GRANULAR

# IMPLEMENTING NEW DATA & METHODS FOR REPLICATION, VALIDATION & UPSCALING

## D3.3

JULY 2026

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).

# D3.3 IMPLEMENTING NEW DATA & METHODS FOR REPLICATION, VALIDATION & UPSCALING

| | |
|---|---|
| **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** | Final version |
| **Dissemination level** | Public |
| **Authors** | McCallum I., Hofer M. (IIASA), Kull M., Vihinen, H. (LUKE), Ysebaert R., Guérois M., Giraud T., Viry M., Lambert N., Laurian, L. (RIATE-CNRS), Lai, S., Voepel H., Steele J. (UoS), Sorichetta A. (UniMi), Tapia C., Berbert, K., Dzhavatova, K., Yantseva, V. (Nordregio), Miller D., Hopkins J., Farinelli V. (HUT), Ulman M., Simek P., Motyckova V. (CZU), Panoutsopoulos H., Chitos A., Fournarakos A., Zafiraki P. (AUA), Berchoux T., Pallavi, J. (IAMM) |
| **Work Package Leader** | IIASA |
| **Project coordinator** | Mediterranean Agronomic Institute of Montpellier (IAMM) |

**Citation:** McCallum, I., Hofer, M., Kull, M., Vihinen, H., Ysebaert, R., Giraud, T., Viry, M., Lambert, N., Laurian, L., Shengjie, L., Voepel, H., Steele, J., Sorichetta, A., Tapia, C., Berbert, K., Dzhavatova, K., Yantseva, V., Miller, D., Hopkins, J., … Pallavi, J. (2026). Implementing New Data & Methods for Replication, Validation & Upscaling. GRANULAR. https://doi.org/10.5281/zenodo.21875481

*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

**Table of contents** .......... **2**

**1. Executive summary** .......... **3**

**2. Introduction** .......... **3**

**3. Background and objectives** .......... **3**

**4. New Data and Methods** .......... **6**

4.1. Accessibility (RIATE) .......... 6

4.2. Mobility (UoS, UMI) .......... 15

4.3. Rural economies (Nordregio) .......... 27

4.3.1. Collecting Data to Monitor Housing Markets In Rural Areas: A Web Scraping Approach Tested in the Övre Norrland Region (Nordregio) .......... 27

4.3.2. High Resolution Analysis of Transport Poverty Risk in Europe's Rural Areas (Nordregio) .......... 30

4.3.3. Using OSM Data to Estimate Economic Diversity in Functional Rural Areas (Nordregio) .......... 34

4.4. Land Use (IAMM) .......... 37

4.5. Citizen Science, Crowdsourcing and Engagement (IIASA) .......... 41

**5. Discussion and Recommendations** .......... **46**

**6. Conclusions** .......... **47**

**7. References** .......... **47**

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).*

# 1. Executive summary

The GRANULAR project has developed, tested, validated, and scaled a suite of novel datasets, indicators, and analytical methods designed to improve understanding of the diversity and dynamics of rural areas across Europe. Recognising that existing official statistics often lack the spatial and thematic granularity required for effective place-based rural policy, the project combined open data, Earth observation, crowdsourced information, social media-derived mobility data, citizen science approaches, and advanced artificial intelligence methods to generate new evidence on accessibility, mobility, rural economies, land use, and rural attractiveness. Validation was conducted through comparisons with official datasets, statistical assessments, local stakeholder engagement, citizen science campaigns, and extensive testing across seven Living Labs and nine Replication Labs representing diverse rural contexts throughout Europe. This iterative process enabled the refinement of methods and ensured that outputs were both scientifically robust and relevant to local decision-makers.

The results demonstrate the significant potential of novel data sources and open analytical frameworks to complement traditional statistics and support rural policy development. Key outputs include European-scale accessibility indicators, mobility and connectivity metrics derived from social media data, innovative approaches to measuring economic diversity and housing market dynamics, AI-enabled land-use monitoring tools based on Sentinel-2 satellite imagery, and new methods for assessing landscape attractiveness and tourism behaviour. Many of the resulting datasets, workflows, visualisation tools, and models have been released through the GRANULAR Zenodo Community to promote transparency, reproducibility, and uptake by researchers, policymakers, and rural stakeholders. Despite challenges related to data completeness, representativeness, privacy constraints, and varying regional contexts, the project demonstrates that integrating novel data sources with participatory validation processes can substantially enhance the evidence base available for supporting Europe's rural transitions and implementing more effective, place-sensitive rural policies.

# 2. Introduction

GRANULAR is a four-year multidisciplinary and transnational project that has developed new datasets, tools and methods to better understand the diversity and dynamics of rural areas across Europe. Validation of the data and methods implemented in GRANULAR was conducted to assess their capability to address the overarching GRANULAR objectives. Results from this validation have provided feedback to the data producers and adjustments and refinements were made to ensure the highest possible quality of the final results.

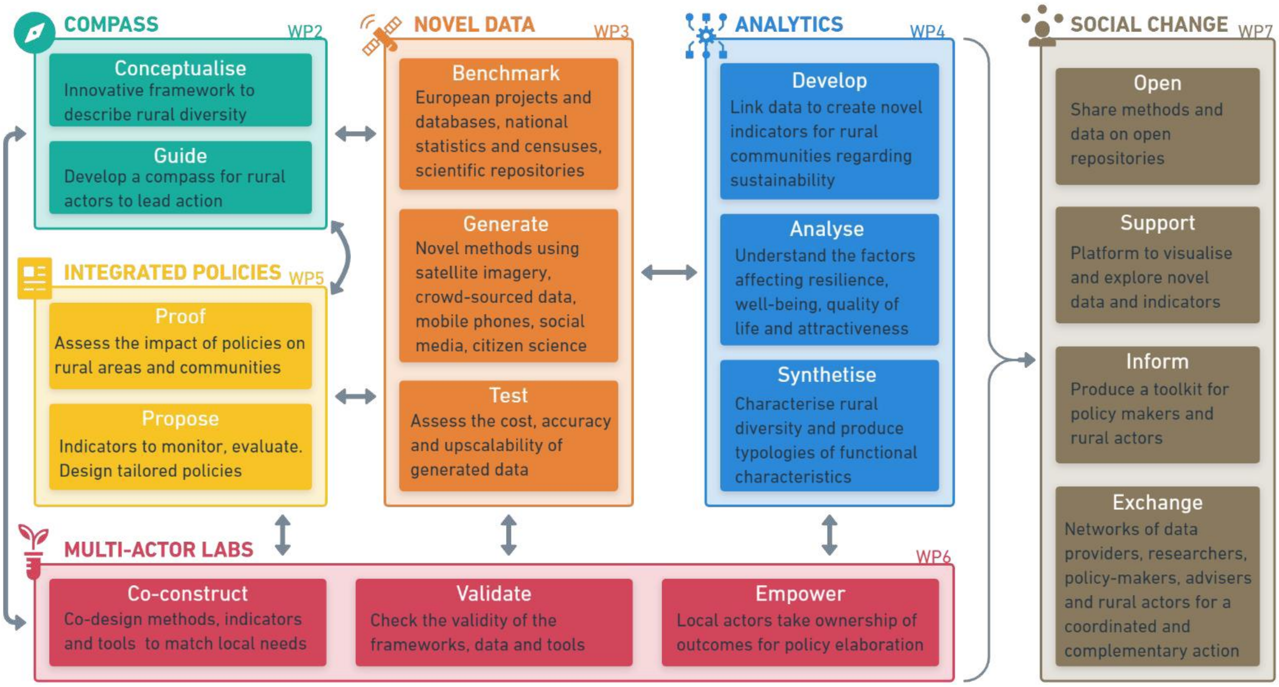
A particular focus was placed on replication and upscaling taking a multi-scalar approach i.e., combining local and regional approaches, cognisant of associated uncertainty. Data and methods failing to meet the objectives were revisited and improved wherever possible. Validation tools included comparison with official datasets (e.g. Eurostat and others), confusion matrices for spatial data, consensus among datasets, Citizen Science campaigns, in-situ sensor measurement comparison and more. Living and Replication Labs have played a crucial role in the validation process using both qualitative and quantitative approaches.

# 3. Background and objectives

Both policy makers and scientists lack data at the appropriate level of granularity to be able to fully grasp rural diversity. They also lack practical tools to develop place-based policies that fit rural specificities and are able to guide rural areas in their multiple transitions: agro-ecological, demographic, industrial and energetic. GRANULARs General Objective is to support digital, economic and ecological transitions in rural areas through integrated place-based evidence and multi-actor processes. At the heart of this approach is the need for robust forms of new and novel data (WP3). Novel data discovered and developed in the context of GRANULAR is used in multiple downstream components of GRANULAR, through the development of supporting indicators, integrated policies, multi-actor labs and various forms of social change, as presented below.

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).




*Key elements and interactions among the proposed GRANULAR Work Packages*

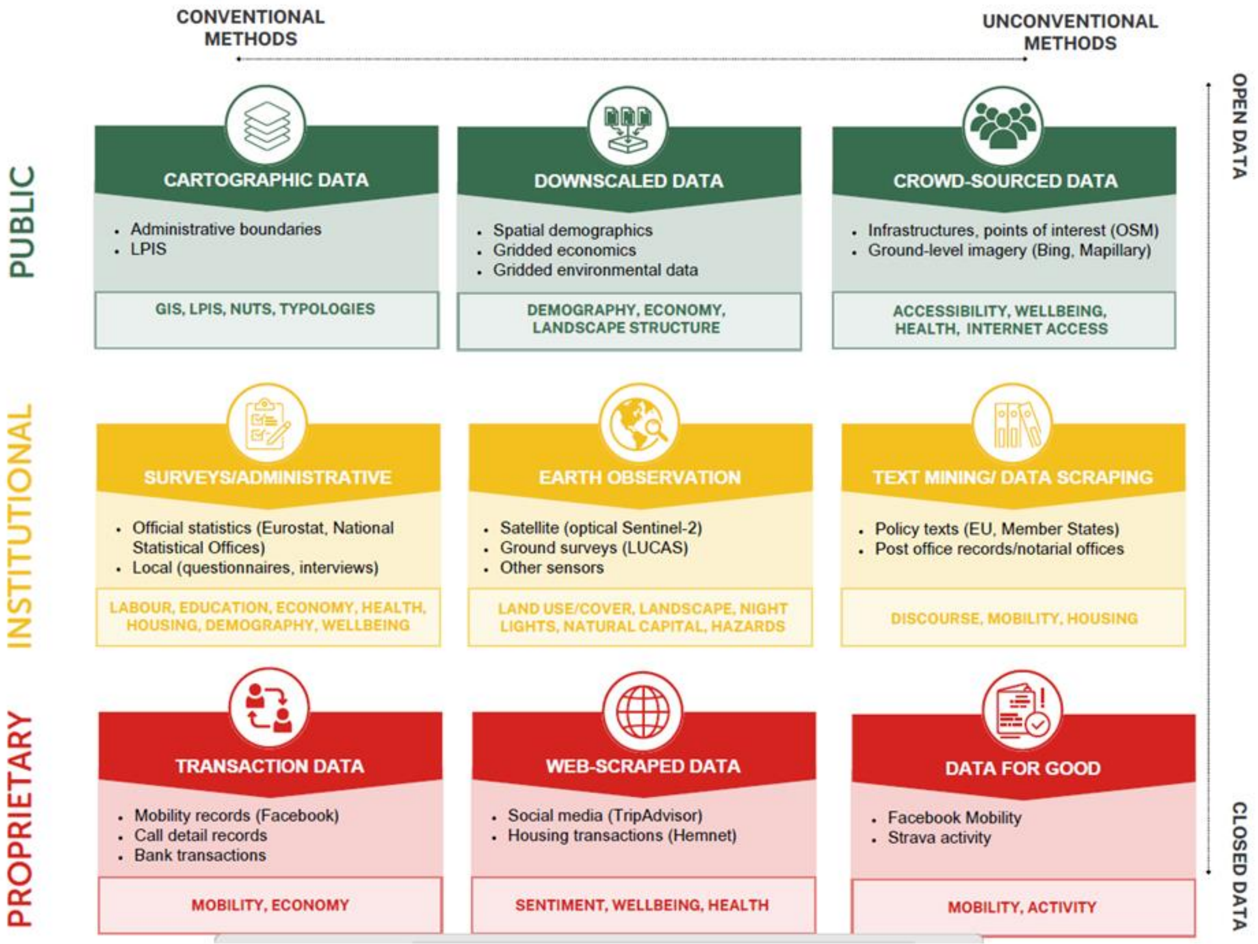
**Baseline rural datasets**

In an effort to underpin the various efforts within GRANULAR, several benchmarking efforts were undertaken, specifically screening potential datasets along with benchmarking performance and costs of such datasets. An initial screening of datasets that are relevant to capture rural diversity and to create novel indicators for rural areas was performed at the onset of the project. Following a semi-structured format of discovery and evaluation, we documented > 90 different datasets which are either already used to characterise rural areas, or could underpin novel indicators. In addition to identifying the datasets themselves and their locations, we provided a suite of associated meta-data. Evaluating the findings of this effort, we demonstrated that the majority of the datasets identified have regional to global coverage, have Local Administrative Unit to gridded (10m - 10km) granularity, are provided annually, are free and open and of moderate relevance in terms of indicator generation for rural areas. For more information see: https://doi.org/10.5281/zenodo.13838524.

Building further upon this work we performed a systematic review of existing datasets. Discovered datasets ranged in type from various forms of social media data (e.g. Facebook, Flickr, TripAdvisor), to crowdsourcing data (e.g. OpenStreeMap), earth observation (e.g. Copernicus services) and low-cost sensors (e.g. air quality sensors). Unstructured datasets such as crowdsourced and web-scraped data present unique challenges in terms of analysis often benefitting from the use of novel techniques, such as ML and AI. The benefit of these methods, particularly AI, is recognized and on the rise but has not yet fully reached its potential for rural policy monitoring and evaluation. Finally, examples of new indicators derived from these novel datasets are already under development, including for example the creation of accessibility indicators from populated 1km EU grid cells to towns and cities at aggregated scales for the entire EU.

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).*




*A broad classification of the data types considered in GRANULAR, ranging from proprietary data through to institutional and public data.*

**A benchmark of performance and costs in the EU and beyond**

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. A 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 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 survey 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. This work is documented here: https://doi.org/10.5281/zenodo.13744714

**Overview of labs and their role in validation**

The GRANULAR Project worked with a network of 7 Living Labs across Europe (France, Italy, Netherlands, Poland, Spain, Sweden, United Kingdom) along with 9 Replication Labs (Albania, Finland, Greece, Italy, Latvia, Lithuania, Moldova, Romania and Serbia). Together, these Labs reflect a significant portion of the diversity of rural Europe e.g., different demographic trends, governance systems, economic structures, climate exposure, and rural–urban relationships. This diversity allows GRANULAR to test whether tools and approaches which might be appropriate in one region remain meaningful across contexts – and where adaptation is required. The role of the multi-actor labs is manifold and includes the following tasks:

- Validate indicators and datasets against local realities;
- Highlight inequalities within rural areas;

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).*

- Identify trade-offs;
- Shape how rural proofing is applied to concrete policies.

**Data fiches**

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. For more information see: https://doi.org/10.5281/zenodo.13744714

# 4. New Data and Methods

Over the course of GRANULAR, a variety of new data and methods have been introduced, tested and improved. Whereby they have reached an acceptable level of maturity and quality, they have been released for public consumption via the GRANULAR Zenodo Repository. Datasets developed under the five main data research themes are described in detail below, namely accessibility, mobility, economics, land use and citizen engagement.

## 4.1. Accessibility (RIATE)

**1. Objectives**

The main objective of this subtask is to **produce accessibility indicators at local scale for all of Europe,** using **crowd-sourced data combined with institutional data**. Much more than just reflecting transportation infrastructure performance, this research theme aims at developing indicators that help **highlight inequalities in accessibility to resources and facilities that meet social needs** (services, shops, schools, workplaces…). While the specific challenges faced by rural areas in relation to this topic (scattered settlements, low density of facilities) have been analysed for several decades, the need to support sustainable mobility in an era of ecological urgency exacerbated by car dependency given the scarcity of alternative modes, especially public transportation, is crucial. At the same time, significant progress has been made recently in granular data related to population and services and facilities location, as well as in routing engines, making it possible to better characterize the high heterogeneity of rural areas in the face of these challenges. In this regard, a key issue of the accessibility subtask is the focus on fully harmonised data, open-source solutions and free routine engines.

**Three sets of results** —some more comprehensive and consolidated, others exploratory— are provided: 1) Indicators of **spatial accessibility by car to towns and cities**, at a European scale, 2) Indicators of **spatial accessibility by car and by bike to secondary schools**, for the vast majority of the project’s Living Labs and Replication Labs, 3) A **reproducible method for comparing accessibility by car to accessibility by public transportation** based on a case study of the French living lab region.

To ensure the transparency and reproducibility of the results, they are disseminated through a **documented framework**, mainly based on the R programming language and on methodological notebooks (input data detail and expertise, metadata, methodological processing, user guidelines, output data), available in several Zenodo repositories. At the same time, **interactive geovisualization tools are developed**, where possible, to provide a user-friendly way to explore and access the output data. The methodological notebooks are compiled in the documentation of four Zenodo repositories, which focus, respectively, on **1)** accessibility indicators to European towns and cities from all 1 km² populated grid cells, 2) the same indicators at aggregated levels (NUTS), **3)** accessibility by bicycle and car to European secondary schools, and 4) **a local case study to explore** the development of **multimodal accessibility indicators**.

**2. General approach/ analytical framework**

**Data**

The calculation of spatial accessibility indicators requires selecting three main categories of data : 1) **origins** (places of residence for the population), 2) **destinations**(services and facilities location) and 3) **transportation networks data** as input for routing engine computing travel time between 1) and 2):

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).*

With regard to **origins**, we built on the significant progress made in recent years by Eurostat to disseminate spatially granular data on the distribution of the population. Since 2023, census data have been provided in a 1km² reference grid, for total population and by gender and age. Compared to administrative data, grid data have a lot of advantages : grid cells have the same size and thus can be assembled to form areas reflecting a specific purpose and study area (NUTS, Functional Areas, mountain regions, water catchments, etc), overcoming the Modifiable Areal Unit Problem effect by using locations that are more precise than the centroid of administrative areas. In 2021, the most recent year for which data is available, it accounts for 2,008,432 cells with at least one inhabitant, for the entire UE (excluding French Oversea territories), Switzerland and Norway and in 2018 for United Kingdom. Based on this source, we defined places of origin as the centroids of the populated grid. In order to specifically target rural areas, it is then possible to cross these centroids with the DEGURBA urban/rural classification (Eurostat / GISCO, 2025 and Copernicus - Global Human Settlement, 2026 for the UK), which relies on a combination of criteria related to geographic proximity and population density.

Compared to origins, the choice of relevant **destinations** raises more substantial issues when viewed from an international perspective. While national databases on the geolocation of services and facilities have proliferated in recent years, international and open source dataset still have significant gaps. Regarding OpenStreetMap (OSM) Points of Interest (POI), data are considered relevant in European dense urban contexts but their completeness has been rated as fairly low in suburban and rural areas. As for institutional databases, there has been a recent development in European geographical layers of basic services for healthcare and education, maintained by GISCO, but there is still a lack of harmonization (in terms of the definition of the service or completeness of attributes) for those data extracted from official national registers. It is especially true for the healthcare data that contain inconsistencies due to the heterogeneity in the definition of a hospital, sometimes requiring a minimum number of beds, which could largely explain the differences in the number of hospitals between France (4,400) and Spain (790). These shortcomings led us to give primary consideration to the availability of urban functions as a relevant proxy for specific services. To do so, we mainly rely on the harmonized cities and towns layer from the European Commission and we define as destinations the 33,000 settlements which were classified into three categories based on their population: 5,606 very small towns of 5,000 to 9,999 inh., 3,724 small town from 10,000 to 49,999 inh. and 866 cities over 50,000 inh. A recent OECD study (OECD, 2024) using the same reference layer showed that larger settlements tend to have a wider variety of services: looking at the number of pharmacies, schools, banks, hospitals and universities in each type of urban object in 25 OECD countries, it showed that on average, towns have at least 3 categories of services, while small cities have more than 4 categories. We also highlighted the relation between these size categories of cities and town and the hierarchy of services through the French study case, for a wider range of facilities hierarchically ordered across several domains in 2023, from an institutional dataset (Table. 1). Lastly, several solutions to transform this polygon layer to destination points layers have then been discussed and experimented. For the second part of this subtask, we also considered the **GISCO dataset about education** in order to target a specific basic service. Referring to the ISCED level to identify schools of interest, we selected the **ISCED level 2** (lower secondary education**,** in average from 12 to 15 years old) that matches with the end of mandatory school in the majority of European countries and corresponds to an age at which students can go to school on their own (including by bike).

**Table 1.** Share (%) of towns and cities with at least one facility, for a set of services in France (sources: European Union 2020, INSEE Base Permanente des Equipements 2023)

| **Facility** | **Towns** 5,000-9,999 inh. | **Towns** 10,000-49,999 inh. | **Cities** 50,000 inh. and above |
|---|---|---|---|
| ISCED 1 (primary schools) | 100.0 | 100.0 | 100.0 |
| ISCED 2 (collèges) | 84.3 | 96.8 | 100.0 |
| ISCED 3 (lycées) | 50.1 | 86.1 | 100.0 |
| ISCED 4-5 (higher education) | 18.0 | 56.6 | 98.7 |
| ISCED 6_8 (universities) | 4.0 | 32.4 | 93.4 |
| General practitioner | 98.6 | 100.0 | 100.0 |
| Optalmologist, dentist | 98.4 | 99.6 | 100.0 |
| Pediatrician | 15.0 | 55.2 | 97.4 |
| Bakery | 99.7 | 100.0 | 100.0 |

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).*

Regarding **transportation networks data**, the vast majority of the analyses relies on the **OSM road network** (for car, bike and pedestrian profiles), whose high accuracy has been noted in numerous scientific publications, in terms of geometric precision, coverage as well as attributes. At administrative scale (country, region, sub-region), data is downloaded on Geofabrik website. In the case of the Valhalla routing tool (see below), a Digital Elevation Model (DEM) is also used and comes from from Tilezen/Joerd project (SRTM, GMTED2010). In addition, in the third part of the accessibility subtask, computations of realistic multimodal travel times are based on **GTFS (General Transit Feed Specifications) tables** for public transport network and time tables. We have listed the available data on Transitland, a platform that gathers GTFS data all around the world, and we have completed it with the National Access Point to data for France. We also selected the Copernicus Digital Elevation Model, in order to take into account the effect of elevation on travel time, particularly for bike and pedestrian profiles. Although the GTFS format is much valuable for developing a standardized approach to accessing public transit data across different regions, it is important to keep in mind that the tables might contain errors (wrong travel time differences between two stops, not exact localisation of a stop, etc.) and that the data validation step is thus essential: different tools were used to visualize potential errors (for instance the Canonical GTFS Schedule Validator developed by MobilityData, or the `get_gtfs_errors()` R function) and correct them.

### Methods

The **choice of spatial accessibility indicators** raises different points of discussion. A first key-point concerns the choice between measures quite simple to calculate and interpret (distance-based access to destinations) and more sophisticated and resourceful approaches with gravity-based and utility-based measures. Following some authors focused on the broader adoption of indicators by local stakeholders, who demonstrated that the results obtained from these different approaches were fairly correlated, we gave priority to more intuitive approaches, considering the emphasis on reusing these indicators within living labs in the Granular project. Another important point relates to the selection of relevant measures from the range of **distance-based access indicators**. We first considered **travel time measures** which is a common estimate of the **travel cost to reach the nearest destination**, for instance in minutes by car. This key indicator has often been given preference in rural contexts since destination choices are limited in a context of low density of facilities. It is here supplemented by a measure of cumulative opportunities based on a total **number of opportunities that can be reached in a given travel time**, which provides additional information on the degree of choice between several destinations (**Fig. 1**).

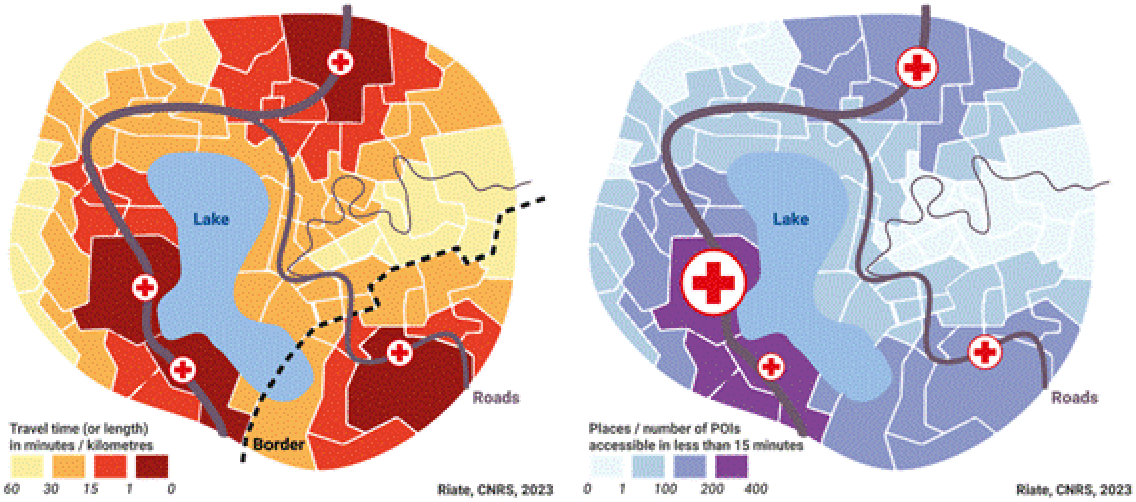



*Figure 1. Two indicators relying on distance-based access to destinations: the distance to the nearest opportunity (left) and the number of opportunities within a distance threshold (right)*

The computation of distances is processed through routing engines based on OSM (OSRM, Valhalla, $R^5$). These tools have gained much popularity in recent years since they offer significant advantages in terms of flexibility (install and hosting of one's own routing service with customisable parameters), transparency (open source code) and price (most are free). First of all, based on the former selection of origins and destinations, we calculated the **matrix of driving times for every European country at a granular level**, using the **OSRM routing engine**. This information is processed through the OSRM R package, based on the OSRM Application Programming Interface (API). This matrix has then been calculated in a buffer area of 100 km around each grid cell to reduce the size of the matrix, this threshold being defined to consider one hour travel time by car (one-way). Secondly, measures of **accessibility by car and by bike to secondary schools at LL and RL scale** provide an opportunity to compare the OSRM routing engine to Valhalla's. The latter, which is also based on OSM road network but refers to different algorithm

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).*

and parameters, is notably interesting to explore when analysing bike routes, since it can take elevation changes into account (**fig. 2**).

| | | OSRM | Valhalla |
|---|---|---|---|
| Computing | | Pre-processed routes (faster)<br>1 instance = 1 transport mode | Graph only<br>1 instance = every transport modes |
| Elevation data | | No | Yes |
| Profiles | | Car, bike, foot | Auto, bicycle, bus, bikeshare (beta), truck, taxi, motor scooter, motorcycle (beta), pedestrian, multimodal (pedestrian + transit) |
| Type | Car | **1** (with size attributes) | **1** (with size attributes) |
| | Bike | **2** : bicycle, cargo bike | **4** : road, hybrid, cross, mountain |
| Default speed (kmph) | Car | Depending on road type (OSM max speed value), smoothness, surface and country | Depending on road type (OSM max speed value), country<br>**Speed reduced on roads in urban area (determined from density)** |
| | Bike | **15** (riding), **4** (walking)<br>Max : road max speed | **25** (Road bicycle), **20** (Cross bicycle), **18** (City bicycle), **16** (Mountain bicycle), **5.1** (*dismount speed*)<br>**Min : 5, max : 60** |
| Penalties (fix or factor) | Car | Maneuvers (turn costs),<br>Reduced speeds on secondary roads<br>Traffic lights, stop signs, etc | Maneuvers (turn costs), with more precision according to the curve degree<br>Reduced speeds on secondary roads<br>**Traffic**<br>**Elevation costs**<br>**Country crossing**<br>Traffic lights, stop signs, etc |
| | Bike | Same as above, but less costful<br><br>Traffic lights, maneuvers (**same values as for cars for traffic light and u-turn, not for turn**), **mode change, unsafe roads**, barriers (roads with limited access) | Same as above, but less costful<br>Return cost for bikshare<br>Speed is modified based on the hilliness of a road section |
| Route selection | Car | Hierarchy according to speed (motorway > trunk > primary road > secondary road, etc)<br>Hierarchy according to legal access permission (motorcar > motor_vehicle > vehicle > access)<br>Avoid tolls & ferries if necessary | Priority to faster roads<br>Influence of real-time traffic |
| | Bike | Priority to bicycle roads over vehicle roads | Priority to cycleways or roads with bicycle lanes<br>Factor to indicate a cyclist's tolerance for riding on roads / hills / bad surface<br>Avoid roads without bicycle access.<br>Stairs possible with dismount mode (?) |

*Figure 2. Comparison of the OSRM and Valhalla routing engines (Realisation: RIATE, 2025)*

Eventually, based on a review of public transport data and of tools handling them, the **R$^5$ routing engine** (Rapid Realistic Routing on Real-world and Reimagined networks) has been used through the **r5r package** for the public transport accessibility study case. Alongside with the OSM network, it supports public transport networks for computing multimodal travel time matrices. It allows the computation of door-to-door trips at many different departure times. The use of this tool raises other methodological questions, particularly regarding the integration of different transportation modes (for instance choice of a time window for departure and arrival times, of the maximum number of public transport rides in the same trip, customization of the transport mode used after reaching or leaving the last public transport -walk, bike or car-).

### Results

#### *Indicators of spatial accessibility by car to towns and cities, at a European scale*

The focus on accessibility by car to towns and cities has resulted in the creation of three main deliverables: 1) a **matrix of driving time** between each centroid of the populated European grid cells (2 million points) and the cities and towns layer (33,000 points), i.e. 512 million travel time by car calculated between origins and destinations; 2) a related **set of 33 indicators at European scale**, which sets out **accessibility measures according to destination size** (towns and cities above 5.000, 10.000 or 50.000 inh.), **to people choice** (first, second or third reachable town or city) and **to different time thresholds** (from 15 to 120 min), from every populated European grid cells to every town above 5.000 inhabitants, in a 100km radius; 3) **for each European country** (European Union and associated countries, namely United Kingdom, Switzerland and Norway)**, a table with a summary of accessibility indicators** (for instance, median travel time to the nearest town), **based on different aggregation levels** (NUTS0 to NUTS3)

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).*

and **according to spatial reference classifications (DEGURBA urban/rural typology)**. This material makes it possible to assess the situation of each area as regard to accessibility to towns and cities (**Fig. 3**), using only open source and non-proprietary data: based on this dataset, for instance, it can be stated that 20 % of the population living in mostly uninhabited areas (53.3 million inhabitants in Europe) is more than 30 minutes by car away from the nearest town above 5000 inhabitants, compared to 23 minutes for population living in dispersed rural areas (72.2 million inhabitants in Europe) and 21 minutes for villages (53.3 million inhabitants). This situation varies considerably between countries, the median travel time to the first town being around 11.5 minutes in Belgium or in the UK, while it exceeds 25 minutes in Bulgaria, Latvia, Sweden, Estonia, Finland and Norway.

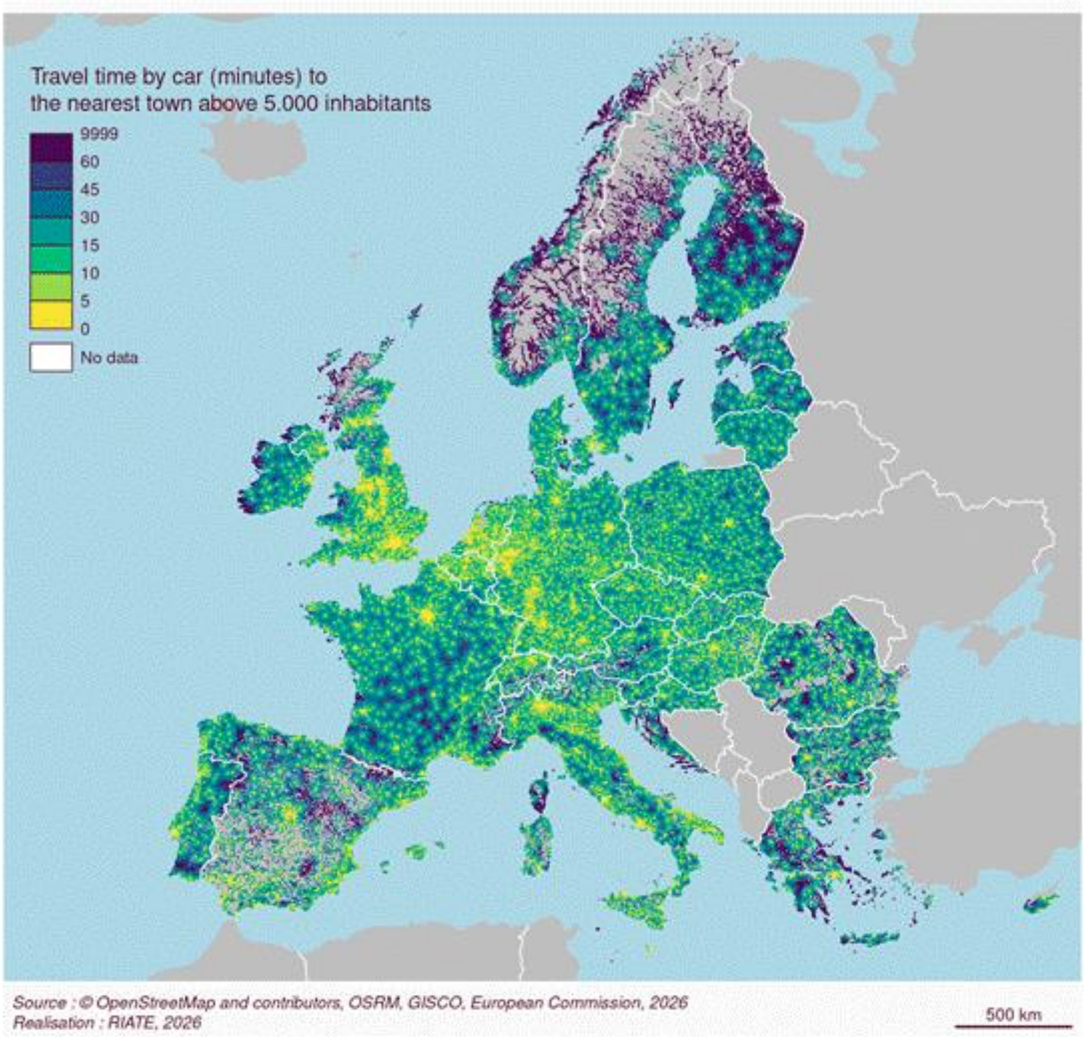



*Figure 3. Travel time to the nearest town above 5000 inhabitants by car from each populated 1km EU reference grid*

***Indicators of spatial accessibility by car and by bike to secondary schools, at the scale of a large set of Living Labs and Replication Labs***

The second part of the subtask contributes to further explore several aspects of accessibility indicators to services of interest, such as 1) Refining accessibility indicators based on origin (young population rather than total population) and destination attributes (public/private schools, ISCED levels), 2) Comparing several transport modes (road and bicycle) and routing engines. An initial set of results focuses on the **impact of choosing one of the two open-source routing engines**, OSRM of Valhalla, on travel time indicators. Based on the GRANULAR's case study Pays-Pyrénées-Mediterranée (FR), it shows that travel times (from populated grid cells to secondary schools) derived from these two routing engines are highly correlated. The main differences observed are due to (1) the default settings of the routing engines (maximum speed, route preferences), easier to implement with Valhalla, and (2) the default algorithm used to reach the OSM network from a given point (grid cell centroid): with Valhalla, it can in some cases cause significant detours, this being observed mainly for car profile. Secondly, **several datasets were produced, covering 13 case studies (7 GRANULAR Living Labs and 6 Replication Labs)** at three scales

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).*

geographic granularities (1km grid, LAU, living-replication labs). Two main categories of indicators were calculated: **accessibility indicators** on travel time to the nearest school by car and by bike using OSRM (**fig. 4**) and Valhalla, and indicators based on **comparison between car and bike transport mode** (**fig. 5**). Even if the diversity in size of the Living / Replication Labs makes it difficult to compare (in particular for indicators most sensitive to a scale effect, such as the travel time to the nearest travel school), some key points can be mentioned: the results show for instance that for every Living Lab but the Dutch one (P10), more than 60% of young people (under 15 years old) have access to a secondary school under 10 minutes by car, and half of the young population has access to a school under 7 minutes by car. By bike, the ranks stay globally the same, but with higher disparity between Living Labs. More than 60% of young people have access to a school under 20 minutes (**fig. 6**), except for Netherlands. Setting the maximum time threshold at 20 minutes, Pays Pyrénées Méditerranée is ranked 7th by car accessibility (97% of young people), and 11th by bike accessibility (60% of young people), probably due to the partially mountainous topography of the region. A more direct comparison between car and bike accessibility can highlight further considerations about transport mode shift, for more localised studies. For instance, mapping the difference in travel time between car and bike at a granular level helps identifying areas where car could be substituted by bike (travel time by bike to the nearest school below 20 minutes and not exceeding time by car by more than 10 minutes) and where new facilities for cycling could be implemented.

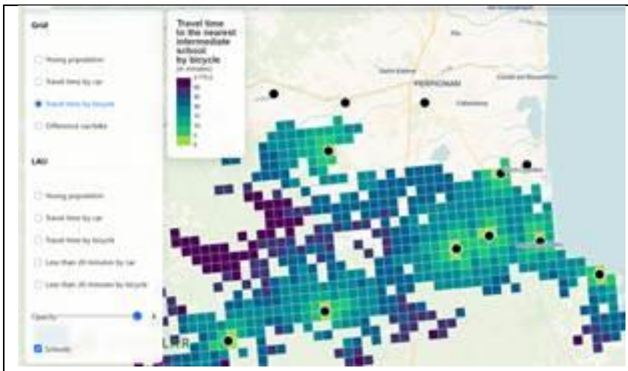


*Figure 4. Travel time to the nearest intermediate school by bicycle, in the Pyrenees-Mediterranean area*

(sources: European Union, 1995 – 2026 © European Commission © REGIOgis © OpenStreetMap and contributors) © data.gov.uk

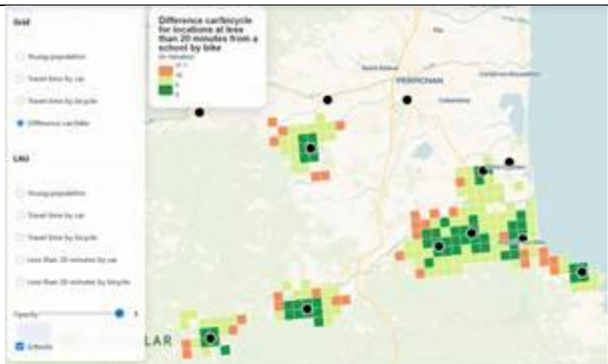


*Figure 5. Difference car / bicycle in travel time to the nearest intermediate school from each grid cell located within 20 min. of a secondary school, in the Pyrenees-Mediterranean area*

(sources: European Union, 1995 – 2026 © European Commission © REGIOgis © OpenStreetMap and contributors) © data.gov.uk

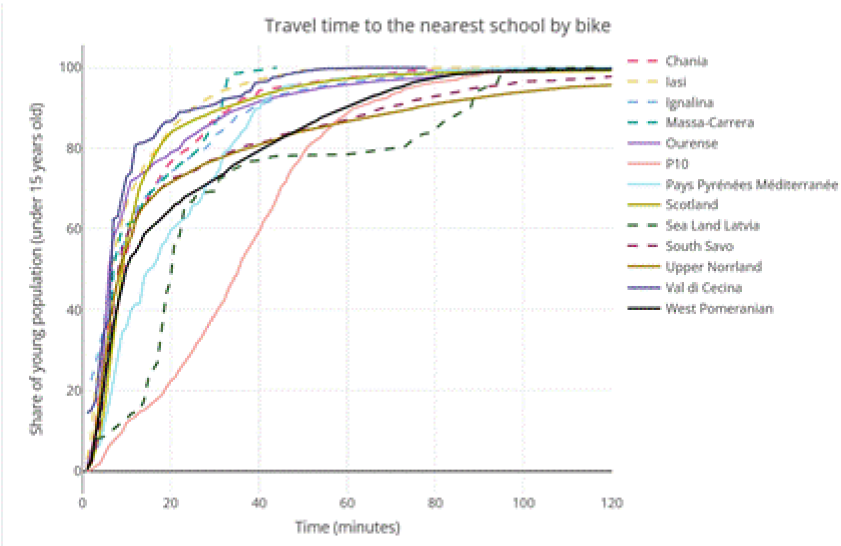



*Figure 6. Share of young people having access to a secondary school under a given time threshold, by bike: comparaison of GRANULAR 7 Living Labs and 6 Replication Labs*
(sources: European Union, 1995 – 2026 © European Commission © REGIOgis © OpenStreetMap and contributors © data.gov.uk)

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).*

***Comparing accessibility by public transportation to accessibility by car: a reproducible methodology tested in the French living lab region***

In this final, more exploratory, section, the focus is on producing test visualizations and indicators that allow for a comparison of accessibility by car with public transportation accessibility, combined with other modes of transport for access to stations. Starting with local measures (i.e. between one location -in our case, the city of Perpignan- and multiple ones) related to isochron maps, preliminary results address one main question: leaving the Perpignan main train station at 8 a.m. on a worked tuesday, where and how far can I go in 60 minutes by car ? by public transport ? by bike ? by mixing both (**fig.7**) ? An interesting variation is to find out where people live who can work an 8-hour day in Perpignan while leaving home after 7 a.m. and returning before 9 p.m. (**fig.8**). In the first case, the combination of bicycle and public transport is quite interesting: in the same amount of time, it is possible to reach destinations that are twice as far away as those reachable by public transportation and walk.

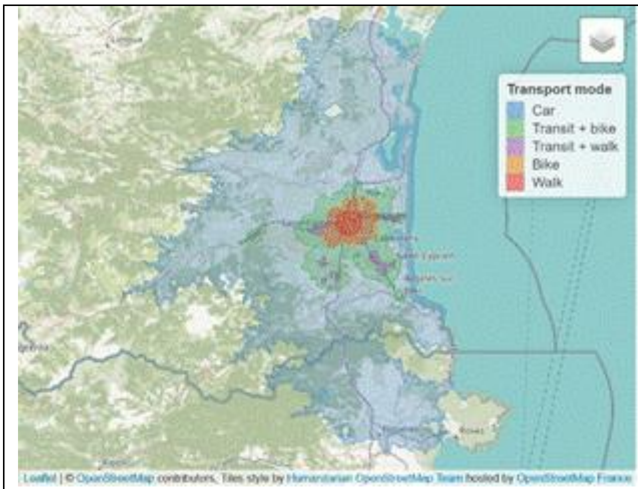



*Figure 7. Comparison of areas accessible within one hour from Perpignan using different modes of transportation*

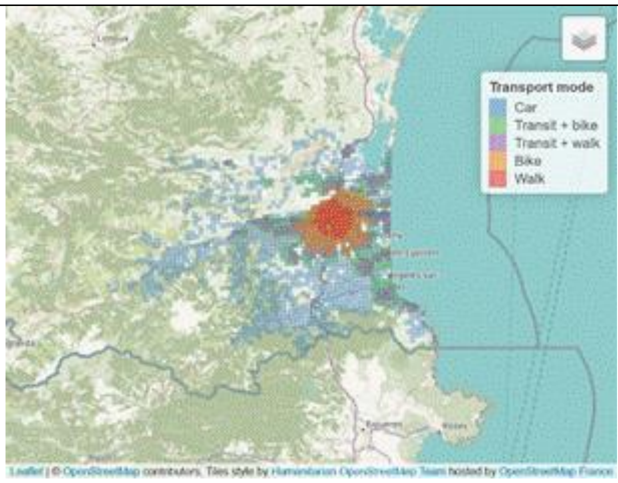



*Figure 8. Comparison of areas where people living in Pyrénées Orientales area (FR) can spend a 8-hour working day in Perpignan, leaving home after 7 a.m. and returning before 9 p.m.*

Further analyses are underway to implement indicators reflecting the point of view of people in rural areas who wish to access a larger set of towns and cities. They aim at answering questions such as: for these people, how long does it take to get to the nearest town/city (above 5,000, 10,000 or 50,000 inh.) by public transportation and how much longer does it take compared to driving a car ? Or, what percentage of the Living Lab's population can reach a small town in less than half an hour by public transportation and how does this compare to accessibility by car ? These analyses can also be broken down by population characteristics, for instance young people (under 15) or the older ones, or people of working age (15-64). The methodology developed, based on the case study of the French LL region, should be adaptable to other contexts.

**3. Implementation and validation at local level (LLs) / Adaptation/upscaling to RLs**

The following describe two examples of validation at the Living Lab level.

| **Attribute** | **Description** |
|---|---|
| Living Lab | **French LL** |
| Context | A seminar (28 February 2024) with local stakeholders (data-oriented technicians mainly) of the Pays-Pyrenees-Mediterranee LL to :<br>- Show and discuss the validity of accessibility indicators implemented within the project.<br>- Discuss the potential of OpenStreetMap as a data source for local monitoring and/or receptacle for data storage/sharing between technicians among the LL. |

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).*

| | |
|---|---|
| Implementation | A presentation highlighting the main indicators/results in the French LL and the potential of OSM for local studies |
| Validation | Global validation of the overall strategy<br><br>A particular interest in the indicator measuring the travel time difference between the first and the second facility/point of interest, reflecting spatial inequalities in terms of the ability to choose different destinations<br><br>Need for indicators at high granularity (1km grid cell), but difficult to use it with policy stakeholders. |
| Outcome | GRANULAR made the link between scientists, policy makers at local scale and data-oriented technicians, demonstrating through concrete examples what can be done about the accessibility issue using open data sources, with some technical skills. |

| **Attribute** | **Description** |
|---|---|
| Living Lab | **Scottish LL** |
| Context | For the Scottish area, GISCO does not provide any data on secondary schools and on demographic age-class at the 1km grid level. The Scottish LL sent us regional data to determine whether the methodology applied to the EU context can be adapted to other ones. |
| Implementation | Adaptation of the methodological data flow to the Scottish data. It did impact the process on OSM data. |
| Validation | It was possible to create output indicators on accessibility to secondary schools for the Scottish context. |
| Outcome | Based on a proposal from the LL, a new visualization tool has been produced at the European level, integrating the Scottish context. This was incorporated into an immersive suite that was screened within the LL. |

## 4. Limitations

**From a theoretical point of view**, it is important to keep in mind that these indicators, which are key to quantifying inequalities in access on a large scale, emphasize the spatial dimension of access, without taking into account other factors affecting rural populations' access to resources, such as economic dimension of access (ticket/gas price), individual components (socio-economic characteristics, physical abilities, etc.) that influence mobility conditions, or even sometimes institutional features (like catchment areas for schools). **As regards to data and methods**, the main limitations relate to remaining gaps in data at the European scale, for the identification of relevant destinations related to services and facilities (see section 2, data). One of the key principle established in this subtask, namely to use only open-source and harmonized data, sometimes narrowed the range of usable data. OSM POI datasets are still rather incomplete in low-density areas. Regarding GISCO database, data attributes seem more consistent for schools than for hospitals, but the spatial coverage is quite incomplete, with a notable lack of data for Germany and Bulgaria. This could thus seem out of step with the expectations of some local stakeholders who would like to be able to feed in their local data to perform accessibility calculations for a specific service. Furthermore, harmonized data about services and facilities still face limitations when it comes to making comparisons over time. Nevertheless, it is a process that is just beginning and the methodology implemented in Granular notebooks could benefit from the delivery of new datasets in the forecoming years. Lastly, the implementation of accessibility measures based on public transportation to a case study could not be implemented across all of Europe within the project's timeframe. But it lays the ground for promising studies, based on a reproducible and shared methodology. Other information could be added about the development of other alternative modes of transportation, for instance car sharing spots

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).*

and city-bike, but r5r does not allow the integration of this kind of data (GBFS tables) for the moment. Other limitations include more practical limits, related to **the skills required for methodological processing and data handling with open science statistical tools**, thereby highlighting the importance of data viewer navigation for less experienced users.

### 5. Contribution to policy/call objectives:

The outputs provided by this subtask contribute to quantify the potential access to resources and allow an improved understanding of its heterogeneity in rural areas, across multiple spatial scales. As such, it offers a significant contribution to the typologies of rural areas, considering driving times. In addition, it gives an important insight into the potential of open-source tools and global data for including public transport and cycling/walking solutions to these indicators.

### 6. How to use this dataset?

There are three different ways to use the outputs of the accessibility research theme, suitable for different audiences:

**Data viewers**

Interactive geovisualisation tools have been produced to disseminate the results and enhance the visibility of the dataset to a wide audience (first and foremost the project's Multi-Actor Labs), outside the community that produced them. Two data viewer tools have been designed as user-friendly portals to make it easier to navigate through the various datasets produced: [the first one](), dedicated to **accessibility to small and medium-sized cities,** displays at the grid level and at European scale either travel times to the nearest towns and cities (**fig. 9**) or the number of towns and cities that it is possible to reach by car within 30 minutes, while providing links to the datasets and to the notebooks' documentation on data processing. [The second one]() concerns **accessibility to intermediate schools** in Granular Living and Replication Labs. It allows to view raw and derived indicators at two different scales: at the grid level, not only travel time by car/bicycle to school but also the time difference between car/bicycle are represented. At the LAU level, the median travel times by car/bicycle are completed by the share of the young population living at less than 20 minutes by car/bicycle.

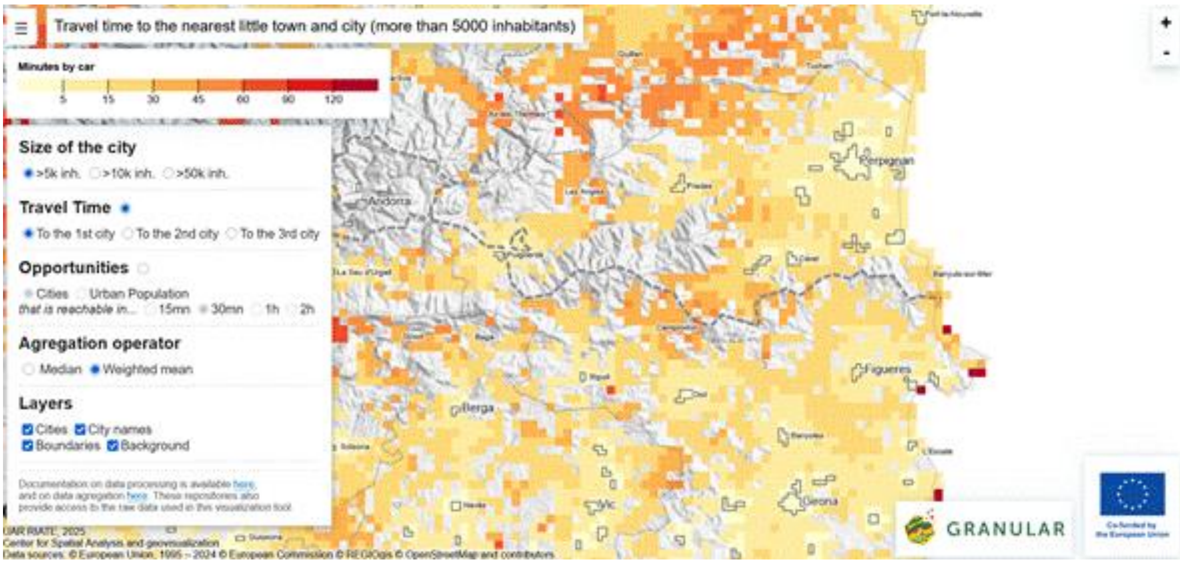



*Figure 9. The data viewer as a user-friendly navigation interface and a portal providing access to indicators and documentation*

**Access to raw output data**

Different options for accessing the output methodology and data are available, depending on the subtasks considered. Regarding **accessibility by car to towns and cities**, a first Zenodo archive allows to retrieve the travel time matrix by car and the related set of accessibility indicators to small and medium-sized cities at 1km grid cells. Data can be downloaded at the European scale or at each country level. The related notebooks describe the input data used, show the R code used to manage the data processing required to build output indicators and provide information and metadata about output indicators. The outcomes related to **accessibility to schools by car and bike** are also accessible from a second Zenodo archive. A specific section shows how to import output indicators


*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).*

(travel time by car and bike to the nearest intermediate schools) and to create some plots and maps in the R programming language, based on the example of the French Living Lab. For the more exploratory subtask related **to public transport accessibility indicators**, there will be a notebook (coming in July) describing the methodology implemented on a test area scale (Pyrénées Orientales, the NUTS3 where the French living lab is located).

**Handling accessibility indicators to characterize rural areas at different scales**

For users with more advanced data handling skills, additional resources are provided to characterize rural areas at different scales, answering to questions such as: at regional scales, what are the NUTS2 where inhabitants of rural areas have to drive more than 20 minutes in average to reach the nearest town above 5,000 inh.? At any scale of analysis, what percentage of rural population can be considered to live far from towns and cities ? Specific sections in the notebooks explain for instance: 1) how to aggregate and summarize the accessibility measures disseminated at the 1km grid level, up to different spatial unit levels, here, or 2) how to use population attributes of grid cells in order to measure how much of the population is located at less than *xx* minutes from a town or city and display these outputs using cumulative frequency plots, here. In addition, both examples illustrating these applications (**fig. 10**) highlight the interest of cross-referencing accessibility indicators with the DEGURBA classification. Further applications of these findings are currently being explored (paper in progress for a submission to the International Journal of Geo-information in progress).

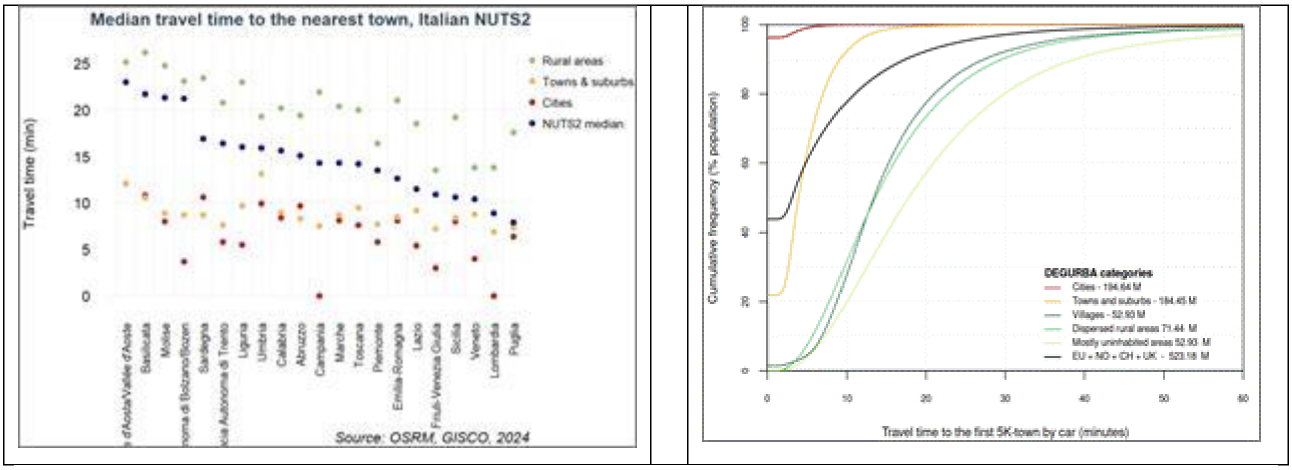



*Figure 10. Two examples of the visualisation of indicators derived from output data: a) aggregation at the NUTS2 level (Italy) of car travel time to the nearest town, broken down by DEGURBA categories (left), b) cumulative frequency plot to visualise how much of the population is living at less than xx minutes from a town above 5,000 inh., according to DEGURBA detailed rural categories (EU-wide).*

## 4.2. Mobility **(UoS, UMI)**

Authors: Hal Voepel, Shengjie Lai, Ale Sorichetta, Michael Kull, Kirsi Korhonen, Anne Vänttinen, Louise Chasset, Ilaria Lusini, Fabiana Stortini, Marco Ricci, Martin Hofer

1. Objectives

The objective of the mobility subtask is to process social media-based mobility data into a format suitable for the GRANULAR's partners and stakeholders. The specific aim is to assemble, integrate, and harmonize the mobility data in an effort to produce within-country comparable metrics, maps, and plots developed with a particular focus on rural areas but also to include their connectivity to urban areas (as most stakeholders consider mobilities between urban and rural areas highly important in order to characterize and improve the latter personal communication with lab stakeholders). **The overall goal is to develop datasets to characterize human mobility and connectivity within, from and to several living and replicate labs (small, mostly rural regions) for supporting policy and decision makers.**

**2. General approach/ analytical framework**

**Description of mobility datasets**

Two Facebook domestic mobility datasets from Meta were used in this work: (1) the Mobility During Crisis (CR) dataset spanning March 2020 to May 2022, and (2) the Activity Space Maps (AS) dataset spanning July 2024 to

Funded by the European Union

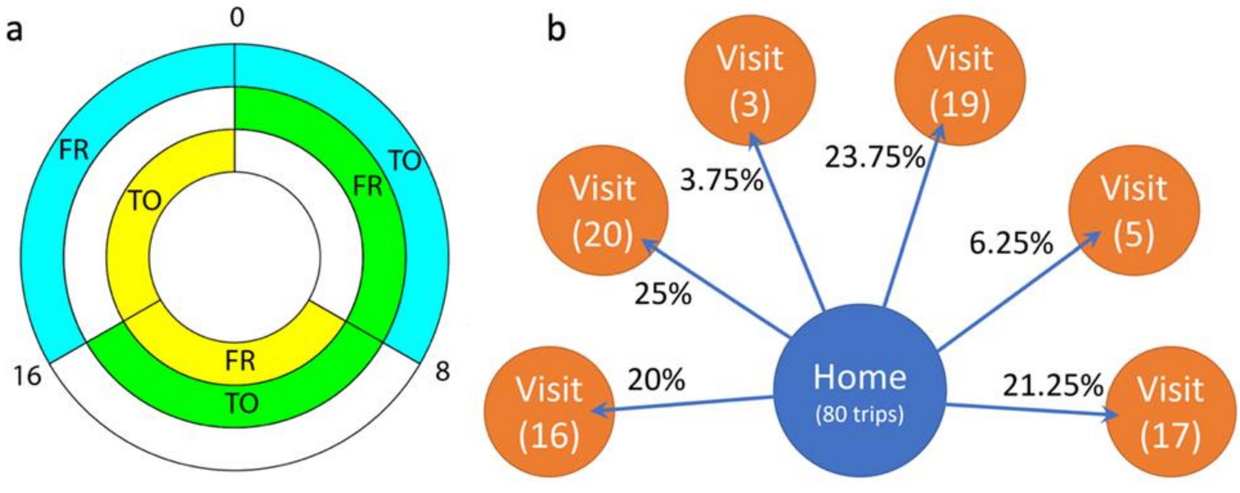
*Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. UK participants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

December 2025. Both datasets are globally available and are provided by country. CR data are provided in two formats: (i) tiled datasets, where the smallest tile size permitted for privacy protection is approximately 600 × 600 meters at the equator, and (ii) CSV files containing point and line coordinates, which can be converted into spatially explicit features (e.g., shapefiles). AS data are available as 0.05° (~5.5 km in isometric coordinates) resolution grids. Mobility measurements for both datasets are sampled movements from Facebook mobile phone apps where users have opted-in to allowing their phone locations to be tracked. Sampling methods used by Meta to derive both datasets legally meet anonymization standards, ensuring that individual Facebook app users cannot be identified or tracked. Specifically, movement counts are only reported when at least 10 individuals contribute to an aggregated movement cluster. Aggregations below the threshold of 10 individual movements are instead recorded as missing data, indicating that mobility did occur but numbers of individual movements are not recorded in the datasets. The sampling designs, aggregation methods, and spatiotemporal resolution for each dataset are, however, quite different.

**CR data** are sampled every 8-hour period at 0000, 0800, and 1600 (in UTC) each day. These samplings are locations where an individual spends the most amount of time during an initial 8-hour period as a “origin” location of movement followed by where they spend most of their time in the next 8-hour period as a “destination” location of movement. A cluster of individual movements from a common origin (FR) location to another common destination (TO) are aggregated counts (plus a small random count) with the summed counts assigned a centre time stamp between the FR 8-hour sample period and the TO 8-hour sample period (see Figure 11a). For example, the 0800-mobility period is an aggregate count with origin movement sampled from 0000 to 7:59 (green FR) and destination movement sampled from 0800 to 15:59 (green TO). Subsequent to 0800 is the 1600-mobility period, which has an origin sampling from 0800 to 15:59 (yellow FR) and a destination sampling from 1600 to 23:59 (yellow TO). The 0000-mobility period (cyan) follows this similar pattern to complete the diurnal cycle for a single day and the sampling pattern repeats over all subsequent days in the period of record. Each of 8-hour measurement periods are sampled over 16 hours total lapsed time, and consecutive measurement periods overlap each other by 8 hours. Hence, the TO sampling period for the current measurement becomes the FR sampling period for the next measurement.

**AS data** have a sampling period of 3 weeks in length, which are updated once per week (52 or 53 samplings per year) across two separate time periods: daytime (8am to 6pm, local time) and night-time (8pm to 6am). AS data are proportions of movements (i.e., visit fractions) from a “home” location to multiple “visit” locations, where the sum of proportions across all visit locations, originating from a single home location on any given weekly time step, is equal to one. The definition of home location is where an individual spends most of their night-time during the 3-week period. Aggregate proportions are made by obtaining summed counts of collected individuals who travel from a single home location to a collection of visit locations over the course of a single 3-week sampling period, where each home-visit location pair has an independent summed count. Then each home-visit paired count is divided by the summed total of all the paired home-visit counts to obtain proportions for that single home location (see Figure 11b). Any grid location can be either a home or visit location on any given day as the samplings for each home location and their paired visit locations are done independently of all other home locations and their associated visit pairs during any given weekly period, and where daytime and night-time periods are also measured independently of one another.

As the University of Southampton has entered into a data sharing agreement with Meta, the raw data obtained for both CR and AS Facebook datasets may not be disclosed to third parties. However, any derived aggregations or modelled outputs, which production is described in the Methods section, will be shared with the project partners and stakeholders.




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).

*Figure 11. Two sampling designs used by Meta to measure human mobility counts via the Facebook mobile app, where users have opt-in to allow their phone location to be tracked: (a) CR mobility is measured counts given in 8-hour increments (0, 8, 16) sampled over a 16-hour period, where a collection of individuals spend most of their time during the leading 8 hours (FR) is followed by most time spent during a concluding 8 hours (TO) for each of the 8 hour time steps of the day; and (b) AS mobility is measured proportions in weekly increments sampled over 3 week intervals as daytime (8am to 6pm) and night-time (8pm to 6am) periods. The "home" location (blue disc) is determined by most time spent by an individual during the night-time period, and the proportion of movement to all corresponding "visit" location (all orange discs) is the aggregate proportion of many individuals with a common home and visit location. The summed total proportion across all home-visit pairs associated with a single home location during a single weekly sampling period is one (here as 100%).*

### Methods

The Methods section is divided into two sub-sections: (1) the first one describing the general processing of mobility data metrics, maps, plots and aggregated tables across all 17 living and replicate labs, and (2) the second one describing the bespoke processing and analysis for specific labs as per their individual requests. The bespoke work was completed for the French, Scottish, Finnish, and Italian labs on an individual basis, with some common analysis for two or more labs.

#### *General Processing: Facebook Crisis Maps and Flow Accumulation Metrics*

From the Facebook CR data for the country in which each living (or replicate) lab resides, line and point feature data were created from CSV tables. Each table contains two sets of coordinates, representing the starting and ending positions of mobility for each of the 8-hour periods: a larger scale in which coordinates are centroid locations of large administrative unites, roughly corresponding to NUTS (Nomenclature of Territorial Units for Statistics) at various administration levels for each country, and GPS centroid locations in which each centroid represents the average GPS locations for the clustered group locations in the sample. As several labs have smaller administrative levels than the corresponding NUTS level administrative units at a given location, average GPS coordinates were used across all living and replicate labs. Indeed, this approach enabled a more spatially refined representation of mobility both within labs (i.e., between administrative units within each lab) and between those units and other administrative units within the same country, rather than relying on a single centroid to represent each lab. For each lab, a separate point feature (derived from start/end line positions) and its corresponding line features were created across the whole country, and then, wherever paired starting-ending points were both outside of the lab, those paired features were dropped from the point feature object. Subsequently, line features were then updated to drop any line features associated with those dropped point pairs. Mobility data were appended to feature dataset representing LAUs (Local Administrative Units) and DEGURBA (Degree of Urbanisation) administrative units, all of which are available through Eurostat. Each feature dataset was then appended with local datetimes, which were derived from default universal (UTC) datetime stamps. From the datetime values, Morning/Midday/Evening "period" attribute was derived to identify each 8-hour time block, and a Weekday/Weekend "weekends" attribute was added. For point features, a start/end "position" attribute was added to note direction of movement, and a IN/OUT "proximity" attribute was added to identify those point features located inside/outside the labs. For line features, NUTS/LAU and DEGURBA levels along with "position" and "proximity" attributes were each assigned to separate attributes for starting and ending movements, which are important to identify flow direction across all point-paired attributes. **Please note** that these datasets are not available to be shared as they contain raw mobility data, which are restricted to WorldPop internal use only as per the data sharing agreement. These spatial feature data are, however, used to derive flow metrics and provide other forms of analysis, for the general and bespoke analytics, which results will be shared with project partners. See Figure 12 for an example CR mapped points and lines in Poland.

The measurement window for the 8-hour time steps of the CR data is quite short, which sometimes limits the number of individual movements observed within both FR and TO sampling periods, particularly in rural areas where movements can be sparse. As such, there are a large number of measured mobility observations that do not reach the minimum threshold for inclusion in the dataset. Indeed, upon initial examination of the data, as much as half of the mobility data across all labs are recorded as missing values (i.e., n < 10). This condition made analysis of rural areas for some living labs somewhat tenuous depending upon which analytical method was used. To deal with this, two approaches were used: (1) gap-filled missing mobility values with a mean value of 5 (halfway between 1 and 9), and (2) aggregate mobilities as daily and periodic (morning, midday, evening) flow accumulation metrics. Flow accumulation metrics are defined for each LAU as: internal (flows start and end within the same LAU), outgoing (flows start in the LAU and end up elsewhere), incoming (flow end in the LAU from somewhere else). Combinations of these three flow accumulation metrics are: starting (internal + outgoing), ending (internal + incoming), and total (internal + incoming + outgoing). Daily metrics are aggregated for each date and periodic metrics for each datetime. **Daily and periodic flow metric data are available as CSV files for all living and replicate labs with data structure shown in Table 2**. The default DEGURBA Level 1 classification (i.e., Cities, Towns, and Rural Areas) was renamed DEGURBA3. Two more DEGURBA classifications were also derived (see Table 2).

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).*

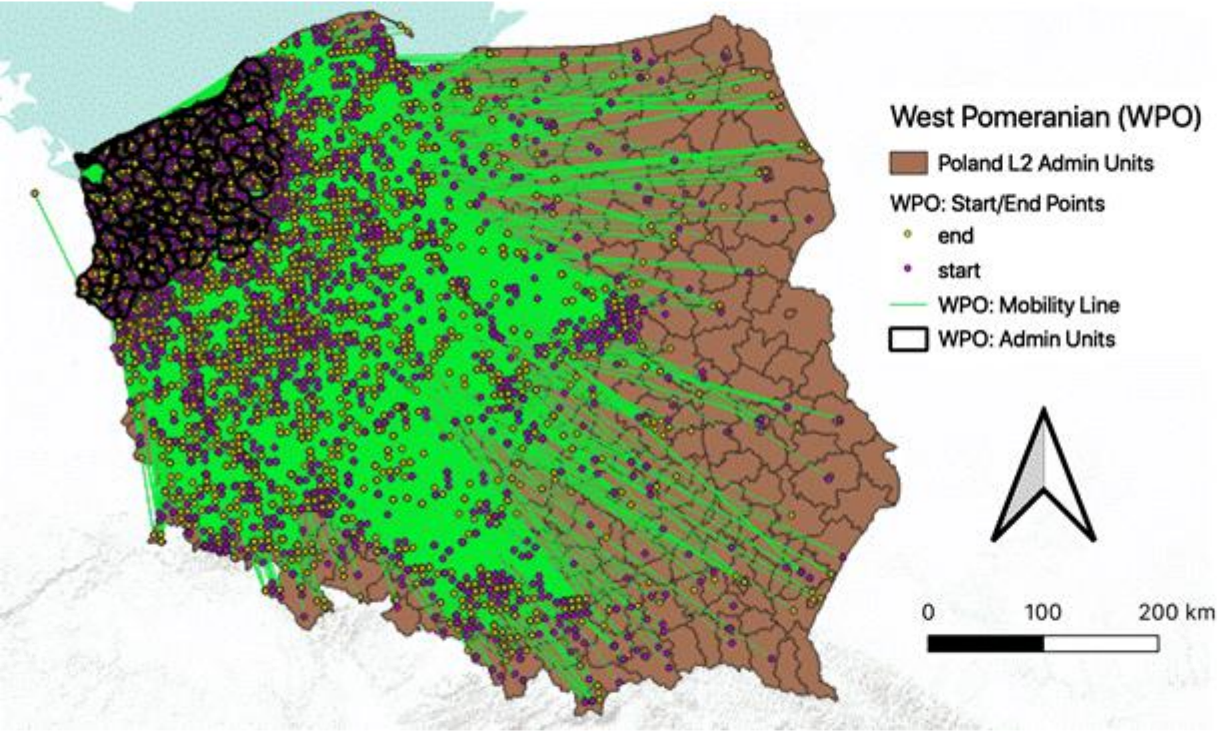



*Figure 12. As an example of Facebook Mobility During Crisis (CR) data developed as spatiotemporal data for West Pomeranian (WPO) living lab in Poland showing living lab admin units (black outline polygons), the starting coordinates (purple points) and ending coordinates (yellow points), and their corresponding mobility lines connecting them (bright green lines).*

**Table 2. Daily and periodic flow accumulation metric data table structure and description.**

| Name | Values |
|---|---|
| metric | internal, outgoing, incoming, starting, ending, total |
| dates† | Daily: YYYY-MM-DD or Periodic: YYYY-MM-DD hh:mm:ss |
| period† | Daily (for daily files), Morning, Midday, Evening (for periodic files) |
| weekends | Weekday, Weekend |
| country | ISO2 country code |
| LAU | Local Administration Unit name |
| DGURBA2* | NonRural, Rural |
| DGURBA3 | Cities, Towns, Rural (default DGURBA values) |
| DGURBA4** | FUA (functional urban area), Cities, Towns, Rural |
| n | Flow accumulation metric |

† Daily and periodic data are tabulated as separate CSV files with similar data types except for date formats.

* DEGURBA2 reclassified NonRural value are combined Cities and Towns from DEGURBA3.

** DEGURBA4 are default DEGURBA3 plus FUA from polygon features which overwrite default DEGURBA values.

***General Processing: Facebook Activity Space animation maps, weekly series plots, and connectivity maps***

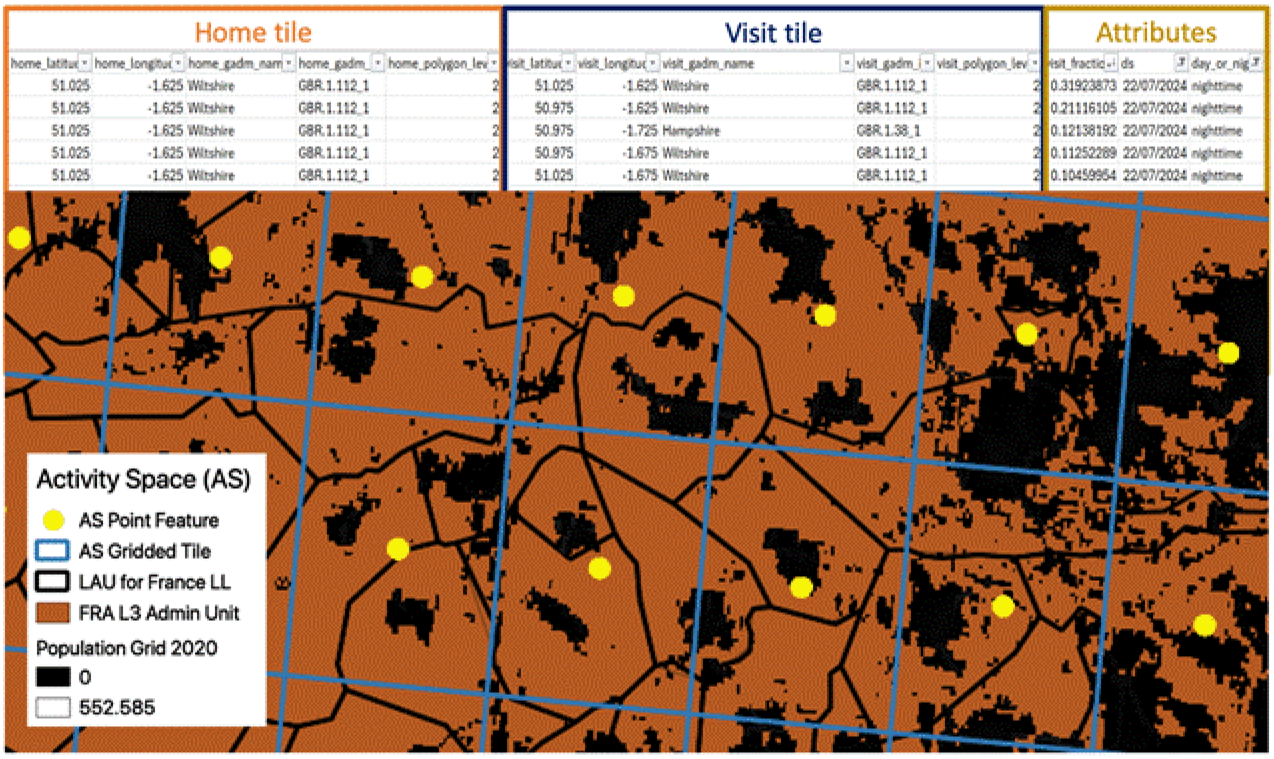
The Facebook AS data were used to develop weekly LAU-level animation maps, time series plots, and seasonal connectivity network maps for all living and replicate labs. Rather than mapping and plotting visit fractions, which would not be useful without knowing the corresponding denominators, a visit fraction of a total population was calculated across AS grid points using 2024 WorldPop 100 m resolution grids (depicting the corresponding spatial distribution of the residential populations), under the assumption that Facebook users and the general population have similar short-term or seasonal mobility patterns. A spatial algorithms was used to develop datasets to produce maps and plots combining the original AS data with a spatial resolution of ~5.5km with the WorldPop population distribution dataset with a spatial resolution of 100m. As derived mobility counts are proportional to population counts, using such metric as a proxy for mobility is a reasonable method to obtain a "maximum possible" mobilization

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).*

count; however, this presumes that the AS data are representative of the mobility of the entire population. For mobility comparison across regions, all resulting mobilities were normalized by a maximum or mean value over an appropriate range of values (e.g., over LAU, by season, etc) depending on the output. The final result yields a relative mobility to either a maximum value (as either proportion or percentage) or to a mean value plotted as a multiple of the average value that typically occurs over the year or a season.

Weekly-sampled Facebook AS data are available in CSV format for daytime and night-time datasets by country, and the daytime datasets were used in the framework of this study. Data for "Home" and "Visit" tiles are given as separate sets of columns in the CSV files along with attributes for the visit fraction, weekly sampling date, and daytime/night-time sampling times (see partial table at the top of Figure 13). Point features for home/visit locations (shown in yellow in Figure 13) were used to create 5 km home/visit tile grids (blue bordered tiles in Figure 13). Visit fraction values for the point features are also assigned to the corresponding tile grids. Assuming populations in each cell with a AS tile has same home-visit fraction, the corresponding 2024 WorldPop 100 m population grid (black cells in Figure 13) was multiplied by the tile grids.



| Home tile | | | | | Visit tile | | | | | Attributes | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| home_latitu | home_longitu | home_gadm_nam | home_gadm_ | home_polygon_lev | isit_latitu | visit_longitu | visit_gadm_name | visit_gadm_i | visit_polygon_lev | isit_fractio | ds | day_or_nig |
| 51.025 | -1.625 | Wiltshire | GBR.1.112_1 | 2 | 51.025 | -1.625 | Wiltshire | GBR.1.112_1 | 2 | 0.31923873 | 22/07/2024 | nighttime |
| 51.025 | -1.625 | Wiltshire | GBR.1.112_1 | 2 | 50.975 | -1.625 | Wiltshire | GBR.1.112_1 | 2 | 0.21116105 | 22/07/2024 | nighttime |
| 51.025 | -1.625 | Wiltshire | GBR.1.112_1 | 2 | 50.975 | -1.725 | Hampshire | GBR.1.38_1 | 2 | 0.12138192 | 22/07/2024 | nighttime |
| 51.025 | -1.625 | Wiltshire | GBR.1.112_1 | 2 | 50.975 | -1.675 | Wiltshire | GBR.1.112_1 | 2 | 0.11252289 | 22/07/2024 | nighttime |
| 51.025 | -1.625 | Wiltshire | GBR.1.112_1 | 2 | 51.025 | -1.675 | Wiltshire | GBR.1.112_1 | 2 | 0.10459954 | 22/07/2024 | nighttime |



*Figure 13. The spatial algorithm to obtain visiting fractions of population combines Activity Space (AS) gridded tiles at about 5.5 km (blue rectangles) with about 100 m population cells from WorldPop.*

Weekly animation maps were created by normalizing the fractional population mobility values by the annual mean value over the period of record (POR: July 2024 to Dec 2025). Weekly map animations were created separately for inbound visitors (see Figure 14a) and outbound residents for each LAU, as single animated GIF files, for all living and replicate labs. Weekly time series plots were created using both maximum- and mean-normalization of mobility population flow values over the POR. Weekly series plots are available as combined inbound-outbound plots (see Figure 14b) and split plots, which plots inbound and outbound separately. Seasonal network connectivity maps were created by seasonal mean-normalizing the fractional population mobility and mapping inbound visitors and outbound residence separately. Maps included span all four seasons and the entire calendar year of 2025. Titles on each map indicate season daytime flows (winter, spring, summer, autumn, whole year), flow direction (inbound, outbound), and LAU.

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).*

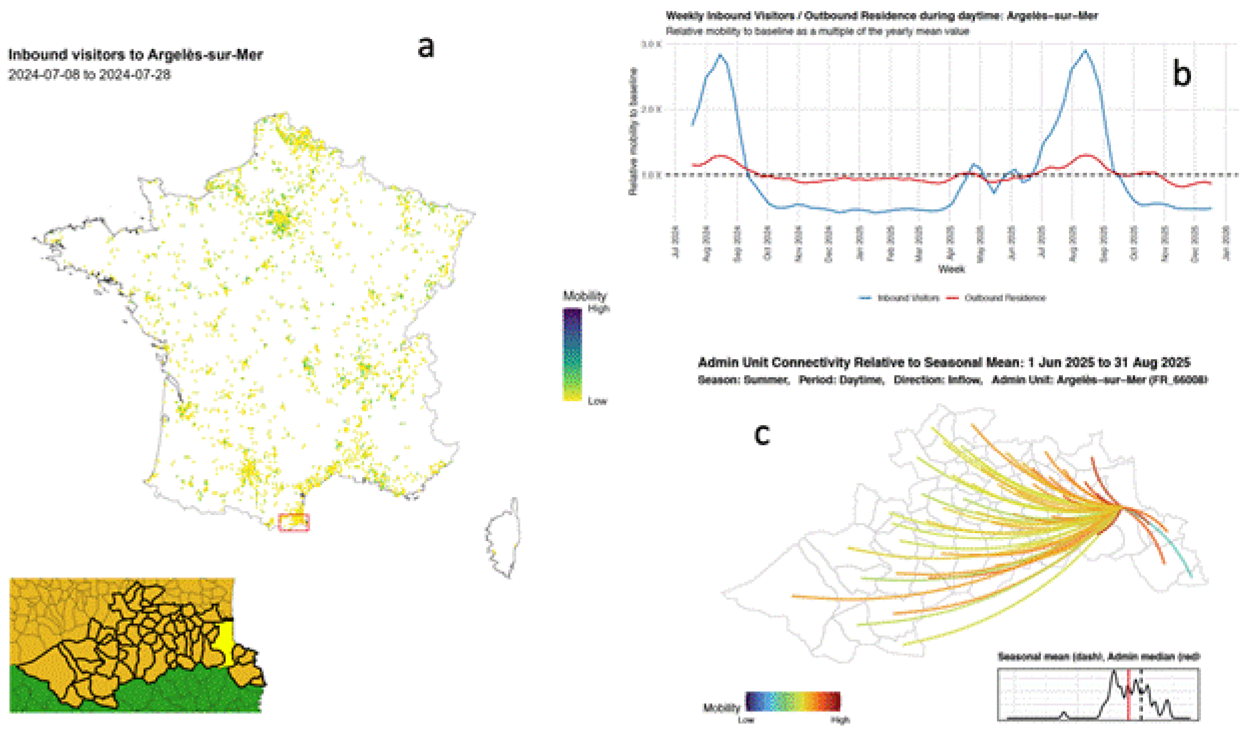



*Figure 14. Visual outputs of Facebook Activity Space Maps (AS) datasets for the coastal LAU Argelès-sur-Mer: (a) weekly animation map showing daytime mean-normalized fractional movement for inbound visitors (outbound residence available); (b) weekly time series plot showing daytime annual mean-normalized fractional movement of inbound visitors (blue line) and of outbound residence (red line) vertical axis scaled as a multiple of the annual mean value (horizonal dashed line) (note: max-normalized series (not shown) were also produced); and (c) connectivity map showing daytime summertime mean-normalized seasonal fractional movements for inbound visitors (outbound residence available).*

***Bespoke Processing: specific requests by France, Finland, Italy and Scotland labs***

Four lab partners provided specific stakeholder needs. These were the French (**FRA**) living lab Pays Pyrénées Méditerranée, the Finnish (**FIN**) replicate lab South Savo, the Italian (**ITA**) living lab Val di Cecina, and the Scottish (**SCO**) living lab Scotland. Through multiple online meetings and email exchanges with these lab partners, and on occasion with stakeholders, it was possible to determine which mobility data and analysis was best suited to their needs. These lab partners have communicated data and analysis outcomes directly with stakeholders and policymakers in their communities, which later resulted in direct feedback used to refine the analysis. This work will result in two manuscripts/papers: (1) a first one based on a common mobility analysis involving all three lab partners regarding mobility and connectivity spatio-temporal variation over holidays, events, and tourist seasons for the French, Finnish and Italian labs; and (2) a second one with only the Scotland lab which details mobility characteristics between urban and rural areas in Scotland. The following sub-sections describe the bespoke analysis performed and show some of the tabulated and visual outcomes of the corresponding results.

**Holidays and Events (FRA, FIN, ITA)**: Partners for these labs provided the information to identify specific periods of time related to holidays, events, and the lockdown period of the COVID-19 pandemic. The CR dataset was best suited to these analyses as most were short periods of a few days although others (e.g., lockdowns) were extended periods of time. These analyses tabulated mean mobility counts of the three primary flow accumulation metrics (internal, incoming, outgoing) during these events and compared those with their overall mean mobility count to obtain a percent change in overall mobility (for France, see Figure 15). This analysis was performed separately for daily flow metrics and for each diurnal periodic morning, midday, and evening.

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).*

**Event mobility for French Living Lab**

| | Metric | Event mean (% change from overall mean) Daily | Morning | Midday | Evening |
|---|---|---|---|---|---|
| Fête de la Saint Vincent | internal | 5033.6 (93.1%) | 1624.6 (47.2%) | 1552.8 (55.4%) | 1856.2 (57.2%) |
| | incoming | 35.9 (-12.2%) | 6.7 (-54.7%) | 48.7 (-2%) | 9 (-65.1%) |
| | outgoing | 35.4 (16.8%) | 10.8 (-64.2%) | 13.6 (-35.2%) | 39 (33.1%) |
| Pellicu-live | internal | 4514.7 (73.2%) | 1494 (35.4%) | 1385 (38.6%) | 1635.7 (38.5%) |
| | incoming | 33 (-19.3%) | 5 (-66.2%) | 30.3 (-39%) | 15.7 (-39.1%) |
| | outgoing | 27.1 (-10.6%) | 17.5 (-42.1%) | 11.2 (-46.7%) | 25.5 (-13%) |

**COVID mobility for French Living Lab**

| Metric | COVID lockdown mean (% change) Daily | Morning | Midday | Evening |
|---|---|---|---|---|
| internal | 2715.4 (4.2%) | 1154.5 (4.6%) | 1087.9 (8.8%) | 1234.5 (4.5%) |
| incoming | 50.4 (23.2%) | 7 (-52.7%) | 56.8 (14.3%) | 52.3 (102.7%) |
| outgoing | 32.1 (5.9%) | 5.5 (-81.8%) | 11.2 (-46.7%) | 40.8 (39.2%) |

**Holiday mobility for French Living Lab**

| | Metric | Holiday mean (% change from overall mean) Daily | Morning | Midday | Evening |
|---|---|---|---|---|---|
| public | internal | 2842.3 (9%) | 1216.7 (10.3%) | 1204.5 (20.5%) | 1481.9 (25.5%) |
| | incoming | 18.7 (-54.3%) | 7.5 (-49.3%) | 19 (-61.8%) | 20.9 (-19%) |
| | outgoing | 19 (-37.3%) | 6.7 (-77.8%) | 7.4 (-64.8%) | 20.6 (-29.7%) |
| school | internal | 3126.6 (19.9%) | 1252.2 (13.5%) | 1154.8 (15.5%) | 1394.7 (18.1%) |
| | incoming | 36.4 (-11%) | 7 (-52.7%) | 38.9 (-21.7%) | 30 (16.3%) |
| | outgoing | 32.4 (6.9%) | 8 (-73.5%) | 11.3 (-46.2%) | 38.2 (30.4%) |

*Figure 15. Images of output tables for holidays, events (as a partial table), and COVID lockdown periods for France. These tables show percent changes in the mean mobilities during these periods to those of the overall means using the three different primary flow accumulation metrics: internal, incoming, and outgoing.*

**Coastal Seasonal Tourism (ITA)**: In addition to the holiday and event mobilities, the Italian living lab partners were also interested in the movements between coastal administrative units of Val di Cecina, Bibbona, and Rosignano Marittimo, and the interior administrative units of the lab. Bibbona, and Rosignano Marittimo were added to the original Val di Cecina living lab at the request of the Italian lab partners as they noted these are important tourist areas to also consider in the analysis. These analyses focused primarily on starting and ending mobilities from and to the coastal administrative units, respectively, along with inland mobilities as a third category. Specific analysis included plotting running average mobilities over time for each category (see Figure 16a), plotting mobility counts vs travel distance (Figure 16b), and mapping mobility lines for starting and ending in a coastal area versus mobilities that occurred elsewhere in the living lab (Figure 16c).

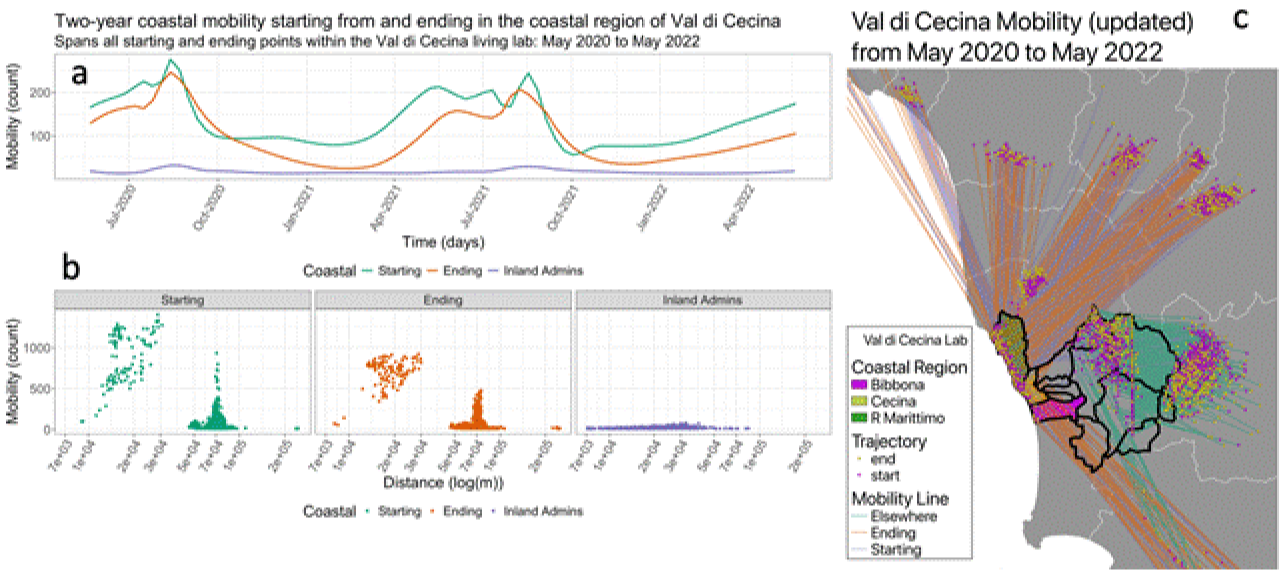



*Figure 16. Visual analysis of coastal mobilities for the Val di Cecina living lab in Italy. Mobilities for the coastal regions of Val di Cecina, Bibbona, and Rosignano Marittimo are categorized into outgoing (starting), incoming (ending), and interior (inland admins) movements. Analyses of LAU categories using raw CR data spanning a two-year period are: (a) running average of mobilities showing seasonal peaks over consecutive summer tourist seasons, (b) mobility counts versus travel distance, and (c) mapping mobilities with starting and ending trajectories (direction between points).*

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).

**Urban/Rural Mobilities (SCO)**: Scotland living lab partners were primarily interested in mobilities that occur between urban and rural areas. Their primary concern was identifying an urban/rural (UR) classification scheme that best reflected mobilities that occurred over differing short time scales, namely diurnal and sub-diurnal periods. Early work primarily focused on variations of the DEGURBA classification, where we used CR flow metrics data to examine weekends vs weekdays differences in mobility between paired-UR classifications, which were tested for their significant difference across all flow metrics (Figure 17). We later performed the same analysis using the Scotland government (SG) urban/rural classifications (Figure 18). Although both figures show 3-fold classifications, we performed analyses across 2-, 3-, and 4-fold classifications comparing DEGURBA and SG classifications.

An additional mobility analysis was to study sub-diurnal stationary mobilities between administrative units to observe patterns between UR classified administrative units or regions of both the DEGURBA and SG classifications. The raw CR dataset was portioned into three tables: weekdays, weekends/holidays, and every day. Then mobility data are used to empirically derive time-varying probability transition matrices for each sub-diurnal period (morning, midday, evening) separately from each of the three data table. A time-varying Markov process model was used to study the periodicity of diurnal mobilities across their stationarity distribution. As noted earlier, using Facebook users as a proxy for mobilized populations is reasonable for use in modelling; however, it is problematic as it assumes that Facebook users are representative of the entire population. Therefore, rather than plotting population movements, the difference in consecutive time steps was taken to obtain first-difference series (see simple one-to-one example for Aberdeen City and Aberdeenshire in Figure 19). These plots show increased population above the zero line for Aberdeen and decreased population below zero. Changes in population in these two plots are a mirror reflection of one another because Aberdeen City is surrounded by the singular rural region of Aberdeenshire – the other urban-rural examples (not shown) are more complicated.

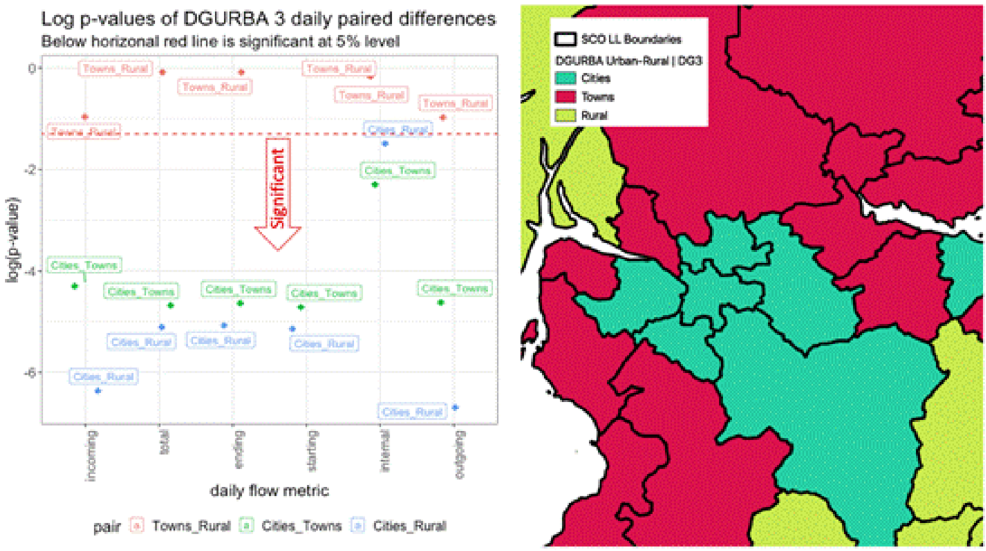
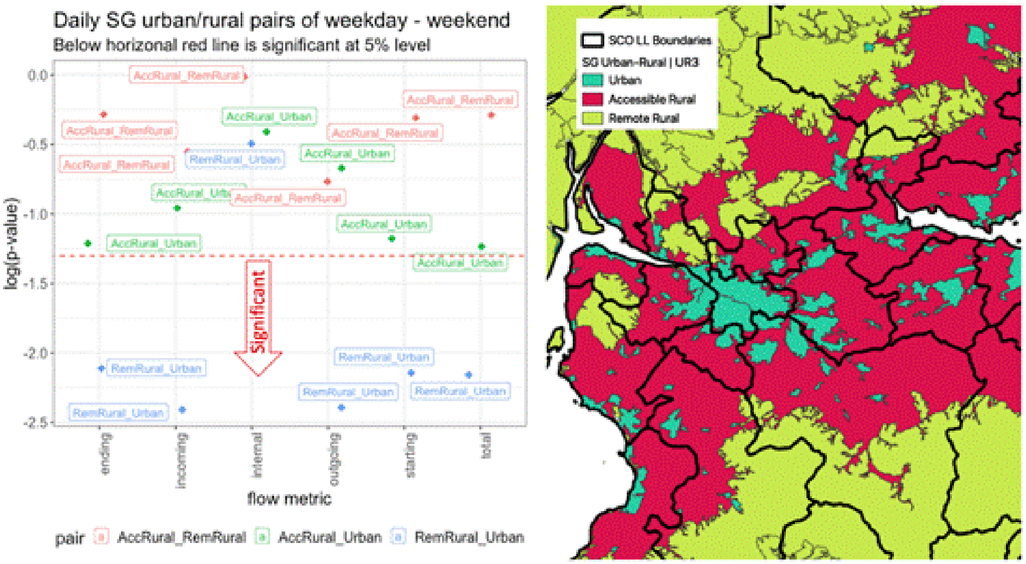



*Figure 17. All pairs comparison of the 3-fold DEGURBA classification between Cities, Towns, and Rural Areas. Significance tests are shown in the plot of log(p-value) over six flow metrics showing consistency*




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).*

*Figure 18. All pairs comparison of the 3-fold SG classification between Urban, Accessible Rural, and Remote Rural. Significance tests shown in the plot of log(p-value) over six flow metrics showing some consistency*

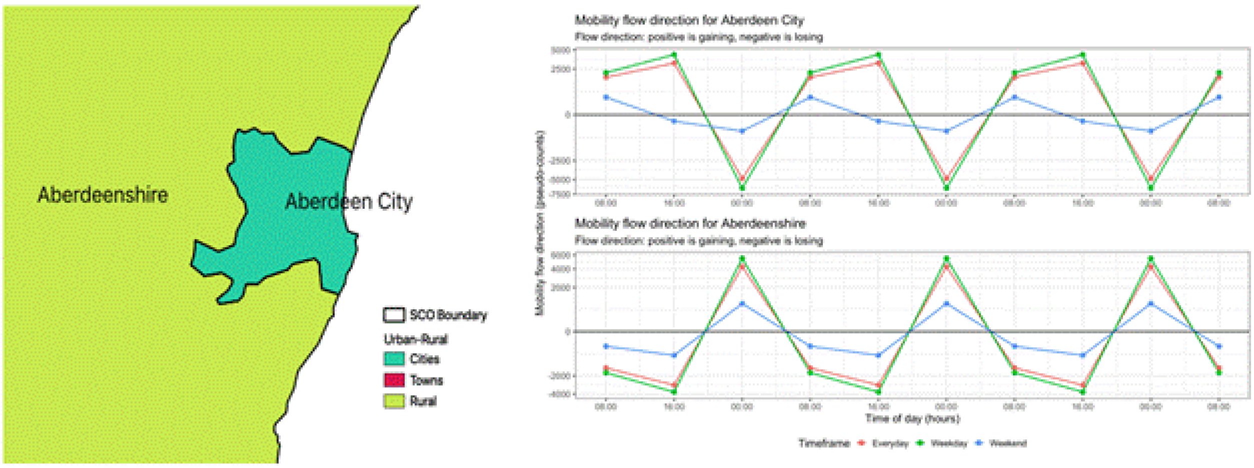



*Figure 19. Modelled stationary distribution of mobility between Aberdeen and Aberdeenshire. These are first-difference plots showing increased population mobility above the zero line, which occurs in Aberdeen City in the morning and midday, and a decreasing mobility during the evening when worker leave the city to go back home to Aberdeenshire. The reverse (or reflection) shown for Aberdeenshire naturally occurs when a single urban area is surrounded by a singular rural area.*

### 3. Implementation and validation at local level (LLs) / Adaptation/upscaling to RLs

| Attribute | Description |
|---|---|
| Living Lab | **French LL** |
| Context | The Pays Pyrénées Méditerranée area is a tourist destination, particularly along the coast during the summer months. Tourism is one of the main economic sectors providing employment. Local stakeholders wish to develop a new, more sustainable model: for instance, several tourist offices are working towards 'green destination' certification. Travel is the main source of energy consumption and CO2 emissions in the living lab area, particularly during the summer months: indeed, the population (locals, tourists and second-home owners) is concentrated on the coast over a short period of time. One-off events such as traditional celebrations or festivals are also case studies. The seasonality of travel is a known phenomenon but not monitored at a local scale. Gaining a better understanding of the volume of people present throughout the year, clarifying their travel patterns, and assessing the temporal variation of mobility and connectivity related to the living lab would help to improve mobility services (particularly public transport) during the summer months, and to anticipate the public services required for the smooth running of the area throughout the year. The research question is therefore: **what is the demand for mobility throughout the year, particularly how does it vary along the coast?** |
| Implementation | Analysis of aggregated mobility datasets (Meta/Facebook Crisis and Activity Space data) combined with TripAdvisor mobility signals to identify flows and seasonal mobility patterns. The analysis focuses on mobility before, during, and after key holiday periods/events and tourism seasons. |

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).*

| | |
|---|---|
| Validation | By the end of the project, the results will be validated using one or more of the following existing complementary datasets in collaboration with French living lab:<br>• Eric Maurense, a freelance consultant, analysed geolocation data from Flux Vision mobile phones on behalf of several coastal municipalities. The results of the analysis show the number of people present (regularly or occasionally) during the day in the municipality throughout the year, and the origin of visitors. These results can be used for research objectives.<br>• INSEE indicates the tourist accommodation capacity (hotels, campsites, tourist residences) for each municipality, without including private rentals such as AirBnB: https://www.insee.fr/fr/statistiques/zones/1405599<br>• The Regional Tourism Observatory publishes annual visitor data, also based on mobile phone geolocation data, available for areas comprising several dozen municipalities: https://pro-tourismeadt66.com/au-niveau-des-6-infra-zones-departementales<br>• As part of GRANULAR, the French Living Lab had conducted a survey in summer 2025 in Argelès-sur-Mer, specifically targeting tourists, to find out about their mobility during their stay, providing information on the places they visited and the modes of transport they used to and from Argelès-sur-Mer. The results of the survey were also discussed with members of the Living Lab's core group, and with several tourist offices and coastal local authorities.<br>• As part of the GRANULAR project, a hackathon is planned for September 2026 to transform the data produced and collected on tourists' mobility patterns during their stay into mobility services and solutions tailored to the specific characteristics of our region. |
| Outcome | So far, the results of the mobility analysis has shown that:<br>• Population variations all along the year across the municipalities within the living lab's area are particularly pronounced in coastal tourist villages (the number of incoming visitors is three times higher in July and August compared to the annual average), in contrast to residential inland villages with fewer tourist accommodation facilities.<br>• School holidays and long weekends (particularly in May) attract tourists throughout the year, especially residents of the Toulouse and Montpellier regions. In July and August, tourists from all over France arrive, particularly from the North and the Paris region.<br>• Increased travel during the summer period, compared to the annual average, across all municipalities in the area. We can hypothesise that this increase is due to the rise in the resident population, but also to a change in the travel habits of the residents themselves during the summer holidays.<br>• Integrated mobility datasets for the French living lab will be provided and a manuscript describing the methodology and findings will be prepared. |
| Living Lab | **Scottish LL** |
| Context | The Scottish living lab group seeks to better understand human mobility between rural and urban areas (including differences in |

| | |
|---|---|
| | terms of mobility characteristics between these areas). There is some concern, however, that the DEGURBA classification does not reflect realities on the ground as this urban/rural classification is defined at the LAU level. As such, the Scottish LL proposes using an urban/rural mapping scheme derived by the Scottish government (SG) as an alternative to the DEGURBA classification. |
| Implementation | Use the CR data for a comparative analysis between DEGURBA and SG rural-urban classifications in terms of how the mobility changes between weekdays and weekends but at the daily and sub-diurnal level. |
| Validation | By the end of the project, comparison of urban/rural classifications will be assessed in collaboration with Scottish living lab using government statistics on travel, tourism, and traffic between urban and rural areas. |
| Outcome | So far, an improved understanding of mobility patterns to and from areas across the rural–urban spectrum, differentiated between weekdays and weekends, in order to support urban and rural transport planning, local tourism management, and seasonal traffic assessment. Integrated mobility datasets for the Scotland living lab will be provided and a manuscript describing the methodology and findings will be prepared. |
| Living Lab | **Finnish RL** |
| Context | The RL South Savo aims to understand mobility patterns between the region and major urban centres in Southern Finland, particularly the Helsinki metropolitan area, as well as Tampere and Turku. Mobility dynamics are strongly influenced by seasonal tourism, holidays, and major cultural events such as the Savonlinna Opera Festival. At the same time, the region is characterised by strong seasonality, a high share of second-home ownership and multi-local living, negative net commuting patterns, and challenges related to accessibility and public services, making it important to better understand how these factors shape rural–urban connectivity and service pressures. |
| Implementation | Analysis of aggregated mobility datasets (Meta/Facebook Crisis and Activity Space data) combined with TripAdvisor mobility signals to identify directional flows and seasonal mobility patterns. The analysis focuses on mobility before, during, and after key holiday periods and tourism seasons. |
| Validation | Triangulation of mobility results with findings from a regional tourism survey conducted in South Savo in summer 2025 and stakeholder knowledge from the RL network (municipalities, tourism actors, and regional development organisations). The results of the survey confirm strong domestic tourism flows from the Helsinki region and highlights heavy reliance on private car mobility and second-home visits. |
| Outcome | Improved understanding of seasonal and event-driven mobility between urban regions and South Savo. Results will be shared and discussed with regional stakeholders during a final on-site workshop to support collective sense-making and identify implications for tourism planning, mobility services, and regional development strategies. Integrated mobility datasets for teh Finnish living lab will be provided and a manuscript describing the methodology and findings will be prepared. |
| Living Lab | **Italian LL** |
| Context | Understanding of tourists' mobility patterns between the municipalities (administrative units) of the Val di Cecina Rural District, specifically between coastal and inland municipalities. In fact, tourism is strongly seasonal and concentrated both in time and |

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).*

| | |
|---|---|
| | in space, with some negative consequences on mobility, infrastructure, natural resources, etc. Although national and regional datasets (e.g., ISTAT, SMART) provide a large amount of data on tourism trends, they do not capture connectivity patterns between municipalities. A more dynamic and integrated perspective is needed to better interpret tourism trends at the district level and support informed local strategies. |
| Implementation | Analysis of mobility datasets (Meta/Facebook Crisis and Activity Space data). The analysis focuses on mobility in the periods 2020-2022 and 2024-2025. |
| Validation | Comparison with regional and national datasets already available, along with interpretation and validation from local administrations and Ambiti Turistici (i.e., territorial tourism organizations that group multiple municipalities and coordinate the planning, management, and promotion of tourism at a higher administrative unit level). |
| Outcome | A more dynamic and comprehensive view of tourism in the Val di Cecina Rural District. The results will be shared with all the administration of the rural district and with the Ambiti Turistici during an in-person technical roundtable, to inform the planning of tourism offerings, mobility services, and local development strategies, in order to contribute to a more balanced distribution of tourism flows both in time and in space. Integrated mobility datasets for the Italian living lab will be provided and a manuscript describing the methodology and findings will be prepared. |

General processing outputs for derived mobility metrics, plots and maps were derived for all living labs (LL) and replicate labs (RL) for both Facebook Crisis (CR) and Activity Space (AS) datasets.

## 4. Limitations

The Facebook Mobility During Crisis (CR) and Activity Space Maps (AS) mobility datasets by Meta were created by utilizing the Facebook mobile app to collect location data of Facebook user's phones for those users who opt-in to allowing their phones to be tracked by Meta. As such, these mobility data sources might not be a representative sample of human mobility from the general population since Facebook users are a subset of the population, and Facebook users who allow their mobile phones to be track are a further subset. Travel surveys with random sampling designs may be more representative of human mobility, but those methods are costly and can be limiting due to smaller sample sizes. Given these limitations, observed mobility through samplings via phone apps are useful in obtaining a general behaviour of human flow direction and timing of movements. Moreover, the comparison and integration of other data sources of mobility (e.g., surveys, Google Location History, Cell Phone Call Detail Records and eXtended Detail Records ) can contribute to improve the understanding of the representations and gaps of human mobility datasets.

CR data have the benefit of high temporal resolution at sub-diurnal time steps; however, this benefit also has a downside as the shorter sampling window can limit the number of individuals included in the aggregate sample to the point of exclusion from a dataset due to not meeting anonymization thresholds. AS data have a much longer sampling window (3 weeks) sampled at weekly intervals, which has the benefit of large sample sizes in the aggregation and provides a more robust dataset; however, the much longer temporal resolution limits the analysis to studies of much longer timeframes (e.g., cannot study workday commune times). Finally, the data sharing agreement that the University of Southampton has with Meta restricts the publishing of raw Facebook datasets to the public. Hence, we are restricted to publishing derived numerical formats such as metrics, aggregate statistics, and model results or visual formats such as maps, plots, and connectivity network maps.

## 5. Contribution to policy/call objectives:

These derived Facebook metrics, together with the associated maps and visualizations, contribute to an improved understanding of the spatio-temporal dynamics of mobility and connectivity in rural areas. In particular, CR data are well suited to capturing diurnal and sub-diurnal movement patterns (e.g., morning and evening commuting flows) between rural areas and other areas along the rural-urban continuum. CR data also enable the assessment of differences in mobility patterns between weekdays, weekends, and holiday periods, providing insights into changes in travel behaviour. Moreover, the higher temporal resolution of CR supports the analysis of short-term events and holidays (e.g., concerts and festivals), including mobility patterns immediately before and after such events.

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).*

The longer sampling period of AS data provides a more robust, contemporary dataset for use with long-term periods such as seasonal or longer school holidays. The larger sample size of AS also provides insights to further understand more spatially detailed movement over weekly periods and to compare differences potentially linked to incoming visitors and outgoing residents.

Overall, these analyses facilitate the quantification of mobility demand and its variability across multiple temporal scales, generating actionable insights to support evidence-based planning, targeted interventions, and investment decisions.

**6. How to use this dataset?**

The output of AS maps and plots are intended to work together, and they are better suited to longer temporal range of mobility study at seasonal length or longer. The CR flow accumulation metrics can be used at the daily and /or sub-diurnal timeframe making them ideal to study commuting flows between rural and urban areas during workdays and flows related to weekend and holiday travels over shorter periods. Flow accumulation metric CSV tables for CR contain categorical attributes (e.g. diurnal periods, weekend/weekday, and DEGURBA at multiple levels) which facilitates analysis of differences in mobilities across any combination of categories. Datetime information included in the flow accumulations metrics data tables for CR allows for temporal analysis to study timings of mobility over a wide range of timeframes such as seasonal, weekly, holiday and weekends, as well as event analysis as described earlier in this report. Documented code scripts that accompany the CR flow accumulation metrics can be used to produce visual and numerical outputs.

Although AS and CR are not temporally coincident, if seasonal or annual mobilities are presumed to be steady state over those periods, then AS maps and plots may be used in conjunction with CR flow accumulation metrics. However, one caveat to consider where AS and CR should not be used together is during the COVID-19 pandemic. Indeed, lockdown periods over early 2020 to mid-2021 occurred over differing timeframes depending on country. While the CR datasets span March 2020 to May 2022 and would be useful in observing lockdown effects on mobility. CR data should not be used in conjunction with AS over lockdown periods and local lockdown policies should be reviewed prior to any analysis involving their simultaneous use. However, the latest, post-pandemic CR data from Meta might be requested for a shorter period of crisis such as floods, storms, heat waves, etc.

## 4.3. Rural economies (Nordregio)

### 4.3.1. Collecting Data to Monitor Housing Markets In Rural Areas: A Web Scraping Approach Tested in the Övre Norrland Region

**1. Objectives**

The work describes a process to collect and analyse real-time data on property transactions in Sweden. The work demonstrates the use of non-conventional data sources for the calculation of indicators usable in a policy context. More specifically, we aimed at illustrating the use of web-scraping tools to derive nearly-real time information on housing markets in areas with limited availability of official statistics. Subsidiarily, we aimed at demonstrating the use of geospatial methods, including distance-based interpolation and kriging techniques, as well as a range of web mapping tools for data-storytelling. The primary goal of this work was hence to document a *non-conventional data collection, processing and use workflow* as a proof-of-concept, rather than producing datasets directly usable for policy analysis.

**2. General approach/ analytical framework**

To the best of our knowledge, in Sweden there is no official database of real estate prices and registrations of title including geocoded information. This prevents a detailed analysis of housing markets in areas outside the big urban centres. In response to this scarcity of information, we scraped data from the Hemnet portal, the largest digital marketplace for housing in Sweden. Scraping was done in June-July 2023, in accordance with the guidance and rules in the "robot.txt" file of the web portal. In addition to the general guidance and rules included defined by the Hemnet in the robot.txt file, the scraping instruction was performed in such away that server overloading was avoided at all costs. Importantly, data scrapping focused on publicly available information rendered by the Hemnet portal. No hidden or back-end data were collected whatsoever. Moreover, the data points collected were further aggregated and anonymized via spatial interpolation.

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).*

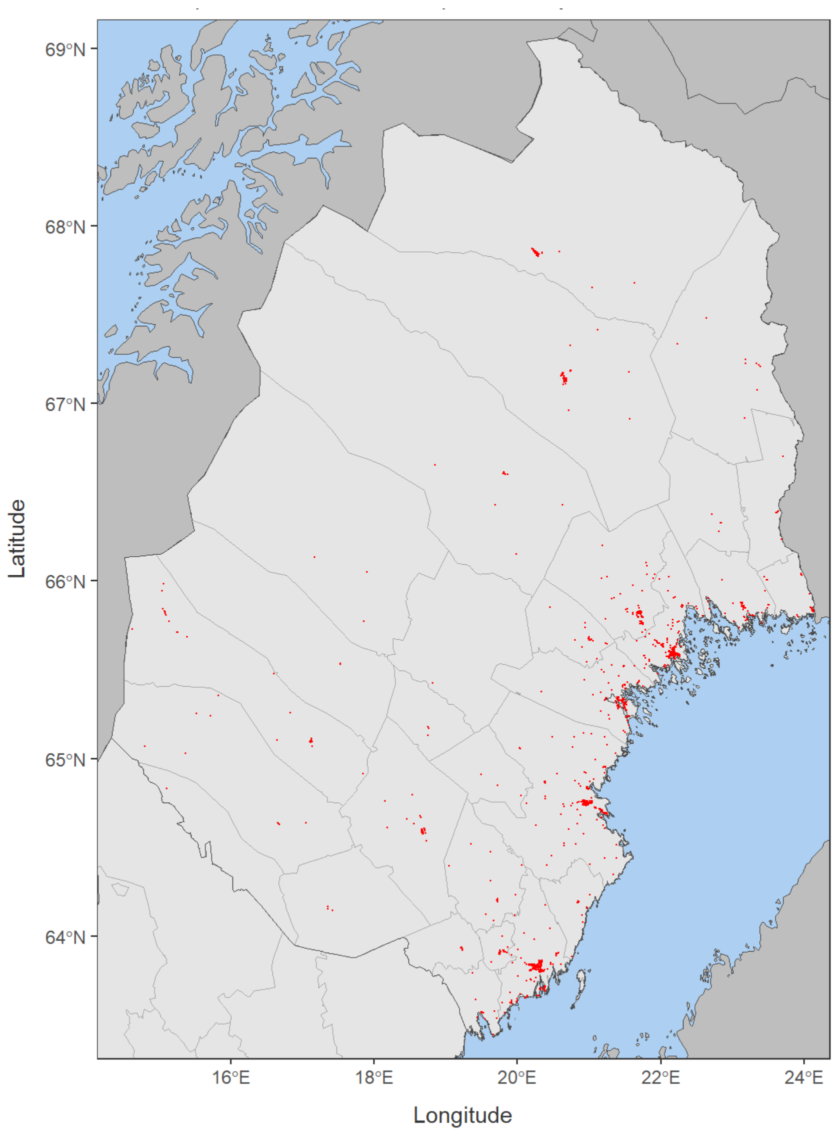



*Figure 20. Map showing the property transactions scraped from the hemnet portal in July 2023. Each point represents one transaction of a housing unit for which sufficient details were available. The data covers the Jan 2023 – Jun 2023 period*

The data were collected for 2022 and the first half of 2023. The data were structured, cleaned, pre-processed and then used to write a data-story documenting the link between (green) reindustrialization processes and housing price spikes in the municipality of Skellefteå.

A static version of the demo, produced with R and R Studio, is included in the supplementary report to this deliverable, available from the Zenodo repository, with DOI: https://zenodo.org/records/21706065 - section titled: "Lab 3: Collecting Data to Monitor Housing Markets In Rural Areas: A Web Scraping Approach Tested in the Övre Norrland Region (Sweden)". A dynamic html version of the demo was generated with Rmarkdown. This format allows to embed text and code in a self-contained html document, ensuring transparency and replicability of the work. This dynamic version can be accessed from the following site: https://nordregio.github.io/granular/housing_demo.html.

**3. Implementation and validation at local level (LLs) / Adaptation/upscaling to RLs**

The demo was developed for the Swedish Living Lab, including the Swedish northernmost counties of Norrbotten and Västerbotten. These two remote and mostly rural regions in Sweden are expected to attract population as a consequence of ongoing processes of green re-industrialization in the area. Housing markets in both regions are hence likely to suffer important changes in the years to come.

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).*

| Attribute | Description |
|---|---|
| Living Lab | North of Sweden |
| Context | In this experimental work, we present a methodology for collecting, cleaning, geocoding, interpolating, and analyzing real-time data on property transactions in rural areas. Our focus lies on Sweden's northernmost counties of Norrbotten and Västerbotten, which have experienced substantial population growth due to the burgeoning process of green re-industrialization. |
| Implementation | The methodology was applied to the Hemnet database covering Norrbotten and Västerbotten counties, producing a dataset of 520 geolocated transactions for year 2022.<br><br>Interpolated maps showed high property prices around urban centers and significant price declines in rural Western parts of Skelleftiå municipality relative to the previous year. |
| Validation | Validation was not possible due to lack of access to microdata about housing market transactions with geocoded information. |
| Outcome | The approach generated an interactive map illustrating the evolution of housing prices during 2023, revealing substantial drops in property values for some rural areas and price increases in urban cores. This data indicates a relationship between population growth from re-industrialization activities and the observed trends in housing market dynamics within Skelleftiå municipality. |

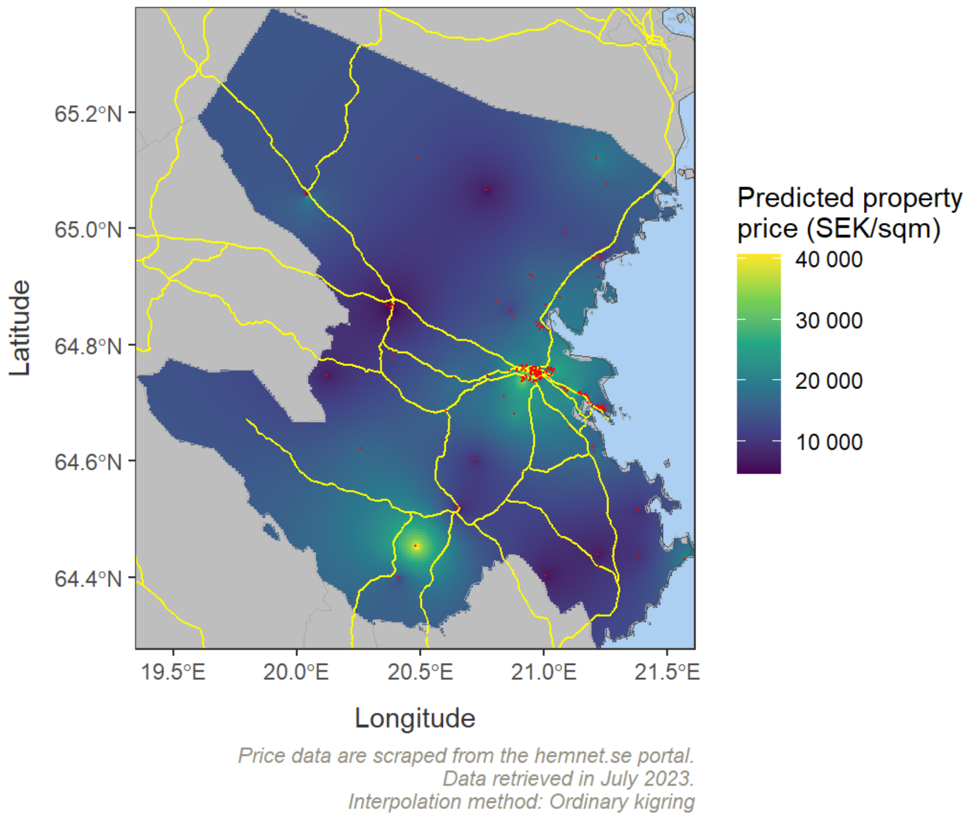



*Figure 21. Average housing prices in Skellefteå municipality, Sweden. Housing prices represent average values in SEK per $m^2$ based on transaction data for the January 2022 - June 2023 period, scraped from the hemnet portal in July 2023. The surface was generated via ordinary kriging on geocoded transaction records. Values have not been validated and are provided for illustrative purposes only*

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).*

**4. Limitations**

Web scraping has well-known limitations related to methodological and ethical constraints that limit its direct applicability in formal policy settings  A primary concern is the inherent lack of representativeness, as scraped datasets often suffer from selection bias driven by platform algorithms, user demographics, and the specific architecture of target websites, thereby failing to capture the full spectrum of the population under study (in this case the entire housing market) and potentially skewing the sample outcomes toward the market targets defined by the housing portal. Furthermore, the collection and utilization of such data raise ethical considerations regarding user privacy, informed consent, and the terms of service agreements of source platforms, which can create legal ambiguities and undermine public trust in evidence-based governance.

**5. Contribution to policy/call objectives:**

The methodology demonstrates potential for delivering high-frequency, fine-grained information on housing prices that could help policymakers monitor rural population dynamics and understand the impact of redevelopment projects on local markets.

**6. How to use this dataset?**

This comprehensive approach aims to provide insights into the performance of these rural housing markets by employing web scraping techniques from the largest digital housing marketplace in Sweden - Hemnet. The importance of this study lies in understanding how housing market dynamics will evolve in these remote regions as they attract population from other areas.

Nonetheless, the work remains highly experimental. In addition to the general limitations and ethical considerations mentioned above, it is important to keep in mind that the price data collected in this demo were not not validated against official sources or alternative private providers. This limits its direct applicability in formal research and policy settings.

Additional details, as well as code used for scraping, data cleaning, interpolation and visualization can be found in the Supplementary report to this deliverable, available from the Zenodo repository, with DOI: https://zenodo.org/records/21706065 - section titled: “Lab 3: Collecting Data to Monitor Housing Markets In Rural Areas: A Web Scraping Approach Tested in the Övre Norrland Region (Sweden)”.

### 4.3.2. High Resolution Analysis of Transport Poverty Risk in Europe's Rural Areas

1. Objectives

This study addresses the need for spatially granular, multi-dimensional measurement of transport poverty risk across European regions. The primary objective is to develop and operationalize a Transport Poverty Risk Index (TPRI) capable of identifying where transport poverty risk concentrates, which populations are most exposed to, and which structural pressures drive risk at sub-regional level. The work responds directly to the growing recognition in EU policy that transport poverty is an unevenly distributed, territorially specific phenomenon that aggregate national statistics are ill-equipped to capture. A secondary objective is to produce a replicable, indicator-based framework that can inform both cross-national comparison and regional responses, with particular attention to rural, remote, and peripheral areas where mobility constraints and socioeconomic vulnerability tend to coincide.

2. General approach/ analytical framework

The analysis adopts an adapted version of the IPCC AR5 Working Group II risk framework, structuring transport poverty risk across three dimensions (IPCC, 2014): Hazard, referring to external cost pressures such as fuel and energy prices; Exposure, capturing the degree to which populations depend on private motorized transport and have limited access to alternatives; and Vulnerability, encompassing the socioeconomic and demographic characteristics indicating the households' capacity to absorb or adapt to mobility-related burdens. This decomposition allows the index to distinguish between regions that face similar transport conditions but fundamentally different risk profiles depending on the characteristics of their resident populations. Principal Component Analysis and Factor Analysis were applied to the indicator set to reduce redundancy, assign data-driven weights, and ensure that each indicator contributes statistically independent information to the composite score. The resulting TPRI is computed at grid-cell level, also enabling sub-regional resolution analysis.

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).*

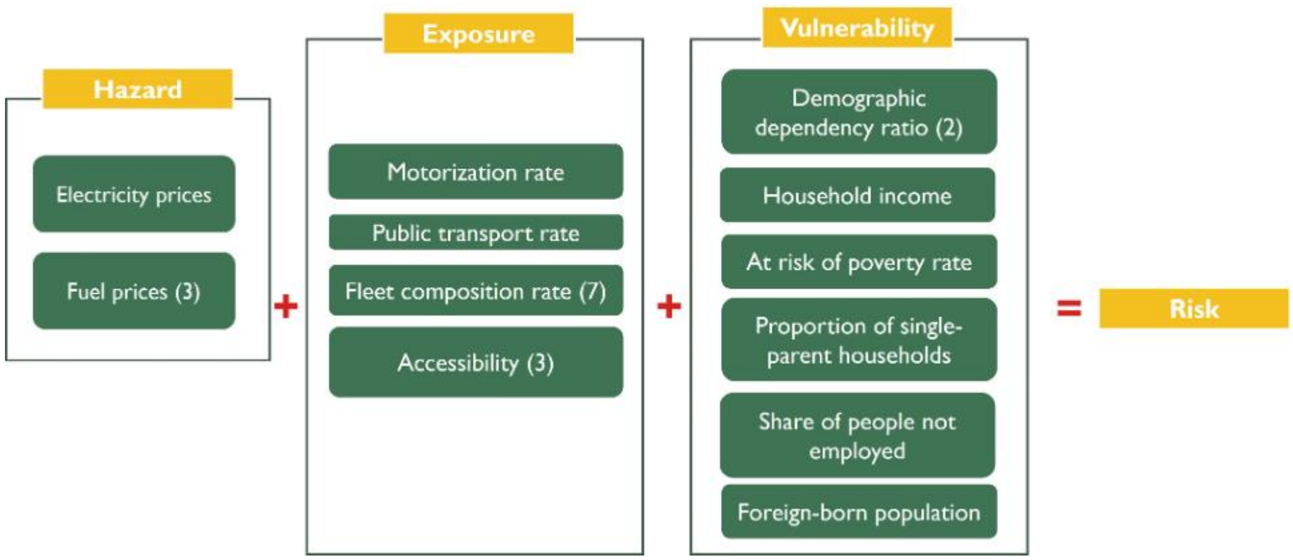



Figure 22. Data model with the dimensions of the risk framework considered and the respective indicators. Fuel price indicators are subdivided into: i) petrol prices, ii) diesel prices and iii) LPG prices. The fuel fleet composition rate is subdivided into: i) diesel vehicle rate, ii) LPG vehicle rate, iii) petrol vehicle rate, iv) hybrid diesel plug-in vehicle rate, v) hybrid petrol vehicle rate, vi) electric vehicle rate and vii) fuel-fleet weighted indicator. The demographic dependency ratio is subdivided into i) young dependency ratio and ii) old dependency ratio. The accessibility indicators are subdivided into i) accessibility to small cities, ii) accessibility to medium cities, and iii) accessibility to large cities.

3. Implementation and validation at local level (LLs) / Adaptation/upscaling to RLs

The index was implemented at European scale, covering EU regions across the full temporal window from 2022 to 2026, and further analyzed through a set of eight Living Lab regions representing diverse territorial contexts — including Nordic remote areas, Mediterranean rural peripheries, and Central and Eastern European rural zones. The Living Labs served both as a finer scale analysis, allowing the spatial and temporal behavior of the index to be compared against locally known conditions, and as demonstration cases for the interpretive value of sub-regional diagnostics. Validation in regions such as Pays Pyrénées confirmed that the index captures expected spatial gradients in accessibility and motorization dependence, while also revealing the distinct temporal dynamics driven by the fuel-fleet weighted indicator. The EU-wide visualization was used to observe the full scale of the indicator consistency and methodological transparency, ensuring that the index remains interpretable and comparable across the full range of European territorial contexts.

| Attribute | Description |
|---|---|
| Living Lab | Six regional Living Labs across Europe: Ourense - Spain, Pays Pyrénées - France, Val di Cecin - Italy, Netherlands, West Pomeranian - Poland, Upper Norrland - Sweden, representing diverse territorial contexts in Europe scale. |
| Context | Each Living Lab constitutes a sub-national rural or peripheral territory characterized by varying degrees of transport accessibility constraints, car dependence, socioeconomic vulnerability, and demographic pressure. Together, they span the main territorial typologies relevant to transport poverty risk in Europe, allowing the index to be tested against locally known conditions across contrasting geographical and socioeconomic backgrounds. |
| Implementation | The TPRI was computed at grid-cell level for each Living Lab region using the same indicator set, dimensional structure, and weighting approach applied at EU scale, ensuring methodological consistency across scales. Spatial outputs were produced for each year in the 2022–2026 period, enabling both cross-sectional and temporal analysis within and across Living Lab territories. |
| Validation | The spatial and temporal behaviour of the index was examined in each Living Lab to assess whether the composite score and its dimensional components reflected expected patterns of transport poverty risk. Validation drew on the |

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).

| | |
|---|---|
| | spatial correspondence between the TPRI and key indicators — particularly accessibility to large cities, motorization rate, and the fuel-fleet weighted composite — as well as on the interpretability of temporal dynamics against known regional conditions. In addition a sensitivity analysis of the index results was perfomed. |
| Outcome | The Living Lab analysis confirmed that the TPRI captures territorially meaningful risk gradients at sub-regional scale, with spatial patterns broadly consistent with known accessibility deficits and demographic vulnerabilities in each area. The analysis also demonstrated that the index's temporal dynamics are primarily governed by the fuel-fleet weighted indicator, while slower-moving structural indicators, such as motorization rate and old-age dependency, shape the baseline risk level rather than month-to-month variation |

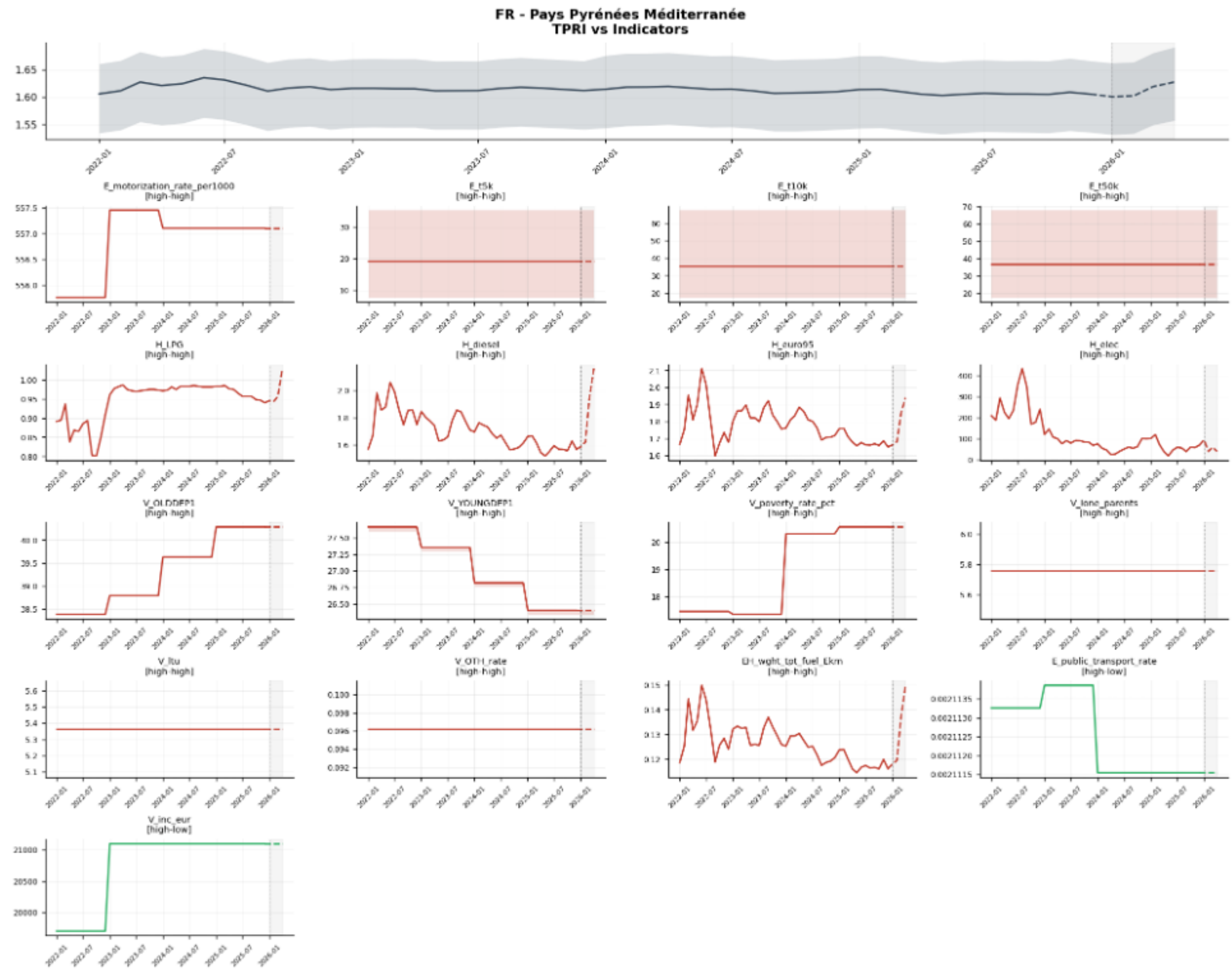



*Figure 23. Multi-panel time-series chart for the Pays Pyrénées region, covering the years 2022 to 2026. The top graph illustrates the TPRI over time, featuring a median line, a band that represents the 10th to 90th percentiles, and a dashed* line indicating data for the year 2026. The panel below displays a separate graph for each of the 17 indicators.

4. Limitations

Several limitations should be acknowledged. First, some input indicators — most notably fuel prices — are available only at national or energy bidding-zone level, introducing spatial smoothing that may understate cost heterogeneity within regions and countries. Second, the index measures risk conditions rather than observed outcomes: high-scoring areas represent elevated structural vulnerability, but household-level transport poverty may be mediated by local adaptive capacity, informal support networks, or targeted policy interventions not captured in the data. Third, the statistical weights derived from PCA and Factor Analysis reflect the informational structure of the available indicator set, which means that dimensions with richer data coverage — such as accessibility — may carry higher

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).*

model weight not only due to their real-world importance but due to their greater representation in the input data. Finally, the index does not directly capture transport behaviour or actual mobility deprivation, and should be interpreted as a proxy for risk exposure rather than a direct measure of experienced poverty.

5. Contribution to policy/call objectives:

The TPRI contributes to European policy objectives at multiple levels. It operationalizes the just transition commitment expressed in recent European Commission and Committee of the Regions instruments on transport poverty, providing a spatially explicit tool for identifying territories where targeted intervention is most needed. By decomposing risk into Hazard, Exposure, and Vulnerability dimensions, the index offers policymakers diagnostic clarity on whether observed risk is primarily driven by cost pressures, infrastructure deficits, or socioeconomic vulnerability — distinctions that are essential for designing effective and proportionate responses. The focus on rural and peripheral areas directly addresses the territorial dimension of the European Green Deal and cohesion policy, where ensuring equitable access to mobility is framed as a condition for sustainable and inclusive regional development.

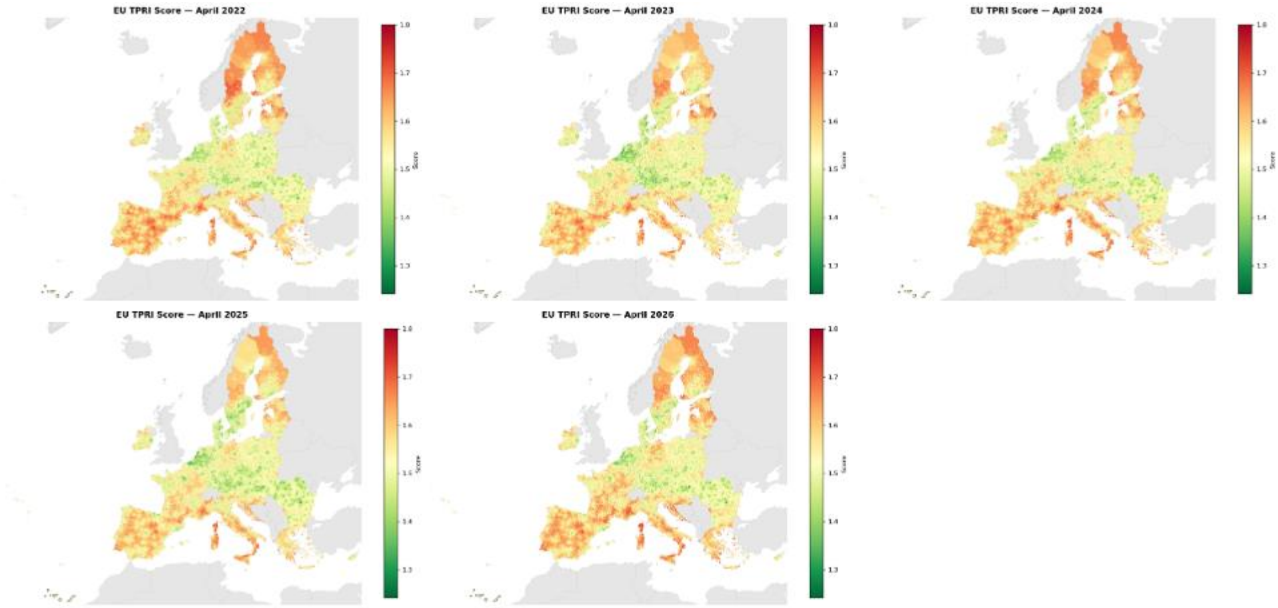



*Figure 24. Transport poverty risk score evolution over time (April 2022–April 2026) in Europe. The shares of lone-parent households, foreign-born populations, and accessibility indicators remained constant over the time series.*

6. How to use this dataset?

The TPRI dataset is designed for use at multiple analytical scales. At European level, it supports cross-national comparison of transport poverty risk profiles and the monitoring of temporal trends across the 2022–2026 period. At regional and sub-regional level, it enables the identification of high-risk grid cells within specific territories, facilitating the targeting of policy interventions and the prioritization of investment. The dimensional structure of the index — with Hazard, Exposure, and Vulnerability scores available alongside the composite TPRI — allows users to diagnose the underlying drivers of risk in any given area rather than relying solely on the aggregate score. The dataset can be integrated with existing regional planning and cohesion policy frameworks, and the methodology allows adaptation to an updated indicator data or extension to additional territorial contexts as new data becomes available.

The TPRI data and software code are openly available through the GRANULAR Zenodo repository, doi: 10.5281/zenodo.21352799. Additional details of the approach and results for the Living labs can be found in the Supplementary report to this deliverable, available from the Zenodo repository, with DOI: https://zenodo.org/records/21706065 - section titled: “Lab 1: High Resolution Analysis of Transport Poverty Risk in Europe's Rural Areas”.

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).*

### 4.3.3. Using OSM Data to Estimate Economic Diversity in Functional Rural Areas (Nordregio)

**1. Objectives**

The primary goal of this research was to harness the potential of OpenStreetMap (OSM) data and Large Language Models (LLM) for estimating business diversity in Functional Rural Areas (FRAs). This study aimed to address an overlooked limitation in conventional economic analysis by transcending administrative boundaries and delving into the functional relationships between economic actors within rural contexts. By doing so, we sought to offer a more comprehensive understanding of economic dynamics in these regions, paving the way for evidence-based policymaking focused on enhancing long-term regional resilience.

**2. General approach/ analytical framework**

In this research, we devised a sophisticated multi-step methodology to estimate business diversity in Functional Rural Areas (FRAs) using OpenStreetMap (OSM) data and Large Language Models (LLMs). Specifically, we made use of Encoder-based sentence embedding models (ESEM), including Bertopic and Cosine similarity, and Generative Large Language Models (GLLMs), including Gemma, and Llama, to estimate business diversity in Functional Rural Areas (FRAs). Our approach was designed to overcome the limitations of traditional economic analysis by harnessing the extensive information embedded within OSM datasets.

The first step involved employing GLLMs for text classification tasks, specifically focusing on translating OSM key-value pairs into standardized NACE (Nomenclature des Actes Economiques) categories. This innovative application of GLLMs capitalized on the richness and detail contained within OSM data, allowing us to capture intricate functional relationships between economic actors in rural contexts that had previously gone unnoticed by conventional methods.

To ensure the reliability and reproducibility of our estimates, we integrated two critical components into our methodology: bootstrap resampling and model-based bias correction. Bootstrapping, a non-parametric statistical technique, allowed us to construct confidence intervals for business diversity metrics based on the data at hand. By employing this approach, we were able to quantify the uncertainty surrounding our estimates, providing valuable context for decision-makers interpreting our findings.

Model-based bias correction was another essential element of our methodology. This technique enabled us to adjust for potential biases inherent in our LLM-based text classification process, thereby enhancing the accuracy and robustness of our estimates. By combining these two components, we created a comprehensive framework that not only captured functional relationships between economic actors but also provided reliable uncertainty measurements.

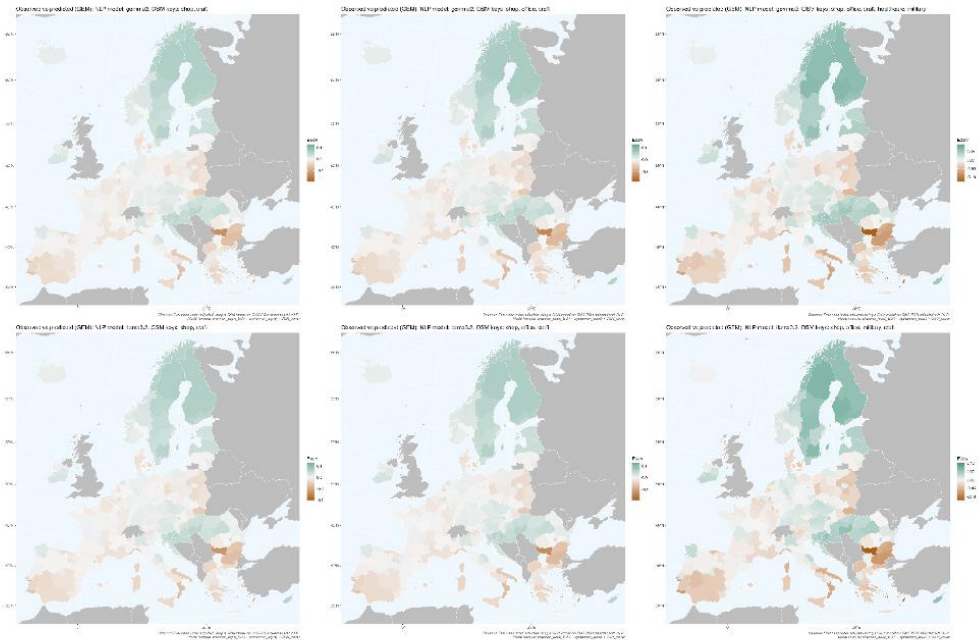


*Figure 25. Comparison of prediction accuracy of different GLLM on specific user-defined POI descriptions in OSM database*

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).*

To further refine our methodological workflow, we engaged in an exhaustive grid search to identify optimal combinations of data filters, classifiers, and regression models. This meticulous process ensured that our approach was both rigorous and adaptable, capable of accommodating diverse datasets and varying research requirements. Through this iterative refinement, we fortified the integrity and versatility of our methodology, making it well-suited for a wide range of applications within rural economic analysis.

By synthesizing these components into a coherent and robust methodological framework, we were able to generate accurate estimates of business diversity in FRAs using GLLMs and OSM data. This research not only expanded the boundaries of traditional economic analysis but also provided valuable insights for decision-makers and policymakers focused on fostering long-term regional resilience through evidence-based interventions.

**3. Implementation and validation at local level (LLs) / Adaptation/upscaling to RLs**

| **Attribute** | **Description** |
|---|---|
| Living Lab | France, Italy, Netherlands, Poland, Scotland, Spain, Sweden Analysis performed at the Functional Rural Area level |
| Context | Existing economic data has a significant gap in the scarcity of comprehensive and standardized metrics for business diversity within functional regions, namely non-administrative geographical units defined by functional criteria, such as commuting zones. While official statistics primarily focus on well-defined administrative boundaries, functional areas, such as FRAs, often encompass a more nuanced understanding of economic landscapes that transcend these confines.<br><br>By leveraging OSM data and LLMs, this study aims to address the dearth of reliable information on business diversity within FRAs. This research is crucial for policymakers, academic institutions, and businesses seeking evidence-based insights into rural economic dynamics. The inability to capture these essential aspects of economic activity through conventional means underscores the need for advanced analytical tools capable of transcending administrative boundaries and uncovering functional relationships between economic entities within FRAs. |
| Implementation | To implement this novel approach at a local level, we began by extracting OSM data relevant to economic actors within FRAs. We then applied the LLMs to classify user-defined OSM descriptors into standardized NACE categories. This process enabled us to capture functional relationships between economic entities in rural settings. |
| Validation | To validate our methodology, we compared our estimates with those generated on official statistics at the administrative level (NUTS2), ensuring that our findings were reproducible across different geographical contexts. The application of bootstrap resampling and model-based bias correction allowed us to estimate business diversity metrics with confidence intervals, providing robust insights into the economic landscapes of FRAs. |

**4. Limitations**

By engaging this methodology, we discovered that OSM data can indeed offer valuable insights into economic diversity in non-standard geographical units. However, uneven coverage remained a persistent challenge. This limitation underscored the need for strategies to address gaps in data availability and improve model performance in under-represented regions.

**5. Contribution to policy/call objectives:**

This research significantly contributes to policy objectives by providing more accurate and comprehensive assessments of business diversity in non-standard geographical units, particularly functional rural areas. This enhanced understanding empowers policymakers to make evidence-based decisions aimed at fostering long-term regional resilience through targeted interventions that consider the unique economic dynamics within FRAs.

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).*

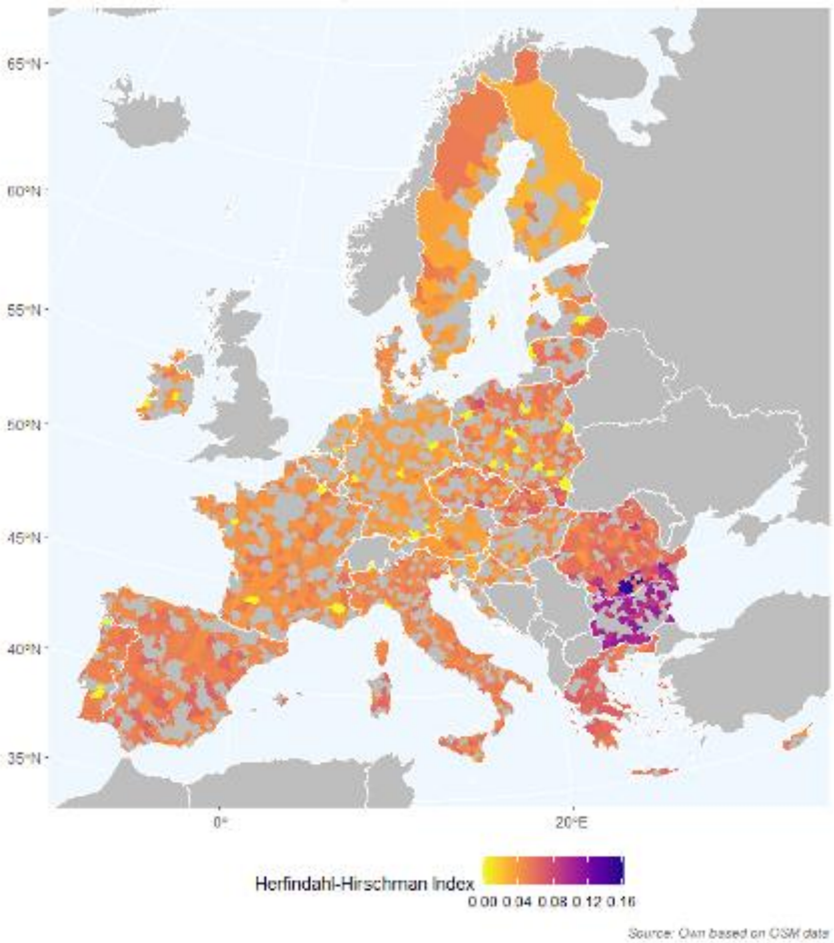



*Figure 26. Predicted Herfindahl-Hirshman Index in Functional Rural Areas (FRAs). Estimates based on modelled OSM data. FRAs defined as catchment areas with 50,000 inhabitants 30m-60m drive. Predictions generated with multi-level GAM model fitted on an classification of economic activities in OSM POIs performed with BERTopic modeling*

### 6. How to use this dataset?

Ultimately, this research fills a critical gap by offering accurate estimates of business diversity in FRAs, empowering stakeholders to make well-informed decisions that promote long-term regional resilience and sustainable growth in rural economies. The insights derived from this study and the analytical workflow can be utilized in various ways for decision-makers and researchers:

- Classified NACE categories obtained from OSM data and LLMs serve as a robust foundation for further analysis of business diversity in rural settings, complementing existing datasets.
- Comparing the results with official statistics or previous research on business diversity in similar contexts allows for validation and cross-verification of findings, strengthening confidence in our estimates.
- The analytical workflow can be adapted for the calculation of economic diversity metrics in other settings.

The data and software scripts underpinning this analysis are openly available through the GRANULAR Zenodo repository, doi: [10.5281/zenodo.21355241](). Additional details of the approach and results for the Living labs can be found in the Supplementary report to this deliverable, available from the Zenodo repository, with DOI: [https://zenodo.org/records/21706065]() - section titled: “Lab 2: Using OSM Data to Estimate Economic Diversity in Functional Rural Areas: A Premier”.

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).*

## 4.4. Land Use (IAMM)

### 1. Objectives

**Objective 1 – Develop a harmonized Sentinel-2 land-use dataset for Europe**

Develop and curate a large-scale dataset linking Sentinel-2 satellite imagery with field observations collected through the LUCAS survey. The dataset provides a standardized reference for land cover and land use monitoring across Europe and supports the development, validation, and benchmarking of AI methods for rural landscape analysis.

**Objective 2 – Develop AI tools for natural language-based land-use monitoring**

This work develops a deep learning framework that enables users to search and analyse satellite imagery using natural language queries for land cover and land use monitoring. The system allows policymakers, urban planners, and environmental managers to find relevant satellite observations by typing simple descriptions (e.g., "agricultural fields with irrigation systems", "forested areas showing seasonal change"), making earth observation data accessible without requiring specialized remote sensing expertise.

### 2. General approach/ analytical framework

The work builds on three complementary developments developed in collaboration between CIHEAM-IAMM, INRIA, INRAE and CIRAD:

**Sen2LUCAS Dataset (Jain et al., 2024):** A large-scale benchmark dataset linking Sentinel-2 satellite imagery with field observations that is ground-level photographs from LUCAS (Land Use and Coverage Area Frame Survey) programme across Europe. The dataset combines multispectral satellite observations with detailed land cover and land use information, providing a harmonized reference for training, validating and benchmarking AI models. By covering diverse European landscapes and agricultural systems, Sen2LUCAS supports the development of transferable methods for land-use monitoring at regional to continental scales.

**SenCLIP Framework (Jain et al., 2025):** SenCLIP is an AI model that bridges natural language and satellite imagery, enabling users to search large archives of Sentinel-2 data using simple text descriptions such as "agricultural fields with centre pivot irrigation", "urban areas with dense vegetation", or "coastal areas affected by erosion".

Rather than relying on manually annotated image-text pairs, the model learns from nearly one million geotagged ground-level photographs collected through the LUCAS survey. Each survey location includes photographs acquired in four directions (north, south, east, and west), providing a detailed ground-level view of the landscape.

The approach aligns Sentinel-2 observations with photographs acquired at the same geographic location, allowing the model to connect how landscapes appear from the ground with how they are observed from space. This enables users to describe landscapes using intuitive, real-world concepts rather than technical remote-sensing terminology.

**TimeSenCLIP Extension (Jain et al., 2026):** Building on SenCLIP, TimeSenCLIP incorporates the temporal dimension by analysing monthly Sentinel-2 observations throughout the year. By capturing seasonal patterns, crop growth cycles, and vegetation dynamics, the framework provides a richer representation of landscapes and improves the discrimination of land uses that may appear similar in single-date imagery.

The model can operate directly on Sentinel-2 time series at the pixel level by leveraging the temporal evolution of multiple spectral bands. This makes the framework computationally efficient while remaining scalable to continental-scale monitoring applications.

**Training approach**: Both SenCLIP and TimeSenCLIP use a self-supervised learning strategy that does not require manual text annotations. The models learn by associating Sentinel-2 observations with ground-level photographs acquired at the same geographic locations.

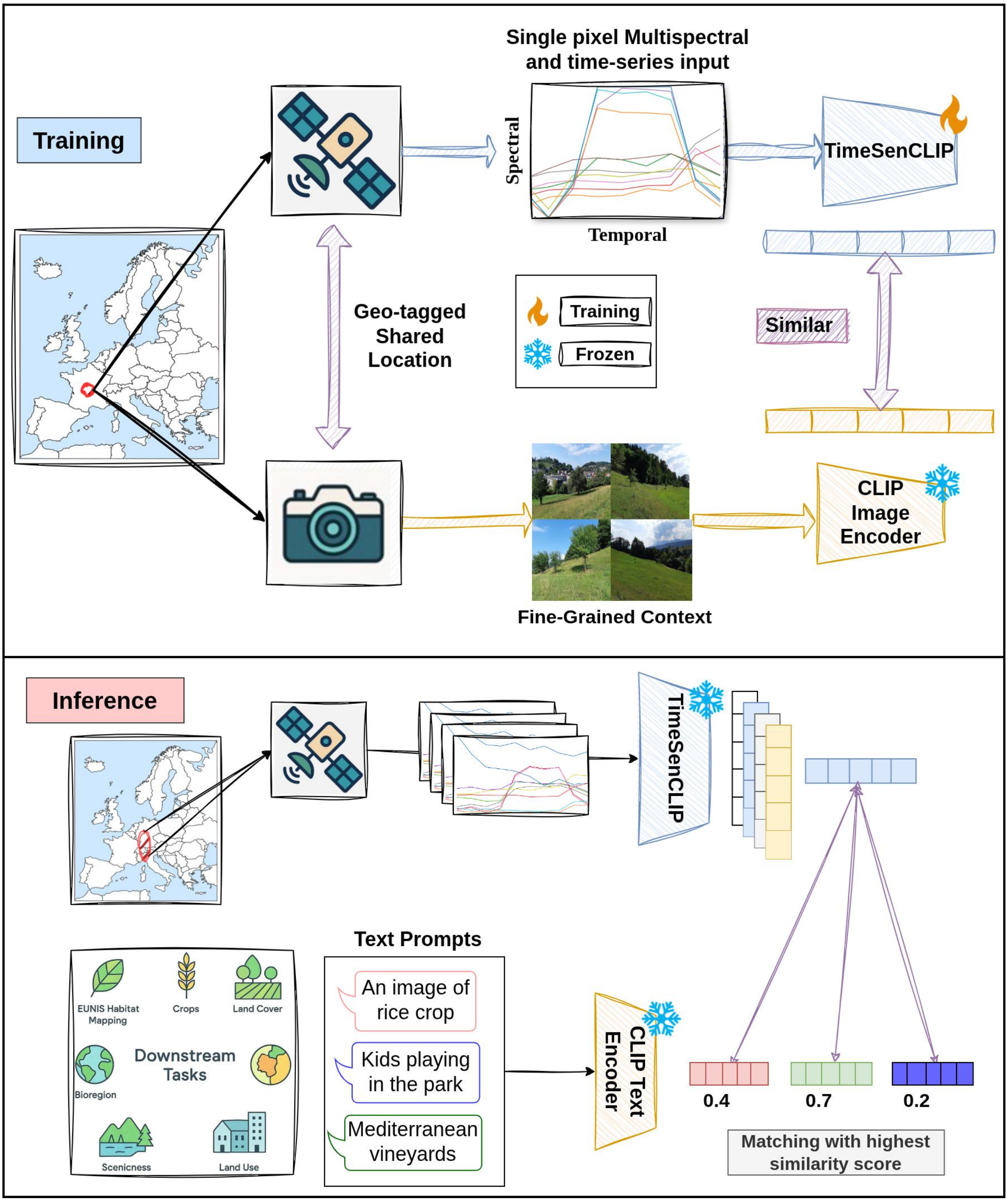
Through contrastive learning and the CLIP vision-language framework (Radford et al., 2021), the models automatically learn relationships between landscape characteristics visible from the ground and their corresponding satellite signatures. Once trained, users can interact with the models through natural-language queries. Satellite images and text descriptions are represented within a shared semantic space, enabling image retrieval, semantic search, and zero-shot classification, including categories that were not explicitly included during training.

#### Evaluation Framework

The developed dataset and AI models were evaluated to assess their robustness, transferability, and suitability for large-scale land-monitoring applications. Experiments were conducted across benchmark datasets representing diverse environmental conditions, agricultural systems, and geographic regions within Europe.

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).*




The Sen2LUCAS dataset served as a reference benchmark for developing and comparing machine-learning methods for land-cover and land-use analysis.

The SenCLIP framework was assessed on semantic retrieval and classification tasks, including:

- Natural-language retrieval of Sentinel-2 imagery;
- Cross-view retrieval between ground-level photographs and satellite observations;
- Land-cover and land-use classification;
- Zero-shot classification of previously unseen categories.

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).*

The TimeSenCLIP framework was evaluated on applications benefiting from seasonal and temporal information, including: Land-cover and land-use classification; Crop type mapping; Habitat mapping; Scenicness prediction; Natural-language retrieval of Sentinel-2 time-series imagery.

Performance was assessed using standard retrieval and classification metrics, complemented by qualitative analyses of retrieval results and spatial prediction maps.

Across tasks, the incorporation of temporal information improved the discrimination of agricultural and natural landscapes exhibiting distinct seasonal dynamics. The results further demonstrated strong transferability across regions, highlighting the potential of the framework for scalable landscape characterization and monitoring.

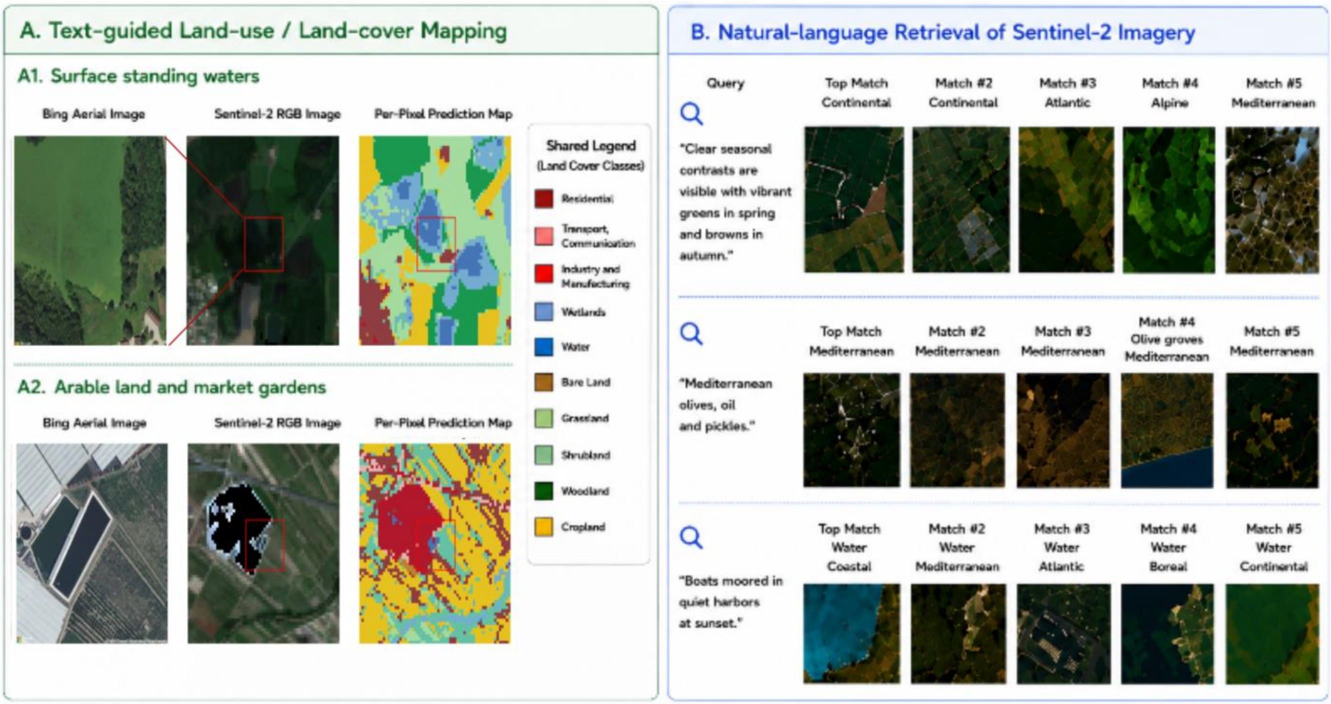



*Figure 1Pixel classification and retrieval of images using text prompts*

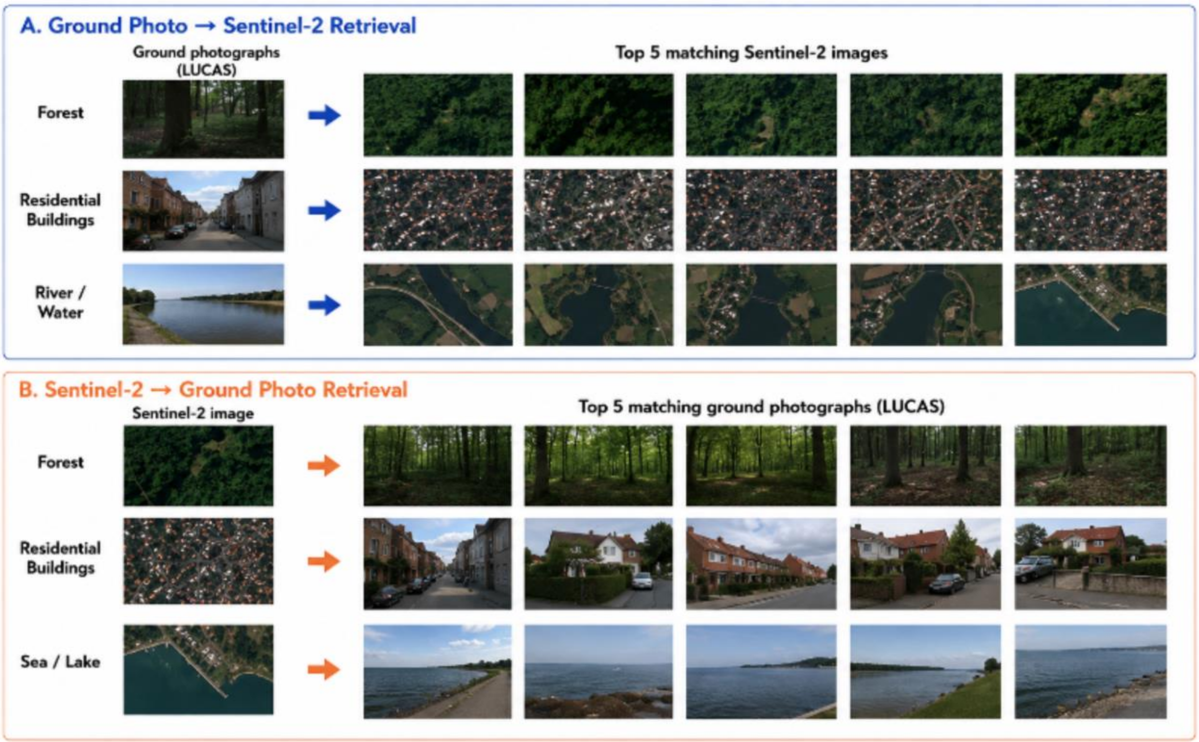



*Figure 2 Image to Image Retrieval*

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).*

**3. Implementation and validation at local level (LLs) / Adaptation/upscaling to RLs**

| Attribute | Description |
|---|---|
| Living Lab | France, Italy, Netherlands, Poland, Scotland, Spain, Sweden |
| Context | Need for scalable and transferable methods for monitoring land use, agricultural systems, and landscape characteristics across diverse rural territories. |
| Implementation | Development of the Sen2LUCAS benchmark dataset and AI models capable of linking Sentinel-2 observations, temporal information, and ground-level photographs within a shared semantic framework. |
| Validation | Evaluation across multiple benchmark datasets and applications, including land-cover classification, land-use classification, crop mapping, habitat mapping, scenic prediction, and semantic retrieval. |
| Outcome | Openly available datasets and AI models supporting landscape characterization, land-use monitoring, image retrieval, and transfer learning across European regions. |

**4. Limitations**

While the proposed framework demonstrated strong performance across a range of land-use and landscape monitoring tasks in Europe, several limitations remain.

First, the models were developed and evaluated using Sentinel-2 satellite imagery and LUCAS field observations collected within the European Union. As a result, their performance in regions with different climatic conditions, vegetation dynamics, agricultural systems, and land management practices remains to be fully assessed. Future work should evaluate the approach in other geographic regions and explore its applicability to global monitoring scenarios.

Second, the semantic understanding learned by SenCLIP and TimeSenCLIP is inherited from the CLIP vision-language model, which was originally trained using photographs and textual descriptions collected from the internet. Consequently, some landscape characteristics that are only observable through temporal evolution, such as crop development stages, phenological cycles, or long-term environmental changes, may not be fully represented in the learned semantic space.

Finally, the availability and geographic coverage of ground-level observations remain an important factor influencing model performance. Expanding the framework through additional sources of georeferenced photographs, citizen science initiatives, and textual geographic descriptions could further improve representation quality, transferability, and global applicability. These limitations provide clear directions for future research, including multi-sensor integration, expansion to new regions, incorporation of additional sources of ground-level information, and improved modelling of temporal landscape dynamics.

**5. Contribution to policy/call objectives:**

The work contributes directly to the objectives of the GRANULAR project by improving the availability, accessibility, and usability of Earth observation data for rural monitoring and land-use assessment.

The Sen2LUCAS dataset provides a harmonized European benchmark linking Sentinel-2 satellite observations with field-based land-use information collected through the LUCAS survey. By establishing a common reference dataset, it supports the development, comparison, and validation of AI methods for rural landscape characterization across different European regions.

The SenCLIP and TimeSenCLIP frameworks contribute to the digital transformation of rural monitoring by enabling users to interact with satellite imagery through natural-language descriptions rather than requiring specialized expertise in remote sensing or machine learning. This lowers technical barriers and facilitates the use of Earth observation data by policymakers, rural observatories, environmental agencies, and land managers.

By combining satellite observations, ground-level information, and advanced AI methods within a common framework, the work contributes to improved landscape understanding at local, regional, and continental scales. The open availability of the dataset, models, and evaluation framework further supports FAIR data principles, reproducible research, and the long-term sustainability of the GRANULAR knowledge infrastructure.

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).*

**6. How to use this dataset?**

The Sen2LUCAS dataset is openly available through the GRANULAR Zenodo repository (https://zenodo.org/records/14259723) and can be used for training, benchmarking and evaluating machine learning models for land-use and land-cover analysis. The associated SenCLIP and TimeSenCLIP models provide reusable tools for text query and transfer-learning applications using Sentinel-2 imagery and can be adapted to a broad range of rural monitoring tasks. Code and model are available at :

SenCLIP

TimeSenCLIP

## 4.5. Citizen Science, Crowdsourcing and Engagement (IIASA)

**1. Objectives**

In close connection with the development of rural GRANULAR indicators (WP4) and various Living and Replication Labs (WP6), selected forms of citizen science, crowdsourcing and engagement methods were applied across the GRANULAR project including a variety of data collection campaigns e.g. analysis by citizens of on-ground imagery, web-scraping of social media data (i.e. TripAdvisor) and analysis of citizen collected photos (i.e. Flickr) applied via various stakeholder groups at the local levels including those in the Labs to address identified data gaps documented in Task 3.1.

**2. General approach/ analytical framework**

**European Landscape Attractiveness**

This raster layer represents a machine learning-based prediction of perceived attractiveness across Europe, extending a German model based on surveys and landscape imagery. Values indicate how people are likely to rate the attractiveness of landscapes based on visible features like land cover, infrastructure, and elevation. Predictions are derived from a model trained on the output of a German reference model (Roth et al., 2021), which was based on a survey of German residents rating landscape images. The extrapolation was done by replicating and retraining a model using pan-European features.

In 2021, Roth et al. published a dataset, assigning a value of landscape attractiveness to each cell of a 1 x 1 kilometer grid in Germany. The basis for this map were user-rated landscape photos of Germany. These photos were taken as part of a photo documentation project in 30 representative sampling areas, each approximately 150 km² in size, across Germany. The sampling areas represent the diversity of German landscapes, from coastal regions to the Alps, and from forested areas to rural regions and urban centers. An expert-selected set of over 800 landscape photos was evaluated through an online survey in collaboration with a citizen science panel of over 3,500 participants.

Only the final output maps are open to the public unfortunately. In order to achieve our goal of extending these models to the rest of Europe, we first need to replicate the model as best as we can, based on the documentation, and then apply it to datasets that are available for the whole of the EU.

These images are analysed using GIS tools to extract the visible landscape features. These features include land-cover, land-use and relief. The features are chosen, such that a connection to a countrywide dataset can be drawn. This feature extraction also notes the distance of said features from the camera. The link between user ratings and landscape features are then modelled using ordinary least square regressions. The final maps are then the prediction of these models using the countrywide available datasets (Figure 29). The data is available here: https://zenodo.org/records/18618619

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).*

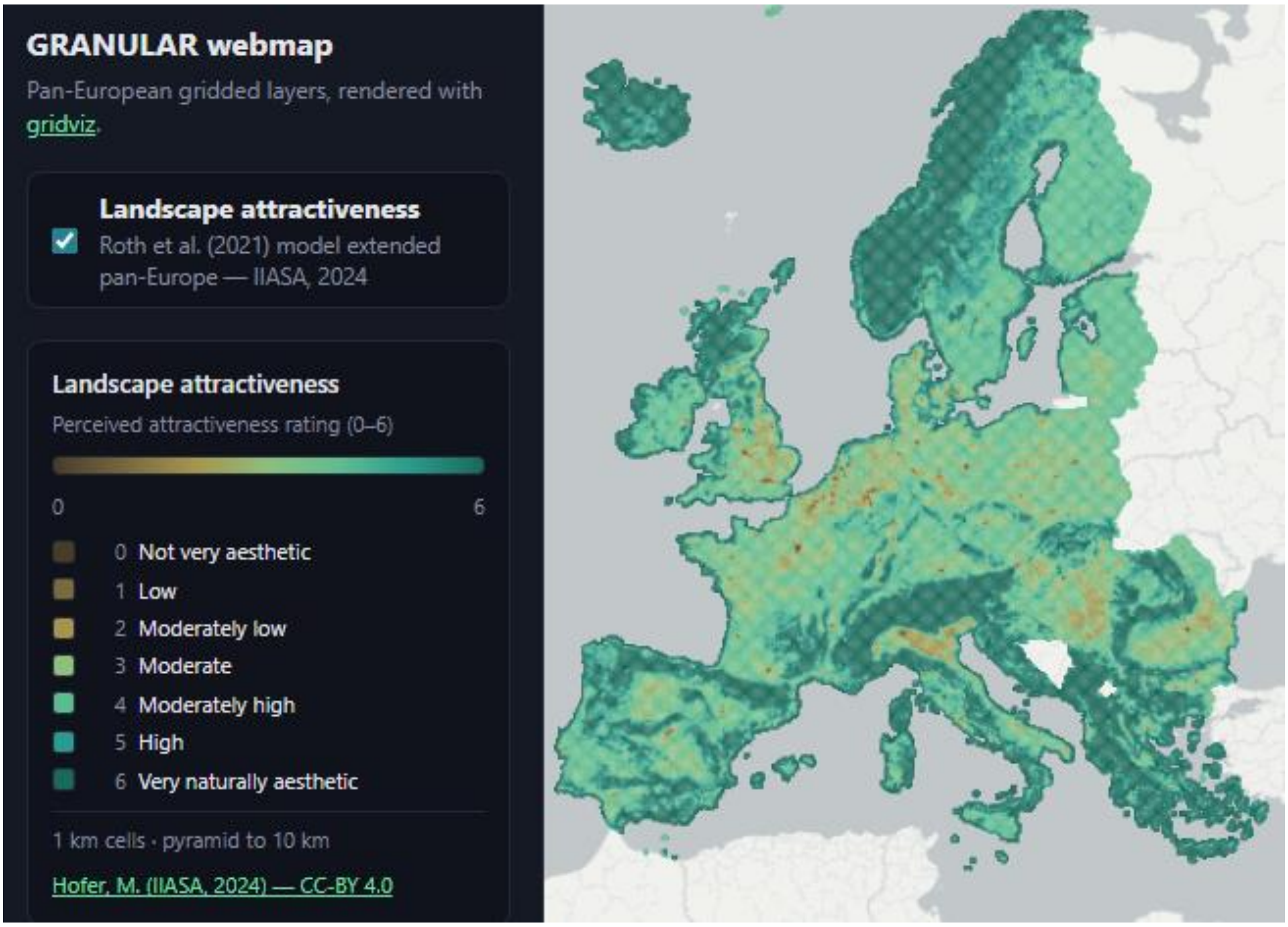



*Figure 29. Map of Landscape Attractiveness.*

**TripAdvisor and Tourism Behaviour**

With the recent proliferation of big data, social media in particular has been shown to provide additional insight into the flows and activities of people, in particular tourists. Here we analyze TripAdvisor data across the EU in four regions (within Spain, France, Finland and Italy), extracting all rural attractions and associated numbers of visitors and countries of origin for the available TripAdvisor record (2012-2023). TripAdvisor data is advantageous in that it records the frequency and timing of visits to specific attractions and provides this data via an API while also lending itself to web scraping. Numerous platforms record accommodation levels (e.g. booking.com, Airbnb, etc.) but these do not inform on where tourists are going beyond nights of accommodation.

In addition to the location and attraction type, we extracted TripAdvisor records from the regions of the number of visitors who frequent each attraction, along with the exact month and year of their visit (Figure 30). For the period 2015-2020 across the four regions we see specific preferences for certain categories of tourism along with seasonal fluctuations with the expected peak of visits within the European summer vacation months of July and August. The other categories display much less seasonal flux providing a steadier stream of visits. In recent years, TripAdvisor usage appears to decrease, likely as a result of a combination of TripAdvisor policies and competitive platforms offering similar services, e.g. Google.

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).*

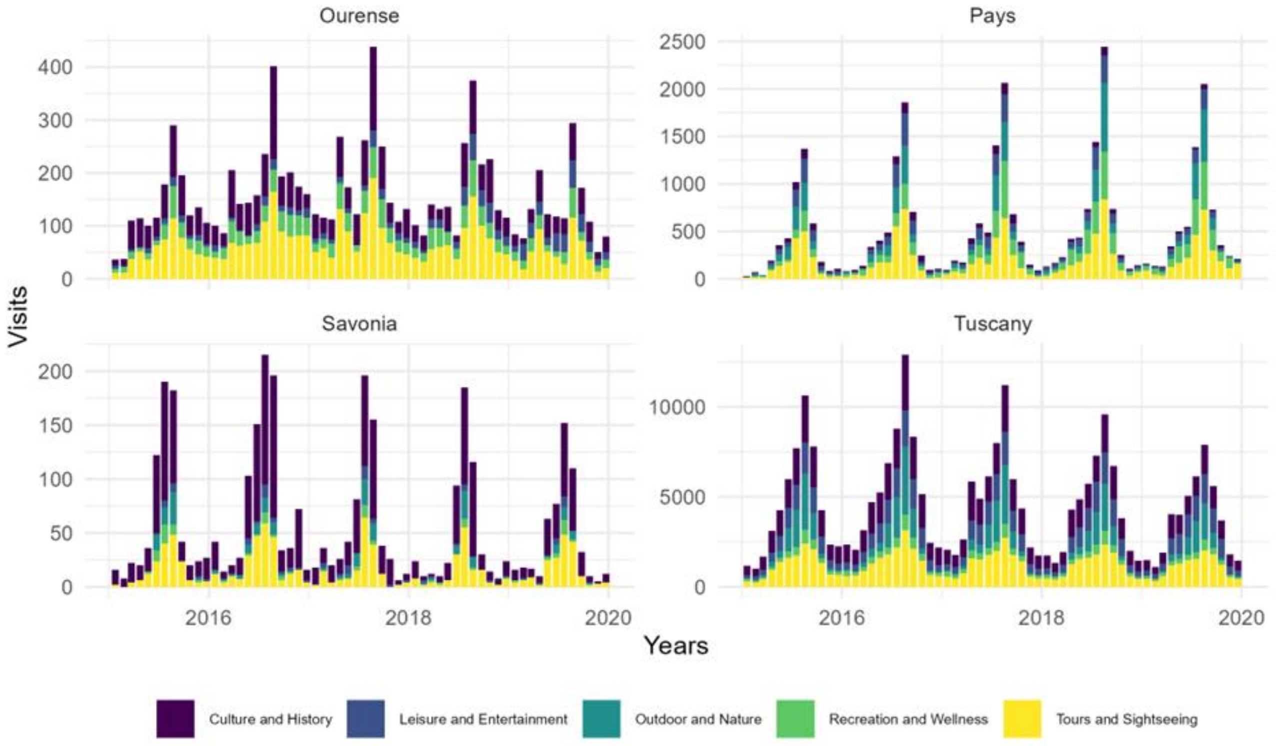



*Figure 30. Total monthly TripAdvisor visits across the four regions namely Spain, France, Finland and Italy, recorded by major attraction type for the period 2015 – 2020.*

**Rural semantic mapping with crowdsourced photos**

Focusing on crowdsourced Flickr imagery from rural areas within Europe, this study applies embedding models for semantic mapping (i.e. automatically identifying thematically coherent spatial areas), essentially clustering images by similarity. Categorizing these clusters through a rural policy lens (i.e. the rural diversity compass with residential, productive, environmental and recreational functions) we can compare regions and countries across Europe in terms of rural attractiveness, resilience, and recreational qualities. Finally we generate new indicators of rural resilience based on the Flickr embeddings.

In a first step we obtained the Flickr archive (do-me/Flickr-Geo · Datasets at Hugging Face). We selected only geo-tagged images with a free and open license. We then filtered these images to select only images which are located within rural Europe, using the Degree of Urbanization dataset - Degree of urbanisation - GISCO - Eurostat (we can use either vector or raster data). We then took the following steps:

- Create embeddings for each Flickr image obtained over rural EU (using this multi-modal model jinaai/jina-clip-v2 · Hugging Face). This takes into account any labels on the images if they exist, otherwise it just works with the imagery itself.
- Create topic clusters with tSNE/HUMAP/UMAP/PCA or similar
- Use kmeans or similar for clustering, removing outliers
- Creating areas from points, e.g. kernel density estimation or similar
- Label these areas with an LLM (“Relaxing in the green”, “Bicycle riding”, "Farming landscape", "Industrial zone"…)

Initial results appear in Figure 31, where the database is queryable via text prompts e.g. Scotland. This results in the selection of all images in the archive which appear to have a Scottish appearance (Figure 31).

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).*


*Figure 31. Top 12 results when querying the resulting Flickr database for "Scotland".*

Finally we are developing an algorithm to turn coordinates + embeddings into a "geospatial semantic cohesion map" or put differently to identify somewhat "semantically coherent areas". The results of this work will appear on the GRANULAR zenodo community and the GRANULAR website once complete.

## 3. Implementation and validation at local level (LLs) / Adaptation/upscaling to RLs

### Landscape Attractiveness

The landscape attractiveness data was designed to be used at the EU level, hence is available at the local level for use across Europe. To validate our results we implored several methods. There first approach is a standard holdout set from the original German training data. For this we achieve a mean-absolute class error of 0.16. Due to the nature of the input data, values outside of Germany should be scrutinised. Data for landscape attractiveness is hard to come by on a European level, however for the United Kingdom specifically we got access to a dataset called ScenicOrNot. This website crowdsources ratings of landscape images in England, Wales and Scotland. The images and ratings are geocoded which allows us to calculate scores. The mean-absolute class error is 1.41. Expectedly larger, but the confusion matrix shows how we still capture orderings even if exact classes don't match as much. Figure 32 shows the confusion matrices for these two validation sets.

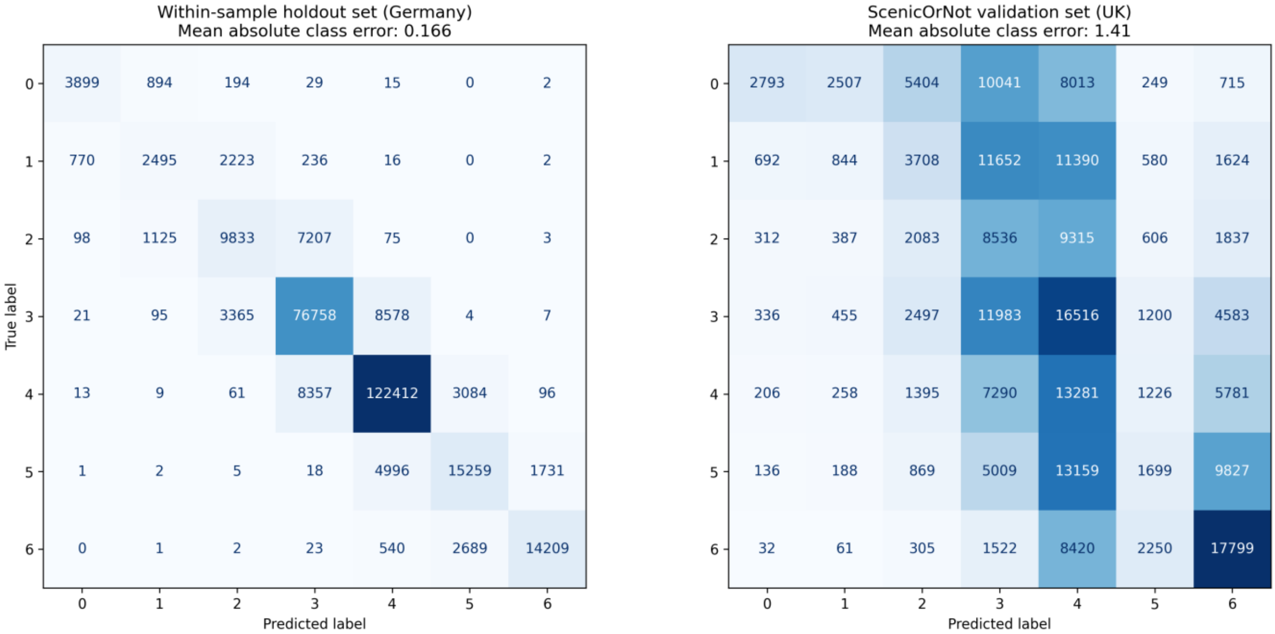



*Figure 32a (left): Confusion matrix of the holdout set. German landscape beauty taken from Roth et al., 2021 on the y axis against the model predictions for the same pixels. Figure 32b (right) shows the the average rating of crowed soured landscape beauty scores for the UK rescaled to 7 classes using quantiles.*

The final validation effort uses images from the LUCAS survey to inspect areas of each class. Currently this is based on the author's subjective interpretation and lacks quantifiable metrics.

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).*

TripAdvisor and Tourism Behaviour

See above section on Mobility, where the TripAdvisor data was applied at the LL/RL level, together with the FaceBook data in the context of local tourism. Validation was carried out at the national level for France, the LL level for Italy, France and Spain, and at the RL level in Finland. Validation was also conducted in Scotland for certain components of the work. Figure 33 compares TripAdvisor-derived visit counts with official tourism statistics for rural NUTS-2 regions in France. Each point represents one region, with the number of visits estimated from TripAdvisor reviews on the x-axis and the number of nights spent in tourist accommodation as reported by Eurostat on the y-axis. Regions with a population density above 800 inhabitants per km² were excluded to restrict the comparison to rural areas.

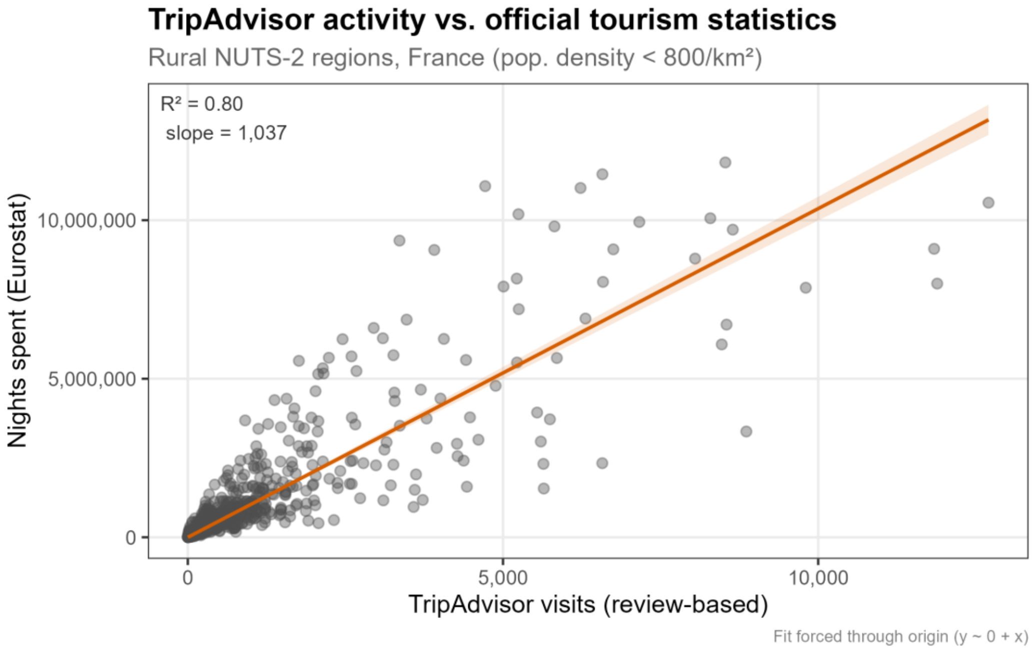



*Figure 33. TripAdvisor visits compared to Eurostat tourism statistics for rural NUTS2 regions in France.*

**Rural semantic mapping with crowdsourced photos**

The mapping of rural europe with crowdsourced photos (Flickr) is still in progress and there are plans to validate this at the LL/RL level via a small citizen science campaign using the Picture Pile application (https://pppui.cloud.geo-wiki.org/).

## 4. Limitations

**Landscape Attractiveness**

The model used here to predict landscape attractiveness across the EU was developed based on training data collected only within Germany. While a significant amount of the continental landscape attractiveness will be captured here (Germany is both central and large and contains many of the realms and biomes found across Europe), it will not capture all of the landscapes across Europe. Furthermore, what may prove attractive to German citizens may not be perceived similarly in other countries and vice versa.

**Trip Advisor**

Due to the numerous challenges in working with the TripAdvisor data (TA), i.e. data volume, costs associated with the API, web scraping limitations, language, etc., along with the noted decline in users, it is not realistic to be able to obtain a complete, long-term and continental or global dataset in perpetuity. This is specifically true for review based data as the volume of sites visited increased enormously. Nonetheless, the TA data available for specific regions is very rich and provides a unique perspective on tourism behaviour as a proof of concept.

**Flickr**

One major limitation which had to be overcome was the sheer size of the data. Furthermore any photo archives such as this will be biased in terms of location and subject matter of the photos. We find much less photos in rural areas versus urban settings, and photos tend to cluster around tourism sites. Furthermore photos tend to be taken during peak holiday times, weekends and often during good weather.

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).*

**5. Contribution to policy/call objectives:**

This work contributes directly to the objectives of the GRANULAR project by improving the availability, accessibility, and usability of data for rural monitoring. In particular, many of the methods described here involve data acquired via citizen engagement - through both active and passive data acquisition.

**6. How to use these datasets?**

The landscape attractiveness dataset is available on the GRANULAR Zenodo community (https://zenodo.org/records/18618619). This dataset provides full coverage of Europe in a grid format, and may be used for both visual analysis via the GRANULAR Viewer, as well as for download and ingestion into GIS software for further spatial analysis. This can also be used for modelling purposes as one predictor of e.g. rural well-being. TripAdvisor data is more challenging to use, however the results of this work have been used to inform assessments at the LL and RL levels – see the above description under Mobility where these results were combined with Facebook data for the Finnish RL. The Flickr data is challenging to work with owing to its size, however we plan to share the resulting embeddings via the GRANULAR viewer for visual analysis.

# 5. Discussion and Recommendations

The analyses undertaken across the five GRANULAR research themes demonstrate that rural areas cannot be understood through a single narrative of decline, remoteness, or dependency. Instead, the results consistently reveal substantial diversity in accessibility, mobility, economic structure, land use, and quality of life, often at very fine spatial scales. Accessibility indicators show that rural populations experience highly unequal access to services and opportunities, both between and within regions, while mobility analyses highlight the strong interdependencies between rural and urban areas through commuting, tourism, seasonal migration, and multi-local living. At the same time, innovative economic and land-use analyses indicate that many rural territories possess under-recognised assets, including diverse business ecosystems, high landscape attractiveness, and strong environmental capital. Collectively, these findings reinforce one of the central premises of GRANULAR: effective rural policy requires granular, place-based evidence capable of capturing local differences that are often obscured within conventional administrative statistics.

A key lesson emerging from the validation process and discussions with Living and Replication Labs is that rural resilience is strongly linked to connectivity, adaptability, and diversity rather than any single economic or demographic characteristic. Regions demonstrating greater resilience were often those able to maintain multiple forms of connectivity: physical connectivity through transport and accessibility, economic connectivity through diverse business networks and tourism flows, digital connectivity through access to data and services, and social connectivity through active stakeholder engagement. The analyses also reveal the growing importance of seasonal and temporary population dynamics in many rural regions, challenging static understandings of population and service demand. Tourism regions, areas benefiting from green industrial transitions, and territories with strong rural–urban linkages frequently exhibited significant temporal fluctuations that are largely invisible in traditional datasets. These findings suggest that rural resilience should increasingly be understood as the capacity to adapt to changing flows of people, services, investment, and environmental pressures, rather than simply the ability to maintain existing structures.

The review process highlighted an important question: beyond developing innovative datasets, what have we learned about rural Europe itself? The evidence generated through GRANULAR suggests that rural areas are undergoing profound but uneven transitions. Climate adaptation, demographic change, tourism pressures, changing mobility patterns, digitalisation, and evolving economic activities are reshaping rural territories in different ways. Consequently, future rural observatories and policy frameworks should prioritise integrated monitoring systems that combine official statistics with novel sources such as Earth observation, crowdsourced information, citizen science, and mobility signals. Investments should focus not only on improving data availability but also on building local capacity to interpret and apply this information. We therefore recommend continued development of open, harmonised, and reproducible data infrastructures; expansion of multi-actor validation processes through Living and Replication Labs; and greater use of high-resolution indicators that can support rural proofing and place-sensitive policy design. The overall experience of GRANULAR demonstrates that novel data sources, when validated and interpreted together with local stakeholders, provide powerful new insights into rural resilience and offer a practical pathway toward more responsive, adaptive, and evidence-based rural governance across Europe.

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).*

# 6. Conclusions

The GRANULAR project demonstrates that combining novel data sources, advanced analytical methods, and participatory validation processes can significantly enhance our understanding of rural territories and their ongoing transitions. Across accessibility, mobility, rural economies, land use, and citizen engagement, the project has shown that rural areas are far more diverse, dynamic, and interconnected than is typically captured by conventional statistics. The validation and upscaling exercises undertaken through the Living and Replication Labs confirmed the value of integrating open data, Earth observation, social media signals, crowdsourced information, and artificial intelligence with local knowledge and stakeholder expertise. While no single dataset or method can fully capture the complexity of rural change, the combined evidence generated by GRANULAR provides a richer and more nuanced picture of rural Europe, highlighting both persistent inequalities and emerging opportunities. Ultimately, the project demonstrates that resilient rural areas are those capable of adapting to change while maintaining strong social, economic, environmental, and institutional connections. By providing openly accessible datasets, reproducible workflows, and practical tools for policy and planning, GRANULAR leaves a lasting legacy that can support more evidence-based, place-sensitive, and future-oriented rural development across Europe.

# 7. References

IPCC (2014) “Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects”. ISBN: 9781107641655

Jain P., Marcos D., Ienco D., Interdonato R., Dhakal A., Jacobs N., & Berchoux T. (2024) “Aligning geo-tagged clip representations and satellite imagery for few-shot land use classification”. *2024 IEEE International Geoscience and Remote Sensing Symposium*. 319-323. https://doi.org/10.1109/IGARSS53475.2024.10641235

Jain P., Ienco D., Interdonato R., Berchoux T., & Marcos D. (2025) “SenCLIP: Enhancing zero-shot land-use mapping for Sentinel-2 with ground-level prompting”. *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 5656-5665. https://doi.org/10.1109/WACV61041.2025.00552

Jain P., Marcos D., Ienco D., Interdonato R., & Berchoux T. (2026) “TimeSenCLIP: A time series vision–language model for remote sensing”. *ISPRS Journal of Photogrammetry and Remote Sensing*, 236, 99–119. https://doi.org/10.1016/j.isprsjprs.2026.03.043

OECD (2024) “Getting to Services in Towns and Villages”. https://doi.org/10.1787/df1e9b88-en

Radford A., Wook Kim J., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., Sutskever I. (2023) “Learning Transferable Visual Models From Natural Language Supervision”. *Proceedings of the 38th International Conference on Machine Learning*, PMLR 139:8748-8763, 2021.

Roth M., Hildebrandt S., Walz U., & Wende W. (2021) “Large-Area Empirically Based Visual Landscape Quality Assessment for Spatial Planning—A Validation Approach by Method Triangulation”. *Sustainability*, *13*(4), 1891. https://doi.org/10.3390/su13041891

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).*