



GRANULAR

KTA webinar on Data for Rural Areas

5 December 2024

GRANULAR has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101061068. UK participants in the GRANULAR project are supported by UKRI- Grant numbers 10039965 (James Hutton Institute) and 10041831 (University of Southampton). 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 (REA). Neither the European Union nor the granting authority can be held responsible for them.

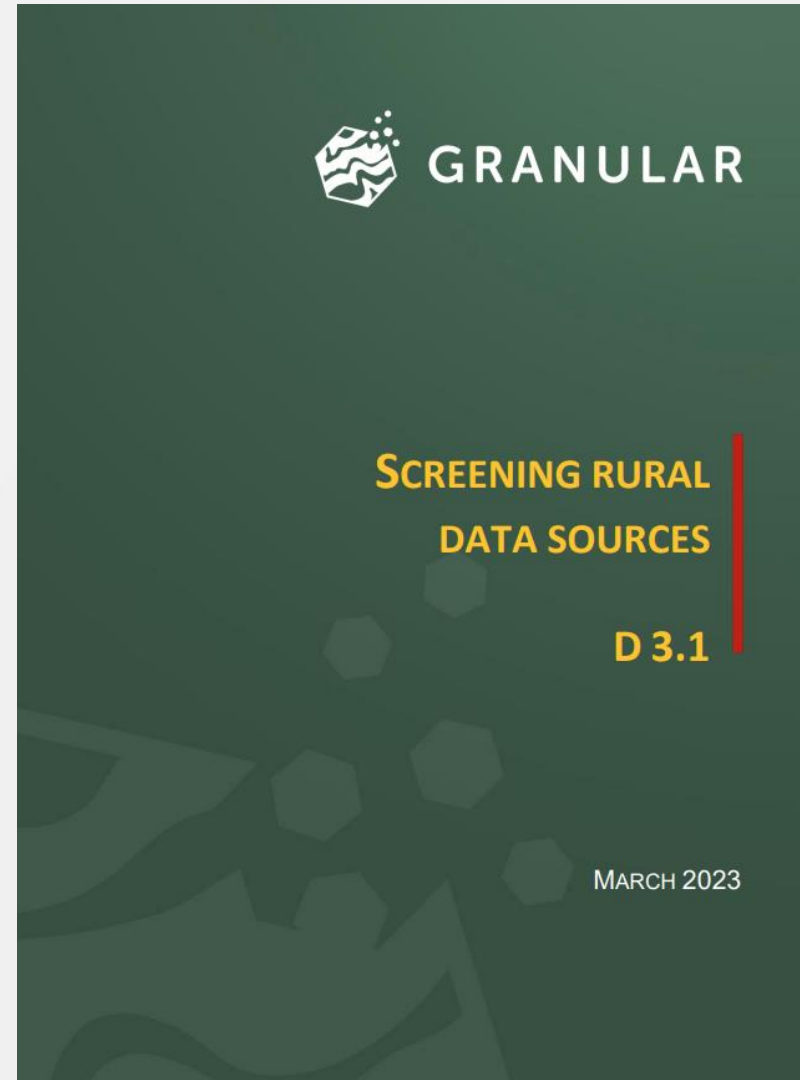
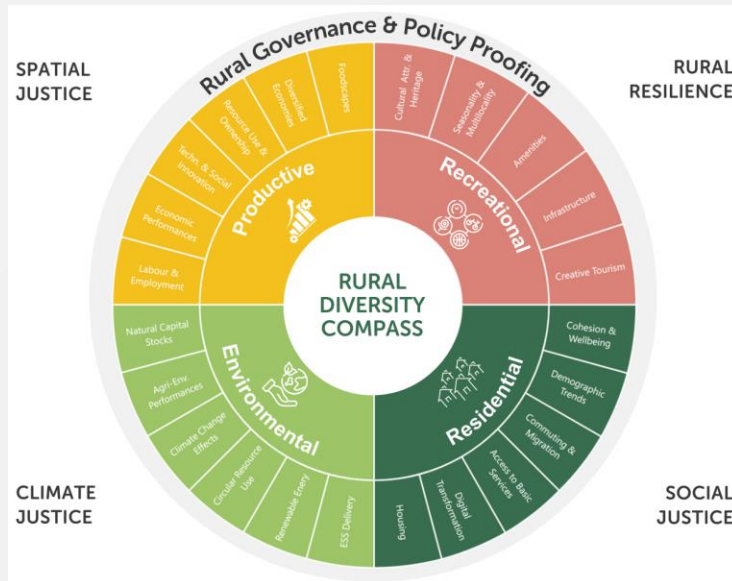


**Co-funded by
the European Union**

BENCHMARK OF RURAL DATA, ISSUES, RESULTS

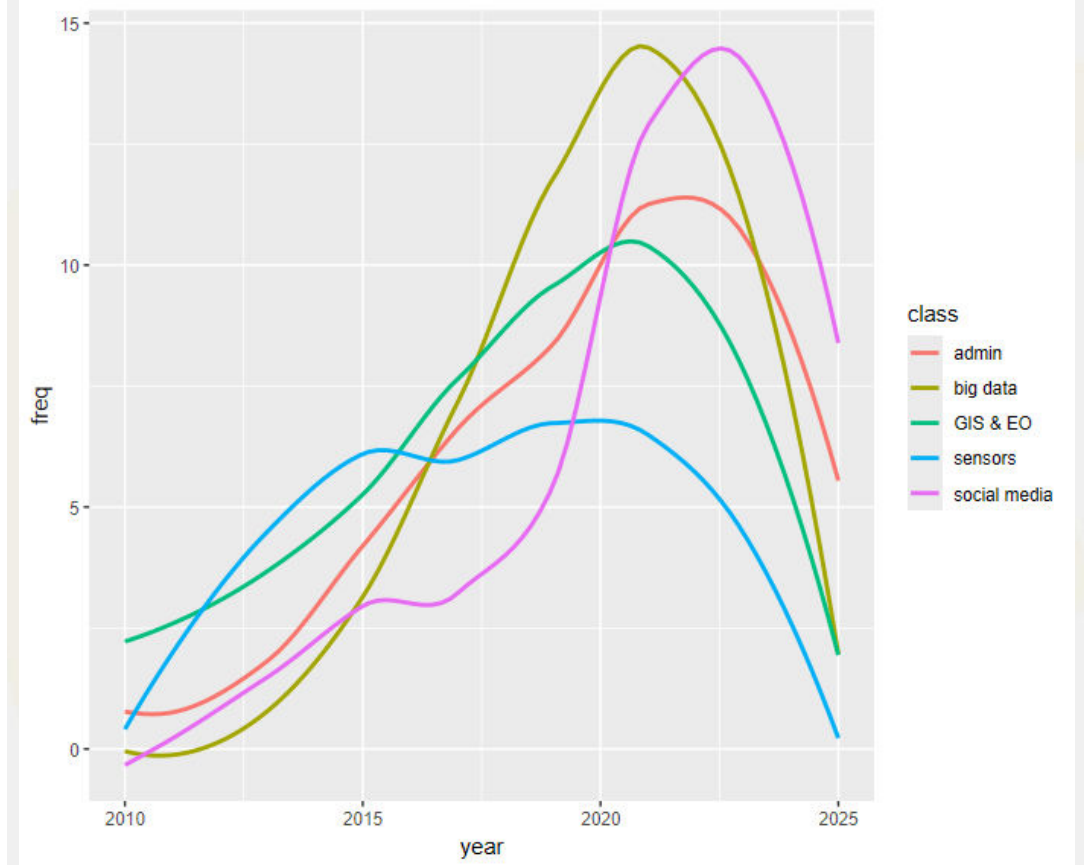
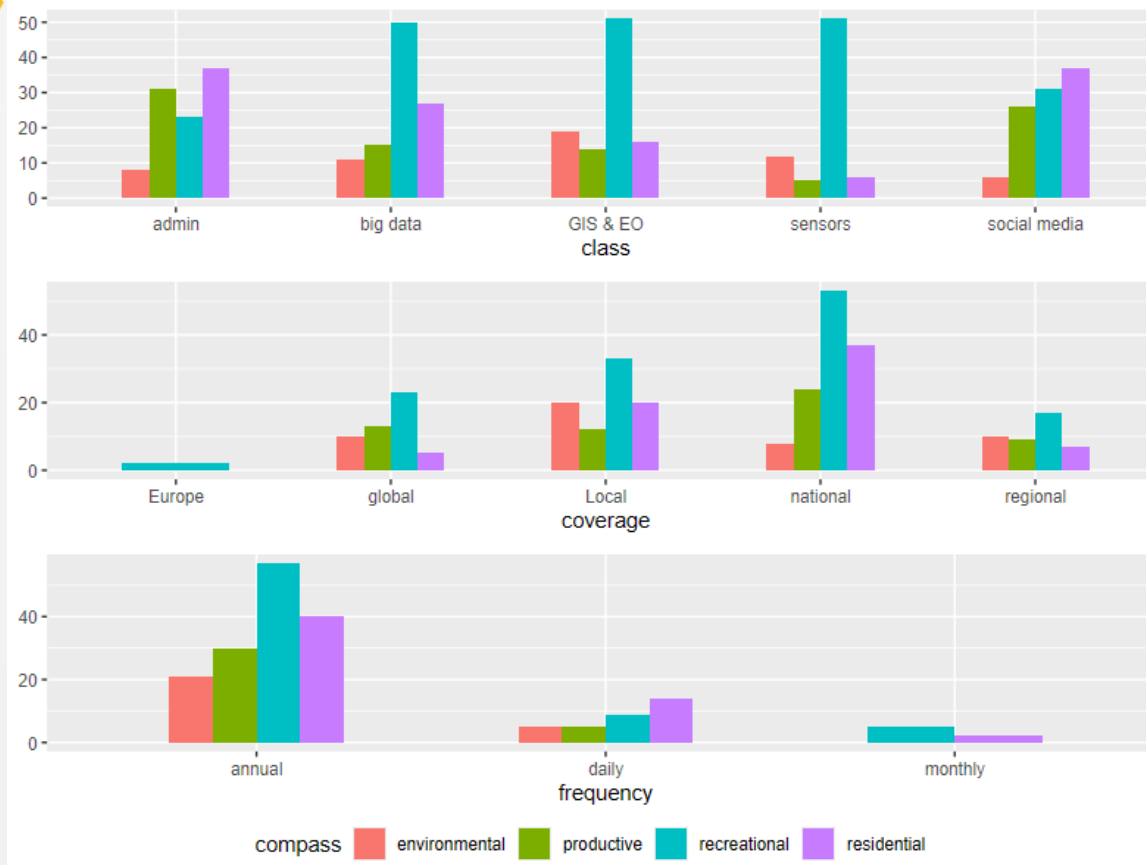
Data opportunities

- **Objective:** Generate and assess new data to better showcase the diversity of rural areas, their functional characteristics, challenges and opportunities.



DATA AVAILABILITY

Data types and trends



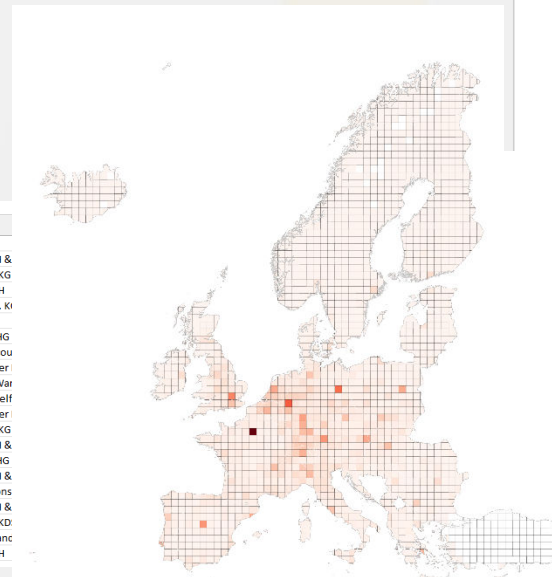
PRODUCTIVE

“labour” OR “employment” OR resource* OR “ownership” OR foodscapes OR “diversified economies” OR “economic performances” OR “technological innovation”

- Data examples
 - Regional economy indicators (Open Street Maps)
 - National Statistics (Eurostat – EU SILC, ...)
 - National Chambers of Commerce

- Methods
 - GIS
 - Surveys
 - Statistical analysis
 - Web-scraping

A	B	C	D	E	F	G	H	I
Zentrale	Region	Markttyp	Partnerfirma	Regie	Organisation	Geschaeftsart	VKF	MA Inhaber
1 Aldi Süd	Aldi-Dormagen	Aldi Süd		R	Aldi	Discounter		ALDI Einkauf GmbH & Lidl Stiftung & Co. KG
1 Lidl	Lidl-Leverkusen	Lidl		R	Lidl	Discounter		PENNY Markt GmbH
1 Penny	Penny-West	Penny		R	Rewe	Discounter		S-Mart GmbH & Co. Ki
1 Rewe Regionen	Rewe-West	REWE City	Rahmati	S	Rewe	Supermarkt (400 bis 799 VKF)		REWE Markt GmbH
1 Rewe Regionen	Rewe-West	REWE City	REWE Regiebetriebe [R R	R	Rewe	Supermarkt (400 bis 799 VKF)		REWE Wintgens OHG
1 Rewe Regionen	Rewe-West	REWE	REWE SEH	S	Rewe	Supermarkt (400 bis 799 VKF)		Netto Marken-Discou
1 Netto-Marken-Discount	Netto Schels	Netto		R	Edeka	Discounter		Markus Hetzenegger
1 Edeka Regionen	Edeka Rhein-Ruh	Edeka	Hetzenegger	S	Edeka	Supermarkt (400 bis 799 VKF)	600	Feldmann-Höner War
1 Edeka Regionen	Edeka Rhein-Ruh	Edeka	Feldmann-Höner	S	Edeka	Supermarkt (400 bis 799 VKF)		Norma Lebensmittel
1 Norma	Norma	Norma		R	Norma	Discounter		Markus Hetzenegger
2 Edeka Regionen	Edeka Rhein-Ruh	nah & gut	Hetzenegger	S	Edeka	Supermarkt (400 bis 799 VKF)	300	Lidl Stiftung & Co. KG
3 Lidl	Lidl-Leverkusen	Lidl		R	Lidl	Discounter		ALDI Einkauf GmbH &
4 Aldi Süd	Aldi-Dormagen	Aldi Süd		R	Aldi	Discounter		REWE Wintgens OHG
5 Rewe Regionen	Rewe-West	REWE	REWE SEH	S	Rewe	Supermarkt (400 bis 799 VKF)		ALDI Einkauf GmbH &
6 Aldi Süd	Aldi-Dormagen	Aldi Süd		R	Aldi	Discounter		ALDI Einkauf GmbH &
7 NN	NN	Alnatura		R	Sonstige	Biomärkte		Alnatura Produktions
8 Aldi Süd	Aldi-Dormagen	Aldi Süd		R	Aldi	Discounter		ALDI Einkauf GmbH &
9 Kaufland	Kaufland	Kaufland		R	Markant	Verbrauchermarkt (2500 bis 4999 VKF)		Kaufland Vertrieb KD
10 Edeka Regionen	Edeka Rhein-Ruh	Marktkauf	EDEKA Regiebetriebe [R	R	Edeka	SB-Warenhaus (ab 5000 VKF)		Marktkauf Einzelhanc
1 Penny	Penny-West	Penny		R	Rewe	Discounter		PENNY Markt GmbH



LOCAL FOOD PRODUCERS – QUESTIONNAIRE

I. GENERAL SECTION:

- Where is your enterprise/farm located? _____ zip code _____
- When did your business first become established? _____
- How long have you been running your business? _____
- What is the legal form of your business?

<input type="checkbox"/> Individual farm	<input type="checkbox"/> Limited liability company	<input type="checkbox"/> Limited partnership
<input type="checkbox"/> Registered business activity	<input type="checkbox"/> Public limited company	<input type="checkbox"/> Association/ Foundation
<input type="checkbox"/> Civil Company	<input type="checkbox"/> General partnership	<input type="checkbox"/> Other, what kind _____
- What is your educational background?

Level: <input type="checkbox"/> Elementary	<input type="checkbox"/> Essential vocational	<input type="checkbox"/> Secondary education/high school	<input type="checkbox"/> Higher
Field: <input type="checkbox"/> Agricultural	<input type="checkbox"/> Economical	<input type="checkbox"/> Artistic	<input type="checkbox"/> Other, what kind _____
- How many people, including yourself, work in the enterprise/farm?

Permanently _____	including family members _____
Seasonally _____	including family members _____

Number of points in each category in the filtered dataset:

```
key
amenity    2800117
shop       1939594
office     325995
tourism    315522
sport      143746
craft      97870
Name: count, dtype: int64
```

RECREATIONAL

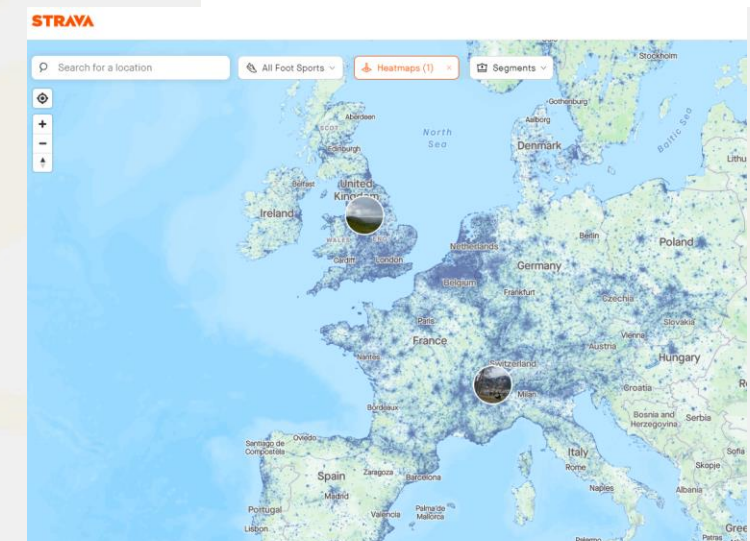
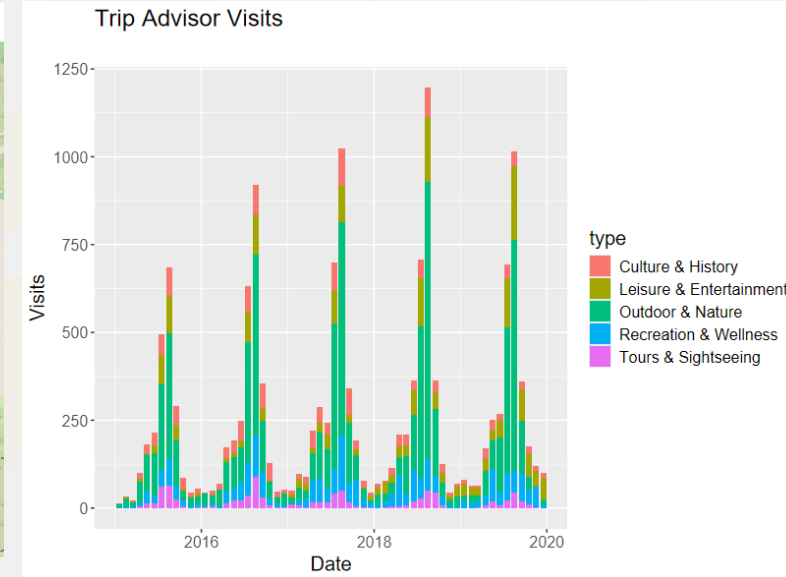
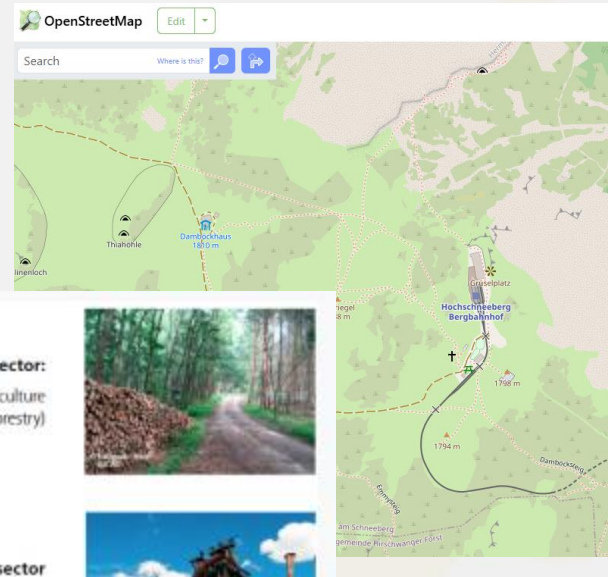
amenities OR infrastructure OR “cultural attractiveness” OR heritage OR tourism* OR seasonality OR “multi-locality”

■ Data examples

- Open Street Maps (OSM)
- Social media (TripAdvisor, Booking, Airbnb, etc.)
- Facebook, Instagram, etc.
- LUCAS - Eurostat
- Strava

■ Methods:

- Web-scraping
- Time series analysis
- GIS/spatial statistics
- Computer vision/VLMs



ENVIRONMENTAL

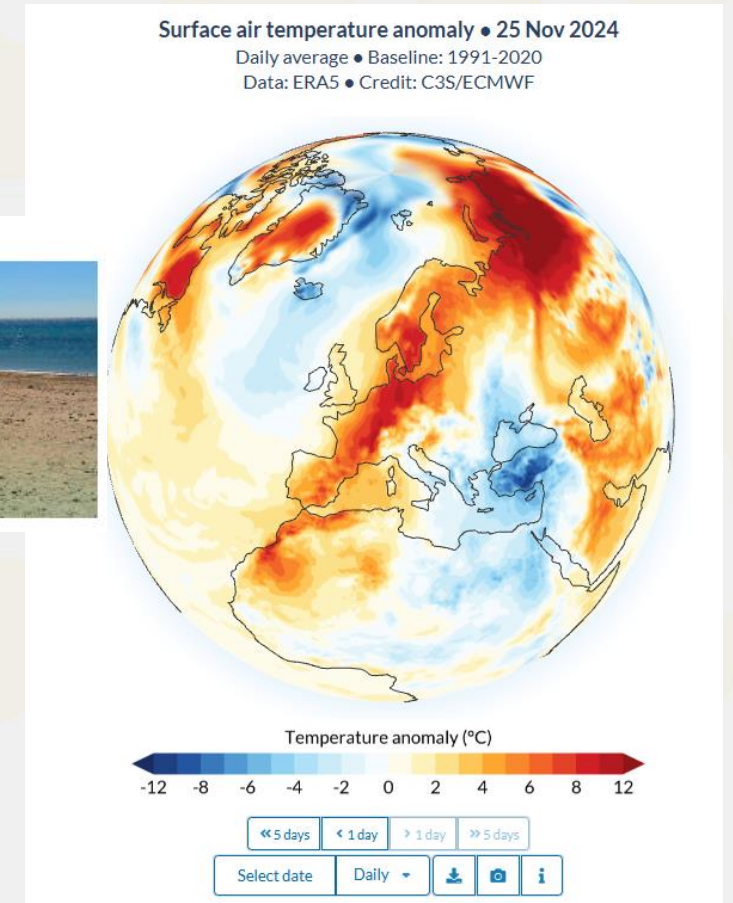
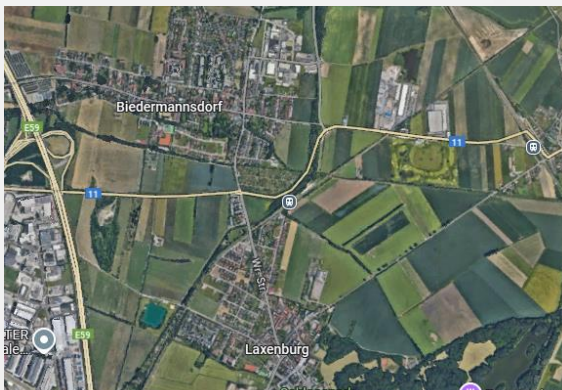
“natural capital” OR “agri-environmental” OR “agriculture” OR “circular resource use” OR “climate change effects” OR “renewable energy” OR “ecosystem services”

■ Data Examples

- Satellite data (Copernicus)
- Land use / Land cover
- Ecosystems (<https://globalecosystemsatlas.org/>)
- LUCAS - Eurostat
- Biodiversity
- WCMC – global protected areas

■ Methods

- AI, ML, Deep Learning
- LLMs, VLMs (SatCLIP, CLIP)



RESIDENTIAL

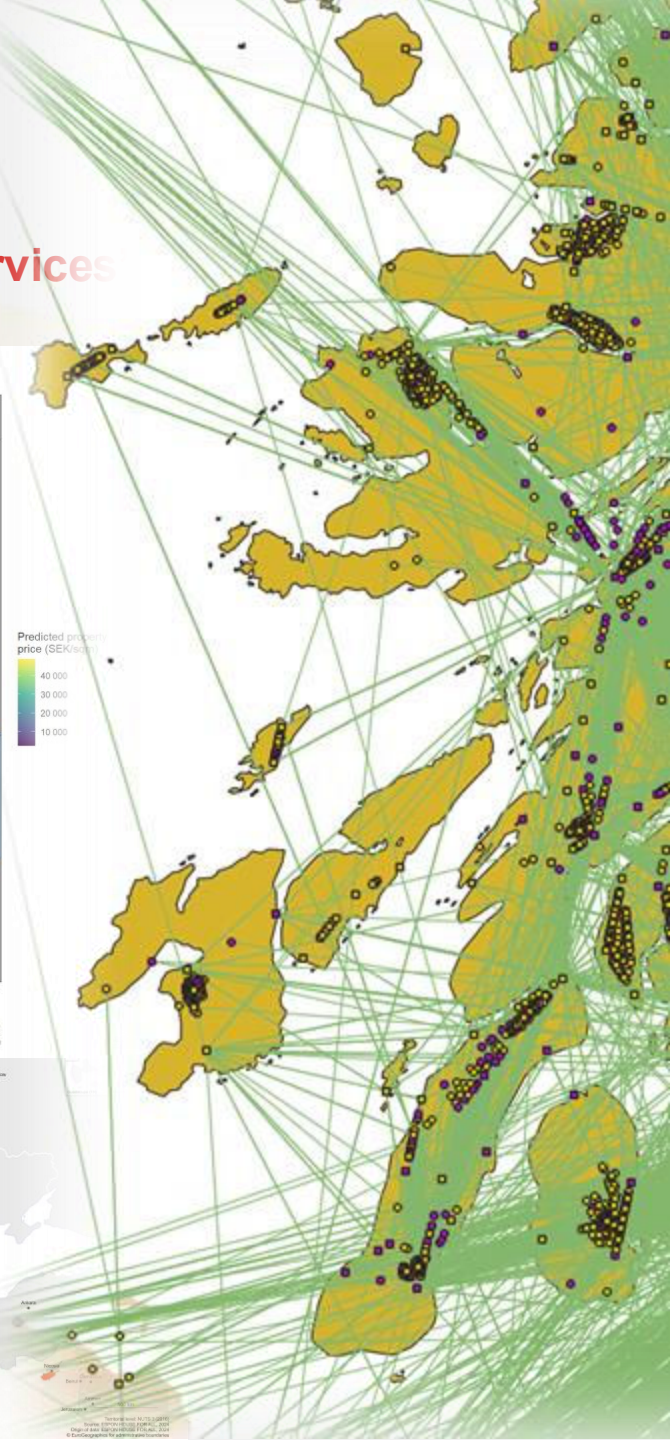
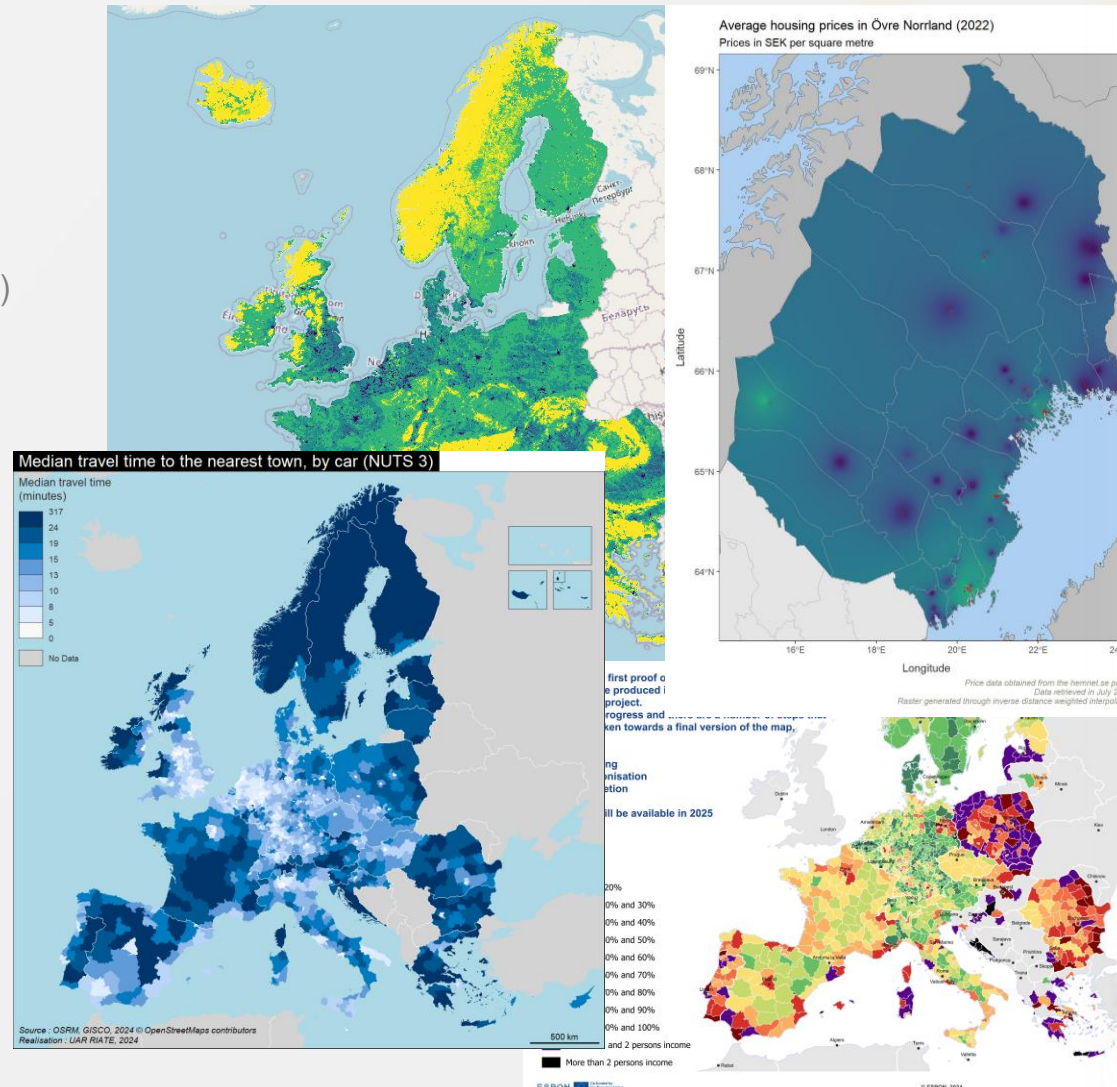
housing* OR “demographic trends” OR commuting OR migration OR “basic services” OR “digital transformation” OR cohesion OR wellbeing OR “well-being”

■ Data examples

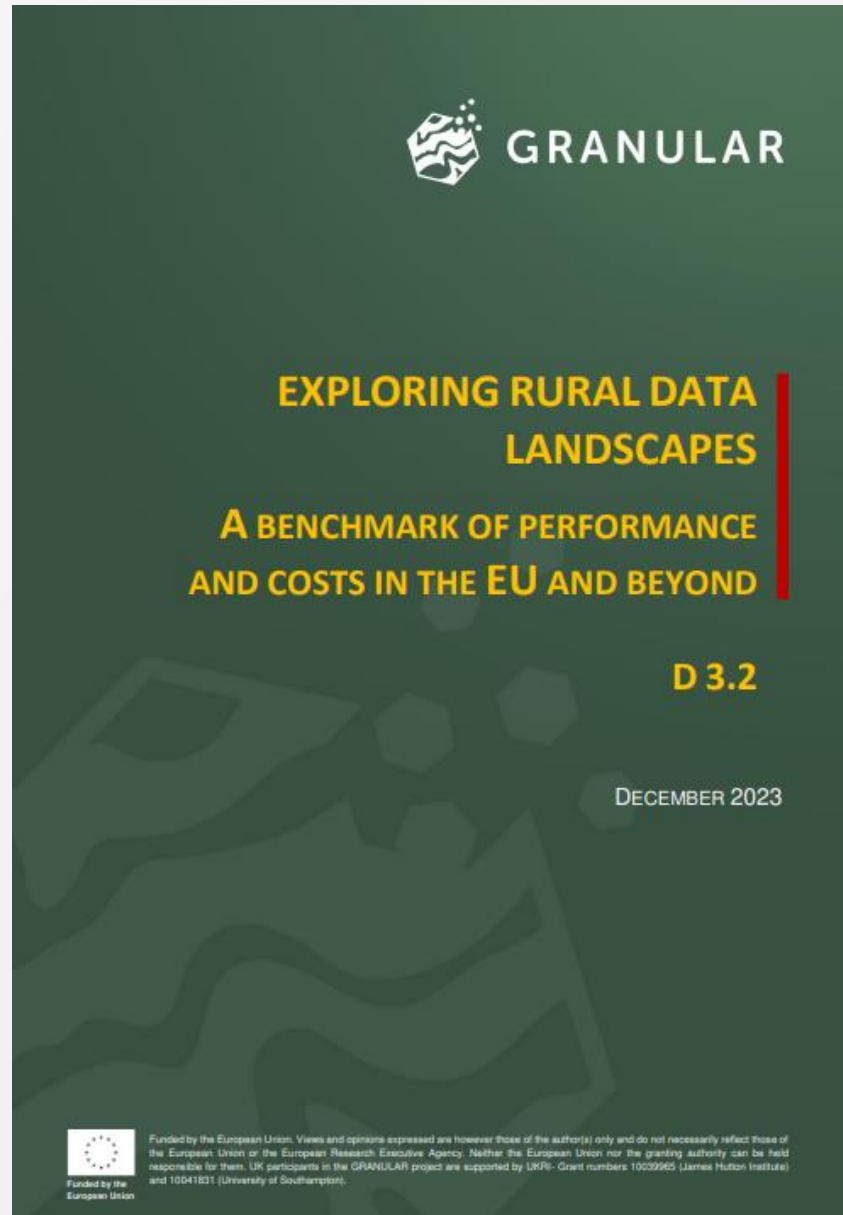
- Housing prices
- Rural attractiveness
- Mobility (Facebook)
- Accessibility (OSRM)

■ Methods:

- Web-scraping
- OSM routing
- Econometric models
- Geo-statistics



TASK 3.2 EXPLORE OPTIONS, PERFORMANCE AND COST (M06-15)



3 key inspirational elements / sections

- 1) The costs of and for rural data – a journey across national statistical offices and authorities in the EU and beyond
- 2) 9 Inspirational examples from Granular partners – Data, tools and approaches
- 3) 27 Data fiches

OVERVIEW OF NATIONAL DATABASES OPENLY AVAILABLE, CONTAINING DATA ABOUT RURAL DEVELOPMENT ISSUES



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

9 INSPIRATIONAL EXAMPLES / GOOD PRACTICES



GRANULAR



Global: Geo-Wiki Earth Observation & Citizen Science

Ivelina Georgieva (IIASA)

The Geo-Wiki platform provides anyone with the means to engage in monitoring of the Earth's surface by classifying satellite, drone or ground-level imagery. Data can be input via desktop or mobile devices, with campaigns and games used to incentivise input. These innovative techniques have been used to successfully integrate citizen derived data sources with expert and authoritative data to address pressing policy-related questions (e.g. European environmental policy, SDG indicators and more). Geo-Wiki was established in 2010 in the Novel Data Ecosystems for Sustainability research group, part of the Advancing Systems Analysis Program at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria.

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

To facilitate the process of visual interpretation, volunteers have access to specifically defined (for the validation task at hand) Geo-Wiki functionalities. The principle set of tools which the platform includes are implemented features like Sentinel Hub time series imagery, an NDVI tool for measuring the Normalized Difference Vegetation Index and Google Earth history imagery. The quality of contributions has been controlled from a group of experts during and after the campaigns and scientific publications are used to share the data, which are later uploaded in public repositories to ensure transparency of the entire process.

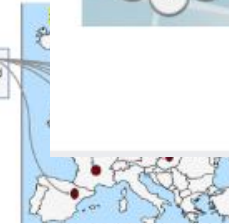


COSTS

Data Collection

The project initially uses LUX ground-level visual context (resolution) and Sentinel-2 (1 collected from Microsoft Bing services, amounting to 230,00

LUCAS 2018 Geo-Tagged Ground Level image



foundational expenses for the tool would encompass creating a user interface, an API, and backend infrastructure, along with utilising online cloud servers like AWS, Google Cloud, or Microsoft Azure for model deployment. Additionally, there might be costs associated with data acquisition, such as purchasing satellite imagery or ground-level data. The costs for purchasing the service point data was around € 12,000.

type of geography for which there was limited data available.

The citizen the re include rural a and its and th statisti views Banon Rural 2020. Couns e.g., fr

Employees a Carwinńska-J collaborated i their research communities at Special Exdu

The primary degraded at assessment i spatial issue diagnostic in created, and was proposed

The R Respo speak (Swed respon The R A plus very i More j T am identifi The c dwelli dream Rural rural p



The Département des Pyrénées Orientales (I system for observing and monitoring tra roads it manages. As a local authority in th the CD66 is responsible for a network of 2 serving an area of over 4,000 km² with a po (source: INSEE 2020).

COSTS

BUDGET OF THE WHOLE INITIA

A total of around €165,000 per year in op whole department. This include is (€100,000/year), costs of replacing and maint 62,000 / year) including ROUTE+ software m / year) and costs for acquisition and training (I

To help manage its road network, CD66 prod several objectives in mind:

- To understand traffic trends and help d long term general studies,
- To design road projects, in particular to structures,
- To adapt measures taken in ten operations, for example by choosing the least impact on traffic during roadside mo

CD66 is responsible for all the stages involve data: from the installation and maintenance a collection, consolidation, and analysis of the information available on open data platfor and one data administrator are working on i relies on a network of permanent, rotating meters, as well as the ROUTE+ software and

France: M scale – Ca Orientales

Louise Chasse Claude Balagu

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

COSTS

The extracted data

The background that allows for a strong territorial analysis of this health care or welfare system is the continuous internal system of data collection. The data collection has both a statistical and a territorial level. On the one hand, a large amount of information is collected for each procedure performed by each user. Depending on the year, the number of users varies between 2000 and 3000 people, and the amount of information is large.



Galicia – Spain: Tele rural areas of Ouren

Maria Isabel Doval Ruiz & Bre

In 2020, researchers at the James contacted by the Scottish Government t analyse the National Islands Plan Surv from residents of Scotland's islands opinions on island life, which were objectives of the National Islands Plan.

The survey implemented a customised s inform both survey sampling and the rep to identify diversity in lived experience local authority regions; these regions developed into an official geography for work by National Records of Scot Government. The survey achieved 43 islands.

COSTS

The project involved a team of six reser on the tool development and publication. skills in R/Shiny programming. Data host currently zero, through a free account at subscription was taken out for that acco of the whole contract for the Survey was



Scotland results

Jonathan I

Nord Mats Stj

Nordic Service Mapper is an intera tool that visualises the proximi across the Nordic Region. It covers Finland, Iceland, Norway, and Swe Islands and Åland. The tool includ services, namely grocery stores, schools. The tool was published in situation in December 2019. The proximity for the population to var different geographical levels.

Various data aggregation and normal in the tool, including access to municipality level, according to Urbanisation at municipality level (cib at a more spatially detailed grid-level selection of areas.

COSTS

The Nordic Service Mapper tool was collaboration with Ubiqu. The tool was the Nordic Thematic Group on Sust (2017–2020) as part of the project R geography of service within the Sternberg and Oskar Penje at Nordi accessibility calculations were carried The work to create the web-mapping j



27 DATA FICHES


GRANULAR
DATA FICHE
Human powered mobility by Strava Metro
<https://metro.strava.com/>
 Ian McCallum & Martin Hofer (IIASA)

INDICATOR CLASS: Residential


Data

Resolution: Class; vectors
 Type: Shapefiles
 License: <https://metro.strava.com/>



**Better communities
for cyclists and
pedestrians.**

[Apply for a metro partnership](#)


Spatial

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


Temporal

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

DESCRIPTION

The **Strava** dataset is the largest collection of human-powered transport information in the world. Metro aggregates, de-identifies and contextualizes this dataset to help make communities better for anyone on foot or a bike.

Strava works with urban planners, trail networks, city governments and safe-infrastructure advocates to understand mobility patterns, identify investment opportunities and evaluate the impact of infrastructure changes free of charge.


USER COSTS

Free & open Access

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


DATA INFRASTRUCTURE

Information not available.


DATA GOVERNANCE & MANAGEMENT

Staff costs: Using the online dashboard and UI would mean costs would be low, and anyone with basic internet skills could query and visualize the data.

Data analysis needed: If a user chooses to download the data, e.g. Shapefiles, this will then require GIS skills to work with and further analyze the data. This should however be fairly basic for any GIS analyst. Onin total would likely be enough to develop basic tools to work with and process the data.

Quality assurance: Strava is performing QA on their end, hence little effort is required for this from the user.


DATA DOCUMENTATION AND/OR ASSISTANCE IN DATA USE

<https://metro.strava.com/> <https://www.strava.com/>

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



Co-funded by
the European Union



UK Research
and Innovation

GRANULAR has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101019188. UK participants in the GRANULAR project are supported by UKRI. Grant numbers: 101019188 (James Hutton Institute) and 101019185 (University of Southampton). 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 Council (ERC). Neither the European Union nor the granting authority can be held responsible for them.

Innovation & use of open data advocated in D3.2.

Inspirational examples => ongoing investments / work with new & open datasets => foster innovation / maximize societal & economic benefits of open data.

Demand for complex & granular data, in line with the Directive's focus on providing valuable information tailored to user needs.

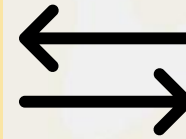
Collaboration, knowledge sharing & peer-to-peer learning among public sector bodies to enhance data availability and reuse. (*Directive's aim to minimize costs for reusing public sector information*).

Capacity building & training in local territories & public sector bodies

Quality assessment & validation to ensure reliable & usable data for decision-making (*Directive's encouragement of high-quality public sector information*).

Sustainability, accessibility => need for capacity-building programs and streamlined data access processes (*Directive's aim of ensuring sustainable and accessible public sector information*).

Cost dimensions: good open data available! YET for utilization, personnel costs, quality assessments, and data management needed



SUGGESTIONS FOR GRANULAR

- **Continue working with presented data & inspirational examples**
- **Living & Replication Labs** may analyze presented data types and survey results for potential replication or further exploration.
- **Explore datasets like the Strava dataset**, possibly through applying for free access, to enhance understanding.
- **Reflect on local needs and circumstances, and engage with authors and contacts** for further discussions.
- **Align actions with the Directive's goals** of improving data availability, promoting collaborative efforts, and tailoring practices to local conditions.