RUSTIK

# Comparative analysis of LL data needs investigations *(D2.2)*

Marta Carbonés
Daniel Pi-Suñer
Harold del Castillo
Camila Álvarez
Oriol Biosca
Andreu Ulied



*This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101061051. The work developed under this project is also supported by Innovate UK through the Horizon Europe Guarantee scheme. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Innovate UK. Neither the European Union nor the granting authority can be held responsible for them.*

Funded by the European Union

UK Research and Innovation

# Comparative analysis of LL data needs investigations (D2.2)

Report containing LL data needs common to all LL and more specifically data needs for each LL. Comparative analysis in relation to the specificities of each LL, in terms of the interest and feasibility (technical, financial, legal) of data.

| Deliverable Number | WP / T |
|---|---|
| D 2.2 | WP 2 |
| **Lead Beneficiary** | **Deliverable Author(s)** |
| MCRIT | Marta Carbonés; Daniel Pi-Suñer; Harold del Castillo; Camila Álvarez; Oriol Biosca; Andreu Ulied |
| **Beneficiaries** | **Deliverable Co-Author(s)** |
| – | – |
| **Planned Delivery Date** | **Actual Delivery Date** |
| 31.12.2023 | 31.12.2023 |

| | | | |
|---|---|---|---|
| Type of deliverable | R | Document, report (excluding periodic and final reports) | x |
| | DATA | Data sets, microdata, etc. | |
| | DMP | Data management plan | |

| | | | |
|---|---|---|---|
| Dissemination level | PU | Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page) | x |
| | SEN | Sensitive, limited under the conditions of the Grant Agreement | |

Funded by the European Union
UK Research and Innovation

INDEX

## TABLES



## FIGURES

# 1. Introduction and approach of the report

This report offers a thorough assessment of data needs and resources within the Living Labs (LLs), presenting a comparative analysis in relation to their similarities in terms of the interest and feasibility of data, but also specificities of each LL. It highlights the RUSTIK Information System and Core RUSTIK Database as essential tools for local, regional, and policy stakeholders, poised to not only inform ongoing research but also guide future research.

First, an in-depth analysis of LLs' data availability and resources is presented. The chapter explains the methodology used for the assessment and presents the results in relation to LLs' shared interests, and their data resources and capacities. The main objective is to identify common needs in the different LL where collective action and collaboration can be beneficial in addressing data-related challenges, and to understand the significant differences between LL that enrich the data experimentation process.

The second section presents the pivotal role of the Core RUSTIK Database in facilitating comparability and data transferability among the 14 LLs. Detailed insights into economic, social, and environmental indicators, alongside data sources from Deliverables 1.1 and 2.1, underscore its significance in informing decision-making processes.

Dedicated to the Living Lab Databases, the third section delineates these repositories' role. Distinct from the Core RUSTIK Database, they encapsulate locally curated datasets tailored to specific challenges, emphasizing local relevance over direct comparability.

In-depth exploration of the RUSTIK Information System constitutes the fifth section, exploring the functionalities of this user-centric interface, and its capacity to provide relevant insights to stakeholders. This chapter also presents a roadmap of its future evolution plans, including enhanced functionalities and integration with the Rural Observatory.

Concluding the report, the final section outlines the comprehensive guidance offered by the WP2 team to LLs across four critical stages, from orientation to active experimentation with data collection and analysis, and the integration of collected datasets into the RUSTIK System.

Funded by the European Union
UK Research and Innovation

# 2. Assessment of LL data availability and resources

## 2.1. Methodology

An evaluation of Living Lab's needs was conducted by using a two-round online survey, based on the outcomes of deliverables 1.1, 1.2 and 2.1. The purpose of the survey was to gather valuable information regarding the availability of data, resources, and capacities within the Living Labs.

The survey was organized in two rounds to gather knowledge and appreciations on two different, but interlinked, matters that need to be taken into consideration when dealing with data. Feedback from WPs 1 and 3 leads, as well as from the project coordination, were considered before the circulation of both surveys. The two surveys were finally circulated in May and June 2023 respectively, and answered by all 14 Living Labs (LLs).

To ensure a comprehensive understanding of the data being utilized by the different Living Labs, the surveys were designed to be answered in a **collaborative effort between the Pilot Region Partners and the Living Lab Coordinators**. Both parties worked together to complete the surveys, aiming to ensure that they were fully aware of the data they were working with. This cooperative approach helped foster a shared understanding and knowledge between the Pilot Region Partners and the Living Lab Coordinators.

*Table 1: Main purposes of the two-round surveys*

| Survey Round | Purpose |
|---|---|
| First-round | Assess data availability and interests |
| Second-round | Assess data resources and capacities |

First, a **first-round survey** about data availability and interests was conducted, which aimed to gauge the specific interests and preferences of Living Labs regarding data, and to determine to which extent the desired data is available within the Living Labs, as well as to help identify the types of data that are missing.

Next, a **second-round survey** was held with the objective of identifying data resources that are available within the Living Labs, and to evaluate the existing capacities within the Living Labs to handle and analyse the data.

### First-round survey: data availability and interests

The first round of the survey aimed to collect information on Living Lab's data availability and interests. More concretely, the main purposes of the survey were:

1) **Assessing data availability**: the survey aimed to determine the extent to which the desired data is available within the Living Labs. It helps identify the types of data that are accessible and can be utilized for various purposes.
2) **Understanding data interests**: The survey seeks to gauge the specific interests and preferences of Living Labs and stakeholders regarding the data. It helps gather information about the kind of data that is most relevant and valuable to them.

To do so, for each of the RUSTIK transitions, a list of topics was provided, along with examples of indicators. The Living Labs were requested to provide the following information for each topic:

- → Is the topic useful for the Living Lab? Assessment whether the specific topic is relevant or not to the Living Lab.
- → Does the data exist in your Living Lab? Assessment whether there is existing data related to the topic in the Living Lab.
- → Can you access the data you need? Assessment whether the data is accessible by the Living Lab actors and can be used throughout the RUSTIK project.
- → What scale is the data available at? Assessment whether the data is provided in a raster dataset (grid, granular) or in local (LAU) or regional (NUTS 3, 2, or 1) administrative units.
- → Other comments in relation to data collection periodicity, major data gaps, or specific relevant indicators on the topic.

### Socioeconomic transition

Based on D1.1 and D2.1, the following topics were provided for the socioeconomic transition. For each of the topics, a relation of example indicators was provided as a mean for helping the different Living Labs to understand the topic:

*Table 2: Topics for the socioeconomic transition*

| Topic | Examples |
|---|---|
| Population ageing | Total population, density, growth, population by age groups, fertility rate, natural population change, major death causes, life expectancy, ageing index, median age, structural dependency |
| Gender imbalances | Gender Ratio, proportion of women in leadership positions and in political decision-making, gender wage gap, female labour force participation rate, gender-based violence rates |
| External migrations from third countries | Migration, immigration, emigration, nationality, racial and/or ethnic distribution of the population by self-identification |
| Urban-rural migrations | Neo-rurals, emigration to urban areas |
| Social inclusion and cohesion issues | Social mobility, poverty rate, share of the population with different degrees of disabilities and labour dependency, max. level of education reached, quality of life |

| Topic | Examples |
|---|---|
| Access to housing. Housing conditions | % of total area under built-up area, % of total area under urban area, % of total area under infrastructures and industrial use, density of rural settlement, share of primary residences, property type |
| Health services availability | Access to healthcare services, hospital beds per capita, prevalence of smoking, obesity rate, disease-specific mortality rates |
| Schools and education | Access to schools and education centres |
| Public equipment and facilities (sports, cultural, leisure....) | Number of cultural amenities; distance to schools, libraries, sportive facilities |
| Local economic development | Hotels, restaurants, touristic areas, no. of overnight stays in hotels and similar accommodation, in holiday and other short-stay accommodation per 1000 inhabitants, no. of available rooms in accommodation establishments |
| Job opportunities | Employment per sectors, VAB per sectors, Gross Domestic Product, median household income, disposable income |
| Support to innovation | Innovation and support to small and medium-sized businesses |
| Citizen involvement, communitarianism | Citizen engagement, trust in neighbours |
| Institutional Governance | Public trust in Government Institutions |

### Environmental transition

Based on D1.1 and D2.1, the following topics were provided for the environmental transition:

*Table 3: Topics for the environmental transition*

| Topic | Examples |
|---|---|
| Vulnerability/Preparedness to Climate change | Availability of adaption plans, monitoring of CC impacts |
| Air, soil, and water quality / pollution | Water pollution levels, waste generation and disposal rates, groundwater contamination, particulate matter (PM) |

| Topic | Examples |
|---|---|
| | concentration, nitrogen dioxide (NO2) levels, ozone (O3) levels |
| Landscape and cultural heritage conservation | Protected areas coverage |
| Food securing. Concurrence of agriculture with other uses | Abandonment of agricultural fields, change in uses, farmers income, soil quality and properties, crop yield, livestock production, irrigated area, number of ecological production certificates |
| Forest management | Forest management |
| Soil erosion | % of agricultural areas and natural grassland affected by moderate or severe soil-water erosion |
| Soil Imperviousness | Soil imperviousness, soil consumption, total soil organic carbon stocks in topsoil (0-20) |
| Sustainable mobility | Collective passenger transport, shared mobility |
| Energy production (wind, solar, other…) | Production, consumption, renewable share and potential, solar potential |
| Biodiversity, ecosystem services, and natural resources | Biodiversity loss, Biodiversity Index |

### Digital transition

Based on D1.1 and D2.1, the following topics were provided for the digital transition:

*Table 4: Topics for the digital transition.*

| Topic | Examples |
|---|---|
| Digital infrastructure | Different networks coverage |
| Digital skills | Basic skills, ICT specialists |
| Digital transformation of businesses | Unicorns, digital companies |
| Digitalisation of public services | Public services, e-health, digital identity |

| Topic | Examples |
|---|---|
| Vulnerability to digital disruptions | Lack of cybersecurity measures, dependence on outdated technology, inadequate data |

### Second-round survey: resources and capacities

This round of the survey aimed to collect information on Living Labs' resources and capacities in relation to data collection and exploitation. Precisely, the aims of this round of the survey were:

1) **Identifying data resources**: the survey helped identify the specific data resources that are available within the Living Labs. This includes understanding the types of data, sources of data, and the type of processes that are pursued when dealing with data.

2) **Evaluating resource requirements**: the survey assisted in identifying the resources used to manage and utilize the available data in the Living Labs. This includes identifying the necessary infrastructure, technologies, tools, and personnel required to handle the data effectively.

3) **Assessing data capacities**: the survey helped evaluate the existing capacities within the Living Labs to handle and analyse the data. It helped identify any gaps or areas where improvements may be needed to enhance data management and analysis capabilities.

The survey was organized around three different sections:

- → Resources and infrastructure
- → Stakeholders and collaboration
- → Future trends and suggestions for improvement

**Regarding the first section about resources and infrastructure**, the survey focused in whether the Pilot Region is using statistical and/or GIS data, and, in the case they do, which are the organizations that are providing data (subregional providers; regional or national providers; European or global providers; or commercial providers).

Also, this section considered data support for policy implementation and assessment in the Pilot Region, by asking which are the roles that data covers in the Pilot Region (the Pilot Region uses data for the design and implementation of public policies; it uses data for the assessment and monitoring of implemented policies and initiatives; it uses data to influence regional or national policy; it uses data for communication/animation activities addressed to local actors, and other options).

There were some questions specifically addressed to georeferenced data (GIS systems), which aimed to get information about the capacities and infrastructures the Pilot Region has when dealing with this type of data (whether the Pilot Region is capable to collect, store and analyse GIS data; it has a GYS system with a georeferenced database; this GIS system is updated real-time using external webservices; it has specialised staff working with GIS systems, and others).

Moreover, the section considered the Pilot Region's capacities in relation to using GIS visualization platforms. It includes questions related to the level of confidence in their abilities to access and interpret GIS data, and also the functionalities the Pilot Region develops in relation to GIS software to visualize GIS data (whether the Pilot Region uses GIS software to visualize GIS data; if the Pilot Region produces online map viewers for internal or external users, between others).

**In relation to the second section, which is about stakeholders and collaborations**, Pilot Regions were asked to select which capacities the Pilot Region Partner and the Living Lab Coordinator hold in relation to data (data collection; data processing; data analysis; data distribution). Also, they were asked whether there are any existing collaborations or partnerships with external stakeholders to identify data resources and collect data.

**Finally, regarding the third section about future trends and suggestions for improvement**, the survey included questions about the need of support or resources to enhance Pilot Region's data capabilities (such as inventory of available data; accessing to geo-spatial data; connecting geo-spatial to other type of data; processing data; interpreting and analysing available data).

Likewise, some new innovative approaches and technologies are presented, and Pilot Regions are asked whether they are aware of this technology and if they think it could be useful.

This section also considers concerns generated by the Pilot Region or acquired or facilitated from other institutions that cannot be freely distributed for copyright, privacy, or security reasons.

## 2.2. Results on data availability and interests in the LL

### Socioeconomic transition

Overall, topics considered most relevant by the Pilot Regions are **population ageing, job opportunities and citizen involvement**, and **social cohesion**. Also considered as important are the topics of social inclusion and cohesion issues, and support to innovation. On the other side, the least relevant topics for the PRs are **access to housing and housing conditions, health services availability** and **schools and education**.

The Austrian Pilot Region suggests specifically the study of vacancy, affordable housing, and secondary residency working remote, quality of life, and local economic development.

Regarding data existence, there is data in relation to barely all the different topics for at least 8 Pilot Regions. There is a relevant lack of data in topics such as **citizen involvement and social cohesion** (only 3 Pilot Regions know about the existence of data around these topics), **support to innovation** (4 Pilot Regions, in this case) and **institutional governance** (6 Pilot Regions). Regarding access to this data, almost for all the topics data is also accessible for the PRs. There seems to be few cases in which data exists, but it is not available and accessible for the Pilot Regions.

When it comes to the data granularity, in general terms the spatial level of detail for most **Pilot Regions data corresponds to the LAU level**. Population ageing, local economic development, and external migrations from third countries are the topics about which more PRs have data at

municipality level. Nonetheless, a significant portion of the topics below NUTS3 suffer from limited data availability in the different PRs. This problem is particularly important for topics such as gender imbalances, urban-rural migrations, and institutional governance.

However, there is a crucial issue in the comparability of data between PRs. This is a big inconvenience towards the work of WP1 on the definition of Functional Rural Areas. The quantity of data available at least LAU level for representing the various transitions and functions is limited, but it still offers possibilities for analysis. Additionally, there is still the possibility of harmonising local datasets produced similarly in the different territories (collected with similar methods, produced on census years…). Despite this inconvenience, it is something that does not directly affect the Data Experiments of the Living Labs, as they are based purely on the information available at a Local Scale and they don't depend on other Living Labs data.

*Table 5: Results of the survey in relation to the socioeconomic transition. The count corresponds to the number of Pilot Regions that have answered positively to each of the questions in relation to the topics.*

| | Useful for the LL? | | | | | Data exists? | | Is accessible? | | At what scale? | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Topic | Very important | Important | Average | Less important | Not relevant | Yes | No | Yes | No | Grid | LAU 2 | NUTS 3 | NUTS 2 | NUTS 1 |
| Population ageing | 9 | 4 | 0 | 0 | 0 | 13 | 0 | 13 | 0 | 1 | 10 | 1 | 0 | 0 |
| Gender imbalances | 2 | 7 | 3 | 1 | 0 | 9 | 4 | 10 | 2 | 1 | 5 | 2 | 2 | 1 |
| External migrations from third countries | 4 | 5 | 3 | 1 | 0 | 12 | 1 | 12 | 0 | 0 | 9 | 2 | 0 | 0 |
| Urban-rural migrations | 6 | 2 | 3 | 2 | 0 | 9 | 4 | 8 | 3 | 0 | 5 | 1 | 2 | 1 |
| Social inclusion and cohesion issues | 5 | 5 | 1 | 3 | 0 | 9 | 5 | 10 | 3 | 0 | 7 | 2 | 1 | 2 |
| Access to housing. Housing conditions | 3 | 4 | 3 | 3 | 0 | 11 | 2 | 11 | 0 | 2 | 7 | 1 | 0 | 0 |
| Health services availability | 2 | 5 | 5 | 0 | 1 | 10 | 2 | 12 | 0 | 0 | 5 | 7 | 0 | 0 |
| Schools and education | 4 | 5 | 1 | 3 | 0 | 11 | 2 | 12 | 0 | 0 | 8 | 3 | 0 | 0 |
| Public equipments and facilities (sports, cultural, leisure….) | 6 | 3 | 3 | 1 | 0 | 9 | 4 | 10 | 2 | 2 | 6 | 1 | 0 | 1 |
| Local economic development | 6 | 2 | 4 | 1 | 0 | 12 | 1 | 12 | 0 | 0 | 10 | 1 | 1 | 0 |
| Job opportunities | 7 | 4 | 2 | 1 | 0 | 12 | 2 | 12 | 1 | 0 | 6 | 5 | 2 | 0 |
| Support to innovation | 5 | 5 | 3 | 1 | 0 | 6 | 7 | 4 | 6 | 0 | 1 | 3 | 2 | 1 |
| Citizen involvment, communitarianism | 6 | 6 | 1 | 1 | 0 | 3 | 10 | 4 | 7 | 0 | 2 | 2 | 2 | 1 |
| Institutional Governance | 4 | 3 | 6 | 0 | 0 | 6 | 6 | 6 | 4 | 0 | 0 | 1 | 1 | 5 |

Regarding population ageing, the basic demographic data is accessible by all, but there is a dependency on census data on more complex indicators. The same happens with the gender imbalances, with gender ratio being accessible but with most PRs facing problems on different indicators.

Regarding migrations, external migration data is accessible for most of the Living Labs, whereas the only data that exists for urban-rural migrations is provided at bigger scales.

Social inclusions and cohesion issues are not well covered, with unequal availability in the different Pilot Regions. The number of datasets at LAU level is reduced, being mainly centred on

the topics of education. 2021 Census appears as an opportunity to solve this lack of data availability. Regarding access to housing data, there is a significant number of datasets available at LAU level.

Concerning health and education, the amount of available data is reduced. It is rather easy to get services of general interest data, being provided by the Rural Observatory at LAU level and Open Street Map (OSM from now on) appearing as a reliable source. Services of General interest (SGI) are basic services which are essential to the lives of most of the general public and where the state has an obligation to ensure public standards. The measurement of their presence is fundamental for assessing quality of life in rural areas.

As regards to local economic development, OSM comes across as a primary data source. Rural Observatory also provides data on the number of accommodation in tourist rooms per LAU. Some touristic institutes in the regions have data on this regard, acquired by various means.

Relating to job opportunities, some PRs have employment per sector data at LAU level. GDP, VAB and other macroeconomic indicators are available only at higher spatial levels for most of the Pilot Regions. Monmouthshire can access census and third party data at postcode scale. In this sense, 2021 census will definitely be a source to take into account.

Regarding support to innovation, Austria has some data available in STATCube. North Karelia suggests that some data is available at NUTS3 level. Monmouthshire is involved in a project that could provide them some data in this regard on the long term. Most of the PRs state they have major data gaps.

In respect of citizen involvement and communitarianism, there is available data in Austria at NUTS2 coming from Voluntary Work (year 2022), with information regarding the voting activity and number of associations and NGOs (ratio per inhabitants). In Poland, NGOs data (number of organisations, type, membership...) is available at NUTS 2 level.

As for institutional governance, in most countries this kind of data is not available at regional level. The OECD Trust Survey provides some data in this sense. Some Pilot Regions suggest considering participation in electoral processes. In Wales, some local information might be used to gather data in this topic.

### Environmental transition

There is some dispersion on the relevance assigned to the environmental transition topics. The different Pilot Regions have given a different degree of relevance to the topics, according to their specific characteristics and needs. Overall, the Pilot Regions highlighted **sustainable mobility, vulnerability/preparedness to climate change, landscape and cultural heritage conservation, and food securing** (and the concurrence of agriculture with other uses and energy production (wind, solar, other...)) as the most important topics. The Pilot Regions do not show a big interest to soil imperviousness, soil erosion or pollution.

The need for studying food loss and food waste has been marked out by the Central Slovenia Pilot Region.

The available data regarding this transition is **reduced when comparing to the socioeconomic transition**. Landscape and cultural heritage conservation data is accessible by nearly all Pilot Regions, and most can access data regarding pollution, food securing, forest management and energy production. Notwithstanding, really few data are available for soil erosion, soil imperviousness or sustainable mobility.

Regarding granularity, landscape and cultural heritage conservation data is the most complete one for this granularity levels (grid and LAU) and still some PRs can access data on food securing and concurrence of agriculture with other uses at LAU level. On the other side, the poorest datasets belong to soil erosion, soil imperviousness, sustainable mobility, and energy production data.

*Table 6: Results of the survey in relation to the environmental transition. The count corresponds to the number of Pilot Regions that have answered positively to each of the questions in relation to the topics.*

| | Useful for the LL? | | | | | Data exists? | | Is accessible? | | At what scale? | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Topic | Very important | Important | Average | Less important | Not relevant | Yes | No | Yes | No | Grid | LAU 2 | NUTS 3 | NUTS 2 | NUTS 1 |
| Vulnerability/Preparedness to Climate change | 3 | 5 | 3 | 1 | 2 | 7 | 7 | 6 | 3 | 0 | 3 | 3 | 1 | 1 |
| Air, soil and water quality / pollution | 2 | 3 | 3 | 4 | 1 | 8 | 5 | 8 | 0 | 2 | 2 | 3 | 0 | 1 |
| Landscape and cultural heritage conservation | 4 | 4 | 3 | 1 | 1 | 13 | 0 | 12 | 1 | 4 | 5 | 3 | 1 | 0 |
| Food securing. Concurrence of agricuture with other uses | 4 | 4 | 2 | 1 | 2 | 8 | 4 | 9 | 0 | 0 | 6 | 1 | 0 | 2 |
| Forest management | 4 | 3 | 2 | 0 | 4 | 8 | 4 | 8 | 1 | 1 | 3 | 2 | 0 | 2 |
| Soil erosion | 2 | 2 | 2 | 3 | 4 | 4 | 8 | 4 | 2 | 3 | 1 | 1 | 1 | 0 |
| Soil Imperviousness | 1 | 2 | 2 | 4 | 4 | 3 | 9 | 4 | 2 | 1 | 2 | 0 | 2 | 0 |
| Sustainable mobility | 3 | 7 | 2 | 1 | 0 | 4 | 8 | 4 | 3 | 0 | 1 | 3 | 0 | 1 |
| Energy production (wind, solar, other…) | 3 | 4 | 3 | 2 | 2 | 8 | 5 | 7 | 3 | 0 | 1 | 3 | 2 | 3 |
| Biodiversity, ecosystem services, and natural | 4 | 2 | 4 | 1 | 2 | 5 | 8 | 2 | 4 | 2 | 1 | 1 | 1 | 1 |

In terms of susceptibility and readiness for climate change impacts, most of the available data is only available at the regional level. The primary sources of data stem from greenhouse gas emissions records and strategic action outlines. Additionally, meteorological data is accessible for many regions. Similarly, data concerning air, soil, and water quality, as well as pollution levels, is primarily accessible on broader regional scales.

Contrary, it appears that high-quality data is readily accessible for the preservation of landscape and cultural heritage across most of the relevant regions.

Concerning agriculture, only fundamental farm statistics are accessible at the LAU level for the majority of areas. As for forest management, there is scarce information available beyond the delimitation of forested areas in most of the pertinent regions. Moreover, data on soil imperviousness and erosion is severely limited for the majority.

Regarding sustainable mobility, limited data on public transportation is available, and not much besides that. On topics such as energy production and biodiversity, ecosystem services, and natural resources, the data is very limited.

Regarding sustainable mobility, there is a dearth of available data, primarily centred on public transportation and lacking information in relation to other modes of transport and areas related to sustainable transport. Following the same trend, when it comes to subjects like energy generation, biodiversity, ecosystem services, and natural resources, the existing data is notably constrained.

### Digital transition

The emphasis on the digital transition primarily revolves around **digital infrastructure**, with lesser significance attributed to digital skills, the transformation of businesses through digital means, and the digitization of public services. Lastly, there is notably less attention directed towards assessing vulnerability to digital disruptions.

On data existence, most of the PRs can access digital infrastructure data, but on the rest of the topics more problems are encountered. This data is accessible by the PRs and 6 can access it at a scale as small as LAU. No other datasets are suggested in this domain.

Regarding the availability of data, the majority of Pilot Regions know about digital infrastructure data, but the accessibility to the rest of the topics' data is much reduced.

In general terms, this data is accessible by the Pilot Regions, with six of them capable of accessing it even at the detailed LAU level. No additional datasets are proposed within this domain.

*Table 7: Results of the survey in relation to the digital transition. The count corresponds to the number of Pilot Regions that have answered positively to each of the questions in relation to the topics.*

| Topic | Useful for the LL? | | | | | Data exists? | | Is accessible? | | At what scale? | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Very important | Important | Average | Less important | Not relevant | Yes | No | Yes | No | Grid | LAU 2 | NUTS 3 | NUTS 2 | NUTS 1 |
| Digital infrastructure | 7 | 4 | 1 | 1 | 0 | 10 | 3 | 10 | 1 | 1 | 6 | 0 | 2 | 0 |
| Digital skills | 3 | 4 | 4 | 1 | 1 | 6 | 7 | 5 | 3 | 0 | 1 | 1 | 2 | 2 |
| Digital transformation of businesses | 1 | 5 | 5 | 1 | 1 | 5 | 7 | 4 | 3 | 0 | 3 | 0 | 2 | 1 |
| Digitalisation of public services | 2 | 5 | 5 | 0 | 1 | 5 | 7 | 6 | 2 | 0 | 2 | 0 | 1 | 3 |
| Vulnerability to digital disruptions | 1 | 2 | 5 | 1 | 3 | 2 | 10 | 1 | 4 | 0 | 1 | 0 | 1 | 1 |

In terms of digital infrastructure, data is available and accessible for most regions at the LAU level through the Rural Observatory or even more granular scales in certain countries. Some pertinent regions have also recognized crowd-sourced initiatives like Nperf, a platform that provides network coverage data collected by its users (but that is not open-sourced).

Concerning the other subjects, as affirmed by the majority of Pilot Regions, data is predominantly accessible at the national level or at larger regional scales like NUTS 1 or 2, surpassing the size of most Pilot Regions.

## 2.3. Living Lab's shared interests

The initial purpose of the first survey was to derive insights into the interests of the Living Labs. Four clusters emerged concerning the socio-economic (red) and environmental (green) transitions, while three clusters emerged for the digital (yellow) transition (refer to Figure 1).

Building upon the findings of survey 1, a cluster analysis was performed using the K-means method in Python. The interest levels in various topics were assigned values from -2 to 2. The resulting clusters underwent manual verification to ensure the coherence of the analysis. The only alteration made was the consolidation of the clusters related to digital transitions, given their negligible differences.

This clustering is not a final resolution but rather a contribution from WP2 to the ongoing efforts of LLs clustering led by WP3. It may have some inconsistences to the data needs stated in latter chapters, as with the first cycle advance, the priorities of the Living Labs have changed, the challenges are more defined, and with them their data needs.

**SOCIO-ECONOMIC TRANSITION**

**Strong interest in social and population components**

Troyan-Apriltsi-Ougarchin (BG)
Garfagnana (IT)
Woj. Mazowieckie (PL)
Galicia (ES)

**Strong interest in social welfare and economic development**

North Karelia (FI)
Rhine-Hunsrück (DE)
Zaječar District (RS)
Sant Miquel de Balenyà – Osona (ES)
Monmouthshire (UK)

**Interest in population, employment and economic development**

Nockregion – Oberkärnten (AT)
Parma and Piacenza (IT)
Woj. Świętokrzyskie (PL)

**Reduced interest in the socio-economic transition**

Osrednjeslovenska regija (SI)
Gloucestershire (UK)

**ENVIRONMENTAL TRANSITION**

**Strong interest in all environmental transition topics**

Garfagnana (IT)
Zaječar District (RS)
Sant Miquel de Balenyà – Osona (ES)
Monmouthshire (UK)

**General interest on the transition, but reduced on soil topics**

Parma and Piacenza (IT)
Woj. Mazowieckie (PL)
Osrednjeslovenska regija (SI)
Gloucestershire (UK)

**Strong interest on energy and sustainability but reduced on soils**

Nockregion – Oberkärnten (AT)
Troyan-Apriltsi-Ougarchin (BG)
Galicia (ES)

**Reduced interest in the environmental transition**

North Karelia (FI)
Rhine-Hunsrück (DE)
Woj. Świętokrzyskie (PL)

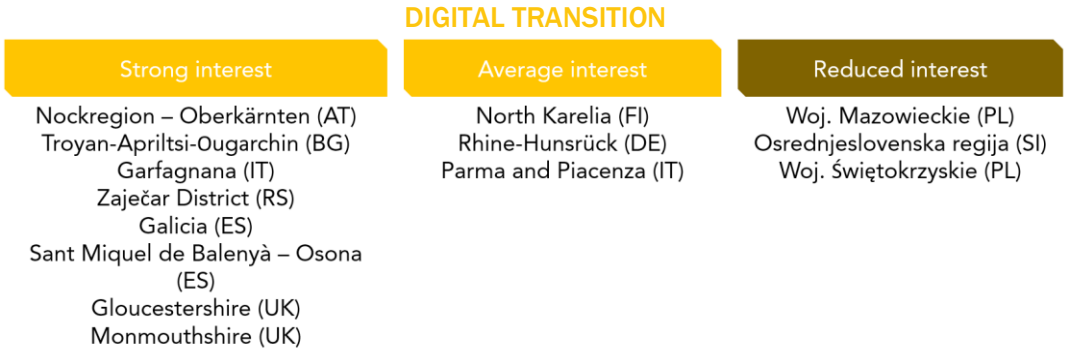


*Figure 1: Living Lab's clustering for each of the transitions based on their data interests, owing to the 1st survey.*

## 2.4. Results on data resources and capacities in the LL

The subsequent findings stem from the second survey conducted in June 2023, encompassing all the Living Labs. This survey assessed various topics concerning resources and capabilities through a series of questions outlined below, along with their respective outcomes.

Regarding the sources of data utilized by the Pilot Regions (refer to Figure 2), it was found that all 14 PRs depend on data generated by public institutions at either national or regional levels. Furthermore, 10 of these regions integrate data from sub-regional entities, while 8 regions extend their data sources to encompass European or global data providers. Only 6 of the Pilot Regions make use of data originating from commercial providers.

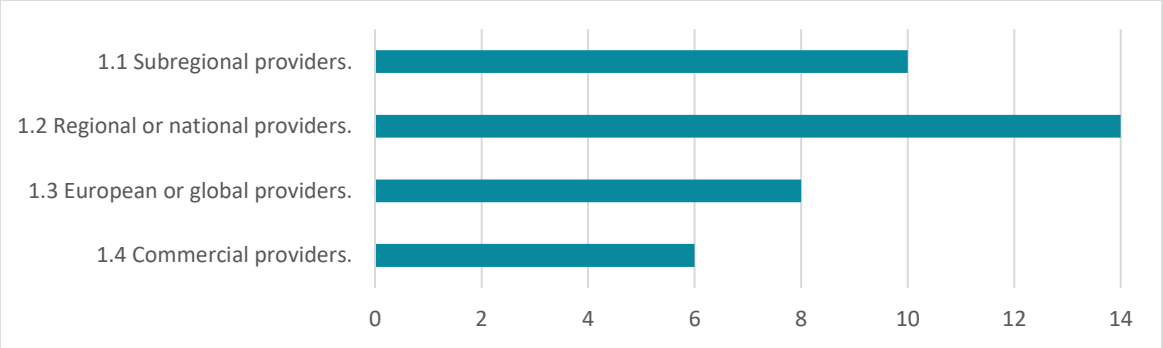


*Figure 2: Organisation providing statistical and/or GIS data to the Pilot Region.*

Concerning the use of data to support policy implementation and evaluation (refer to Figure 3), 11 of the Pilot Regions utilize data for formulating and executing public policies, assessing the effectiveness of implemented policies, and engaging with policymakers. Additionally, 10 of these regions utilize data for communication purposes with local stakeholders. However, two Pilot Regions specify that they are not involved in the design, execution, or evaluation of public policies.

Moving forward, the survey explores the application of data in policy design and evaluation. Key actions related to data include analysing both raw data and data collected by the Pilot Regions. Furthermore, a prevalent practice involves utilizing statistical data obtained from official national repositories. Notably, 3 PR actors integrate data derived from scientific research into their processes.

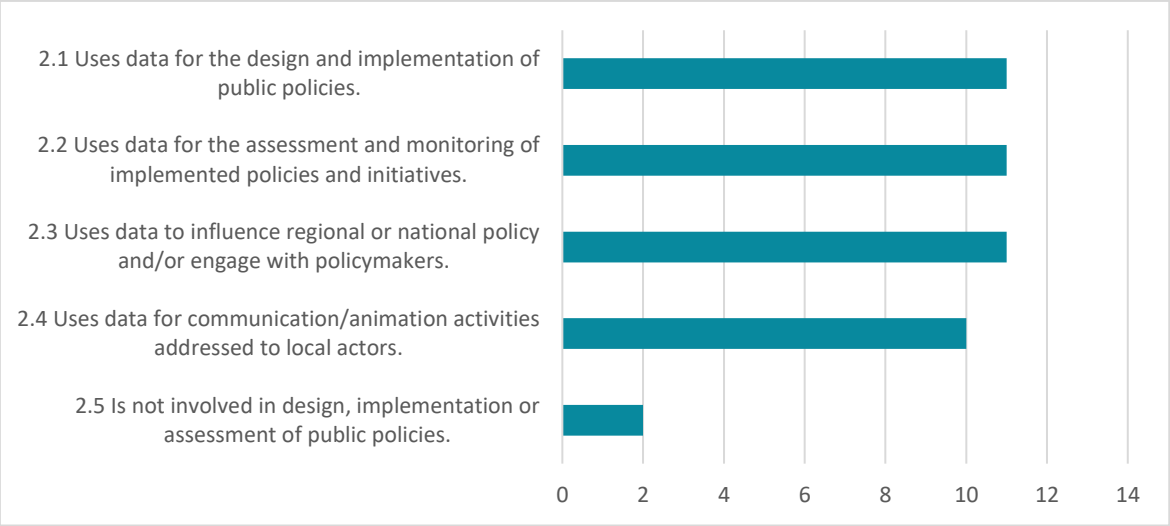


*Figure 3: Statements applying to the Pilot Regions in relation to data support for policy implementation and assessment.*

Data serves various purposes across the PRs: 6 use it for policy design, while an additional 6 employ it to assess implemented policies. Less commonly, it's used to influence policy development and support communication and engagement efforts.

Regarding the adoption of Geographic Information Systems (GIS) in the Pilot Regions (refer to Figure 4), 11 regions exhibit confidence in their ability to collect, store, and analyse GIS data. Among these, 8 already possess a GIS system equipped with a georeferenced database. Furthermore, 10 regions opt for open-source GIS systems. Conversely, 3 Pilot Regions lack an internal system and either depend on external alternatives or operate without one. Notably, 7 regions benefit from specialized personnel proficient in working with GIS systems. Additionally, one of the Pilot Regions utilizes external services to maintain real-time updates to their data.

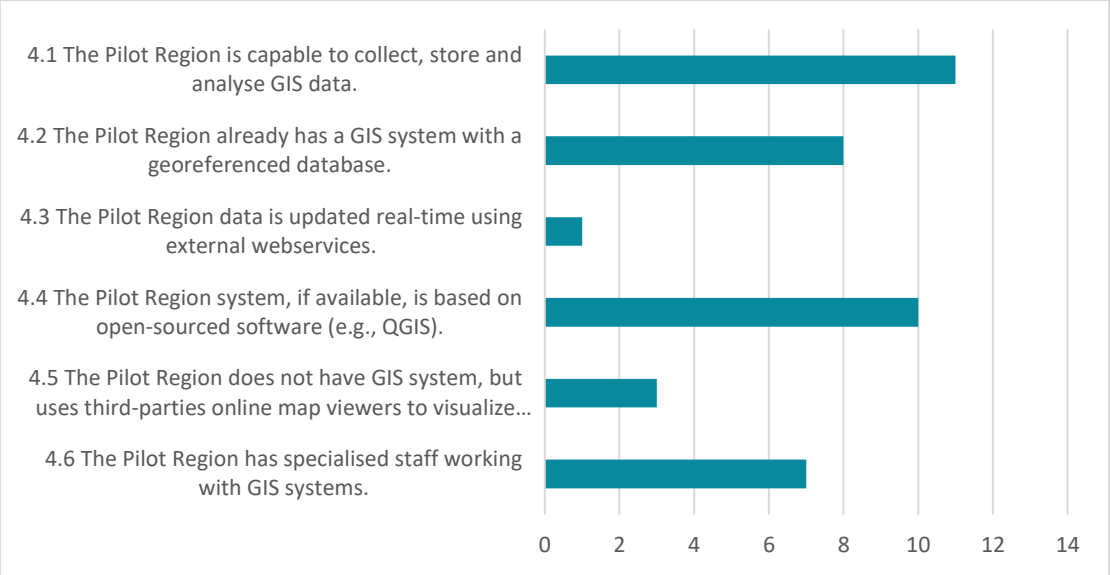


*Figure 4: Pilot Region capabilities in relation to working with georeferenced data (GIS systems).*

In terms of GIS data visualization (refer to Figure 5), 10 PRs utilize GIS software like QGIS, ArcGIS, or GeoMedia for visualizing GIS data. However, only 7 PRs feel confident in their ability to access and interpret GIS data. Among these, 5 have specialized staff capable of handling web GIS (WMS, PostGIS). Additionally, only 3 PRs have developed an online map viewer, while 7 collaborate with external organizations to bolster their data resources and infrastructure.

Regarding the use of GIS data and visualization platforms and the associated benefits across the Pilot Regions, responses vary. Key advantages highlighted include increased transparency, public access, and enhanced accountability. The utilized platforms encompass a wide range, including visualization platforms, GIS software, data collection tools, and gazetteers. Notably, two Pilot Regions have in-house GIS specialists, while two others either outsource this expertise or receive external support for their GIS services.

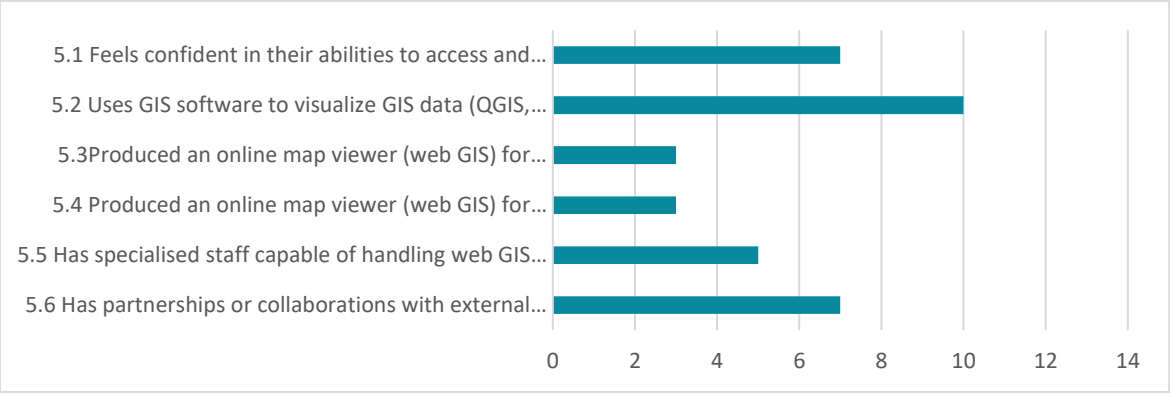


*Figure 5: Statements applying to the Pilot Regions in relation to the use GIS visualization platforms.*

In terms of the capabilities of the PR partners (figure 6), 13 out of the 14 are capable of data collection, while 11 of those can also handle data processing. All 14 PRs express confidence in their ability to analyse data, but only 9 show enthusiasm for data distribution. Similarly, among the 14 Living Labs Coordinators surveyed, all are equipped for data collection and processing, while only 8 possess the capacity for data analysis. Merely one of them is additionally engaged in data distribution. The results are shown separately for Pilot Region Partners and Living Lab Coordinators to assess the level of expertise of each of the parts in the Living Lab.

Regarding the capabilities of PR partners (refer to Figure 6), 13 out of 14 are proficient in data collection, with 11 also adept at data processing. All 14 PRs express confidence in their data analysis abilities, but only 9 show a strong inclination toward data distribution. Among the 14 surveyed Living Labs Coordinators, all are equipped for data collection and processing, while only 8 possess the capacity for data analysis. Additionally, only one is engaged in data distribution. These results are separately presented for Pilot Region Partners and Living Lab Coordinators to evaluate their expertise within the Living Lab framework.

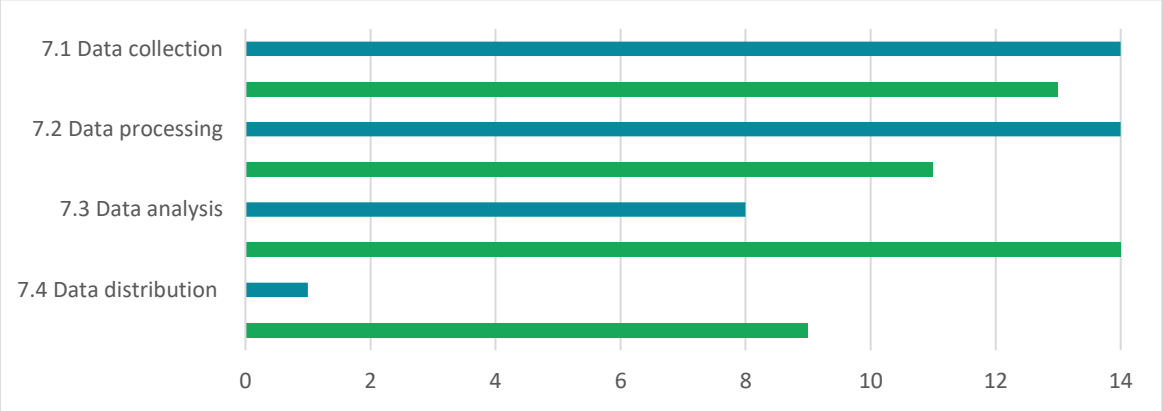


*Figure 6: Capabilities held by Pilot Region Partners (blue) and Living Lab Coordinators (green) in relation to data.*

Among the Pilot Regions, 10 have established collaborations or partnerships with external stakeholders. Typically, these partnerships involve a diverse array of actors, such as public administrations, non-profit organizations, universities, private companies, and research partners. In 5 instances, these cooperative efforts are ongoing and permanent.

Regarding the augmentation of data capabilities within the Pilot Regions, several areas are identified as weaknesses. The table below highlights the most prevalent ones:

*Table 8: Key areas where training is needed in the Pilot Regions.*

| Key areas where training is needed in the PRs | Count |
|---|---|
| Inventory of available data | 5 |
| Interpretation and analysis of available data | 5 |
| Access to geo-spatial data / common data platforms | 4 |

| Key areas where training is needed in the PRs | Count |
|---|---|
| Data collection | 3 |
| Linking geo-spatial data to other types of data | 3 |
| Data processing | 3 |
| Data visualisation | 2 |
| Involvement of private/public data owners | 2 |

Regarding the awareness of innovative data collection approaches (refer to Figure 7), the findings are as follows: 9 PRs are aware of web scraping, with 7 expressing confidence in its potential use within the project; PPGIS is familiar to 10 PRs, among whom 11 find it interesting for implementation; and satellite imagery is known to 10 PRs, with 8 considering it suitable for the project. Inquiring about data-cube technology, 5 out of the 14 PRs are familiar with it, and 4 find it applicable to the project.

Upon soliciting additional data collection techniques, the Slovenian Pilot Region proposed tracking data from FoodWaste prevention applications like TooGoodToGo and PriHrani. Conversely, the Gloucestershire Pilot Region suggested leveraging satellite imagery combined with AI models to interpret extensive datasets.

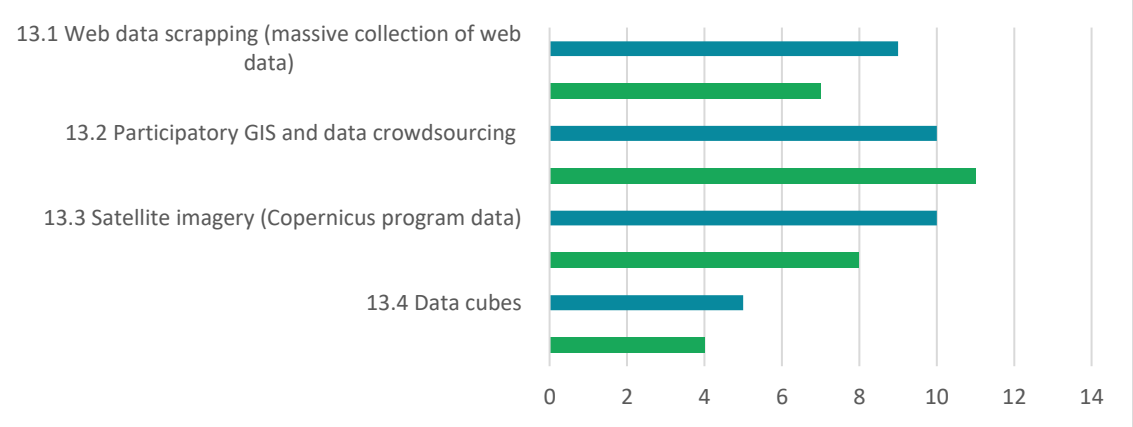


*Figure 7: Pilot Regions' familiarity with selected innovative approaches and technologies for data collection (blue) and assessment of their suitability for the project (green).*

The final inquiry pertains to data security, privacy, and distribution concerns. Half of the PRs express apprehensions in this domain. The predominant issue revolves around data granularity, necessitating the sharing of highly precise datasets in an aggregated manner. Additionally, there are specific concerns related to GDPR compliance for certain datasets, ongoing studies involving multiple institutions, and privately acquired datasets by the PRs that cannot be freely distributed.

# 3. Core RUSTIK database

The Core RUSTIK Database originates from the initial analyses in Deliverables 1.1 and 2.1. The first set of indicators, suggested in the WP1 deliverable, underwent an expansion in the subsequent deliverables.

Functioning as a repository of datasets, the Core RUSTIK Database facilitates data comparability and transferability among the 14 Living Labs. Its primary objectives include supporting the redefinition of Functional Rural Areas led by WP1, furnishing detailed insights at a granular level, and aiding in the analysis of transition challenges across PRs. As a result, the Living Labs stand as the primary beneficiaries of this database, utilizing the information layers as valuable inputs for their Data Experiments.

Comprehensively, this database describes and monitors various regions from economic, social, and environmental perspectives. It encompasses economic aspects like sectors, employment, and income; social dimensions such as quality of life and well-being; and environmental factors covering climate mitigation, adaptation, and energy.

Continual enhancements to the database occur as new pertinent data becomes accessible. The current composition of the database is detailed in ANNEX 3. Anticipated additions include datasets identified by the Universitat Autònoma de Barcelona (UAB) team, responsible for European-scale data collection within the project. Additionally, forthcoming data from 2021 census, spanning from LAU level to small grid resolutions (1 to 10 km), will be incorporated once the data is released (expected for March 2024).

## 3.1. Data collection

There are several data sources and types that have been explored in order to overcome the data gaps identified in the previous steps.

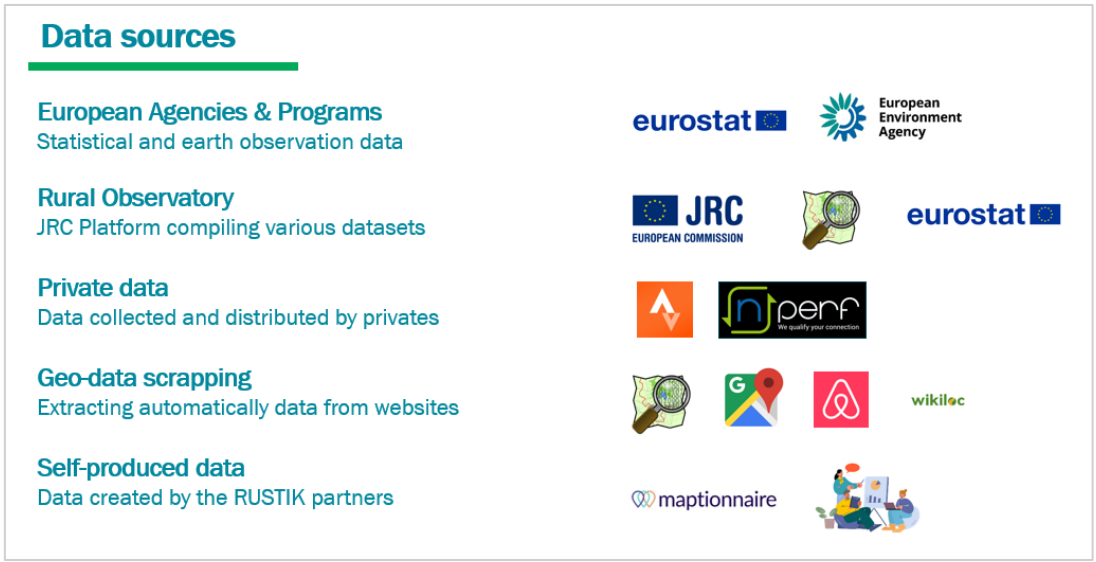


*Figure 8: Data sources and types of the RUSTIK database.*

The different data collection methods and sources are explored in this section:

### Data from European agencies and programs

These entities serve as robust repositories of diverse information crucial for our analysis and understanding. Within this category, a multitude of agencies and programs contribute a spectrum of data, ranging from demographic statistics to economic indicators and environmental metrics. The most used are Eurostat and the European Environmental Agency.

Programs initiated by the European Commission play a pivotal role in augmenting the available data. Initiatives focusing on research, innovation, and social development generate invaluable datasets, and an example for this are the ESPON projects. The data from the Rural Observatory, led by the JRC, has received special treatment within our project owing to their efforts to include datasets at LAU level.

### Data from private actors

Private sector is an open possibility for gathering data beyond what governments and institutions offer. Companies, NGOs, and research institutions create unique datasets capturing specific areas, like consumer behaviour or specialized research. Collaborating with these private entities provides crucial insights for shaping policies and business strategies. These datasets complement broader public analyses, adding depth and specialization. Working with private data demands attention to ownership, privacy laws, and proprietary constraints.

Remarkable examples in the domain can be: Strava, an internet service for tracking physical exercise which incorporates social network features, and that allows the analysis of its data or Nperf, a platform that provides network coverage data collected by its users.

### Geo-data scraping

Geo-data scraping automates the extraction of geospatial details from online sources like websites, databases, and APIs. This method gathers geographically referenced data from various places. By scraping this publicly available geo-data, we build extensive spatial databases, which offer powerful insights into urban growth, environmental changes, infrastructure planning, and more.

Commercial platforms and niche websites often host geospatial datasets tailored to specific fields. Scraping this data enriches public datasets, providing a broader perspective for diverse applications. Challenges arise with geo-data scraping concerning data accuracy, quality, and ethical considerations. Different methods have been explored, including gathering data from Open Street Map and Google Maps, which serve as a foundational base for various analyses in Living Labs. This list of sources can expand to the Living Lab Databases, based on specific Living Lab needs.

### Open Street Map

Open Street Map (OSM) is a collaborative project that creates a free, editable map of the world. It's a platform where individuals contribute geographic data, much like a Wikipedia for maps. Users, including volunteers and organizations, add and edit information about roads, trails, cafes, buildings...

OSM data is easy to scrap and provides a powerful sample of data. Various types of geospatial data can be scraped from OSM, the more relevant and complete are the following: Natural features (rivers, lakes, parks…), Points of Interest (locations of businesses, shops, restaurants, hospitals, schools…), addresses and buildings (building outlines or house numbers), and transportation: data related to infrastructure (roads, highways, footpaths…) and to public transportation routes, stops, and schedules.

### Google Maps

Google Maps is a widely used online mapping service provided by Google. It offers maps, satellite imagery, street views, and route planning for users worldwide. From this service, the Points of Interest can be extracted, with more features than the OSM ones, but with more concerns regarding data privacy. The data available on Google Maps includes information about business concurrence, reviews, images and related searches.

# 4. Living Lab Databases

The Living Lab Databases serve as the foundation for the Data Experiments conducted within the Living Labs. They consist of two types of datasets: those inherited from the Core RUSTIK Database (CRD) with a European scope, and datasets specifically gathered and produced for individual Living Labs.

The composition of these databases varies significantly across the 14 different contexts. This diversity is rooted in the specificity of the addressed topics, with each Living Lab focusing on a distinct challenge, taking into account the particularities of national or regional contexts.

Unlike prioritizing comparability among Living Labs, the primary goal of Living Lab databases is to furnish pertinent information for local Data Experiments. This approach ensures relevance to local contexts and challenges without mandating strict comparability across Living Labs.

## 4.1. Data needs

This section is derived from the First Living Lab Report, submitted as a draft by the Living Labs on November 20th. An initial overview of their needs is presented next, followed by a preliminary analysis in the subsequent table.

| Living Lab | Challenge | Data Needed | Data Availability | Limitations | Capacities |
|---|---|---|---|---|---|
| Galicia | Reconciling land use and ownership | Land ownership, housing data, existing policy instruments | Detailed land ownership data, model settlement insights | Model settlement data may lack full representativeness, qualitative data approach | Quantitative and qualitative analysis, additional support for interviews |
| Garfagnana - Montagnappennino | Sustainable use of forest for productive and environmental purposes + Role of community projects | Social capital, quality of relationships, trust between people, reciprocity... (qualitative) | Lack of up-to-date population and industry census data on a municipal basis | No limitations identified up to this moment. | Agreement with University of Pisa to handle data. Sufficient skills in the LAG team |
| Gloucestershire | Development of a resource and a strategy for enhanced digital inclusion | Assessing rural community digital access and confidence. | Comprehensive, integrated data amidst diverse sources | Not enough hardware to handle a data lake. | Limited IT expertise of the LLC team, but strong PRP in the field. |
| Mazowieckie voivodeship - Szydłowiecki powiat | Transitioning economy | Detailed employment structures, local resources analysis, entrepreneurial sector insights | Gaps in data at LAU 1 & 2 levels, limited qualitative data | Insufficient data at specific geographical levels, especially in qualitative aspects | Strong partnership building, digital competencies, need for expertise |
| Monmouthshire | Demographic change and youth retention | Reasons for youth migration, housing and educational/employment data, income limitations | Demographic, educational, housing market data | Lack of local business recruitment data, limited sharing between departments | Data analysis team, GIS expertise, potential for departmental collaboration |
| Nockregion-Oberkärnten | Identifying needs of Small Rural Businesses | Demographics, economic status, challenges faced by SRBs, employment data | Statistical data available, spatial data limited, | Spatial view of economic activities uncertain, gaps in qualitative data | Socio-economic research expertise, geodata analysis, |

| Living Lab | Challenge | Data Needed | Data Availability | Limitations | Capacities |
|---|---|---|---|---|---|
| | | | qualitative data uncertain | | OpenStreetMap database |
| North Karelia | Integration of immigrants in North Karelia | Population trends, immigrant background, job numbers, integration statuses | National quantitative data sources, local government reports | Difficulty obtaining immigrant experience data, language barriers | National data source access, surveys, potential support for survey setup |
| Osrednjeslovenska regija | Excess food management and food access in rural areas | Real-time excess food data, social exclusion stats, subjective wellbeing metrics | Lack of current & real-time data on excess food and social exclusion | Scarce real-time data, privacy concerns, lack of poverty-related information | Exploring software solutions, collaborating with institutions, respecting privacy concerns |
| Parma and Piacenza | Water availability and management for irrigation and production | Real-time data for water availability and demand | Much data already available | Data is not analysed at this stage and is not useful for management | IBO has skills and access to the data |
| Rhein-Hunsrück | Making the region an attractive place to live and work for young people | Real-time data on people seeking jobs and apprenticeships and data on open positions. Qualitative surveying | Available at NUTS 2/3 level, and thus not suitable for a local analysis | Less experience handling qualitative data | Skills to analyse qualitative data and surveys. Alliances with a company in case of big data |

| Living Lab | Challenge | Data Needed | Data Availability | Limitations | Capacities |
|---|---|---|---|---|---|
| | (unfilled jobs and apprenticeships) | | | | |
| Sant Miquel de Balenyà | Improving quality of life via territorial and urban planning | Demographics, surveys, environmental, infrastructure data | Scarcity of data at EMD scale, lack of comprehensive big data | Insufficient granularity at the EMD scale, limited big data | Foundational knowledge, fieldwork, adaptability to unconventional sources |
| Świętokrzyskie | Reversing demographic decline through rural tourism development | Insights on spatial distribution, tourist dynamics, rural area conditions | Limited, scattered data on rural tourism and demographic specifics | Scarcity of comprehensive data, especially on rural tourism and demographic trends | Compensating data gaps through research methods, survey capacity |
| TAU – Troyan-Apriltsi-Ugarchin | Strengthening the Rural Food System | Local population access, farming cooperation, vocational schools, visitor habits | Limited access to certain registers, absence of a family gardens register | Challenges accessing registers and gaps in data regarding family gardens | Expertise in source identification, data collection, processing, potential need for more expertise |
| Zaječar district | Development of tourism and short food supply chains | Data on local food system participants, consumer preferences, tourism relevance | Limited documentation and datasets regarding food | Insufficient data on tourist demands, local food system capacities | Access to national statistical databases, expertise in research design |

| Living Lab | Challenge | Data Needed | Data Availability | Limitations | Capacities |
|---|---|---|---|---|---|
| | | | system participants | | |

## 4.2. Data needs analysis

Some general remarks can be drawn from the analysis of the Living Labs data needs, on the analysed topics of data needs, availability, limitations and capacities:

**Data needs**: There is a general need of contextual data and a shared interest on human wellbeing. Of course, there is a huge dependence on the challenge addressed and the data experiment.

- Local Context: Detailed specifics about local conditions, demographics, employment, and economic structures.
- Quality of Life Indicators: Metrics on subjective wellbeing, social exclusion, and demographic trends.
- Sector-Specific Insights: Data on food systems, tourism, land use, and business needs.

**Data Availability**: Challenges persist in obtaining real-time, comprehensive, and ready to analyse data, with specific registers and qualitative aspects often being restricted or absent.

- Scarcity of up-to-date Data: Most regions lack real-time data, this lack of data is severe in topics such as food surplus, social exclusion or demographic shifts.
- Fragmented and Limited Data: Datasets are scattered, often lacking specificity, and are sometimes non-machine-readable.
- Specific Registers and Indicators: Registers such as farmers' data, family gardens, and qualitative aspects are either restricted or absent.

**Limitations**: Privacy concerns, data sensitivity, and gaps in finer geographical or qualitative data levels pose significant barriers.

- Privacy and Sensitivity: Challenges in accessing poverty-related information due to privacy concerns and data sensitivity.
- Data Gaps at Low Levels: Lack of comprehensive data at finer geographical levels.
- Insufficient Real-time Data: Issues with accessing current and real-time information on excess food, social exclusion, and specific demographic trends.

**Capacities**: Despite challenges, strengths lie in partnership building, adaptability to explore unconventional data sources, and expertise in data collection; however, identified gaps necessitate additional expertise to bridge existing data limitations.

- Partnership Building and Expertise: Strong capabilities in data collection, processing, and partnerships; an openness to innovative approaches.
- Expertise and Capacity Gaps: Identified needs for additional expertise in certain domains (e.g., short food supply chains, qualitative data analysis).
- Adaptability and Exploration: Ability to explore unconventional data sources, software solutions, and adapt to limitations.

These observations highlight the need for enhanced data inventory and collection methodologies, strategies to address privacy concerns, and the importance of strengthening expertise to bridge existing data gaps. Collaboration, adaptability, and innovation emerge as critical strategies for addressing these challenges across different Living Labs.

# 5. RUSTIK system's solution

The RUSTIK Information System is a reliable tool designed for all Living Labs to gather relevant data for their experiments. Additionally, external users can utilize it to track the project's progress. The system is built to provide a centralized platform for accessing data from the RUSTIK Database. It aims to facilitate comprehensive insights and streamline information retrieval for various stakeholders.

The RUSTIK Information System is constructed to serve as an analytical interface, providing valuable insights through datasets from the RUSTIK Database. It operates on multiple layers, linking a central EU system with regional, Pilot Region-level systems.

The user interface prioritizes user-friendliness with a 'user-centric' design philosophy, emphasizing anticipation, visible navigation, intuitiveness, and self-explanation for ease of use. The system supports the generation of user-specific knowledge by suggesting, combining, and visualizing key data tailored to different user types and their information needs.

Its primary aim is to help Living Labs to centralise the information relevant to their challenges. Furthermore, it should benefit local, regional, and policy stakeholders, while also serving as a resource for further research endeavours. The system incorporates indicator sets defining functional rural areas and rural transitions identified within the project.

This open-source system is available to Living Labs for data collection and analysis. It is also accessible to external users interested in monitoring project progress. The system serves the purpose of providing valuable insights and supporting informed decision-making.

The system is fully documented and seamlessly integrated into existing systems for widespread accessibility and usability. Data is added to the system by MCRIT, which is also responsible for its ongoing maintenance.

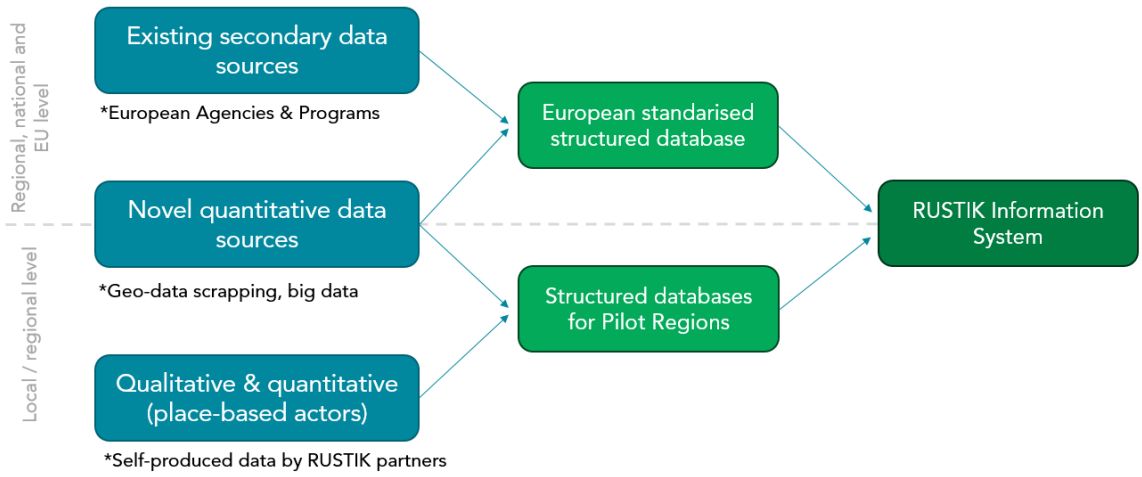


*Figure 9: Structure of the data bases and RUSTIK information system.*

This system facilitates the creation of user-specific, pertinent knowledge by suggesting, combining, and visualizing key data tailored to different user types and their information needs. It's accessible across all devices (desktops, tablets, mobile phones). An advanced indexing system enables intelligent and integrated searches. Cost-effective methods are implemented to harmonize data and ensure quality checks.

Beyond its utilization by local, regional, and policy stakeholders, the database serves as a resource for further research. There's potential to merge this database with big data modelling across various types - environmental, climate-energy, socio-economic, and digital - to explore different transition pathways.

Regarding its functionality, the system has a navigation bar on the left side and a layer's menu that emerges from it. The different possibilities offered by the sidebar are, from top to bottom:

- Layers menu: The essential function of the System. It allows the navigation through the European Core and the 14 Living Lab Cores. More information on the layers and capabilities will be provided in the following sections.
- Area draw tool: This feature enables users to create custom shapes or areas on the map, to know their extension.
- Measure tool: This tool allows measuring distances on the map.
- Information tool: Provides details about specific elements or locations on the map, offering the attributes associated with the selected area.
- Zoom all: Adjusts the view to encompass the entirety of the EU extension.
- Zoom to layer: Quick focus to the active layer.
- Zoom in: This feature allows users to magnify a particular section of the map.
- Zoom out: Enables users to decrease the zoom level, offering a broader perspective of the mapped region.

More capabilities will be added in the near future, as discussed in the 5.3 section. Besides this basic functionalities, the System's layers menu contains two different sections, referring to the two different databases of the RUSTIK project:

- European Core Database (1)
- Living Labs Databases (2)

The layers can be accessed as follows: Below you can see an example of data visualization of the Natura 2018 network for Europe. The layer is accessed through the menu, and activated through the multiple selection checkbox. Each of the layers has an information button that provides the legend and basic metadata of the layer, and a configuration button, that nowadays allows the regulation of the layer's opacity.

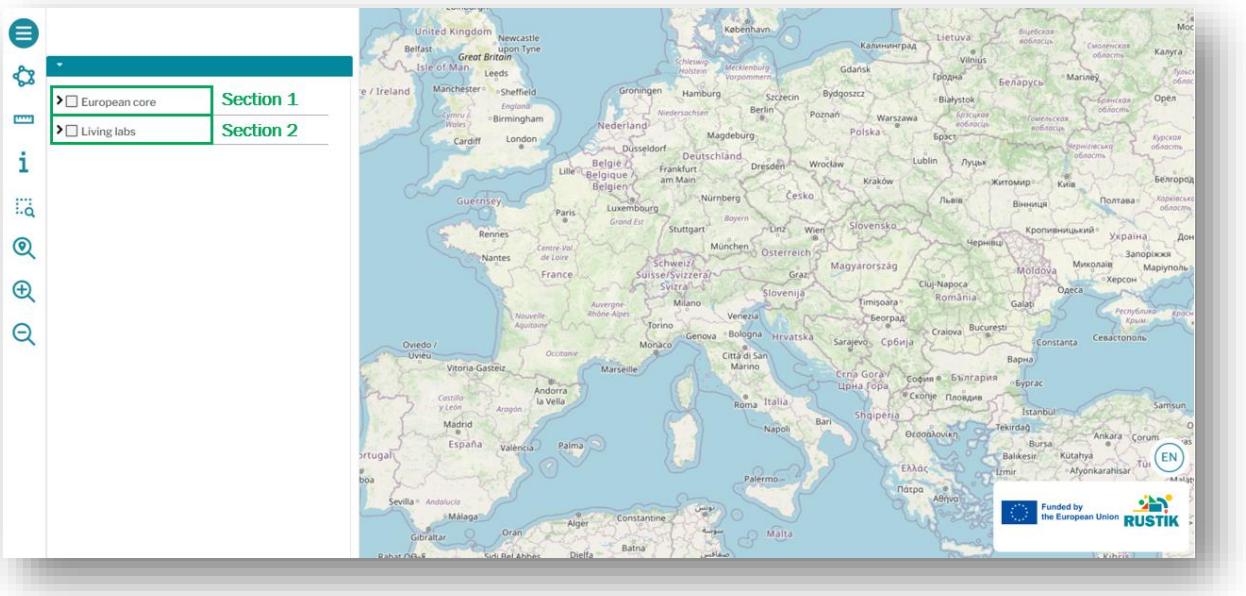


*Figure 10: The two sections of the RUSTIK System.*

## 5.1. RUSTIK system's solution at European scale

Inside of the **European Core section (1)**, four different subsections have been designed, which correspond to basic administrative information and the RUSTIK transitions:

- Administrative units (1.1)
- Socioeconomic transition (1.2)
- Environmental transition (1.3)
- Digital transition (1.4)

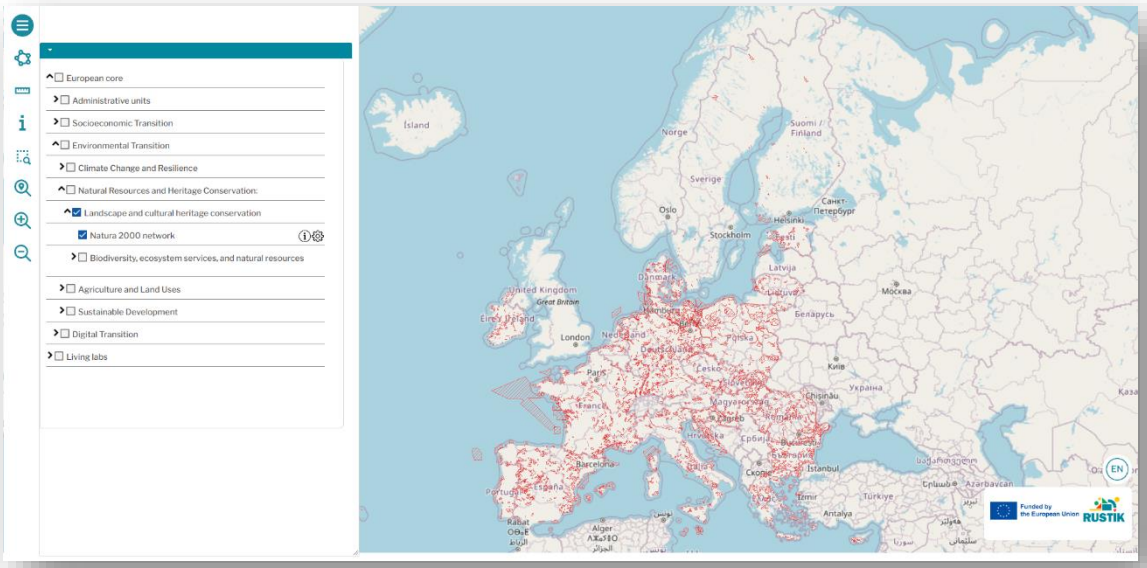


*Figure 11: Navigation through layers in the RUSTIK System.*

For each of these sections, different categories have been added. These categories were identified based on the key area topics present in D1.1 and D2.1, as well as other relevant European policies and action plans. As we continue to integrate the data on the different topics, the categories may vary and some of them will disappear when data is not available. The categories present in the system at the present date is:

*Table 9: Topics for the RUSTIK Transitions in the European Core.*

| Socioeconomic Transition | Environmental Transition | Digital Transition |
|---|---|---|
| Demographics and population | Climate change and resilience | Digital infrastructure |
| Employment and economic development | Natural resources and heritage conservation | Digital skills |
| Community engagement and governance | Agriculture and land uses | Digital transformation |
| | Sustainable development | |

## 5.2. RUSTIK system's solution for all 14 LL

Inside of the **Living Labs section**, an independent section has been created for each of the 14 Pilot Regions, as can be seen in the following figure. For each of the Living Labs, the user can zoom in to the selected one (by pressing to option A in the following figure) or open the layers dropdown list (by pressing option B). The category dropdown list will feature all the local/regional-level data from the Living Labs relevant to the Data experiment.

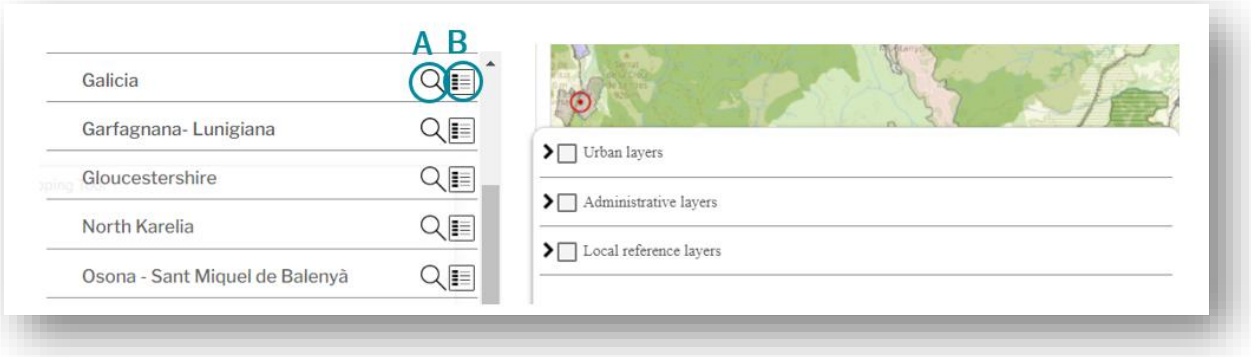


*Figure 12: Living Labs section of the RUSTIK System.*

A process to ease the data integration from all the Living Labs has been envisaged by the WP2 team. It runs through Sharepoint/Teams, a familiar tool for all the project partners, and it is compound of a static set of administrative boundaries and excel files matching them by an ID where the data is published. All the administrative units are provided only one time by the Living Labs, and their data is updated through this process. Additionally, raster datasets and points, lines and minor polygon layers can be uploaded through the same Sharepoint. All this data is uploaded to the System by the MCRIT team and it is handled in PostGIS.

## 5.3. Improvements to the System

The System is now running in a beta version, which is not linked to the rustik-he.eu domain, but hosted by MCRIT in the following link: https://gis.geovincles.com/rustik/visor.php. This will remain the same until the start of the experimentation phase, in March 2024. By then, the System will be improved and updated, and at the publishing date, more EU-level data will be uploaded, and the Living Labs will have populated populate their Databases with all the state-of-the-art data.

### Improving the RUSTIK System interface

As an improvement to the current beta version, the current layers menu is being restructured to enable a more efficient layer navigation. Also, the responsiveness (the ability of a website of being accessible across different devices such as desktops, tablets, and mobile phones), will be enhanced during this process.

Additional features to be implemented are:

→ The capability of downloading the visible datasets for each Living Lab

→ Allow the comparability of at least two datasets

In relation to the displaying of the outputs, it is envisaged that the system will be capable to handle formats like indices, dashboards, figures, tables, and interactive media. This will enable the integration of qualitative and non-geospatial data.

## 5.4. Connection to the Rural Observatory

The Information System is envisaged as a tool that serves primarily the Living Labs, but also has the potential to draw beyond the scope of our project. In this regard, the collaboration with the Rural Observatory becomes strategic.

As detailed in the 2.1 deliverable, in 2022, it was launched this flagship initiative based on the European Commission's Long-Term Vision for Rural Areas (LTVRA) and developed by the JRC in coordination with DG AGRI and DG REGIO. This platform is intended to provide data about rural areas, including data and knowledge produced by EU-funded projects such as GRANULAR and RUSTIK (SHERPA, 2022). It offers statistics, indicators and analyses based on data from multiple sources, such as JRC data, ESPON, Eurostat as well as Horizon Europe projects.

Collaboration with the JRC on regard of the Rural Observatory has already started, with the celebration of two meetings during summer 2023. These discussions have been productive and the main outcomes can be summed up as follows:

→ There is little data available at European level on the LAU level. Of the data produced by the JRC, about 90% is for NUTS3, while only 8% is for LAU. The JRC produces data at LAU level on population by age and sex, and also produces experimental indicators, such as broadband speed and distance to services (already available in the Rural Observatory).

→ Data from the 2021 national census will be available around March 2024 at LAU level, information can be found here: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_and_housing_census_2021_%E2%80%93_overview#Statistical_output_of_the_2021_EU_census_programme.

→ The JRC is currently working on the development of renewable energy indicators at grid level. Grid data is not published in the Rural Observatory but stored in the JRC catalogue. The World Bank will estimate poverty rates at NUTS-3 or LAU level, in work that will not start until 2024, with data not being available presumably until 2025.

→ There is a consensus between the two parts on the difficulties of finding data to feed socio-economic indicators at such low scales. JRC suggests to the project to focus on natural indicators and infrastructures (transport, internet…), and also point out some concerns in relation to data scraping, especially in terms of privacy when revealing information about individual places or activities. When using this kind of data, some degree of harmonisation and aggregation of data would be necessary.

→ The NUTS boundaries change every three years, which can be chaotic for the time series. They are currently working mainly with 2021 boundaries but expect to update all their

datasets to 2024 when they are published. They use 2018 boundaries for LAU, even though these are updated every year. This is because they have a better resolution.

- The JRC suggests to keep a reference year and use these same boundaries for the whole project, as switching from one to another during the course of the project would be resource intensive. The same JRC is using 2018 boundaries for the Rural Observatory and has no plans for switching to more up-to-date versions of LAUs.

- Regarding data integration between the RUSTIK System and the Rural Observatory, the JRC suggests us to download the data instead of pointing to the JRC server from the RUSTIK System, as the 11 available LAU datasets are not expected to be updated in the near future. To optimise this process, it is suggested to download the data in the Rural Observatory from the "Trends" section instead of "My Place".

- ARDECO is an experimental project that aims to fill the remaining gaps in the statistical data. It is linked to AMECO. It will provide aggregated information on fixed capital at NUTS 3 and 2, according to Eurostat data. NUTS3 datasets are expected to be online before the end of the year.

- Asked about the availability of JRC data for the UK and Serbia, it is said that the situation is different: for the UK less and less data will be published as part of the current situation, but the JRC expects to produce more data for Serbia, as it is a candidate country.

- Finally, the JRC states that is also working on refining land use maps, adding components for accessibility, residential areas, and population, but no deadline is provided for this to happen. This is already an improvement to the Corine Land Cover, using the same dataset (100 metres resolution for 2018 data).

# 6. Guidance to Living Labs

The guidance to the Living Labs in terms of data will be led by the WP2 team, with two main areas of expertise.

The first expertise area lies on the shoulders of MCRIT, with guidance on the use of the RUSTIK System, the addition and download of data to this system and its actualisation and maintenance. Furthermore, there will be provided support in Data collection methods, such as surveying, volunteered data, data scraping, remote sensing… and in the use of Data Collection Tools (KoboToolbox, Qfield, LocusGIS, Forms, Mentimeter, Sentinel Hub…). At a later stage, also, support on Data Preprocessing and Analysis Methods and Tools (GeoDa, QGIS, Excel…).

The second expertise area is the domain of Mapita and their software: Maptionnaire. This will run in parallel to the above mentioned and it will be specifically oriented to the Living Labs using the Maptionnaire software, which will be engaged through the specifically designated Learning Hub.

Guidance to the LL is envisaged in 4 steps, following the WP3 distribution of Cycle 2. The timeline is tentative at this stage, but reference dates are provided in the following schema (Figure 13). There is a set of questions aimed to be resolved in this phase that are presented in each of the stages and that should be of guidance for the Living Labs.


*Figure 13: WP3 timeline of the data experiment.*

## 6.1. Envision a Data Experiment

The main objective of this phase for the Living Labs is to answer the following question:

> Which is the Challenge of my Data Experiment? Which are the topics of interest?

This question is solved by the Living Labs in the context of the Deliverable 3.1, but it is fundamental to keep this in mind when starting to search for data for the experiment. In this phase, the guidance offered from the WP2 is introductory to the data side of the experiment, coming from two sides:

### Introduction to the RUSTIK System

The System is presented during November's Living Lab Coordinators' Meeting (LLCM). A detailed explanation on how to upload data to the System is provided on December's Living Labs Meeting (LLM, as renamed since this month). It is requested to all the Living Labs (LLs) to follow the given guidelines and upload a first dataset, ensuring all LLs are aligned with the process and that it is functional.

### First Maptionnaire Steering Group Meeting

The first meeting organized is all about Maptionnaire practicalities: how and when the LLs will get their access to Maptionnaire, what to expect from the coming Learning Hub sessions and what are the capabilities of Maptionnaire. LLs are encouraged to participate actively in discussion during the session so that it can be ensured that the Learning Hub will answer to the needs and wishes the LLs have.

## 6.2. Co-design of the Experiment

### Inventory of available data

Having an inventory of available data is a common concern amongst most of the Living Labs, as mentioned in the project meetings and especially in the 1st Survey. To overcome this issue, the following process should be followed:

| Who is producing data in my region? Which data is being produced? |
|---|

First of all is fundamental to understand who is producing what for my region. The WP2 team will be of great help to the living Labs identifying relevant datasets at European level or collected with innovative methods, but the Living abs should step in with their local knowledge for searching for local sources.

In this phase, Living Labs start to envisage which data will be used as they co-design their experiment by, first, looking for **who produces data** and then, getting to know **which data do they? produce**. More particularly, it will be fundamental getting to know not only what is produced but **what data is available for their Challenges and Data Experiments**.

The RUSTIK System appears as a crucial tool for grouping the relevant datasets for the different Living Labs at different scales or in different formats. During this phase, all the Living Labs identify which data is being produced in their Pilot Regions and upload data relevant to their Challenges and Data Experiments to the RUSTIK System. At this stage, the data that is envisaged is the one provided by regional institutes and the one already in use by the local partner institutions. Some fundamentals should be kept in mind:

#### Granularity and scale

Local and regional data often require a more granular approach compared to broader European datasets. There is no must-follow rule to follow in the definition of the analysis scale, but the following considerations must be followed:

- → It is fundamental to find the right scale: not too narrow to avoid getting lost in unnecessary details, nor too broad to miss relevant information.
- → The data availability gets reduced as we scale down.
- → Data can be easily scaled-up, but when splitting features, we are in risk of falling into incorrect generalisations.

### Local Stakeholder Involvement

Engaging local stakeholders in the Living Lab activities can be the key to proximity data. Comprehensive and accurate representation of local realities. Their insights not only aid in identifying relevant data sources but also in understanding the context and interpreting collected data effectively.

### Contextual relevance

Data collection methods must be contextually relevant to the specific needs and challenges of the local or regional area. This might involve adapting surveys, interviews, or data sources to reflect the cultural, linguistic, or socio-economic diversity of the region.

Training will be provided on efficient methods to organise data: best practices for organizing files and folders, including naming conventions and version control, as well as on clear and intuitive directory structures. Also, training on organizing and structuring data within a database.

Special attention will be given to documentation practices to underline the importance of documenting all pre-processing and processing steps to create a clear and reproducible documentation, both for internal LL purposes and for sharing experiences.

## Training on data pre-processing and collection

Once Living Labs have identified what data they have at their disposal, they must ask themselves:

| Are there any data gaps? Can I fill them? |
|---|

A Data Experiment is nothing but using data to fill a knowledge gap, but too many data gaps will make impossible bridging from data to knowledge. In this context, innovative collection methods will be presented.

Data collection is envisaged for the LLs in a similar approach to the one used for the European Core. However, data collection for local and regional contexts has peculiarities that are worth mentioning. Leveraging existing local databases, community records, or collaborating with regional institutions becomes crucial to fill in data gaps. It might also involve innovative approaches such as participatory data collection through community engagement initiatives.

## Quantitative data collection

Training can be offered by MCRIT on the use of the following methods and tools.

### Data Collection Methods

Training on various data collection methods is essential for professionals working in fields such as research, analytics, and data science. Here are some tips for training on different data collection methods:

Surveying: Survey design (how to design effective surveys, to set clear objectives, well-defined questions, and appropriate response formats), sampling techniques, data quality and validity (minimize biases and errors) and survey data analysis.

**Volunteered Data Collection**: Emphasize the real-time nature and diverse sources of volunteered data, discuss the concept of volunteered data and its potential biases, and discuss strategies for handling noisy or incomplete volunteered data.

**Data Scraping**: Provide practical training on popular web scraping tools and libraries (e.g., BeautifulSoup, Scrapy) and teach participants how to clean and transform scraped data for analysis.

**Remote Sensing**: Overview of remote sensing technologies, including satellite and aerial imagery, training on image processing techniques for extracting valuable information from remote sensing data and integration of Geographic Information System (GIS) tools for spatial analysis.

### Data Collection Tools

There are a set of useful applications for the experiment analysis that range between different levels of expertise and purposes. A selection of them is showcased here:

**Maptionnaire**: Created by our project partners Mapita, it is a community engagement platform that allows data collection and enables inclusive and engaging decision making, it contains an easy-to-use editor for creating map-based and conventional online surveys. The Maptionnaire Learning Hub, led by Mapita, will be leading this training.

**KoboToolBox**: An open-sourced data collection, management, and visualization platform used for form development, offline data collection, and project management.

**QField or LocusGIS**: Geo-data collection tools, capable to edit and create vector features and attach basic fields or forms, images and sounds.

**Google Forms**: Easily create and share online forms and surveys and analyse responses in real-time.

**Mentimeter or Kahoot**: Survey tools that enable real-time interaction both in on-site and online meetings.

**Sentinel Hub**: An EEA platform that allows to instantly visualize satellite data from numerous satellites and data collections. The process in the background takes care of the selection of appropriate scenes, download and processing of data, as well as mosaic creation.

**Google Earth Engine**: A planetary-scale platform for Earth science data & analysis to detect changes, map trends, and quantify differences on the Earth's surface.

## Qualitative data collection

Living Lab partners will also receive qualitative training, as a significant number of them will incorporate qualitative data into their experiments. Qualitative research aids in hypothesis generation, validates quantitative findings, and often reveals unanticipated patterns, contributing to a more comprehensive and nuanced interpretation of research outcomes. Integrating qualitative and quantitative data enhances the robustness and applicability of research, ensuring a more holistic understanding of the studied phenomena.

Living Lab partners were surveyed about their preferences for exploring specific aspects of qualitative data, and the respondents ranked their interests in the following order of importance:

1. Different approaches to analysing qualitative data
2. How to organize qualitative data for analysis
3. How to assess validity in qualitative research
4. Drafting an Interview Protocol
5. Ethics
6. Interviewing Technique
7. Conversational Interviewing

The training program will be led by our project partner FEUGA, responsible for customizing it to prioritize Living Lab partners' preferences, focusing on diverse approaches to qualitative data analysis, effective data organization, ensuring research validity, crafting interview protocols, understanding ethics, refining interviewing techniques, and embracing conversational interviewing.

### Maptionnaire Learning Hub meetings

Three meetings are envisaged during this period: A first inspirational session about good Maptionnaire examples, a training session about getting started and implementing data collection with Maptionnaire and a reflection session for sharing learnings from different Living Lab case study experiences.

## 6.3. Experimentation stage

Living Labs work on Data Collection and Analysis. Support will be provided to all of them from WP2. Working examples will be presented to illustrate the different methodologies and tools.

### Data analysis

Once data is available, the Living Labs should start answering to the following:

| Are there any trends over time or within the territory? |
|---|

At this stage, it will be fundamental to identify if the selected indicators are improving or getting worse, and if are there spatial patterns guiding this changes. In this regard, training will be offered by three different actors, as FEUGA will be leading the training on qualitative methods, MCRIT will be leading the quantitative methods side and MAPITA will offer formation in regard of the Maptionnaire tool.

For the qualitative methods, resources will be provided to the Living Labs in methods for Data Preprocessing: Discussion of common issues in raw data (missing values, outliers) and methods for handling them, train on techniques for imputation, outlier detection, and data validation. Also formation and recommendations will be provided for Data Analysis Tools, with a set of useful

applications for the experiment analysis ranging between different levels of expertise and purposes. A selection of them is showcased here:

GeoDa: A free and open source software designed to facilitate new insights from data analysis by exploring and modeling spatial patterns.

QGIS: A free and open source geographic information system that enables to create, edit, visualise, analyse and publish geospatial information.

Excel (Google Spreadsheets or LibreOffice): Spreadsheet software that allows to organize data, perform calculations, create charts amd automate tasks through formulas and macros.

On the Maptionnaire side, it is envisaged a training session on how to analyse the data collected with this software. The training will introduce the possibilities of Maptionnaire's inbuilt analysis tool for traditional survey questions and the map-based questions. Furthermore, the trainings will show how to export the collected data from Maptionnaire to Excel and GIS datasets.

## 6.4. Experiment reflection

The produced and collected datasets will be integrated to the RUSTIK System to facilitate the experiment reflection. Examples from Osona's Living Lab will be presented to illustrate the different methodologies and tools.

# Annex 1. First-round Survey

The following survey aims to collect information on Living Labs' data availability and interests.

For each of the RUSTIK Transitions, a list of topics is provided, along with examples of indicators.

*The questions marked with an asterisk are to be answered only if there is existing data in the Living Labs.

This survey is intended to be completed in cooperation between the Pilot Region Partners and the Living Lab Coordinators to ensure that both actors are aware of the resources and capacities stated and that they are working together.

Each of the Living Labs is requested to submit the completed survey by Friday, June 9th.

General Information

Name of the Pilot Region: Please, complete.

Country: Please, complete.

Contact person (email): Please, complete.

The Living Labs are requested to provide the following information for each topic:

- Is it useful for your Living Lab? (Whether the specific topic is relevant or not to the Living Lab)
- Does the data exist in your Living Lab? (Whether there is existing data related to the topic in the Living Lab)
- Can you access the data you need?* (Whether the data is accessible by the Living Lab actors and can be used throughout the RUSTIK project)
- What scale is the data available at?* (Whether the data is provided in a raster dataset [grid, granular] or in local [LAU] or regional [NUTS 3, 2, or 1] administrative units)
- Comments (Information is expected on data collection periodicity, major data gaps, or specific relevant indicators on the topic).

The following list of topics and examples is provided for the different RUSTIK transitions.

## Socioeconomic Transition

| | Topic | Examples |
|---|---|---|
| 1 | Population ageing | *Total population, density, growth, population by age groups,* |
| 2 | Gender imbalances | *Gender Ratio, Proportion of Women in Leadership Positions and in* |
| 3 | External migrations from third countries | *Migration, immigration, emigration, nacionality, racial and/or* |
| 4 | Urban-rural migrations | *Neo-rurals, emigration to urban areas* |
| 5 | Social inclusion and cohesion issues | *Social Mobility, Poverty Rate, Share of the population with* |
| 6 | Access to housing. Housing conditions | *% of total area under built-up area, % of total area under urban* |
| 7 | Health services availability | *Access to Healthcare Services, Hospital Beds per capita,* |
| 8 | Schools and education | *Access to Schools and education centers* |
| 9 | Public equipments and facilities (sports, cultural, leisure, | *Number of cultural amenities; distance to schools, libraries,* |
| 10 | Local economic development | *Hotels, restaurants, touristic areas, No. of overnight stays in hotels* |
| 11 | Job opportunities | *Employment per sectors, VAB per sectors, Gross Domestic Product* |
| 12 | Support to innovation | *Innovation and support to small and medium-sized businesses* |
| 13 | Citizen involvment, communitarianism | *Citizen Engagement, trust in neighbors....* |
| 14 | Institutional Governance | *Public Trust in Government Institutions* |
| | Others (indicate which ones) | |

## Environmental Transition

| | Topic | Examples |
|---|---|---|
| 1 | Vulnerability/Preparedness to Climate change | *Availability of adaption plans, monitoring of CC impacts...* |
| 2 | Air, soil and water quality / pollution | *Water Pollution Levels, Waste Generation and Disposal Rates,* |
| 3 | Landscape and cultural heritage conservation | *Protected Areas Coverage* |
| 4 | Food securing. Concurrence of agricuture with other use | *Abandonment of agricultural fields, change in uses, farmers* |
| 5 | Forest management | *Forest management* |
| 6 | Soil erosion | *% of agricultural areas and natural grassland affected by moderate* |
| 7 | Soil Imperviousness | *Soil imperviousness, soil consumption, total soil organic carbon* |
| 8 | Sustainable mobility | *Collective passenger transport, shared mobility* |
| 9 | Energy production (wind, solar, other...) | *Production, consumption, renewable share and potential, solar* |
| 10 | Biodiversity, ecosystem services, and natural resources | *Biodiversity loss, Biodiversity Index* |
| | Others (indicate which ones) | |

## Digital Transition

| | Topic | Examples |
|---|---|---|
| 1 | Digital infrastructure | *Different networks coverage* |
| 2 | Digital skills | *Basic skills, ICT specialists* |
| 3 | Digital transformation of businesses | *Unicorns, digital companies...* |
| 4 | Digitalisation of public services | *Public services, e-health, digital identity* |
| 5 | Vulnerability to digital disruptions | |
| 6 | Others (indicate which ones) | |

# Annex 2. Second-round Survey

The following survey aims to collect information on Living Labs' resources and capacities in relation to data collection and exploitation.

For some of the questions, a multiple-choice list is provided. Select the statements by adding an "x" inside the [x].

The term "Pilot Region" is used throughout the survey and includes the Pilot Region Partner organisation AND the Living Lab coordinator. If the question is addressed specifically to either the Pilot Region organisation OR the Living Lab coordinator, it will be clearly specified (questions A & B on Page 4).

This survey is intended to be completed in cooperation between the Pilot Region Partners and the Living Lab Coordinators to ensure that both actors are aware of the resources and capacities stated and that they are working together.

Each of the Pilot Regions is requested to submit the completed survey by Wednesday, July 5th.

General Information

Name of the Pilot Region: Please, complete.

Country: Please, complete.

Contact person (email): Please, complete.

## Resources and Infrastructure

a. When the Pilot Region is using statistical and/or GIS data, which organization is providing it (more than one answer is possible)

[ ] The Pilot Region uses statistical and/or GIS data produced by subregional providers.
[ ] The Pilot Region uses statistical and/or GIS data produced by regional or national providers.
[ ] The Pilot Region uses statistical and/or GIS data produced by European or global providers.
[ ] The Pilot Region uses statistical and/or GIS data produced by commercial providers.

b. Select the statements that apply to your Pilot Region in relation to data support for policy implementation and assessment.

[] The Pilot Region uses data for the design and implementation of public policies.
[] The Pilot Region uses data for the assessment and monitoring of implemented policies and initiatives.
[ ] The Pilot Region uses data to influence regional or national policy and/or engage with policymakers.

[ ] The Pilot Region uses data for communication/animation activities addressed to local actors.
[ ] The Pilot Region is not involved in design, implementation or assessment of public policies.

c. If you have selected in question b any option but the last one (not involved), explain how the Pilot Region is using data for the design and/or assessment of policies. Please explain as well if they analyse data themselves (raw data) or if they use already analysed data (e.g., from scientific projects).

Please, answer here.

d. If the Pilot Region is working with georeferenced data (GIS systems), select the statements that apply to your PR.

[ ] The Pilot Region is capable to collect, store and analyse GIS data.
[ ] The Pilot Region already has a GIS system with a georeferenced database.
[ ] The Pilot Region data is updated real-time using external webservices.
[ ] The Pilot Region system, if available, is based on open-sourced software (e.g., QGIS).
[ ] The Pilot Region does not have GIS system, but uses third-parties online map viewers to visualize data produced by external providers.
[ ] The Pilot Region has specialised staff working with GIS systems.

e. Select the statements that apply to your Pilot Region in relation to using GIS visualization platforms.

[ ] The Pilot Region feels confident in their abilities to access and interpret GIS data.
[ ] The Pilot Region uses GIS software to visualize GIS data (QGIS, ArcGIS, GeoMedia...).
[ ] The Pilot Region produced an online map viewer (web GIS) for internal use of Pilot Region officers and technicians.
[ ] The Pilot Region produced an online map viewer (web GIS) for external users (citizens, economic agents, NGOs...).
[ ] The Pilot Region has specialised staff capable of handling web GIS (WMS, PostGis...).
[ ] The Pilot Region has partnerships or collaborations with external organizations to enhance data resources and infrastructure.

f. If you have selected one or several options in (d) and/or (e), explain how the Pilot Region is using GIS data and/or visualization platforms and what benefits does it bring.

Please, answer here.

### Stakeholders and Collaboration

a. Select which capabilities the Pilot Region Partner holds in relation to data (more than one is possible):

[ ] Data collection
[ ] Data processing
[ ] Data analysis
[ ] Data distribution (e.g., public viewer)

b. Select the capabilities of the Living Lab Coordinator holds in relation to data (more than one is possible):

[ ] Data collection
[ ] Data processing
[ ] Data analysis
[ ] Data distribution
(e.g., public viewer)

c. Are there any existing collaborations or partnerships with external stakeholders to identify data resources and collect data in the Pilot Region?

Please, answer here.

d. Are these collaborations permanent or occasional?

Please, answer here.

e. In which phase have these collaborations been implemented more frequently? (programme design, specific studies, monitoring and evaluation activities, other activities)

Please, answer here.

### Future Trends and Suggestions for Improvement

a. Are there any support or resources needed to enhance your Pilot Region's data capabilities? What are the key areas where capacity building and training are needed? (i.e., inventory of available data, common platforms of data, involvement of private/public owners of data, accessing to geo-spatial data, connecting geo-spatial to other type of data, processing data, interpreting and analysing available data, etc.)

Please, answer here.

b. Is the Pilot Region familiar with the following innovative approaches and technologies for data collection? If so, state which ones may enhance the data capabilities of your LL.

| | I am aware of this technology | I think it could be useful |
|---|---|---|
| Web data scraping (massive collection of web data) | [ ] | [ ] |
| Participatory GIS and data crowdsourcing | [ ] | [ ] |
| Satellite imagery (Copernicus program data) | [ ] | [ ] |
| Data cubes | [ ] | [ ] |
| Others (Please, specify here) | [ ] | [ ] |

c. Are there any concerns with data generated by the partner or acquired or facilitated from other institutions that cannot be freely distributed for copyright, privacy or security reasons?

Please, answer here.

### Pilot Region extension and administrative units' definition

a. To which administrative unit does your entire Pilot Region correspond?

[ ] NUTS 2

[ ] NUTS 3

[ ] LAU 1

[ ] LAU 2

[ ] Other (e.g. Aggregation of Lau 2)

b. Which administrative level is the most relevant for accessing data that is useful for characterizing your entire Pilot Region?

[ ] NUTS 3

[ ] LAU 1

[ ] LAU 2

[ ] Census / electoral divisions

[ ] Other (define)

c. Please attach a map of your entire Pilot Region, preferably in a georeferenced vector format such as Shapefile or Geopackage. This map should ideally also feature the relevant administrative units mentioned in question b.

Funded by the European Union
UK Research and Innovation

# Annex 3. Current composition of the CORE RUSTIK Database

The following tables display the different datasets available for each of the functions and transitions as defined in the previous deliverables. Only data with a granularity of LAU administrative units or comparable raster sizes is considered for this purpose.

### Production function

| Topic | Subtopic | Name |
|---|---|---|
| Agri-industrial | Agriculture | Agricultural Area |
| | Land use | Land Use per municipality |
| Forestry | Forest | Forest typology |
| | Land use | Forestal Land Use |
| Diversified economy (secondary sector) | Land use | Land Use per municipality |
| Diversified economy (tertiary sector) | Tourism | Tourism capacity in rooms per LAU2 |

### Production function

| Topic | Subtopic | Name |
|---|---|---|
| Regulating water availability and quality | Water | River network system |
| Soil protection and functionality | Imperviousn ess | Imperviousness |
| Landscape and cultural heritage conservation | Land Use | Forest and semi-natural areas land use |
| Maintaining/ increasing biodiversity | Protected Areas | Natura 2000 areas |
| | | Nationally designated areas (CDDA) |
| | | Emerald Network (UK Only) |
| | Land Use | Arable crop land use |
| | Air Quality | European air quality data (interpolated data) |

Funded by the European Union
UK Research and Innovation

| Topic | Subtopic | Name |
|---|---|---|
| Climate change mitigation and adaptation | Extreme Weather | Areas burnt by wildfires |
| | | Drought Frequency change |
| | | Projected changes in heavy precipitation in winter and summer |
| | | Projected change in meteorological forest fire danger |
| Sustainable production of energy | Facilities | Count of OSM POIS: windmills |

### Consumption function

| Topic | Subtopic | Name |
|---|---|---|
| Leisure activities, tourism and activities related to cultural heritage | Tourism | Tourism capacity in rooms per LAU2 |
| | Leisure infrastructure | Strava Metro & Heatmap |
| | Land Use | Green leisure, forest and semi-natural areas land use |
| | Protected areas | Protected natural areas: Natura 2000 areas, Nationally designated areas (CDDA), Emerald Network (UK Only) |
| Residential | Population | Degree of Urbanisation (DEGURBA) |
| | | Population Grid |
| | | Population per LAU2 |
| | Settlement | European Settlement Map |
| | Land Use | Infrastructures and industry land use |
| Provision of services of general interest (SGI) | Road infrastructure | Road infrastructure and hierarchy |
| | SGIs | Average Distance to SGI |

Funded by the European Union
UK Research and Innovation

### Socio-economic transition

| Topic | Subtopic | Name |
|---|---|---|
| Population ageing & demographic challenge | Population | Population Grid |
| | | Population per LAU2 |
| | Settlement | European Settlement Map |
| Public equipment and facilities (sports, cultural, leisure....) | Leisure infrastructure | Strava Metro & Heatmap |
| | Facilities | Count of OSM POIS in the categories leisure & sports |
| Local economic development | Touristic assets | Hotels, restaurants, touristic areas |
| | | No. of available rooms in accommodation establishments |
| Schools and education | Facilities | Count of OSM POIS: university, school, kindergarten, college |
| Access to housing. Housing conditions | | % of total area under built-up and under urban area |
| Health services availability | Facilities | Count of OSM POIS & service areas: Access to Healthcare Services |

### Environmental transition

| Topic | Subtopic | Name |
|---|---|---|
| Sustainable mobility | Infrastructure | Count of OSM bus and train stations and stops. |
| Landscape and cultural heritage conservation | Protected Areas | Protected natural areas: Natura 2000 areas, Nationally designated areas (CDDA), Emerald Network (UK Only) |
| Vulnerability/Prepared ness to Climate change | Extreme Weather | Areas burnt by wildfires |
| | | Drought Frequency change |
| | | Projected changes in heavy precipitation in winter and summer |

Funded by the European Union
UK Research and Innovation

| Topic | Subtopic | Name |
|---|---|---|
| | | Projected change in meteorological forest fire danger |
| Food securing. Concurrence of agriculture with other uses | Agriculture | Agricultural Area |
| | | Abandonment of agricultural fields |
| | | Corine Land Cover Change 2012-2018 |
| Energy production (wind, solar, other...) | Facilities | Count of OSM POIS: windmills |
| Forest management | | Forest typology |
| Air, soil and water quality / pollution | Air quality | European air quality data (interpolated data) |
| | Noise | Noise Exposure |
| | Water | River network system |
| Soil Imperviousness | | Soil imperviousness |

### Digital transition

| Topic | Subtopic | Name |
|---|---|---|
| Digital infrastructure | Infrastructure | Broadband speed for fixed and mobile connection per LAU |

Funded by the European Union
UK Research and Innovation

# Bibliography

Carbonés, Biosca and Ulied (2023). D2.1 Report on guidelines for LL. RUSTIK HE Project. Available at: https://rustik-he.eu/wp-content/uploads/2023/10/RUSTIK-D2.1.pdf

European Environmental Agency. (2023). "DataHub". Retrieved May to October 2023. From: https://www.eea.europa.eu/en/datahub

Joint Research Centre. (2023). "Trends". Rural Observatory. European Union. Retrieved July to November 2023. From: https://observatory.rural-vision.europa.eu

Mantino (2023). D1.2 Synthesis report based on 14 narrative descriptions from the academic partners responsible for the Pilot Regions. RUSTIK
HE Project. Available at: https://rustik-he.eu/wp-content/uploads/2023/10/RUSTIK-D1.2.pdf

Mantino, Forcina and Morse (2023). D1.1 Report on the methodological framework to define Functional Rural Areas and rural transitions.
RUSTIK HE Project. Available at: https://rustik-he.eu/wp-content/uploads/2023/04/RUSTIK_D-1-1_Methodological_Framework_31.03.23.pdf

nPerf SAS. (2023). "Mapa de cobertura 3G / 4G / 5G, España". Retrieved June to October 2023. From: https://www.nperf.com/en/

Open Street Map contributors. (2023). "Open Street Map". Retrieved July 2023. From: https://www.openstreetmap.org

Organisation for Economic Co-operation and Development (2020). "Public governance > Trust in Government". Retrieved November 2023. From: https://www.oecd.org/governance/trust-in-government/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(85), 2825-2830. Available at: http://jmlr.org/papers/v12/pedregosa11a.html

SHERPA, Sustainable Hub to Engage into Rural Policies with Actors (2022). "Interfaces Linking Science, Society and Decision-Makers to Shape Future Rural Policies". Retrieved September 2023. From: https://rural-interfaces.eu/

Strava. (2023). "Strava Metro: Better communities for cyclists and pedestrians". Retrieved September 2023. From: https://metro.strava.com/

Funded by the European Union
UK Research and Innovation