istat.it/databrowser/#/en. It makes use of the open-source tools “Data Browser” and “Meta & Data Manager” developed by Istat following the international SDMX (Statistical Data and Metadata eXchange) standard for exchanging and sharing statistical data and metadata. Currently, six themes are covered: National Accounts, Population and Households, Household Economic Conditions, Agriculture, Enterprises, Welfare and Pension. Time Series (https://seriestoriche.istat.it/index.php?id=18&L=1) contains over 1,500 time series organized into 22 thematic areas made available to inform abot the environmental, social and economic changes in Italy.	See left.	Tabular and vector data.	Agricultural plot level.
Moldova	The statistical databank of Statistics Moldova is available at https://statbank.statistica.md/. It contains data on environment, population and demographic processes, social statistics, economic statistics, gender statistics, and regional statistics.	Accessibility, agriculture, demography, economy, energy, environment, health, housing, infrastructure, mobility, tourism / recreation, transport.	Raster & tabular data.	Currently the office disseminates at the level of communes (with a commune being formed of one or several villages) – both in the Statistical databank and as static maps in its publications. In 2024, the office plans to carry out a Population and Housing Census and will then have available spatial data both at vector and grid-level. Currently, the office does not provide spatial data against a fee.
Poland	The Knowledge Database (https://dbw.stat.gov.pl/en) by Statistics Poland contains 31 domain areas including Demography, Education, Energy, Social economy, Municipal and housing infrastructure, Agriculture, Labor Market, Transport, Tourism, Living conditions, Health and healthcare. For some indicators data is available by voivodships, in the case of demographic data and of local government unit budgets also for lower levels of territorial division. The Knowledge Database is a publicly available and free of charge. Furthermore, there is also a Centre for Rural	N.a.	N.a.	NUTS 2.

	Statistics (https://olsztyn.stat.gov.pl/en/) and a Small Areas Statistics Centre (https://poznan.stat.gov.pl/en/) in addition to regional statistics centers throughout the country. (https://stat.gov.pl/en/regional-statistics/). EUROSTAT Economic accounts for agriculture - values at current prices and Economic accounts for agriculture - values at n-1 prices are available, too.			
Portugal	The main database openly available, containing data about rural development issues of Statistics Portugal is available here https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_bdc_tree&contexto=bd&selTab=tab2. Themes also include Agriculture, forest and fisheries.	Agriculture, demography, economy, energy, environment, health, housing, mobility, tourism / recreation, transport in addition to culture, prices, living conditions.	Grid-level & tabular data.	For the Population and housing Census grid of 1km2 is free of charge. For some of the other domains, data is at parish-level.
Scotland	The main database openly available, containing data about rural areas are the 1) National Performance Framework (https://nationalperformance.gov.scot/measuring-progress/national-indicator-performance) with 26 out of 81 indicators providing data for Rural Scotland and the 2) Rural Scotland Key Facts 2021 (https://www.gov.scot/publications/rural-scotland-key-facts-2021/documents/) and a considerable number of sources. There are several public sector open data portals, most allow for publication of data at the 3* level of openness (csv or equivalent).[14] Many of Scottish Government statistics is available online, for free and without restrictions. It contains around 300 open datasets and reference material, mainly at the 5* level – the highest level of openness, with associated metadata.	N.a.	N.a.	Grid-level data.
Serbia	The main database openly available, containing data about rural development issues of the Statistical Office of the Republic of Serbia is available here: https://data.stat.gov.rs/?caller=SDDB&languageCode=en-US.	Accessibility, agriculture, demography, economy, energy, environment, health, housing, infrastructure, tourism / recreation, transport.	N.a.	Spatial resolution depends on the coverage of statistical survey, from which the dataset is produced. In the annual plan of statistical surveys, the spatial resolution of available datasets which are free of charge is defined.
Slovakia	The main database openly available, containing data about rural development issues of Slovak Statistics is called DataCube and is available at https://datacube.statistics.sk/. It contains multidimensional tables for indicators of economic and socio-economic development, for the following areas: demographic and social statistics, macroeconomic statistics, business statistics,	Accessibility, agriculture, demography, economy, energy, environment,	N.a.	Municipalities - especially demographic data. All data in the database are free of charge.

[14] According to Tim Berners-Lee, open data can be published at various levels of openness (see 5-star Open Data (5stardata.info).

	sector statistics, environment, multi-domain statistics and selected tables of the Eurostat database. 1 km x 1 km grid-level data is available from the 2011 and 2021 census: https://slovak.statistics.sk/wps/portal/ext/themes/demography/census/indicators/.	health, housing, infrastructure, tourism / recreation, transport.		
Slovenia	The main database openly available, containing data about rural development issues of Statistics Slovenia is available at https://pxweb.stat.si/SiStat/en/Podrocja/Index/583/regionalni-pregled.	Agriculture, demography, economy, energy, environment, housing, infrastructure, mobility, tourism / recreation, transport.	N.a.	Grid-level data available at the STAGE pages at https://gis.stat.si/#lang=en. All data at Statistics Slovenia is free of charge.
Sweden	The majority of Statistics Sweden's statistics are openly available at *"Statistikdatabasen"* (Statistical database) (https://www.statistikdatabasen.scb.se/pxweb/en/ssd/). In addition, other data openly available from Statistics Sweden includes geospatial data (https://www.scb.se/en/services/open-data-api/open-geodata/). Furthermore, there are 29 different public agencies, which publish official statistics, and partly cover other topics than Statistics Sweden: https://www.scb.se/en/About-us/official-statistics-of-sweden/government-agencies-responsible-for-official-statistics/. Furthermore, in addition to Statistics Sweden, other agencies also publish similar data (including geospatial data/GIS layers). In addition to the database mentioned above, there are also data at the following spatial resolution, which fully or partly go below municipal level: Grid statistics, preschools and agency and municipal offices, localities and small localities, holiday-home areas, retail trade areas, activities zones, and RegSO (Regional Statistical Areas). For accessing the GIS layers, please see: https://www.scb.se/en/services/open-data-api/open-geodata/	Agriculture, demography, economy, energy, environment, housing, infrastructure, mobility, tourism / recreation, transport.	Grid-level, point, tabular & vector data.	Grid-level data in Sweden is free of charge. In the database mentioned previously, some statistics are made available according to DeSO, "Demografiska statistikområden" (in English: "Demographic Statistical Areas"). There are also corresponding GIS layers available. For more information, please see https://scb.se/hitta-statistik/regional-statistik-och-kartor/regionala-indelningar/deso---demografiska-statistikomraden/ or https://www.scb.se/en/services/open-data-api/open-geodata/deso--demographic-statistical-areas/.

12. Appendix 2 - Data Fiches

12.1. Accessibility

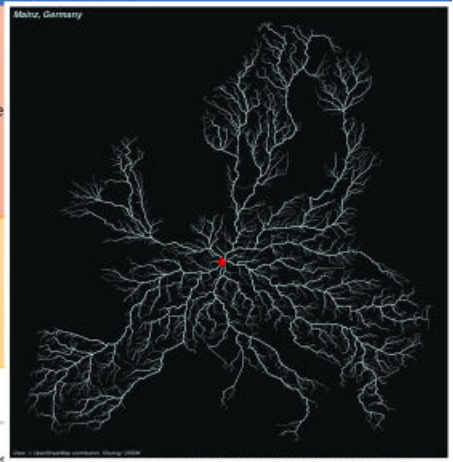
Routing engines based on OpenStreetMap
https://www.openstreetmap.org/

Indicator class: Accessibility

Data
Class: Crowdsourcing – Routing engines
Type: Origins – Destinations matrices / itinaries (lines) / Isochrones
License: Data Commons Open Database license

Spatial
Resolution: Itinaries from origin/ destination points
Extent: World
Granularity: Roads included in the OSM database and its attributes (max speed).
Coordinate System: WGS 1984 (EPSG: 4326)

Temporal
Resolution: Up-to-date (depend on OSM)
Extent: Not relevant
Frequency: Continuously updated



Description:
Several open-source routing engines based on the OpenStreetMap database are available to compute travel time matrices between origins and destinations (OSRM, Valhalla, Graphhopper, Openrouteservice, pgRouting). All these engines use the OSM road network. It requires as input a set of origin/destination points. Outputs are usually routes, trips, isochrones and travel distance matrices (travel time or kilometric distance). For our concerns, the travel distance matrices will make possible the computation of accessibility indicators.
Routing engines supports profiles, representing routing behavior for different transport modes like car, bike and foot. For the car profile, the OSM tag *maxspeed* is used to calculate distance and travel time between two points. If the maximum speed information is missing, which is frequently the case (only 7.4 % of all road elements stored this value in 2019), predefined speed limits for each country are applied.
Bike profile avoids road without bicycle access. Some routing engines, such as Valhalla, provides also digital elevation model (DEM) data as the result of the query. Penalties are applied to roads based on elevation change and grade.
Multimodal models (using several transport profiles for a same route) are not implemented at the moment, but associated documentation of some routing engines (Valhalla, namely) specifies that the implementation of the feature is strongly envisaged for the future.

How to cite
© OpenStreetMap contributors

Data File Size
europe-latest.osm.pbf: 27.6 GB (input) / depending on O/D (output)

File format: pbf (input) / csv (output)

USER COSTS

Free & open Access

Data Infrastructure

Infra needed (software, hardware): A modern computer with 32 GB RAM or more.

Data repositories / storage needed: Routing outputs may be usufully stored (O/D matrices) in a csv format. Size depending on the number of O/D points considered.

Data Governance & Management

Staff costs: Advanced computer / GIS skills (Docker). Advanced knowledge of the OSM database.

Data analysis needed: No. Require relevant Origin/Destionation points as an input.

Quality assurance: The road network available in OpenStreetMap can be considered of good quality : A completeness analysis led by Barrington-Leigh (2019) based on satellite imagery, multilevel regression and postratification model argued that globally, OSM is ~83% complete, and more than 40% of countries-incuding several in the developing world have a fully mapped street network, concluding that in many places, researchers and policy makers can rely on the completeness of OSM, or will soon be able to do so.

Outputs of these routing engines deliver theoretical travel time, without traffic congestion. This should be taken into account, especially in urban contexts.

The choice of the routing engine depends on several aspects: the activity of the project, hardware requirement, features of the routing engine (turn restrictions, elevation, time awareness), costing options (influence of the route finding with many custom factor/penalties/costs. e.g. avoid highways).

For the needs of the GRANULAR project, requiring to cover all Europe (heavy calculations) and several transport profiles, OSRM and Valhalla routing engines appear the most suitable solutions.

Data documentation and / or assistance in data use

Luxen, 2011, Real-time routing with OpenStreetMap data
OSRM github repository
Valhalla github repository
Giraud, 2023, How to Build a European-Wide OSRM Server on a Desktop Computer
Valhallr: A tidy Interface to the Valhalla Routing Engine

Compilation of the fiche by: Ronan Ysebaert & Marianne Guérois (CNRS)

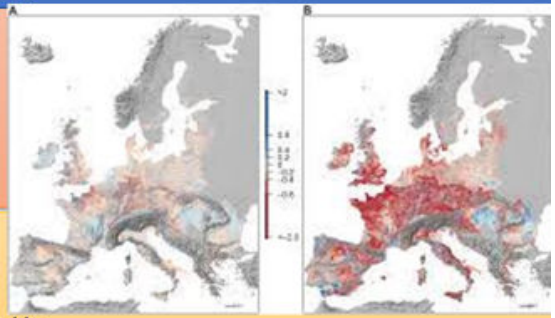
12.2. Agriculture

Carbon budget in the EU agricultural soils — By JRC
URL: https://esdac.jrc.ec.europa.eu/content/carbon-budget-eu-agricultural-soils

Indicator class: Regional

Data
Class: Model
Type: Raster
License: under NO CIRCUMSTANCES are these data passed to third parties. They can be used for any purpose, including commercial gain.
DOI: 10.1111/ejss.13315

Spatial
Resolution: 1 X 1 km
Extent: EU
Granularity: Gridded
Coordinate System: ETRS_1989_LAEA_L52_M10



Temporal
Resolution: Yearly
Extent: 2016-2100
Frequency: not applicable

Description:
C budget in the EU agricultural soils including lateral C fluxes
Soil play a significant environmental role in balancing the climate as it currently acts as a carbon sink, sequestering CO2 from the atmosphere into soil organic carbon. This dataset tracks the possible transformations of the organic carbon across the landscape by using a biogeochemistry-erosion model to quantify the impact of future climate on the carbon cycle. Taking into account all the additional feedbacks and C fluxes due to displacement by erosion, it is estimated that there will be a net source of 0.92 to 10.1 Tg C year−1 from agricultural soils in the European Union to the atmosphere over the period 2016–2100. These ranges represented a weaker and stronger C source compared to a simulation without erosion (1.8 Tg C year−1), respectively, and were dependent on the erosion-driven C loss parameterization, which is still very uncertain.

How to cite
Dataset:
Lugato, E., Smith, P., Borrelli, P., Panagos, P., Ballabio, C., Orgiazzi, A., Fernandez-Ugalde, O., Montanarella, L., Jones, A. 2018. Soil erosion is unlikely to drive a future carbon sink in Europe. *Science Advances*. **4**, eaau3523.

Data File Size
700 MB
File format: GeoTIFF

USER COSTS

Free of charge & Access limited (request form)

Data Infrastructure

Infra needed: GIS (QGIS for example) or software environment for statistical computing (R)

Data repositories / storage needed: Datasets are hosted on ESDAC and made available for free.

Data Governance & Management

Staff costs: Data can be easily downloaded. Its use requires a certain expertise in the world of geographic information and the structure of LPIS data (low level GIS Technician ~25,000€ + charges per year)

Data analysis needed: None for simple visualisation, requires running a R code (supplied) to run simulations.

Quality assurance: Data are model-based and might contain errors. Any error or omissions should be noted and reported to the JRC.

Data documentation and / or assistance in data use
https://esdac.jrc.ec.europa.eu/content/carbon_-budget-eu-agricultural-soils

Compilation of the fiche by Tristan Berchoux (IAMM)

Cover Crops across Europe — By JRC

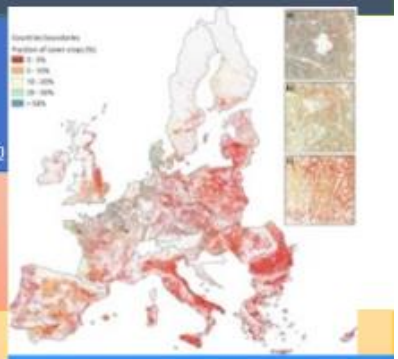
URL: https://esdac.jrc.ec.europa.eu/content/cover-crops-accross-

Indicator class: Regional

Data
Class: Satellite + Survey
Type: Raster
License:
DOI: https://doi.org/10.1016/j.scitotenv.2023.162300

Spatial
Resolution: 100 m
Extent: EU + UK
Granularity: Gridded
Coordinate System: EPSG:3035

Temporal
Resolution: Yearly
Extent: 2016
Frequency: n/a



Description:Disaggregated map of cover crops occurrence for Europe and the UK

Despite the growing importance given to cover crops as a sustainable agricultural practice, the availability of spatial data about them is scarce. The best information available is regionally aggregated survey data, which, although indicative, hinders the development of spatially accurate studies. Using a disaggregation model, this dataset combines satellite data (Sentinel-1) with aggregated survey data to generate a high-resolution map of cover crops for Europe and the United Kingdom for the reference year of 2016. The map was validated with parcel-level data in France, where overall good results were found, with regional variations. This dataset includes the median, standard deviation, 5th and 95th percentiles of the predicted cover crop fraction. Such a fraction is assumed to vary between 0% (i.e., no cover crops in the pixel) to 100% (i.e., pixel completely covered by cover crops). The values in the raster files are multiplied by 100 to facilitate compression. For instance, a raster value of 1234 corresponds to a 12.34% cover crops fraction. The dataset aims to replace the practice commonly found in the literature of assigning aggregated values to agricultural pixels by randomly sampling them in space. It can be useful for researchers and practitioners requiring spatially explicit knowledge of cover.

How to cite
Dataset:
Fendrich, A. N., Matthews, F., Van Eynde, E., Carozzi, M., Li, Z., d'Andrimont, R., Lugato, E., Martin, P., Cialis, P., Panagos, P., 2023. From regional to parcel scale: A high-resolution map of cover crops across Europe combining satellite data with statistical surveys. Science of the Total Environment, pp.162300. https://doi.org/10.1016/j.scitotenv.2023.162300

Data File Size
~1 GB
File format: GeoTIFF

USER COSTS

Free & open Access (Registration needed)

Data Infrastructure

Infra needed: GIS (QGIS for example) or software environment for statistical computing (R)

Data repositories / storage needed: Datasets are hosted on ESDAC and made availablefor free upon request.

Data Governance & Management

Staff costs: Data can be easilydownloaded after completing the request form. Its use requires a certain expertise in the world of geographic information (low level GIS Technician ~25,000€ + charges per year)

Data analysis needed: None for simplevisualisation

Quality assurance: Data has good results (validationwith parcel-level data in France) but no other detailsare given. No layer of qualityassessmentif given.

Data documentation and / or assistance in data use
https://esdac.jrc.ec.europa.eu/content/cover -crops-accross -europe

Compilation of the fiche by Tristan Berchoux (IAMM)

EuroCrops

By EU

URL: https://github.com/maja601/EuroCrops

Indicator class: Regional

Data

Class: Compilation of government datasets

Type: Vector

License: Creative Commons Attribution-ShareAlike 4.0 International License

DOI: https://doi.org/10.5281/zenodo.7476474

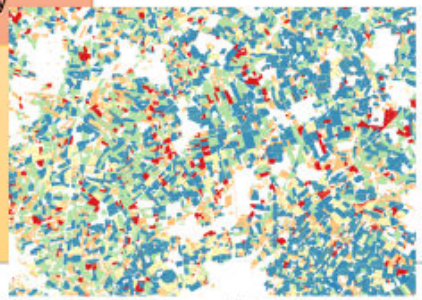
Spatial

Resolution: not applicable

Extent: 17 Member States across EU (in green)

Granularity: Parcel level (sub-communal)

Coordinate System: Projected (country specific)



Temporal

Resolution: Yearly

Extent: 2015-2023

Frequency: Annual

Description:

Land Parcel Identification Systems (LPIS) data cover the essential part of the spatial information used to support Integrated Administration and Control System (IACS) applications under the common agricultural policy (CAP). It is individually collected from the member states and contains georeferenced blocks of agricultural parcels that have been identified and are eligible for EU aid application. This data usually shows the main crop for a certain year as the subsidy is granted with respect to that.

The EuroCrops data combines all publicly available self-declared crop reporting datasets from countries of the European Union. As the raw data obtained from the countries does not come in a unified, machine-readable taxonomy, the EuroCrops data rely on a Hierarchical Crop and Agriculture Taxonomy (HCAT) that harmonises all declared crops across the European Union.

How to cite

Dataset: Schneider, M., Chan, A., & Körner, M. (2023). EuroCrops [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10118572

Data File Size

9.3 GB (for the 17 countries listed below), between 100 MB and 2.6 GB for one country

File format: zip file (containing shp) and csv

USER COSTS

Free & open Access

for Austria, Belgium, Czech Republic, Germany, Denmark, Estonia, Spain, France, Croatia, Lithuania, Latvia, Netherlands, Portugal, Romania, Sweden, Slovenia, Slovakia

Data Infrastructure

Infra needed: GIS (QGIS for example) or software environment for statistical computing (R for example)

Data repositories / storage needed: Datasets are hosted on Zenodo and made available for free.

Data Governance & Management

Staff costs: Data can be easily downloaded. Its use requires a certain expertise in the world of geographic information and the structure of LPIS data (low level GIS Technician ~25,000€ + charges per year)

Data analysis needed: None for simple visualisation of crops

Quality assurance: Data and annotations are derived from self-declarations by farmers receiving subsidies under the *common agricultural policy (CAP)* of the *European Commission (EC), with the usual caveat of possible errors*. Moreover, this data does not represent the complete state of agricultural crops that are cultivated in each member state as fields from farmers that do not receive subsidies from the CAP are not represented.

Data documentation and / or assistance in data use

Github

Wiki

Zenodo

Compilation of the fiche by Tristan Berchoux (IAMM)

12.3. Climate

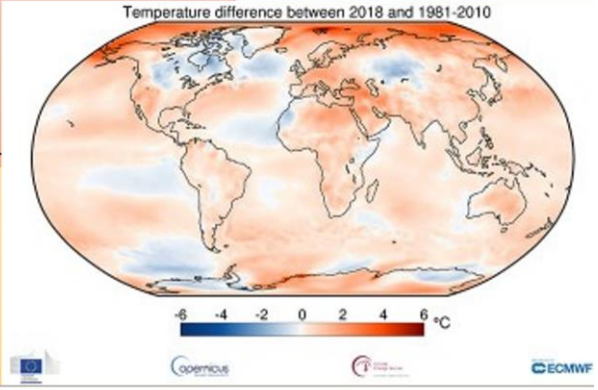
ERA5 historic climate variables By ECMWF
URL: https://www.mapillary.com/

Indicator class: Residential;

Data
Class: gridded
Type : API or download
License: Licence to use Copernicus Products (available here)
DOI: https://doi.org/10.24381/cds.143582cf

Spatial
Resolution NA
Extent: Global
Granularity: 32km
Coordinate System: Decimal Degrees Lat/lor

Temporal
Resolution: hourly
Extent : 1940 -
Frequency: Daily



Description :
ERA5 is the fifth generation ECMWF atmospheric reanalysis of the global climate covering the period from January 1940 to present. It is produced by the Copernicus Climate Change Service (C3S) at ECMWF and provides hourly estimates of a large number of atmospheric, land and oceanic climate variables. The data cover the Earth on a 31km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80km. ERA5 includes an ensemble component at half the resolution to provide information on synoptic uncertainty of its products.

How to cite
Dataset:
For ERA5-complete:

Hersbach , H., Bell, B., Berrisford , P., Hirahara, S., Horányi, A., Muñoz-Sabater , J., Nicolas, J., Peubey , C., Radu , R., Schepers , D., Simmons , A., Soci, C., Abdalla, S., Abellan, X., Balsamo , G., Bechtold , P., Biavati, G., Bidlot , J., Bonavita , M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming , J., Forbes, R., Fuentes, M., Geer, A., Haimberger , L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu , C., Radnoti , G., de Rosnay , P., Rozum , I., Vamborg , F., Villaume, S., Thépaut , J-N. (2017): Complete ERA5 from 1940: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service (C3S) Data Store (CDS). DOI: 10.24381/cds.143582cf (Accessed on DD-MMM-YYYY)

Data File Size
Global dataset: The whole historic dataset for all the multiple variables is in the order of petabytes. Single variables for one time for the globe are in the order of few gigabytes.

File format: GRIP, NETCDF

USER COSTS

Free & open Access

Data Infrastructure

GIS software, depending on the size – more or less sophisticated data infrastructure is needed.

Data Governance & Management

Data extraction and handling for large amounts. For smaller areas and time frames, no particular hardware is needed.

1 PM to cater the data for a specific local purpose.

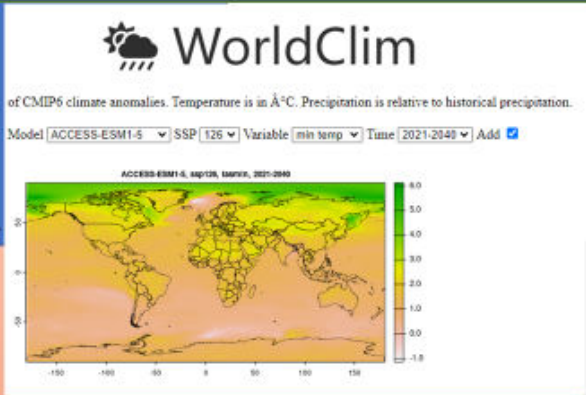
Data documentation and / or assistance in data use
https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation

Compilation of the fiche by Ian McCallum & Martin Hofer (IIASA)

Global climate and weather data — By Worldclim
URL: https://www.worldclim.org/data/index.html

Indicator class: Environmental;

Data
Class: Model-based
Type: Geo-tiff
License: Data freely available for academic and other non-commercial use.



Spatial
Resolution: 10, 5, 2.5 minutes, 30 seconds
Extent: Global
Granularity: Gridded
Coordinate System: Geographic WGS84

Temporal
Resolution: Historical climate data, Historical monthly weather data, Future climate data,
Extent: 1970 – 2000, 1960-2018, The monthly values were averages over 20 year periods (2021-2040, 241-2060, 2061-2080, 2081-2100).
Frequency: version 2.1 released in 2020.

Description:
WorldClim is a database of high spatial resolution global weather and climate data. These data can be used for mapping and spatial modeling. The data are provided for use in research and related activities; and some specialized skill and knowledge is needed to use them (here is some help).

How to cite
Fick, S.E. and R.J. Hijmans, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37 (12): 4302-4315.

Data File Size
Global dataset zipped: The largest files for the 1km2 global products on the order of 4GB. The 10 minute global grid is by compariosn only 40 MB.

File format: .zip, .tif

COSTS

Free & open Access

Data Infrastructure
Infrastructure needed: Minimal infrastructure needed. These datasets are relatively small, and can simply be downloaded directly, or can be found e.g. via R. Packages (geodata)

Data Governance & Management
Data analysis needed: The files come ready to use with little reprocessing required. Nonetheless, basic GIS and climate expertise is required to fully undertand and work with the files.

Staff costs: with basic data handling skills this data can be acquired and processed for ingestion into models etc. in a short time – meaning costs are negligible.

Data documentation and / or assistance in data use
https://www.worldclim.org/data/worldclim21.html
Good documentation is available at worldclim.org.
Furthermore, the pre-processing effort that has gone into producing this dataset ensures that it is easy to work with.

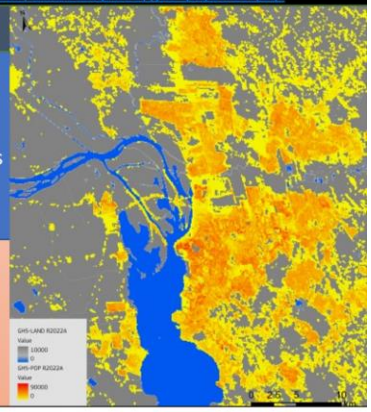
Compilation of the fiche by Ian McCallum & Martin Hofer (IIASA)

12.4. Demography

GHS-POP

By Joint Reseach Centre (JRC)
URL: https://ghsl.jrc.ec.europa.eu/ghs_pop2023.php

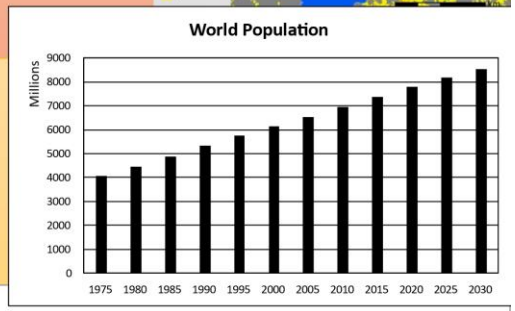
Indicator class: Demography, residential



Data
Class:Census-based + Remote -sensing
Type: Raster, GeoTIFF
License: Full Open Access (Creative Commons Attribution 4.0 International License)
DOI: 10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE

Spatial
Resolution: 100m, 1 km, 3 arcsec, 30 arcsec
Extent: Global
Granularity: Gridded
Coordinate System: World Mollweide (ESRI: 54009) and WGS 1984 (EPSG: 4326)

Temporal
Resolution: Every 5 years
Extent: 1975-2030
Frequency: >1 years



Description:
The spatial raster dataset (GHS-POP – R2023A) depicts the distribution of residential population, expressed as the number of people per cell (float64: [0, Nodata: -200). Residential population estimates between 1975 and 2020 in 5-year intervals and projections to 2025 and 2030 derived from CIESIN GPWv4.11 were disaggrega from census or administrative units to grid cells, informed by the distribution, vol and classification of built-up as mapped in the Global Human Settlement Layer (GHS global layer per corresponding epoch. The main characteristics of this dataset ar listed in the citations below.

How to cite
Dataset:
Schiavina M., Freire S., Carioli A., MacManus K. (2023) :
GHS-POP R2023A - GHS population grid multitemporal (1975 -2030). European Commission, Joint Research Centre (JRC)
PID: http://data.europa.eu/89h/2ff68a52 -5b5b -4a22 -8f40 -c41da8332cfe
doi: https:// doi.org /10.2905/2FF68A52 -5B5B -4A22 -8F40 -C41DA8332CFE
Concept & Methods :
Freire S., MacManus K., Pesaresi M., Doxsey -Whitfield E., Mills J. (2016)
Development of new open and free multi -temporal global population grids at 250 m resolution. Geospatial Data in a Changing World; Association of Geographic Information Laboratories in Europe (AGILE) , AGILE 2016

Data File Size (100m resolution)
Global grid: 5.41GB ZIP file (8.93GB uncompressed folder of TIFF, XLSX, and PDF)
Tile grid: up to 100+MB ZIP (up to 100+MB uncompressed of TIFF, XLSX, and PDF)
File format: .zip, .tif, .xlsx, .pdf

USER COSTS

Free & open Access

Data Infrastructure
Infra needed (software, hardware): Any raster capable GIS software
Data repositories / storage needed: ~9GB for 100m global grid

Data Governance & Management
Staff costs: None.
Data analysis needed: None.
Quality assurance: See §2.5 in the report linked below

Data documentation and / or assistance in data use

The complete information about the GHSL main products can be found in the GHSL Data Package 2023 report.

Compilation of the fiche by Hal Voepel (University of Southampton)

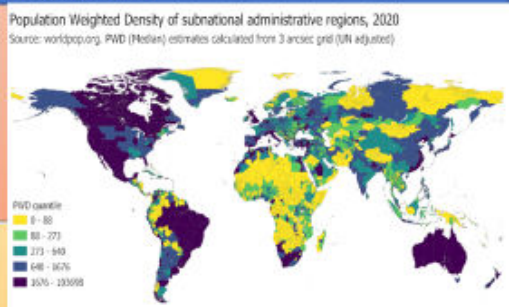
WorldPop – Population Weighted Density

By University of Southampton (UoS)
URL: https://hub.worldpop.org/project/categories?id=20

Indicator class: Demography, residential

Data
Class: Census-based + Remote-sensing and GIS reference data
Type: Raster
License: Full Open Access
DOI by resolution and coverage:
10.5258/SOTON/WP00701 (100m national)
10.5258/SOTON/WP00702 (1km national)
10.5258/SOTON/WP00703 (100m subnational)
10.5258/SOTON/WP00704 (1km subnational)

Spatial
Resolution: 100m, 1 km
Extent: Global
Granularity: Gridded
Coordinate System: World Mollweide (ESRI: 54009) and WGS 1984 (EPSG: 4326)



Temporal
Resolution: Every 5 years
Extent: 2005-2020
Frequency: >1 years

Description: Population Weighted Density (PWD) is an alternative to conventional approaches to population density that is arguably better suited to some types of research in fields of social science and epidemiology. In this release WorldPop publishes what we believe may be the first set of global estimates for PWD, which we offer at national and subnational levels since 2000. Data are available at both national and subnational level. Please make sure you read the Population Weighted Density overview page before choosing and downloading a PWD dataset (see link to Methods at the end of this document).

How to cite to be updated
Dataset:
Robin Edwards, Maksym Bondarenko, Andrew J. Tatem and Alessandro Sorichetta. Unconstrained subnational Population Weighted Density in 2000, 2005, 2010, 2015 and 2020 [100m resolution]. WorldPop, University of Southampton, UK. doi:10.5258/SOTON/WP00703
Concept & Methods:
Freire, Sergio; MacManus, Kytt; Pesaresi, Martino; Doxsey-Whitfield, Erin; Mills, Jane (2016): Development of new open and free multi-temporal global population grids at 250 m resolution. Geospatial Data in a Changing World; Association of Geographic Information Laboratories in Europe (AGILE). AGILE 2016.

Data File Size
National grid: < 1MB compressed file
Subnational grid: < 1MG compressed file
File format: CSV, SHP, GeoJSON

USER COSTS

Free & open Access

Data Infrastructure
Infra needed (software, hardware): Any raster capable GIS
Data repositories / storage needed: < 1MB

Data Governance & Management
Staff costs: None.
Data analysis needed: None.
Quality assurance: See cited references and methods below

Data documentation and / or assistance in data use

Methods: https://www.worldpop.org/methods/pwd/

Compilation of the fiche by Hal Voepel (University of Southampton)

WorldPop – Population Density

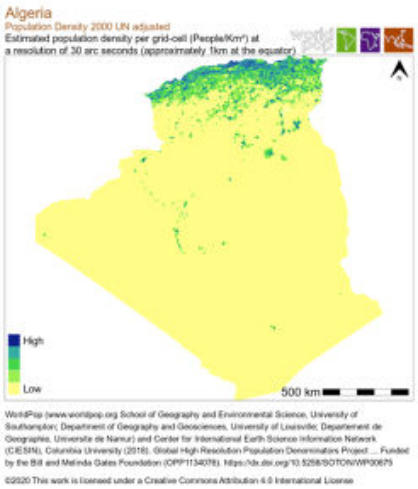
By University of Southampton (UoS)
URL: https://hub.worldpop.org/project/categories?id=18

Indicator class: Demography, residential

Data
Class: Census-based + Remote-sensing and GIS reference data
Type: Raster
License: Full Open Access
DOI by resolution and method:
10.5258/SOTON/WP00674 (1km individual countries)
10.5258/SOTON/WP00675 (1km individual countries, UN adjusted)

Spatial
Resolution: 1 km
Extent: Country-level
Granularity: Gridded
Coordinate System:
WGS 1984 (EPSG: 4326)

Temporal
Resolution: Annually
Extent: 2000-2020
Frequency: >1 years



WorldPop (www.worldpop.org School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by the Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00675

©2020 This work is licensed under a Creative Commons Attribution 4.0 International License

Description: Population Weighted Density (PWD) is an alternative to conventional approaches to population density that is arguably better suited to some types of research in fields of social science and epidemiology. In this release WorldPop publishes what we believe may be the first set of global estimates for PWD, which we offer at national and subnational levels since 2000. Data are available at both national and subnational level. Please make sure you read the Population Weighted Density overview page before choosing and downloading a PWD dataset (see link to Methods at the end of this document).

How to cite to be updated
Dataset:
WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076).
Concept & Methods:
Sorichetta, A., Hornby, G., Stevens, F. *et al.* High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Sci Data* **2**, 150045 (2015). https://doi.org/10.1038/sdata.2015.45

Data File Size
National grid: < 1MB compressed file
Subnational grid: < 1MG compressed file
File format: GeoTIFF, ASCII XYZ

USER COSTS

Free & open Access

Data Infrastructure
Infra needed (software, hardware): Any raster capable GIS
Data repositories / storage needed: < 1MB

Data Governance & Management
Staff costs: None.
Data analysis needed: None.
Quality assurance: See cited references and methods below

Data documentation and / or assistance in data use

Methods: https://www.worldpop.org/methods/pwd/

Compilation of the fiche by Hal Voepel (University of Southampton)

WorldPop – Age & Sex Structures By University of Southampton (UoS)

URL: https://hub.worldpop.org/project/categories?id=8

Indicator class: Demography, residential

Data

Class: Census-based + Remote-sensing and GIS reference data

Type: Raster

License: Open Access (Creative Commons Attribution 4.0 International License)

DOI by group, method, and resolution:

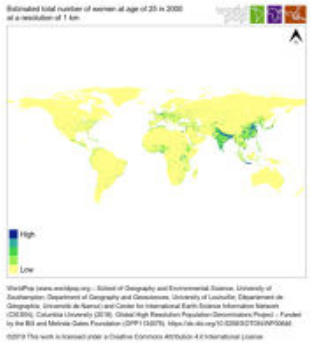
10.5258/SOTON/WP00696 Constrained countries in 2020 (100m)
10.5258/SOTON/WP00699 Constrained countries in 2020 UN adjusted (100m)
10.5258/SOTON/WP00732 School age population (1km)
10.5258/SOTON/WP00654 Unconstrained global mosaic 2000-2020 (1km)
10.5258/SOTON/WP00646 Unconstrained single countries 2000-2020 (100m)
No DOI assigned to data: Women of childbearing age (15-49yrs) in 2015 (1km)

Spatial

Resolution: 100m, 1 km
Extent: Global and country-level
Granularity: Gridded
Coordinate System: WGS 1984 (EPSG: 4326)

Temporal

Resolution: Every 5 years
Extent: 2000-2020 (unconstrained), 2020 (constrained and school age), and 2015 (women of childbearing age)
Frequency: >1 years



Description: WorldPop produces a variety of gridded age & sex structures. A description of the modelling methods used for age and sex structures can be found in Tatem et.al. (2013) and Pezzulo et.al. (2017). Both top-down constrained and unconstrained versions of the dataset are available, and the difference between the two methods are outlined here. The datasets represent the outputs from a project focused on construction of consistent 100m resolution population count datasets for all countries by male/female in 5 year classes with a final class for infants defined as less than one years old.

How to cite

Concept & Methods:

Pezzulo, C., Hornby, G., Sorichetta, A. *et al.* Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. *Sci Data* **4**, 170089 (2017). https://doi.org/10.1038/sdata.2017.89

Tatem, A.J., Garcia, A.J., Snow, R.W. *et al.* Millennium development health metrics: where do Africa's children and women of childbearing age live?. *Popul Health Metrics* **11**, 11 (2013). https://doi.org/10.1186/1478-7954-11-11

Data File Size (100m)
Global grid: 3.05GB TIFF
Tile grid: 5.38MB TIFF
File format: GeoTIFF

USER COSTS

Free & open Access

Data Infrastructure

Infra needed (software, hardware): Any raster capable GIS
Data repositories / storage needed: Up to 3+GB global tile

Data Governance & Management

Staff costs: None.
Data analysis needed: None.
Quality assurance: See population methods below

Data documentation and / or assistance in data use

Methods: Population count methods

Compilation of the fiche by Hal Voepel (University of Southampton)

WorldPop – Population Counts

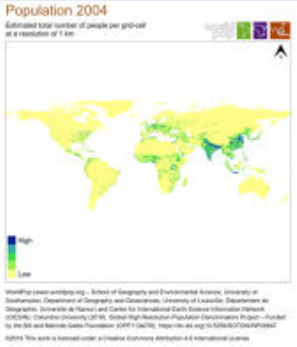
By University of Southampton (UoS)
URL: https://hub.worldpop.org/project/categories?id=3

Indicator class: Demography, residential

Data
Class: Census-based + Remote-sensing and GIS reference data
Type: Raster
License: Full Open Access (Creative Commons Attribution 4.0)
DOI by resolution and method:
10.5258/SOTON/WP00682 (100m constrained countries)
10.5258/SOTON/WP00683 (100m constrained countries, UN adjusted)
10.5258/SOTON/WP00647 (1km unconstrained global mosaic)
10.5258/SOTON/WP00645 (100m unconstrained countries)
10.5258/SOTON/WP00670 (1km unconstrained countries)
10.5258/SOTON/WP00660 (100m unconstrained countries, UN adjusted)
10.5258/SOTON/WP00671 (1km unconstrained countries, UN adjusted)
wopr.worldpop.org/ (Bespoke methods for individual countries)

Spatial
Resolution: 1 km
Extent: Country-level
Granularity: Gridded
Coordinate System:
WGS 1984 (EPSG: 4326)

Temporal
Resolution: Annually
Extent: 2000-2020 (Unconstrained)
2020 (Constrained)
Frequency: >1 years



Description: Bespoke methods used to produce da
are available through the WorldPop Open Population Repository (WOPR) link above under Data. These are 100m resolution gridded population estimates using the bottom-up or top-down methods described here. The remaining datasets linked above use the top-down method, with either constrained or unconstrained top-down aggregation methods. Be sure to read the Top-Down estimate modelling overview page to see which dataset is appropriate for your purposes. Datasets are available as GeoTIFF or ASCII XYZ at 100m and 1km resolution.

How to cite
Dataset: Bondarenko M., Kerr D., Sorichetta A., and Tatem, A.J. 2020. Census/projection-disaggregated gridded population datasets for 51 countries across sub-Saharan Africa in 2020 using building footprints. WorldPop, University of Southampton, UK. doi:10.5258/SOTON/WP00682 (see citations for each individual link in blue Data section above)
Concept & Methods: *Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLoS ONE 10(2): e0107042. https://doi.org/10.1371/journal.pone.0107042*

Data File Size Range
National grid: < 1.08GB TIFF
Country grid: < 13MG TIFF
File format: GeoTIFF, ASCII XYZ

USER COSTS

Free & open Access

Data Infrastructure
Infra needed (software, hardware): Any raster capable GIS
Data repositories / storage needed: < 1MB

Data Governance & Management
Staff costs: None.
Data analysis needed: None.
Quality assurance: See cited references and methods below

Data documentation and / or assistance in data use

Methods: Population Counts

Compilation of the fiche by Hal Voepel (University of Southampton)

HiRes Population Density

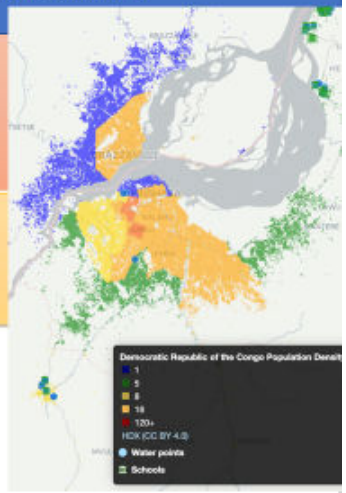
By Meta, Data for Good
URL: https://dataforgood.facebook.com/

Indicator class: Demography, residential

Data
Class: Census-based + Remote-sensing
Type: Raster, GeoTIFF
License: Full Open Access (Creative Commons Attribution 4.0)
DOI: N/A

Spatial
Resolution: 30m
Extent: Global
Granularity: Gridded by country
Coordinate System: WGS 1984 (EPSG: 4326)

Temporal
Resolution: Varies
Extent: Varies
Frequency: ~annually



Description:
These high-resolution maps estimate the number of people living within 30-meter grid tiles in nearly every country around the world. Additionally, our datasets provide insights on the distribution of certain populations within each country, including the number of children under five, the number of women of reproductive age, as well as young and elderly populations, at unprecedentedly high resolutions. These maps aren't built using Facebook data and instead rely on combining the power of machine vision AI with satellite imagery and census information. These high-resolution maps can provide the necessary insights for health organizations to allocate resources and control outbreaks. There are many use cases for understanding the demographics of various populations – demographics can help organizations target vaccination campaigns, plan infrastructure, and distribute resources.

How to cite
Dataset:
Varies by location.
Concept & Methods:
Tiecke, T.G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E.B., & Dang, H.H. (2017). Mapping the world population one building at a time. ArXiv:1712.05839.

Data File Size (30m resolution)
Tile grid: Varies, depending on country areal size
File format: CSV, GDAL, GeoTIFF, JSON, ZIP

USER COSTS

Free & open Access

Data Infrastructure

Infra needed (software, hardware): Any raster capable GIS software

Data repositories / storage needed: <100MB for 30m UK data

Data Governance & Management

Staff costs: None.

Data analysis needed: None.

Quality assurance: See individual dataset documentation

Data documentation and / or assistance in data use

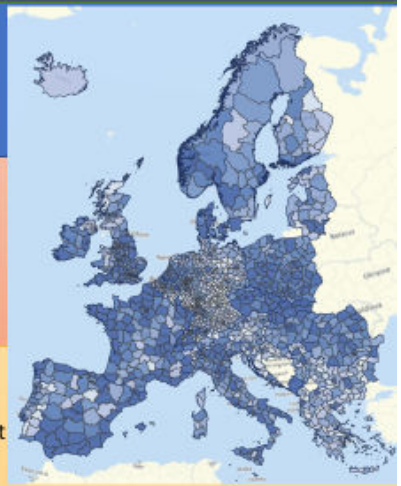
See documentation for each dataset here:
https://data.humdata.org/organization/meta?q=population%20density&sort=if(gt(last_modified%2Creview_date)%2Clast_modified%2Creview_date

Compilation of the fiche by Hal Voepel (University of Southampton)

Population Series	By ARDECO URL: https://urban.jrc.ec.europa.eu/ardeco?lng=en

Indicator class: Demography, administrative level

Data
Class: Census-based + growth model
Type: CSV with NUTS/LAU primary key
License: Full Open Access
DOI: N/A



Spatial
Resolution: NUTS / LAU
Extent: EU27 + others, 50 countries
Granularity: NUTS/LAU
Coordinate System: None

Temporal
Resolution: Annual
Extent: 1960 to current + 2 years forecast
Frequency: updated biannually

Description: ARDECO is the "Annual Regional Database of the European Commission". ARDECO is maintained by the Joint Research Centre, in close collaboration with the Directorate General for Regional and Urban Policy. Data are a set of long time series based on the ESA 2010 system for the last period and on ESA 95 and ESA 79 for the earlier years, where growth rates are used to gap-fill missing values. The database covers, at an annual scale, 27 EU plus 23 Non-EU countries for a total of 50 countries (see metadata for details). Data are in tabular form with NUTS/LAU primary keys. ARDECO is updated biannually with major updates released early March (after the release of updated regional accounts by Eurostat) and in May and November, following the release of Directorate General ECFIN's short-term economic forecasts. ARDECO data are publicly available and free of charge.

How to cite
Dataset: Metadata
https://economy-finance.ec.europa.eu/system/files/2022-10/Reference%20Metadata%20AMECO_September%202022.pdf
Concept & Methods:

Data File Size
Tabular: 1.5MB
File format: CSV

USER COSTS

Free & open Access

Data Infrastructure

Infra needed (software, hardware): Excel or GIS software

Data repositories / storage needed: 1.5MB

Data Governance & Management

Staff costs: None.

Data analysis needed: None.

Quality assurance: see documentation

Data documentation and / or assistance in data use

See documentation for each dataset here: https://economy-finance.ec.europa.eu/economic-research-and-databases/economic-databases/ameco-database_en

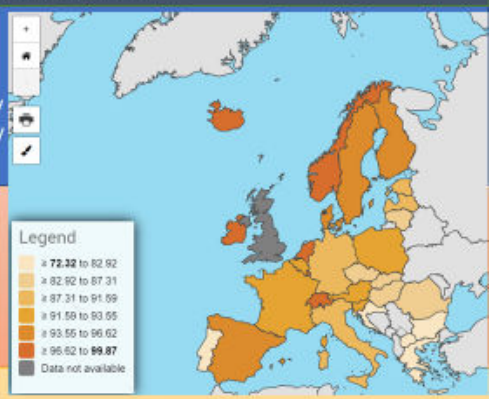
Compilation of the fiche by Hal Voepel (University of Southampton)

12.5. Digitalisation

Rural household internet access — By eurostat
https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_in_

Indicator class: Residential, Digitalization

Data
Class: Statistics
Type: Tabular
License: https://ec.europa.eu/eurostat/web/main/about-us/policies/copyright



Spatial
Resolution: not applicable
Extent: EU
Granularity:
Coordinate System: Projected (country specific)

Temporal
Resolution: Yearly
Extent: 2002 - 2022
Frequency: Annual

Description:

The EU survey on the use of Information and Communication Technologies (ICT) in households and by individuals is an annual survey conducted since 2002 aiming at collecting and disseminating harmonised and comparable information on the use of ICT in households and by individuals.

Data presented in this domain is collected on a yearly basis by the National Statistical Institutes and are based on Eurostat's annual model questionnaire. This questionnaire is updated each year to reflect the evolving situation of information and communication technologies

How to cite
Dataset:
Eurostat. ICT usage in households and by individuals (isoc_i): reference metadata in Euro SDMX Metadata Structure (ESMS).
"This study/report/paper is based on data from Eurostat, EU Statistics Rural household internet access in [reference year(s)]."
Concept & Methodology:
https://ec.europa.eu/eurostat/cache/metadata/en/isoc_i_esms.htm

Data File Size
Dataset: 20,7 kt

File format: .xlxs, tsv., cvs.

USER COSTS

Free & open Access
Full open access

Data Infrastructure

Infra needed: Software environment of statistics in XLS, CVS, TSV and other format

Data repositories: FTPStore

Data Governance & Management

Staff costs: Basic statistical skill or more advanced depending on usage

Data analysis needed: Depending on the analysis. Charts and graphs can be viewed directly.

Quality assurance: The content of metadata and quality reports followsthe Commission Implementing Regulation (EU) 2019/2180 of 16 December 2019 specifying the detailed arrangements and content for the quality reports pursuant to Regulation (EU) 2019/1700 of the European Parliament and of the Council.

Data documentation and / or assistance in data use

https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_in_h__custom_8667150/default/table?lang=en

https://ec.europa.eu/eurostat/cache/metadata/en/isoc_i_esms.htm

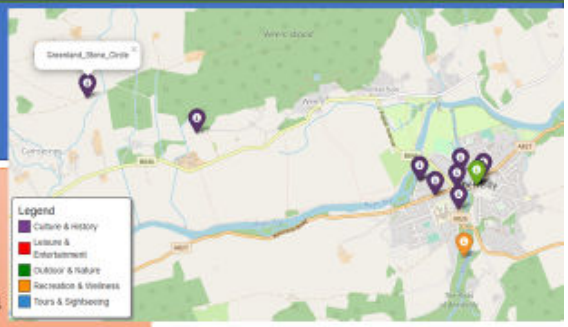
Compilation of the fiche by Lauri Niskanen (LUKE)

12.6. Economic Development

Hotel, Restaurant, Attractions and visitors statistics (reviewers) By Tripadvisor
URL: https://www.tripadvisor.com/developers

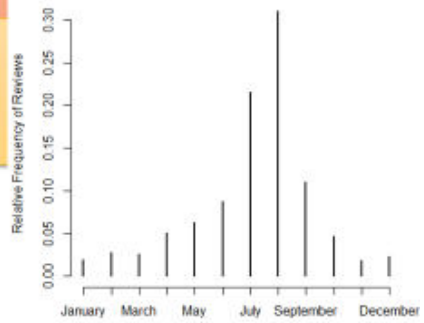
Indicator class: Residential;

Data
Class: GIS & tabular
Type: API
License: Proprietary API
Use Licence (see here)



Spatial
Resolution: NA
Extent: Global
Granularity: street-level
Coordinate Sys.: WGS 84

Temporal
Resolution: Monthly
Extent: 2015 -
Frequency: realtime



Description:
TripAdvisor is a popular travel website that offers user-generated reviews and comparisons for hotels, restaurants, and attractions worldwide. Launched in 2000, it has become a go-to platform for travelers to plan and book their trips, as well as share experiences, helping others make informed travel decisions. It's one of the largest travel communities globally, impacting travel industry dynamics significantly.

How to cite Dataset:-

Data File Size
100 Bytes per review
Europe has 150 million reviews (estimate)

File format: geojson, SQL

USER COSTS

Proprietary API

Data Infrastructure

Database for all attractions and reviews for Europe would be around 20 GB. A relation database (SQL) with locations, reviews users and tags should suffice.

Data Governance & Management

The tags for the locations need to be consolidated into rural functional areas. Manually or through the use of large language models.

Getting location data of activities/hotels via the API 1PM.

With additional analysis the reviews can inform tourist statistics. The data needs to be calibrated against other tourism statistics first. (2PM)

Data documentation and / or assistance in data use
https://tripadvisor-content-api.readme.io

Compilation of the fiche by Ian McCallum & Martin Hofer (IIASA)

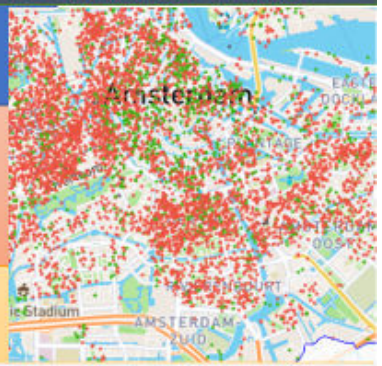
Airbnb/Hotel locations and capacity	By Airbnb via insideairbnb.com URL: http://insideairbnb.com/data-requests

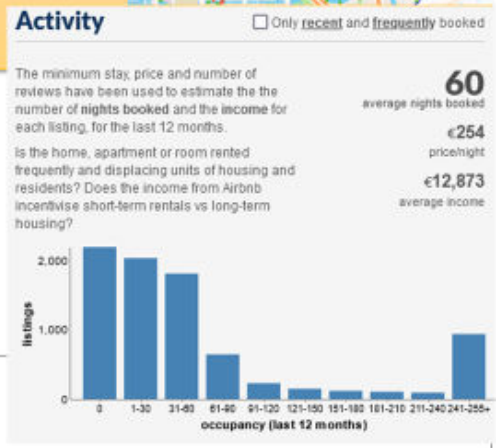
Indicator class: Residential;

Data
Class: GIS points & tabular
Type: table

Spatial
Resolution: NA
Extent: Global
Granularity: street-level
Coordinate System: WGS 84

Temporal
Resolution: 30, 60, 90, 365 days
Extent: 2015
Frequency: annual





Description:
From their website:
"Inside Airbnb is a mission driven project that provides data and advocacy about Airbnb's impact on residential communities." Insideairbnb scrapes Airbnb data and gives access to the public. They provide the location (long/lat), room and accommodation type, info on the accommodation, number of available rooms in the next 30, 60, 90 and 365 days, from the last scraping.
Access to the data varies. Many cities are available for free. Archived data or new locations require a budget, depending on how much the requesters mission aligns with Inside Airbnb's mission.

How to cite
Dataset: We need to discuss how/if we can publish their data.

Data File Size
The city of Amsterdam for 2023 has 23MB.
File format: csv/SQL, geojson

USER COSTS

Sold or shared on request

Data Infrastructure

A GIS capable database for all of Europe would be fairly easily manageable. E.g. Postgres with a GIS plugin.

Data Governance & Management

GIS and dataThe data comes very well cleaned and prepared. Not much work needed.

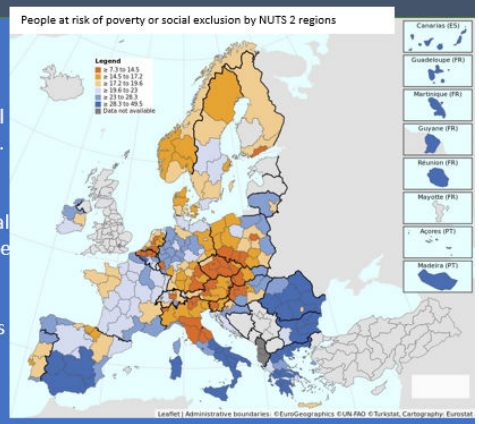
base knowledge needed. 1PM.

Data documentation and / or assistance in data use
Data Dictionary

Compilation of the fiche by Ian McCallum & Martin Hofer (IIASA)

EU-SILC microdata By National Statistical Institutes of EU Member

Indicator class: Productive



Data

Class: Alphanumeric

Type: Tabular

License:EU-SILC data belongs to National Statistical Istitutes of EU Member States. Licensing varies between countries. As coordination body, Eurostat performs a first evaluation of each research proposal consults with NSI before the data is made accessible to the researchers. The datasets contain partially anonymised microdata delivered as scientific use files that can be dowloaded and used by authorised researchers under restrictive Terms of Use. There is also a public version of microdata files.

Temporal

Resolution: Annual, 5-annual (see note)

Extent: 2004-2022

Frequency: Yearly

Spatial

Extent: EU (27 countries) + UK, CH, IS, NO, RS.

Granularity: IESS Regulation (EU:2019/1700) states that reliable statistics relating to persons and households based on individual samples should be provided at national and at regional level (NUTS-2). However, the availability of data at NUTS-2 level is limited by effective sample sizes. As regards the estimated ratio of people at risk of poverty or social exclusion to the population in each NUTS-2 region, these requirements are not compulsory for NUTS-2 regions with fewer than half million inhabitants. Similarly, NUTS-1 regions with under 100 000 inhabitants are exempted from the requirement. Even in most recent EU-SILC surveys, MS continue to report at-risk-of-poverty or social exclusion data at NUTS-1 level, in some cases, only at national level.

Description:

The EU Statistics on Income, Social Inclusion and Living Conditions (EU-SILC) is a cross-sectional and longitudinal sample survey, coordinated by Eurostat, based on data from the EU MS. The EU-SILC survey covers objective and subjective aspects of these themes in both monetary and non-monetary terms for households and individuals. The EU-SILC provides two types of annual survey data:

- Cross-sectional data with variables on income, poverty, social exclusion and other living conditions
- Longitudinal data pertaining to individual-level changes over time, observed periodically over a 4-year period

How to cite

Researchers must cite the source of EU-SILC data as follows: "This study/report/paper is based on data from Eurostat, EU Statistics on Income and Living Conditions [reference year(s)]." The following disclaimer must be added: "The responsibility for all conclusions drawn from the data lies entirely with the author(s)."

Data File Size: Variable and depends on the type of the requested format (cross-section vs panel version), as well as on the number countries and number of years requested. Tentatively, a complete microdata file for the EU can be up 3GB in size and is delivered as stand-alone 7z-compressed files including the csv tables.

USER COSTS

Free & open Access

Data Infrastructure

The processing of Eurostat microdata does not require a very sophisticated / particularly powerful computing capacity. However, the strict security and confidentiality requirements set by Eurostat may increase data management costs. In order to fulfil the Terms of Use set by Eurostat, the following requirements need to be fulfilled:

- *The confidential data for scientific purposes must be stored on a password-protected computer.*
- *Access to the data must be restricted to authorised researchers named in the research proposal.*
- *The intermediate results of analysis containing confidential data must be stored in a protected environment.*
- *The confidential data for scientific purposes must be solely used on the premises of the research entity.*

These terms not only prevent a distributed use of the microdata, but they also encourage the establishment of a secured IT environment for microdata manipulation, with devoted working stations with physical access restrictions.

Data Governance & Management

No specific staff costs related to the use of EU-SILC microdata files. Still, the EU-SILC SUFs provide access to complex data survey. Adequate use of these files is only possible for researchers with a robust knowledge of statistics and data management capacity.

Data documentation and / or assistance in data use

Upon download of the EU-SILC data, the users are provided with several documents describing the SUF files.

Additionally, number of specialised forums and conferences gather social researchers who use EU-SILC in research:

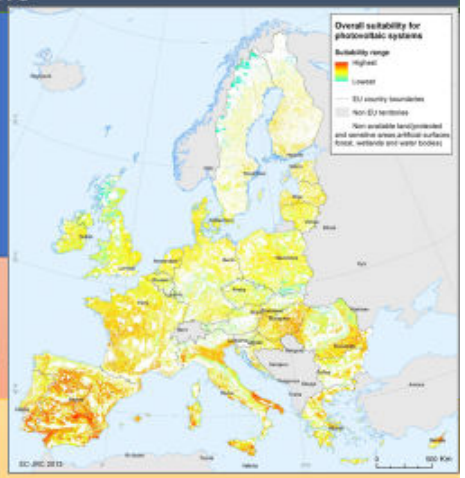
- Additionally, the EU Platform "Collaboration in Research and Methodology for Official Statistics" curates a Forum where users can pose questions related to EU-SILC scientific use files: https://cros-legacy.ec.europa.eu/EU-SILC-SUF/forum_en
- EU-SILC user conference: https://www.statistik.at/en/about-us/events/eu-silc-user-conference-2023

Compilation of the fiche by: Carlos Tapia (Nordregio)

12.7. Energy

Suitability map for solar energy (PV)s — By JRC
URL: https://energy-industry-geolab.jrc.ec.europa.eu/

Indicator class: Energy,Productive



Data
Class: Heterogenous
Type: Raster
License: The data in the Energy & Industry Geography Lab are free forpublic use
DOI:https://doi.org/10.1016/j.enpol. 2015.10.004

Resolution: 1 km
Extent: EU
Granularity:
Coordinate System: Spherical/Web Mercator EPSG:3857

Temporal
Resolution: Every five years
Extent: 2010-2050
Frequency: Irregular

Description:

SOLAR (PV and CSP) data set of ENSPRESO - an open data, EU-28 wide, transparent and coherent database of wind, solar and biomass energy potential. Solar radiation can be converted into sustainable-produced electricity by using photovoltaic (PV) technology. Large-scale photovoltaic (PV) systems provide significant environmental benefits and advantages when compared to conventional, non-renewable energy sources, the reduction of greenhouse gas emissions, and the reuse of marginal lands being two key examples (IPCC, 2011)

- A European suitability map for the solar energy (PV) systems deployment is created.
- PV systems can contribute to sustainable energy production in many regions in EU.
- There is no correlation among the EU investment and the suitability in solar energy.
- Using marginal lands to place PV systems might avoid the uptake of agricultural land.
- Validation of the EU suitability map demonstrated a satisfactory degree of accuracy.

How to cite
Dataset:
European Commission, Joint Research Centre (JRC) (2019): ENSPRESO - SOLAR - PV and CSP. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/18eb348b-1420-46b6-978a-fe0b79e30ad3
Concept & Methodology:
Castillo, C. P., e Silva, F. B., & Lavalle, C. (2016). An assessment of the regional potential for solar power generation in EU-28. Energy policy, 88, 86-99.
https://doi.org/10.1016/j.enpol.2015.10.004

Data File Size
Not specified

USER COSTS

Free & open Access
Full open access (Creative Commons 4.0)

Data Infrastructure
Infra needed: GIS, statistical or remote sensing software
Data repositories / storage needed: Yes

Data Governance & Management
Staff costs: Basic GIS or statistical skills, more arvanced depending on usage
Data analysis needed: No
Quality assurance: None specicified

Data documentation and / or assistance in data use
https://data.jrc.ec.europa.eu/dataset/18eb348b-1420-46b6-978a-fe0b79e30ad3#dataaccess

Compilation of the fiche by Lauri Niskanen (LUKE)

12.8. Health

Healthcare services location & number of beds in Europe — By EuroStat — URL: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/healthcare-services

Indicator class: Health, Residential

Data
Class: Point locations, In-situ
Type: Vector
License: Open Access (Eurostat's general copyright notice and license)

Spatial
Resolution: Point locations
Extent: EU + Switzerland/Norway
Granularity: National
Coordinate System: WGS 1984 (EPSG: 4326

Temporal
Resolution: Annual
Extent: 2017-2022
Frequency: >1 years

Healthcare service locations in Europe

Description: The dataset contains information on main healthcare services considered to be 'hospitals' by Member States. The definition varies slightly from country to country, but roughly includes the following: "*'Hospitals' comprises licensed establishments primarily engaged in providing medical, diagnostic, and treatment services that include physician, nursing, and other health services to in-patients and the specialised accommodation services required by inpatients. Hospitals may also provide out-patient services as a secondary activity. Hospitals provide in-patient health services, many of which can only be provided using the specialised facilities and equipment that form a significant and integral part of the production process. In some countries, health facilities need in addition a minimum size (such as number of beds) in order to be registered as a hospital.*" In some cases, facilities without in-patient services may be included.
A variety of data sources are available at Member State level with different quality criteria: level of detail, timeliness and update frequency, etc. This methodology aims to progressively build pan-European datasets from Member State official data and maintain them in a sustainable manner. Different data sources could be selected depending on the needs and the cases but priority is given to data sourced from Health Ministries, or other official outlets indicated by Ministries. Data is semantically harmonised with a common (simple) schema. Production relies on automated processes: Member State data are retrieved, combined, harmonised and updated automatically, as far as possible, using up-to-date techniques.

How to cite
Dataset: Eurostat. (2023). Healthcare services locations. Retrieved from Eurostat website: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/healthcare-services
Concept & Methodology: Eurostat. (2020). Healthcare services in Europe. Retrieved from Eurostat website: https://gisco-services.ec.europa.eu/pub/healthcare/metadata.pdf

Data File Size: 2.2MB

File format: .csv, .geoJson, .GPKG

USER COSTS

Free & open Access
Open access

Data Infrastructure
Infra needed (software, hardware): GIS or statistical software
Data repositories / storage needed: No, .json API available

Data Governance & Management
Staff costs: Basic GIS skills
Data analysis needed: No
Quality assurance: Data quality varies from country to country

Data documentation and / or assistance in data use

https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/healthcare-services

https://gisco-services.ec.europa.eu/pub/healthcare/metadata.pdf

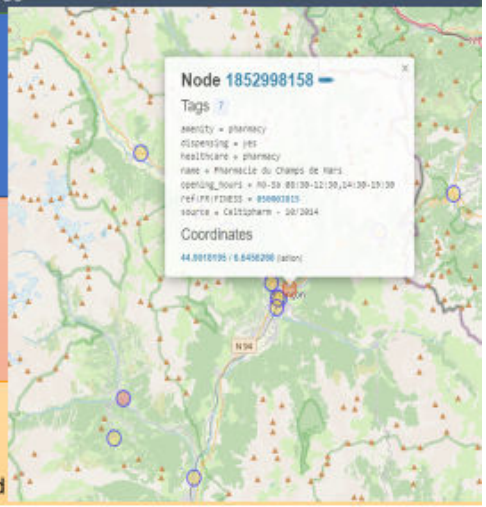
Compilation of the fiche by Antti Hiltunen (LUKE)

12.9. Infrastructure

OpenStreetMap for Points of Interest

https://www.openstreetmap.org/

Indicator class: Infrastructures



Data
Class: point locations
Type: vector (nodes, ways, relations)
License: Data Commons Open Database license

Spatial
Resolution: Not relevant
Extent: World
Granularity: Not relevant
Coordinate System: WGS 1984 (EPSG: 4326)

Temporal
Resolution: Up-to-date
Extent: Not relevant
Frequency: Continuously updated

Description:
OpenStreetMap (OSM) is a well-known crowdsourced project which aims to produce a free vector geographic database of the world. The OpenStreetMap database is one of pioneering and best-known source of Volunteered Geographic Information (VGI). Free and open source since its creation in 2004 by Steve Coast, a variety of data can be found in OSM including buildings, land use, road network. Many use cases can be made from OSM (map OSM tiles, routing, etc.). Extracting Points of Interest derived from this database is one of them.
Indeed, OpenStreetMap represents physical features on the ground (e.g., roads or buildings) using tags attached to its basic data structure. Each tag defined by a key-value combination describes a geographic attribute of the feature being shown by that specific node, way or relation. OpenStreetMap's free tagging system allows the map to include an unlimited number of attributes describing each feature. The community agrees on certain key and value combinations for the most commonly used tags, which act as informal standards. These key-value combination can be used to query the OSM database and extract specific geographical objects of interest, covering services (schools, pharmacies, hospitals, etc.), shops (restaurants, banks, bakers, etc.), leisure activities (sport, music schools, sport playground, etc.) or cultural aspects (museum, archeological sites, historic buildings, etc.)

How to cite
© OpenStreetMap contributors

Data File Size
europe-latest.osm.pbf: 27.6 GB (input), depending on the selection (output)

File format: pbf (input) / csv (output)

USER COSTS

Free & open Access

Data Infrastructure

Infra needed (software, hardware): GIS or software environment for statistical computing (R, Python, among others).

Data repositories / storage needed: Geofabrik server makes available the OSM dataset for free. Several software packages are available to transform pbf format in more standard geographical format, by filtering the content of the OSM database regarding OSM tag of interest. Storing and documenting these extracts (database version, OSM id of geographical objects) may be considered for reproducibility purpose.

Data Governance & Management

Staff costs: Data can be extracted using various tools linked to the OSM project. Its use requires a certain expertise in the world of geographic information and the structure of OSM data.

Data analysis needed: Managing OSM geometries variety: data can be contributed through points (nodes), polygons (ways) or relations (lines) depending on the tag of interest.

Quality assurance: It requires important consistency checks due to a lack of international harmonization, poor completeness, multiple database entries, errors and unclear standards. Its use also implies to identify the relevant key-values to catch all the possibilities of contribution referring to a specific service. Most of the studies concluded that OSM can efficiently complete authoritative or institutional data sources. It underlines also that POI completeness is very heterogenous, depending on the POI (banks or a pharmacies are significantly more complete than hairdressers) or the area of interest (rural areas are significantly less contributed than urban areas).

Data documentation and / or assistance in data use

OSM wiki
OSM taginfo
Geofabrik – OpenStreetMap Data Extracts
Overpass turbo
Introducing osmextract (R)
Pyrosm (Python)

Compilation of the fiche by: Ronan Ysebaert (CNRS)

Street level imagery

By Mapillary

URL: https://www.mapillary.com/

Indicator class: Residential;

Data
Class: images/video
Type: API
License: Creative Commons Share Alike (CC BY-SA) license (available here)

Spatial
Resolution: NA
Extent: Global
Granularity: street-level
Coordinate System: World Mollweide (ESRI: 54009) and WGS 1984 (EPSG: 4326)

Temporal
Resolution: Daily
Extent: 2013 - present
Frequency: Random – based on when volunteers upload their data

Description:
Mapillary brings together a global network of contributors who want to make the world accessible to everyone by visualizing the world and building better maps. Anyone can join and collect street-level images, using simple tools like smartphones or action cameras. With computer vision, we connect images across time and space to create immersive street-level views and extract map data.

How to cite
Dataset:
López-Antequera, M., Gargallo, P., Hofinger, M., Bulò, S.R., Kuang, Y., & Kontschieder, P. (2020). Mapillary Planet-Scale Depth Dataset. *European Conference on Computer Vision*.

Data File Size
Global dataset zipped: 20 GB

File format: .zip, jpg, png, an API is also available.

USER COSTS

Free & open Access
Since joining Meta, Mapillary no longer charges for access to the Mapillary platform and Mapillary data. You are welcome to use the platform free of charge, as long as you comply with our Terms of Use.
All imagery that you upload to Mapillary can be used by yourself in any form. Public imagery with a CC-BY SA licence is also available for use.
Plugins come for ArcGIS which is proprietary software. However the imagery is also available for OSM. In addition derived products are available in shapefile and GeoJSON formats for use in open software.

Data Governance & Management

Staff costs: working with this data will require some expertise in handling GIS data and possibly computer vision. For a person with these skills, one person month would likely be needed to investigate, extract and process the data.

Data analysis needed: Computer vision, data segmentation, and data extraction. This would be dependent on what information one wanted to extract.

Quality assurance: this would have to be done by the user on a case by case basis.

Data documentation and / or assistance in data use
https://help.mapillary.com/hc/en-us
https://forum.mapillary.com/

Compilation of the fiche by Ian McCallum & Martin Hofer (IIASA)

Tourism

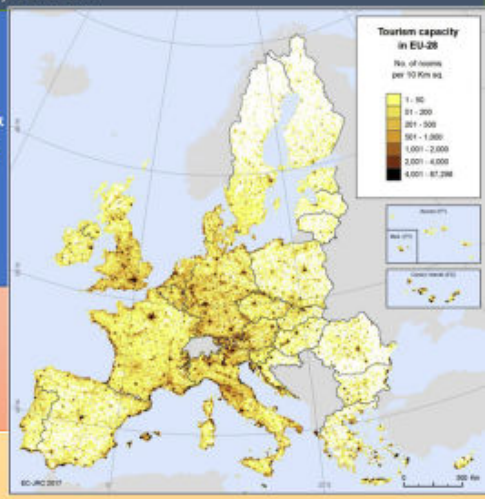
By JRC
URL: https://tourism-dashboard.ec.europa.eu/?lng=en&ctx=tourism

Indicator class: Infrastructure, Tourism

Data
Class: Webscraping
Type: Raster
License: The data in the Energy & Industry Geography Lab are free for public use
DOI: https://doi.org/10.1016/j.tourman.2018.02.020

Spatial
Resolution: 100m
Extent: EU
Coordinate System: Spherical/ Web Mercator EPSG:3857

Temporal
Resolution: Extent: 2019-2022
Frequency: Annual



Description:

Tourism capacity and density based on booking.com, TripAdvisor and Eurostat data . This descriptor measures the number of bed-places at tourism accommodation establishments available at a destination (country or region) in a given year. It provides an account of the absolute dimension of a tourism destination from the supply perspective. Higher values indicate a higher tourism supply. This descriptor is provided in the EU Tourism Dashboard as complementary information to characterize tourism supply. UNIT OF MEASURE: number of bed-places. RESOLUTION: NUTS0, NUTS2. COMPLETENESS: Missing countries for 2020: CH, CZ, DE, DK, HU, IE, LV, PL, SE, SI. DATA SOURCES: tour_cap_nuts2 [ESTAT], TripAdvisor.

How to cite
Dataset:
Ricardo Barranco (2022): UDP - Tourism capacity. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/659a45dd-5bc2-4aaf-8bcd-bb337ba03f92
Concept & Methodology:
Batista e Silva, F., Herrera, M. M., Rosina, K., Barranco, R. R., Freire, S., & Schiavina, M. (2018). Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management*, *68*, 101-115. https://doi.org/10.1016/j.tourman.2018.02.020

USER COSTS

Free & open Access
Full open access (Creative Commons 4.0)

Data Infrastructure

Infra needed: None for viewing. Analyses requires GIS or statistical software

Data repositories / storage needed: Yes

Data Governance & Management

Staff costs: Basic GIS or statistical skills needed, more arvanced analyses require higher skills set

Data analysis needed: No

Quality assurance: None specicified

Data documentation and / or assistance in data use

http://data.europa.eu/89h/659a45dd-5bc2-4aaf-8bcd-bb337ba03f92

Compilation of the fiche by Lauri Niskanen (LUKE)

12.10. Mobility

Human powered mobility — By Strava Metro
URL: https://metro.strava.com/

Indicator class: Residential;

Data
Class: vectors
Type: Shapefiles
License: https://metro.strava.com/
Terms

Spatial
Resolution: NA
Extent: Global
Granularity: street-level
Coordinate System: WGS 1984 (EPSG: 4326)

Temporal
Resolution: Annual
Extent: 2015 - present
Frequency: Updated monthly by users who agree to opt in to making their data public

Description:
The Strava dataset is the largest collection of human-powered transport information in the world. Metro aggregates, de-identifies and contextualizes this dataset to help make communities better for anyone on foot or on a bike. We work with urban planners, trail networks, city governments and safe-infrastructure advocates to understand mobility patterns, identify opportunities for investment and evaluate the impact of infrastructure changes – all completely free of charge.

How to cite
Dataset: https://metro.strava.com

Data File Size
Dependent upon area requested – smallest area is a county.

File format: Shapefile, or online dashboard

USER COSTS

Free & open Access
Since 2020, Metro is free to any organization that shares the mission to make cities better for cyclists and pedestrians. COVID-19 has accelerated the world's need for Metro – for smarter and more sustainable design of our cities, and for giving back to the communities that support millions of Strava athletes around the world. If you believe in human-powered transport and think you can make an impact, apply now. If your application is successful, you will be granted free access.

Data Governance & Management
Staff costs: Using the online dashboard and UI would mean costs would be low, and anyone with basic internet skills could query and visualize the data.
Data analysis needed: If a user chooses to download the data, e.g. Shapefiles, this will then require GIS skills to work with and further analyze the data. This should however be fairly basic for any GIS analyst. One person month in total would likely be enough to develop basic tools to work with and process the data.
Quality assurance: Strava is performing QA on their end, hence little effort is required for this from the user.

Data documentation and / or assistance in data use
https://metro.strava.com/
https://www.strava.com/

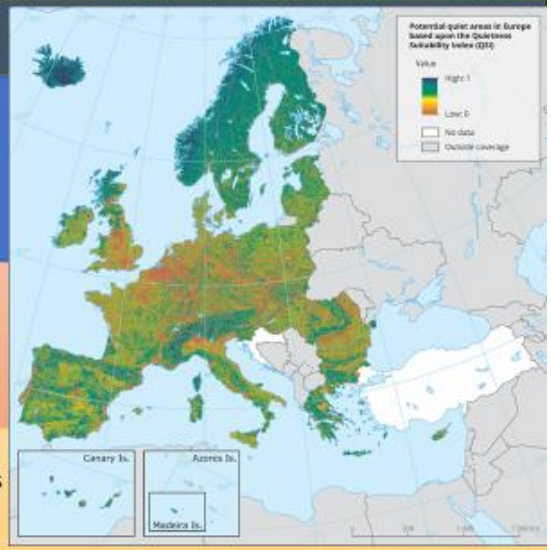
Compilation of the fiche by Ian McCallum & Martin Hofer (IIASA)

12.11. Recreation

Potential quiet areas in Europe based upon the quietness suitability index

By European Environment Agency

URL: https://www.eea.europa.eu/data-and-maps/figures/quietness-suitability-index-qsi-and-1



Indicator class: Health, Environmental

Data
Class: Model
Type: Vector + Raster
License: EEA standard re-use policy

Spatial
Resolution: 100m, 3 arcsec
Extent: Europe
Granularity: Gridded
Coordinate System: ETRS89 (EPSG: 4258)

Temporal
Resolution: ~Every 5 years
Extent: 2006, 2011, 2016
Frequency: >1 years

Description: Based on the multidimensional character of the notion of quietness, it can be defined according to objective criteria (noise levels), which are measured by quantitative data, but also according to a subjective component linked to perception. In this way, and beyond noise exposure, quietness is described in the Quietness Suitability Index or QSI as a combination of noise limit values (contour maps delivered under END requests) and land use and land cover elements that is perceived as positive and usually related to human cultural construction of naturalness.
The QSI is composed of two main elements:
• Noise disturbance as a result of distance to noise sources (objective criteria, quantitative data): threshold distances are determined considering noise levels determined by noise contours maps (END areas exposed to less than 55 dB Lden).
• The perceptive dimension of quietness by human beings (subjective criteria, qualitative data): this dimension is related to the importance given to natural elements and to landscape configuration. This dimension has been introduced in the QSI formula as a reclassification of the Corine Land Cover database based on the hemeroby index (Jalas, 1955; Blume and Sukopp, 1976).
To establish the distance values to the different noise sources, the noise contour maps for the main transport infrastructures and also the location of the major noise sources have been used as the main input information. These data have been provided by the Member States following the END requirements, accounting for main transport networks at European level and urban areas above 100 000 inhabitants.
Text: EEA (2016). Quiet areas in Europe. Technical report No 14/2016. p. 12-13.

How to cite dataset: European Environment Agency (2016). Potential quiet areas in Europe, based upon Quietness Suitability Index (QSI). Retrieved from EEA's website: http://data.europa.eu/88u/dataset/e9151c34-da65-48b9-a2ca-b9b835480812
How to cite concept & methodology: EEA (2016). Quiet areas in Europe. Technical report No 14/2016 https://www.eea.europa.eu/publications/quiet-areas-in-europe

Data File Size
.TIF: 2,8Gb
.gdb: 126Mb
File format: .TIF, .gdb

COSTS

Free & open Access
No limitations to public access

Data Infrastructure
Infra needed (softw, hardw): GIS/Remote sensing software
Data repositories / storage needed: Yes, the files are quite large and require download to access

Data Governance & Management
Data analysis needed: No
Staff costs: Basic GIS skills
Quality assurance: Very high

Data documentation and / or assistance in data use

European Environment Agency (2016). Potential quiet areas in Europe, based upon Quietness Suitability Index (QSI). Retrieved from EEA's website: http://data.europa.eu/88u/dataset/e9151c34-da65-48b9-a2ca-b9b835480812

EEA (2016). Quiet areas in Europe. Technical report No 14/2016 https://www.eea.europa.eu/publications/quiet-areas-in-europe

Compilation of the fiche by Antti Hiltunen (LUKE)

12.12. Transversal

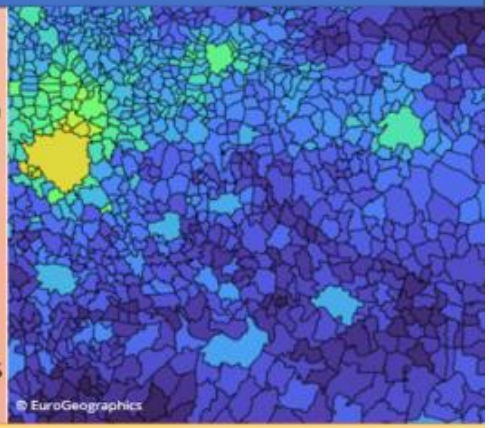
Local Administrative Units geometries (LAU)

https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/lau

Indicator class: Transversal

Data
Class: geometries
Type: vector (polygons)
License: Creative Commons Share Alike (CC BY-SA)
DOI: Not relevant

Spatial
Resolution: LAU
Extent: EU + EFTA + Candidate Countries (Serbia, Albania, North Macedonia) + the UK (until the 2020 version)
Granularity: Municipalities and communes of the European Statistical System.
Coordinate System: ETRS89-extended / LAEA Europe (EPSG:3035), Web Mercator projection (EPSG:3857) and WGS 1984 (EPSG: 4326)



Temporal
Resolution: Annual
Extent: 2011-2021
Frequency: Annually since 2011

Description:
LAUs are the building blocks of the NUTS (Nomenclature of territorial units for statistics) and statistical regions. It covers municipalities and communes of the European Statistical System (ESS). Data is available annually and currently described by total resident population, where available. This is the core layer to fit with EU national censuses.

How to cite
Dataset:
© EuroGeographics for the administrative boundaries

Data File Size
LAU_RG_01M_2021_3035 zipped: 31,3 MB

File format: SHP, TopoJSON, GeoJSON, PBF, SVG

USER COSTS

Free & open Access

Data Infrastructure

Infra needed (software, hardware): GIS or software environment for statistical computing

Data repositories / storage needed: No, data is moreover distributed via API: Several software API packages helps to retrieve Eurostat – GISCO data such as giscoR in R or happygisco in Python: LAU units can be uploaded directly via computer programming.

Data Governance & Management

Staff costs: Basic GIS skills

Data analysis needed: No, just requires to find the appropriate LAU version to fit with data provided by Eurostat/National Statistical Institutes.

Quality assurance: Very High. Seamless geometries for all Europe are a real added value to map results at local scale. However, this territorial nomenclature is quite heterogeneous and highly subject to MAUP effects (35 000 units for France, 290 for Sweden). It must be considered when proposing indicators at this territorial scale.

Data documentation and / or assistance in data use

Eurostat, 2023, Local Administrative Units (LAU)

Eurostat, 2019, Merging statistics and geospatial information, Experiences and Observations from National Statistical Authorities

Compilation of the fiche by: Ronan Ysebaert (CNRS)

GEOSTAT 1km population grid

https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/grids

Indicator class: Transversal

Data

Class: geometries

Type: vector (polygons, points), raster and tabular (centroids)

License: Creative Commons Share Alike (CC BY-SA)

Spatial

Resolution: 1 km, also aggregated at 2-5-10-20-50-100 km

Extent: EU + Liechtenstein + Switzerland + Norway

Granularity: Grid cell

Coordinate System: ETRS89-extended / LAEA Europe (EPSG:3035), Web Mercator projection (EPSG:3857) and WGS 1984 (EPSG: 4326)

Temporal

Resolution: Total population for year YYYY.

Extent: 2006 – 2011 – 2018 - 2021

Frequency: Certainly more frequent In the future.

Description:

The grid data consist of statistics that are geographically referenced to a system of square grid cells in a common European reference grid net with Cartesian coordinates. The use of a common European-level grid allows analyses to be undertaken for areas that cross national boundaries. The net of 1 km^2 grid covering the territory of the EU contains 4 448 981 grid cells and 1 824 619 populated cells. In march 2024 the 2021 population grid will be updated with new indicators: sex (males, females), age (under 15, 15 to 64, 65 and over), employed persons (as far as possible), place of birth (in the reporting country, in another EU country, outside the EU), usual residence 12 months before the census date (unchanged, within the reporting country, outside of the reporting country).

How to cite

Dataset:

© EuroGeographics for the administrative boundaries

Data File Size

Eurostat_Census-GRID_2021_V1-0 zipped: 181 MB

File format: GeoPackage, csv, parquet

USER COSTS

Free & open Access

Data Infrastructure

Infrastructure needed (software, hardware): GIS or software environment for statistical computing

Data repositories / storage needed: No, data is moreover distributed via API: Several software API packages helps to retrieve Eurostat – GISCO data such as giscoR in R or happygisco in Python: grid cells can be uploaded directly via computer programming.

Data Governance & Management*Staff costs*: GIS skills needed

Data analysis needed: Data handling requires smart approach, in particular when manipulating vector data (computing time / cost).

Quality assurance: High in general, and growing over time. Three methodologies are used to attribute a number of inhabitants to each square kilometre cell: 1/ Aggregation method, using geo-referenced micro data (a method used by 25 Member States and 3 EFTA countries for the 2021 product). 2/ Disaggregation method is used in the absence of geocoded micro data, using statistical data for the lowest available administrative / territorial units in combination with auxiliary spatial data (also called top-down approach). None of the Member States used this method for the 2021 population and housing census, but it was the case for previous grid products. 3 / Hybrid method that combines aggregation and disaggregation techniques and represents a compromise between accuracy and availability of data. It is used for France and Greece. Analysis using evolution over time must consider these aspects (methodologies evolving over time accross EU countries).

Data documentation and / or assistance in data use

Eurostat, 2023, Statistics Explained, Population and housing census 2021, population grids

Eurostat, 2023, Statistics Explained, the degree of urbanisation manual - Constructing a population grid

European Forum for Geostatistics, 2012, ESSnet project GEOSTAT, Reprenting Census data in a European Population grid, Final Report GEOSTAT 1A – Representing Census data in a European population grid

Compilation of the fiche by Ronan Ysebaert & (CNRS)

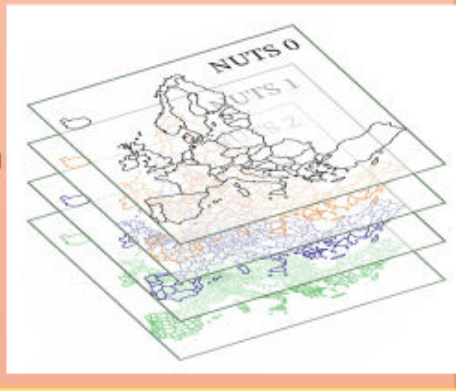
NUTS geometries

URL:https://ec.europa.eu/eurostat/fr/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts

Indicator class: Transversal

Data
Class: geometries
Type: vector (polygons, lines, points)
License: Creative Commons Share Alike (CC BY-SA)
DOI: Not relevant

Spatial
Resolution: NUTS 0-1-2-3
Extent: EU + EFTA + Candidate Countries + the UK.
Granularity: 5 levels of detail (scales): 01M, 03M, 10M, 20M, 60M
Coordinate System: ETRS89-extended / LAEA Europe (EPSG:3035), Web Mercator projection (EPSG:3857) and WGS 1984 (EPSG: 4326)



Temporal
Resolution: Depending on territorial reforms
Extent: 2003, 2006, 2010, 2013, 2016, 2021
Frequency: every 3-4 years in general

Description:
The NUTS are a hierarchical system divided into 3 levels. NUTS 1: major socio-economic regions, NUTS 2: basic regions for the application of regional policies, NUTS 3: small regions for specific diagnoses. The NUTS legislation is periodically amended; therefore multiple years are available for download. The NUTS geometries are therefore the basis to map statistical information made available by Eurostat.

How to cite
Dataset:
Administrative boundaries: © EuroGeographics, © TurkStat. Source: European Commission – Eurostat/GISCO

Data File Size
NUTS 2021 20M Polygons zipped: 0,6 MB

File format: SHP, TopoJSON, GeoJSON, PBF, SVG

USER COSTS

Free & open Access

Data Infrastructure

Infra needed (software, hardware): GIS or software environment for statistical computing

Data repositories / storage needed: No, data is moreover distributed via API: Several software API packages helps to retrieve Eurostat – GISCO data such as giscoR in R or happygisco in Python: NUTS units can be uploaded directly via computer programming.

Data Governance & Management

Staff costs: Basic GIS skills needed

Data analysis needed: No, just require to find the appropriate NUTS version to fit with Eurostat data.

Quality assurance: Very High

Data documentation and / or assistance in data use
https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts

Compilation of the fiche by: Ronan Ysebaert (CNRS)