GRANULAR

GUIDELINES AND TRAINING MATERIALS FOR DATA COLLECTION

D6.2

SEPTEMBER 2024

Funded by the European Union

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

D6.2 Guidelines and Training Materials for Data Collection

Project name	GRANULAR: Giving Rural Actors Novel data and re-Useable tools to Lead public Action in Rural areas
Website	www.ruralgranular.eu
Document type	Deliverable
Status	Version 2 (Final)
Dissemination level	Public
Authors	F. Stortini, M. Moretti, T. Trinchetti (UNIPI), L. Chasset (PPM), D. Miller (HUT), M. Bahta (VAS), H. Huiskamp (P10), M. Doval, B. Martins (UVIGO), M. Ricci (VdC), A. Kurdys-Kujawska (KUT), I. McCallum (IIASA), A. Kirsch, Y. Damari, T. Berchoux (IAMM)
Work Package Leader	University of Pisa (UNIPI)
Project coordinator	Mediterranean Agronomic Institute of Montpellier (IAMM)

Citation: Stortini, F., Moretti, M., Trinchetti, T., Chasset, L., Miller, D., Bahta, M., Huiskamp, H., Doval Ruiz, M. I., Martins, B., Ricci, M., Kurdyś-Kujawska, A., McCallum, I., Kirsch, A., Damari, Y., & Berchoux, T. (2024). Guidelines and Training Materials for Data Collection. GRANULAR. https://doi.org/10.5281/zenodo.13884364

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

Funded by the European Union

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

HISTORY OF CHANGES

VERSION	DATE	CHANGES
v1 D6.2	02/10/2024	Initial version submitted
v2 D6.2	22/01/2026	*Addition of a "How to use this tool" section:* • Page 6: New section 1.1 added with the description of the necessary steps which can guide the actors on data planning activities from what the main appropriate type of data are to use for catching the reality to the process of analyse it. *Changes made to the 'In Practice' boxes to improve formatting consistency:* • Page 12: Improved images formatting, added titles for images, improved readability, changed text in line with improvements. In general, defined as better formatting with other boxes. • Page 14: Inserted bold in the main points of the box. • Page 16: Added image and related title, added image explanation in line with the topics' box. • Page 18: Improved image formatting, added title for image, improved readability, changed text in line with improvements. In general, defined as better formatting with other boxes. • Page 25: Modified image title, text formatting. • Page 31: Added formatting for image titles and in text. • Page 33: Improved image and text formatting. *Provision of GDPR and ethics considerations in local data collection:* • Page 40: Annex 1 added (reference to this Annex added page 22 in the introduction to Section 2.4 "Primary data collection") detailing all processes that must be implemented to ensure the protection of research participants' personal data, in compliance with the EU General Data Protection Guidelines (GDPR) or equivalent that pertain in each country. *Improved text formatting* • Page 26-28: Added uniform formatting in orange paragraphs and boxes. • Page 34-35: Included in the text reference to the "In Practice" boxes.

Table of contents

EXECUTIVE SUMMARY **5**

1. INTRODUCTION **6**

1.1. How to use it? 6
1.1.1. Who should read this? *6*
1.1.2. The core process: a five-step policy cycle *7*
1.1.3. Which section should you read? *8*
1.1.4. Key principles *9*

2. GENERAL APPROACH **10**

2.1. Identifying priorities: the need for data 11
2.1.1. Criteria for Selecting Indicators *12*
2.1.2. Generic Quality Criteria for Indicators *13*
2.1.3. Feasibility and Clarity of Indicators *13*
2.2. Identifying data needs 15
2.2.1. How much data is needed? *15*
2.2.2. Quantitative or Qualitative data? *15*
2.2.3. How objective should the data be? *16*
2.2.4. How should the data be disaggregated? *16*
2.2.5. What scope for the data? *16*
2.2.6. When should data be collected? *17*
2.2.7. Primary or secondary data? *17*
2.3. Inventory of secondary data 18
2.3.1. Data Banks *19*
2.3.2. Demographic data *20*
2.3.3. Agricultural surveys *20*
2.3.4. Local official records *20*
2.3.5. Other spatial data *21*
2.4. Primary data collection 22
2.4.1. Surveys *22*
2.4.2. Rapid Rural Appraisals *26*
2.4.3. Sensors and observation *26*
2.4.4. Citizen science *27*
2.4.5. Scraping *28*
2.5. Analysis and visualisation 30
2.6. Towards policy formulation 32

3. PRACTICAL APPLICATION: RESULTS FROM LIVING LABS **34**

4. CONCLUSION **39**

ANNEX 1: GDPR AND ETHICS CONSIDERATIONS IN LOCAL DATA COLLECTION **40**

EXECUTIVE SUMMARY

This document advocates for more detailed rural data to support accurate, useful, and impactful decision-making. It aims to provide rural stakeholders with **practical insights into data requirements and offers methodological support for data collection and usage at every stage of the planning process**. Developed in collaboration with the seven GRANULAR Living Labs, the document is organized as a set of **guidelines for local practitioners**. These guidelines provide a step-by-step approach to different phases, from assessing territorial policy priorities and identifying data needs to making more informed decisions. Practical examples from the Living Labs, which cover a wide range of territorial situations with unique characteristics, needs, and challenges, are included. These examples showcase data, tools, and methods that can be replicated at the local level.

Data is essential for effective planning, particularly in rural planning, as it plays a critical role throughout the entire process. When collecting data, several key factors need to be considered: **identifying minimum data needs, defining the scope of indicators, balancing quantitative and qualitative information, ensuring objectivity, deciding on the level of data disaggregation, setting appropriate collection timing, specifying accuracy standards, and prioritizing between primary and secondary data sources.** Secondary data should be prioritized when available and appropriate, as it is typically more cost-effective. Useful secondary data sources for local rural planning include national censuses, surveys, government records, spatial data, and local knowledge. Primary data collection methods can vary in resource intensity. While traditional surveys are reliable, they are resource-intensive. Rapid rural appraisal techniques, though less comprehensive, can provide sufficient data with fewer resources. Participant observation, using local staff, can also be an efficient and effective method for gathering local data. Ultimately, data is critical for rural policy and planning, as informed decision-making requires access to sufficient, relevant, and accurate information.

The general approach is structured into five main sections. Chapter 2.1 introduces key methods for identifying priorities, focusing on the use of the Rural Diversity Compass. This tool helps stakeholders navigate the complexities of rural planning by highlighting key areas of concern and guiding them toward more effective outcomes through the identification of relevant indicators for monitoring and evaluation processes. Chapter 2.2 covers essential considerations that should be addressed before data collection begins, such as the quantity, quality, type, format, and source of data. Chapter 2.3 provides an overview of relevant open data banks, drawing on the inventory of available datasets for rural development conducted by GRANULAR (Deliverable D3.1 Screening Rural Data Sources). It also includes an evaluation of the performance and costs associated with acquiring rural data at various geographical scales (Deliverable D3.2 Exploring rural data landscapes – A benchmark of performance and costs in the EU and beyond). This section highlights five key types of secondary data crucial for the social aspects of rural area planning: demographic data, agricultural household surveys, official government records, maps, and general knowledge. Chapter 2.4 discusses various methods for collecting primary data, while Chapter 2.5 examines techniques for data analysis and visualization. Finally, Chapter 2.6 explores how to formulate effective rural policies, using three sector-specific examples that illustrate the challenges and importance of relying on reliable data.

1. Introduction

Existing data often fail to accurately reflect the realities of rural Europe and are plagued by spatio-temporal resolution issues. Factors such as mobility, accessibility, biodiversity, demographics, and wellbeing are particularly critical for understanding rural areas, yet they are often inadequately covered by current public data sources.

Varying definitions of "rural" also complicate analysis. In some cases, rural is simply defined as what remains after urban or metropolitan areas are delineated. This can result in more prosperous rural communities—those near urban centres or experiencing population and economic growth on the urban fringe—being classified as metropolitan areas, leaving only the most remote and struggling communities as "rural" (Deliverable D4.6 Scoping of rural typology for the EU). Other definitions consider factors such as population density, travel time, concentrated economic activity, commuting patterns, and local character. While these definitions may offer more granularity and are often available at smaller geographic levels, they often lack the necessary differentiating indicators for local analysis and decision-making. The use of multiple, competing definitions makes it difficult to create a clear picture of rural areas and complicates comparisons.

The small population size of many rural areas also makes data collection and reporting challenging. Surveys and data sources that cover Europe broadly, like the European Social Survey (ESS) or Eurostat datasets, often rely on samples from already small populations, which can result in high margins of error and affect the timeliness of the data. Data collection challenges are compounded by inadequate communication infrastructures, including broadband access. In some datasets, long reporting periods and small sample sizes cause numbers to appear "sticky," with measures remaining unchanged for long periods before suddenly jumping or dropping, masking more gradual changes.

Rural actors thus need more accurate and detailed data to facilitate more precise, valuable, and impactful research on rural communities; empower local practitioners to make more informed decisions; enhance the capacity of local governments to make decisions and set policies, including for operations and long-term planning; and challenge prevailing narratives about rural Europe, which often homogenize rural areas and perpetuate stereotypes.

1.1. How to use it?

Local actors in rural areas are increasingly called to find innovative solutions and support the decision-making process. Their ability to assess appropriate data and information has become crucial. These guidelines help you navigate the data landscape—from identifying what you need to know, to collecting and analysing information, to using findings to shape local policy.

This section shows you how to move through this document efficiently, depending on who you are and what problem you are trying to solve.

1.1.1. Who should read this?

These guidelines serve different users with different needs:

Local government officials and planners: you need data to inform strategic decisions—whether about tourism, agriculture, energy transitions, or social services. You will find this document most useful for understanding what secondary data exists, what gaps remain, and when to invest in primary data collection.

Project managers and research coordinators: you are designing a specific data collection initiative—a survey, interviews, or participatory mapping exercise. You will use these guidelines to plan your approach, choose methods suited to your resources, and ensure your data will actually answer the questions that matter to your community.

Community members, farmers, and local stakeholders: you have knowledge about your area and may be asked to participate in data collection or contribute local insights. You will find this document helpful for understanding why data matters, what kinds of questions researchers ask, and how your input feeds into local planning.

1.1.2. The core process: a five-step policy cycle

Data work follows a cycle. You begin by identifying what matters most to your community, then figure out what information you need, collect or find that information, analyse it, and finally use it to make better decisions. The process often loops—you may return to earlier steps as you discover new questions or gaps. Here is a breakdown of the cycle:

1. Identify your priorities (cf. Section 2.1)

Start by clarifying what territorial challenges or opportunities matter most. Is tourism management your concern? Food system resilience? Healthcare access? Use the Rural Diversity Compass to systematically explore these priorities and identify the dimensions that need data support. This step grounds everything that follows.

2. Define your data needs (cf. Section 2.2)

Once you know your priorities, ask: what specific information do I need? How much detail? What type—numbers, stories, maps, or a mix? How should the data be broken down (by geography, by age group, by farm size)? This section helps you answer these questions before you start collecting, saving time and money.

3. Collect or access data (cf. Sections 2.3 and 2.4)

You have two main choices: use existing data (secondary data—what others have already collected) or gather new information yourself (primary data). Most projects use both. Section 3.3 guides you through existing data sources—national databases, government records, maps, spatial data. Section 3.4 describes methods for collecting new data: surveys, interviews, rapid rural appraisals, sensor observations, and citizen science approaches. Choose based on your needs, resources, and timeframe.

4. Analyse and visualize (cf. Section 2.5)

Once you have data, you need to make sense of it and present it in ways that inform decisions. This might mean statistical analysis, mapping, creating charts, or summarizing qualitative findings as themes. This step ensures your data becomes insight.

5. Inform policy (cf. Section 2.6)

Use your findings to frame policy discussions, set achievable goals, and guide action. Real-world examples from Living Labs show how data informs decisions in agriculture, tourism, and other sectors.

1.1.3. Which section should you read?

Not every section applies to every situation. This quick-reference table helps you find what you need.

Your Situation	Start With	Then Read	Your Main Decision
You have a local challenge but are not sure what data matters	Section 2.1 (Identifying Priorities) + Rural Diversity Compass	Section 2.2 (Identifying Data Needs)	What are the key dimensions of this problem?
You need information urgently and have limited budget	Section 2.2 (Identifying Data Needs)	Section 2.3 (Existing Data Sources)	Can existing data answer our question?
You have checked existing sources but gaps remain	Section 2.4 (Primary Data Collection)	Methods in 2.4.1–2.4.5 (Surveys, interviews, sensors, citizen science, etc.)	Which collection method fits our resources and timeline?
You need to collect data from households or farmers	Section 2.4.1 (Surveys)	ANNEX 1 (GDPR & Ethics)	Sample survey or full census? Online, in-person, or mixed?
You need to involve local people in data gathering	Section 2.4.2 (Rapid Rural Appraisals) or 2.4.4 (Citizen Science)	Methods section + ANNEX 1	How much time and training do facilitators need?
You have collected data; now you need to make sense of it	Section 2.5 (Analysis & Visualization)	Real-world examples in Section 3	What analysis methods match our skill level and data type?
You need to communicate findings to decision-makers	Section 2.5 (Analysis & Visualization) + Section 2.6 (Towards Policy Formulation)	Examples in Section 3	What format (maps, charts, narrative) will persuade?
You are managing data collection across multiple sites	Section 2.2 (Identifying Data Needs)	Sections 2.4 + ANNEX 1 (GDPR & Ethics for multi-site work)	How do we ensure consistency and compliance?

1.1.4. Key principles

Whenever using this document, keep these principles in mind:

- **Start with priorities, not data:** ask "what do we need to know?" before "what data exists?"
- **Secondary before primary:** always check existing sources first. You may find what you need without the cost of primary collection.
- **Mix methods when it makes sense:** combine quantitative data with qualitative data to get the full picture.
- **Iterate if needed:** if initial findings raise new questions, you should loop back to Section 2.2 and refine your data needs.
- **Plan for participation:** involve local people early—not just as survey respondents, but as collaborators in identifying priorities and interpreting results.
- **Respect data protection:** if collecting personal information, follow GDPR or equivalent local regulations (see ANNEX 1).

2. General Approach

This section introduces a structured approach to the local data and policy cycle, emphasizing the importance of data to support effective decision-making. The process begins with agenda setting, where priorities are identified, followed by defining specific data needs aligned with SMART indicators (specific, measurable, achievable, relevant, and time-bound indicators that are used in monitoring and evaluation) to address these goals. In the next phase, existing data is mined for relevant information, and where gaps exist, primary data is collected. The collected data (primary and secondary) undergoes thorough processing, analysis, and interpretation, ensuring that it is accurate and actionable. Ultimately, this approach aims to generate insights that directly inform local planning and development strategies, fostering more impactful and community-driven outcomes.

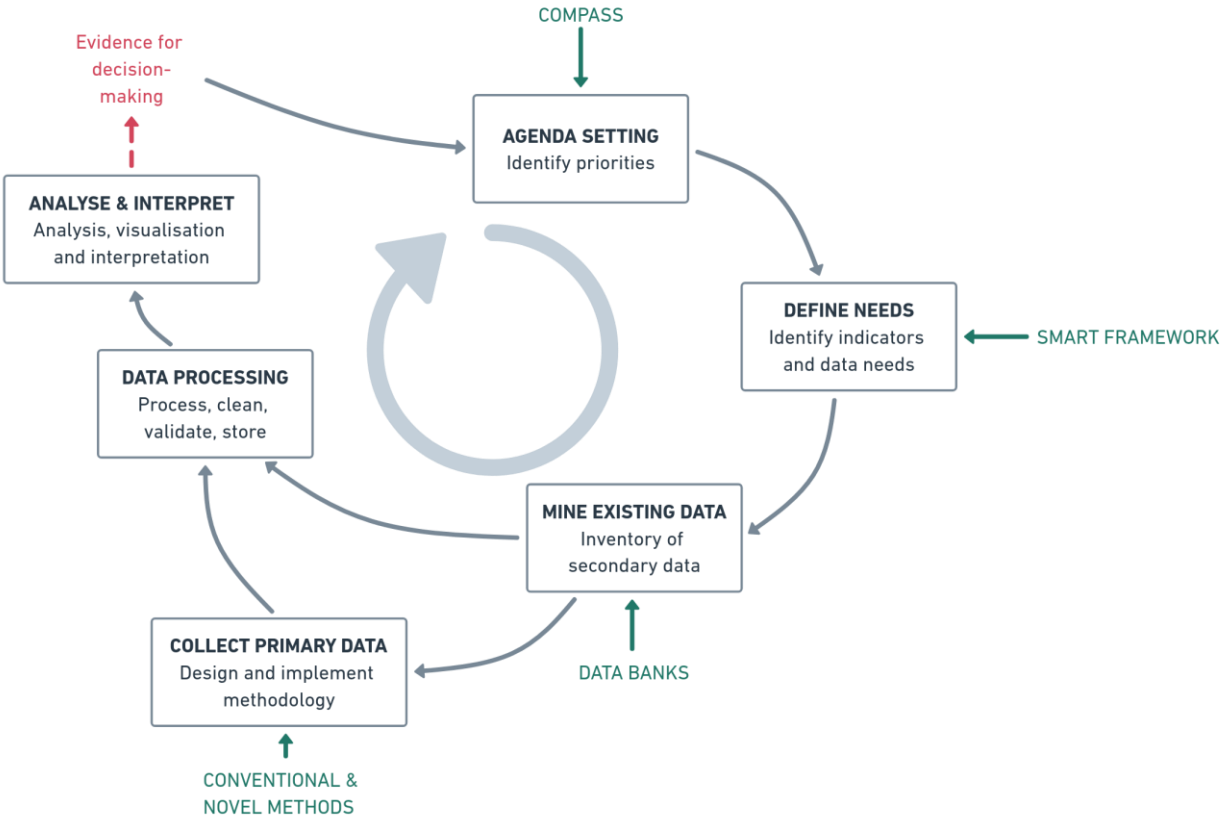


Figure 1: The "Local data for rural planning" cycle.

The following sections provide a detailed breakdown of each step in the local data and policy cycle, offering practical tools to guide the process, accompanied by practical examples from GRANULAR Living Labs. From setting a clear agenda to defining precise data needs, and from leveraging existing data sources to collecting and processing new data, this section aims to equip practitioners with the tools and methodologies necessary for each stage.

By following these guidelines, local actors can ensure that their data-driven efforts are well-structured and aligned with the ultimate goal of producing insightful analyses that inform and improve local planning initiatives.

2.1. Identifying priorities: the need for data

Data is essential for any form of rural policy, as planning involves making informed decisions that must be based on sufficient and relevant information. It is required at every stage of the planning process, from the initial identification of a problem or objective, through the evaluation of alternative policies, programs, or projects, to monitoring the implementation process, and finally assessing the impact of the policy. In rural areas, the type of data required and the challenges associated with collecting and using it can vary depending on the diversity of rural areas and on the specific type of policy.

To assist in the process of identifying priorities, stakeholders may use the Rural Diversity Compass (Fig. 2), which helps to navigate the complex landscape of rural planning by highlighting key areas of concern. The compass enables stakeholders to systematically evaluate the place-specific differentiation tendencies based on particular functionality dynamics of rural areas, considering various perspectives, and thus aligning policy goals with the needs of the community. By using this structured approach, planners can ensure that their efforts are both strategic and responsive to the unique challenges of rural areas, ultimately leading to more effective and impactful outcomes.

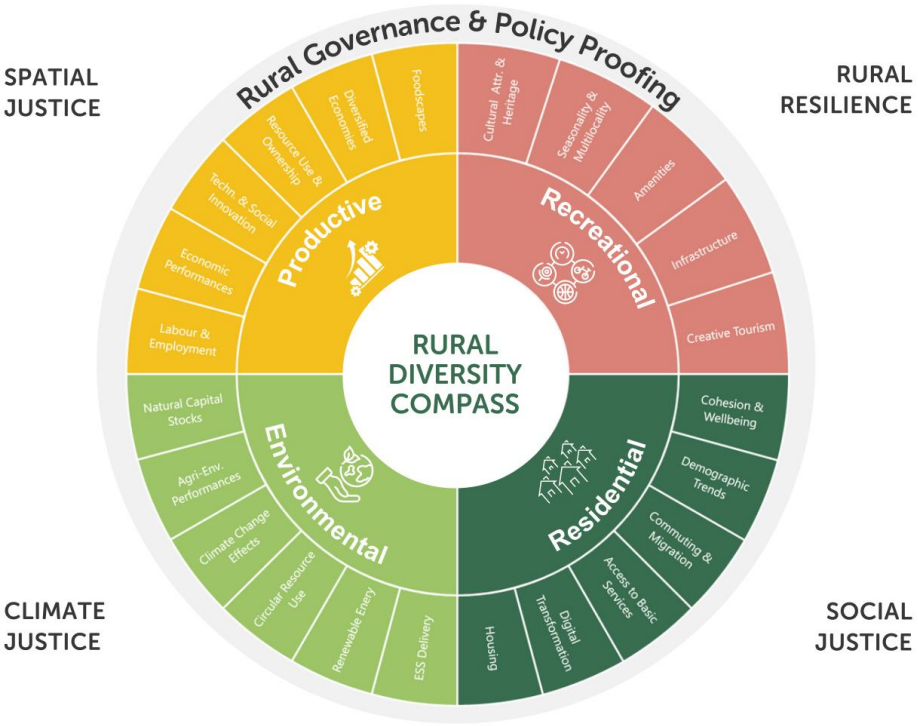


Figure 2: The Rural Diversity Compass.

Once the priorities identified, local stakeholders need to identify relevant indicators. When selecting indicators for monitoring and evaluating processes, it is essential to recognize that the relevance of indicators varies depending on the context in which they are applied. For instance, while gross national product (GNP) is a significant indicator at the national level, it may not be as relevant at the local level, where household income provides a more accurate reflection of economic conditions within smaller rural communities. Conversely, certain issues, such as energy systems, are managed on a national scale, making local data less relevant. Energy infrastructure, supply, and regulation are typically planned and maintained at higher administrative levels, where national data offers a more comprehensive understanding of system efficiency, costs, and resource allocation across the country. This highlights that there is no universal set of indicators that can be applied across all scenarios. Analysts must carefully consider what information will be most meaningful to the users, particularly those involved in designing and funding effective policy interventions or assessing the impact of such policies.

IN PRACTICE

Preliminary assessment: Identification of priorities and data gaps

The Living Lab of **Val di Cecina Rural District** in Italy focuses its data activities on systematic changes in food systems, climate change adaptation, and sustainable tourism. The identification of priorities was based on a **SWOT analysis** (Strengths, Weaknesses, Opportunities, Threats), an extract is shown below,that included the active involvement of stakeholders and led to the preparation of their Territorial Economic Plan. Following a stakeholder mapping exercise, new local stakeholders were engaged through public and ad hoc meetings for facilitate the process of prioritization, scoping of available data and identification of data gaps.

Food system			
Strengths	**Weaknesses**	**Opportunities**	**Threats**
• dissemination of the organic management method for over 30% of the UAA • high number of farms or businesses linked to agriculture • pedoclimatic suitability for quality agricultural production • presence of denominations and quality certifications • consolidated presence of specialized crops of high quality (vines and olive trees in primis) • agricultural economy to a certain extent diversified and significant presence of multifunctional companies	• lack of quantity in the supply of quality products • increase in the average age of farmers and poor generational turnover • insufficient entrepreneurial culture • absence of adequate transformation logistics • lack of a structured territorial platform for distribution logistics • scarce propensity for intercompany or supply chain projects in both the agricultural and forestry sectors • insufficient concentration of supply and strong dependence on market fluctuations • for some sectors, insufficient structuring of origin and quality traceability systems	• increased demand for organic products • renewed interest in setting up young people in agriculture • renewed interest in direct purchases by consumers and tourists • presence of subjects of agricultural cooperation interested in local productions and collaboration, in terms of logistics and marketing • partial attention of large-scale distribution towards branded, typical or tracked products • widespread increase in knowledge of origin and quality certifications • possible construction of the plant production processing centre in the former slaughterhouses of Pomerance	• prevalence of large-scale distribution in the context of supply chain relationships • decline in livestock activity • abandonment of crops in marginal areas or with mechanization problems • growing bureaucratization of the sector, with the imposition of onerous standards of hygiene and food safety, safety in the workplace, animal welfare and environmental compatibility • globalization of markets with exaltation of the phenomena of competition on international ones • political uncertainties and conflicts • climate change and consequent impacts on production and on the territory • difficulty in accessing credit

Extract of the SWOT analysis on Val di Cecina Food system

Based on the SWOT analysis, the Living Lab then identified what data and indicator were **available from official data sources** nationally (Istat) and regionally (Artea) and consequently whatdata were insufficient or **missing** at more appropriate scale (the “wishlist”).

Priority	Indicator and data	Wishlist
	Please specify current data being used by local governments and actors to understand the territory	*Please define preliminary assessment of data/indicators/tools needed for the territory in each of your thematic priority area:*
Food system	• ISTAT (2010 database) trend, composition by age and level of education of the population, number and type of businesses, workers employed by sector, total agricultural area, number and composition of farms by size, number of livestock farms and reared animals, PDO and PGI cultivated areas • ARTEA (annual database) declared agricultural area and types of crops, • Regione Toscana classification of rural areas	Data relating to the needs of the agri-food sector, in terms of: • quantity and type of product; • types of transformation processes of agri-food products present in the LL area; • number of infrastructures for the transformation of agricultural products present in the LL area; • quantification of agri-food consumption in local public canteens as a demand for local products; • quantification of agricultural waste and verification of the possibility of their reuse in the territory of LL

Overview of the food system in Val di Cecina and related indicator, data and wishlist

During this phase, the Living Lab faced challenges due to the **complexity of overlapping databases**, **the multiplicity of providers and the absence of directly available data.**

2.1.1. Criteria for Selecting Indicators

For an indicator to be valid and useful for decision-making and in public discourse, it must meet certain minimum requirements. According to The Wye Handbook (p.148):

"Indicators on rural development need to be based on (1) published strategies that are (2) consistently collected in (3) comparable areas, using the (4) same unit of measurement and based on a (5) clear definition. Indicators should also be (6) sensitive to changes and trends over time that can inform future policy direction."

2.1.2. Generic Quality Criteria for Indicators

Analysts generally agree on several generic quality criteria when selecting indicators. The quality of the underlying data significantly influences the quality of the constructed indicators. Data quality and metadata quality requirements are crucial and include:

- **Relevance:** Indicators should reflect the legal and institutional environment, resource availability, and transparency.
- **Comparability:** Concepts, definitions, scope, and classifications should be consistent across different areas within the country, ensuring that comparisons are valid.
- **Accuracy:** The source data should be reliable, with proper assessment and validation of intermediate data and statistical inputs.
- **Timeliness and Punctuality:** Indicators should be based on data that is current and updated regularly, with appropriate periodicity and geographical coverage, and clear policies for data revision.
- **Accessibility:** Both data and metadata should be accessible to users, ensuring transparency and ease of understanding.

A widely recognized approach for ensuring that indicators are well-defined and effective is the SMART framework, which stands for Specific, Measurable, Achievable, Relevant, and Time-bound. This framework aids in the formulation of robust indicators by ensuring clarity in purpose (Specific), the ability to quantify or qualify progress (Measurable), the feasibility of achieving the targets (Achievable), the alignment with overarching goals (Relevant), and the presence of a clear timeline for evaluation (Time-bound). Applying the SMART framework can enhance the reliability of indicators and ensure they effectively guide decision-making processes, complementing the broader quality criteria of data such as relevance, comparability, and accuracy.

2.1.3. Feasibility and Clarity of Indicators

In any context, it is crucial to consider whether it is feasible to construct an indicator, meaning the necessary data must be available and reliable. Furthermore, the ease with which an indicator's meaning and construction can be explained is important. The simpler and clearer an indicator is, the more likely it is to be understood and accepted by a diverse and potentially broad audience.

By ensuring that selected indicators meet these criteria, policymakers and analysts can create a robust framework for monitoring and evaluating rural development efforts, ultimately guiding more effective interventions and policy decisions.

Once indicators are identified, stakeholders will have to identify the data needed to construct them, as detailed in the following section.

IN PRACTICE

Selection of indicators for catching the diversity and complexity in Northern Sweden

The Living Lab of Northern Sweden encompasses four regions: Västernorrland, Jämtland-Härjedalen, Norrbotten, and Västerbotten. Its goal is to better understand the diverse characteristics of rural areas within this territory. Northern Sweden is known for its vast, cold, and sparsely populated landscapes, rich in natural resources, green energy potential, and innovation capacity. However, challenges such as long distances and demographic shifts pose significant issues. To address these, **the Living Lab explores various dimensions, including housing and property, accessibility, innovation, and rural development**, through a multidimensional approach.

Landscape of Northern Sweden (photos by Andreas Johansson)

For instance, various indexes recognize that rural areas in Northern Sweden are generally rated well in terms of service provision. However, this assessment often overlooks the fact that travel times to the nearest hospital can span several hours, or that some health centres operate only one day a week. To address these gaps, **the Living Lab seeks to develop indicators that provide a more nuanced understanding of service accessibility**, considering factors such as the costs associated with maintaining healthcare facilities in rural areas. This approach aims to create a more accurate picture of service availability by accounting for both accessibility and the financial implications of delivering services in remote locations. The Living Lab aims to develop the following indicators to better capture the realities of service accessibility and associated costs in rural areas:

a) An indicator that measures **service availability**, including travel time to the nearest service point, providing a clearer picture of how long it takes to reach essential services like hospitals or health centres.

b) An indicator that links service availability to the **costs of providing these services** in rural areas, helping to assess the financial burden of maintaining essential services across large, sparsely populated regions.

c) An indicator that estimates the cost for the public sector of supporting an **aging population**, considering factors like healthcare, social services, and infrastructure needs specific to rural demographics.

2.2. Identifying data needs

One of the most fundamental yet challenging tasks in any policy planning process is determining which data to collect. Without this clarity, time and resources may be wasted on gathering unnecessary data, while essential information could be overlooked. This section explores key considerations related to the quantity, quality, type, form, and source of data that should be addressed before data collection begins.

2.2.1. How much data is needed?

The first step is to clearly define the specific data required to characterise the priorities to be tackled. Rather than collecting all potentially relevant data, it is important to focus on what is essential for informed decision-making. This is especially crucial at the local level, where resources for data collection and analysis are often limited. Stakeholders must strike a balance between the ideal and the practical, aiming to collect the minimum amount of data needed to meet the priority goals. Any significant gaps should be acknowledged, as they may affect the accuracy or comprehensiveness of the decisions made.

2.2.2. Quantitative or Qualitative data?

The most effective approach to presenting a clear picture of an issue or situation is often through a combination of quantitative and qualitative data. However, these two types of data are not interchangeable, as each has its own strengths and weaknesses.

When deciding whether to collect quantitative or qualitative data, several factors should be considered:

Purpose: Quantitative data is ideal for precision and statistical analysis, while qualitative data provides detailed insights into complex issues.

Subject Matter: Some topics, like demographic information, are easily quantified, while others, such as cultural beliefs, are better suited to qualitative description.

Data Collection Method: Quantitative data typically involves structured surveys, while qualitative data relies more on detailed observations or interviews.

Data Presentation: Qualitative data can sometimes be converted into quantitative form, such as by categorizing attitudes for statistical analysis (eg. 'high', 'medium', 'low'; 'strongly agree', 'agree', 'neutral', 'disagree', 'strongly disagree').

IN PRACTICE

Combination of quantitative and qualitative data: mixed methods for essential services

Guaranteeing the **accessibility and availability of essential services** (public and private) is crucial for facing contemporary challenges in rural areas and contribute to broader wellbeing. This need is shared by **P10 Living Lab** - partnership of 33 rural municipalities across **the Netherlands** - which has decided to start its investigation with the objective **to have better insight into essential services** in its areas and create a policy guide for municipalities.

The Living Lab adopted step by step approach reviewing data based on **National databases** such as StatLine - CBS. These databases gave various information on basic services availability and accessibility on local level for an extended period. Especially StatLine that provides pictures, regional statistics and afterwards spatial data with the possibility to check proximity to facilities.

Moreover, the Living Lab considered necessary to integrate quantitative data with **qualitative data** through the **use of interviews with civil servants and inhabitants of villages to catch different perceptions** and collect data on the importance of essential services.

P10 Report "Uitkomsten van het inventariserend onderzoek basisvoorzieningen in 3 P10-gemeenten"

The final result was an Official Report (image above and available in Dutch) prepared by researchers from the University of Wageningen and the P10 facilitator highlighting concerns over accessibility to basic services, which is decreasing in rural areas (e.g., loss of bus stops, centralisation of hospitals).

2.2.3. How objective should the data be?

A distinction is often made between objective data, which is independent of the attitudes or biases of those involved, and subjective data, which is influenced by such biases. In practice, however, data is rarely completely objective, as the process of deciding what to collect introduces some level of bias.

Example: social indicators might reflect the priorities of planners rather than the general population. It is important to recognize that quantitative data is not inherently more objective than qualitative data, as both can be influenced by the methods of collection and analysis. The key is to be aware of the limitations of the data and the potential impact on decision-making.

2.2.4. How should the data be disaggregated?

Before collecting data, it is important to consider how it should be disaggregated, as this affects both collection and analysis. Data for rural planning is often disaggregated by geographical area or social group. Geographical disaggregation helps compare characteristics of different areas, which is essential for prioritizing needs or focusing on specific locations and their diversity. Social group disaggregation is crucial for identifying variations and inequalities, which is particularly important in social planning aimed at reducing disparities between groups.

2.2.5. What scope for the data?

In rural policy, the scope of data collection must carefully distinguish between the producer and consumer perspectives. For instance, when analysing electricity in rural areas, it is important to differentiate between the electricity generated by local energy producers, such as wind farms or hydropower plants, and the electricity consumed by rural households and businesses. The choice of scope directly affects policy decisions, as focusing on production highlights the region's role in contributing to national energy goals, while a focus on consumption reveals local energy needs and challenges, such as access and affordability.

Policymakers must select a perspective that aligns with their objectives, whether they are allocating responsibility for emissions, designing infrastructure investments, or improving local energy efficiency. Understanding both production and consumption dynamics is essential for crafting effective rural energy policies that balance regional contributions with local needs.

2.2.6. When should data be collected?

The timing of data collection depends on the type of data and its intended use. Data can be collected as a one-off snapshot, at regular intervals (time series), or before and after a specific event (for evaluation purposes). Each approach serves different purposes, from understanding the current situation to tracking changes over time or assessing the impact of a project.

2.2.7. Primary or secondary data?

Data can be classified as primary or secondary. Primary data is collected specifically for the planning purpose at hand, while secondary data has been collected for other purposes but can be re-purposed. It is advisable to explore the availability of secondary data before embarking on a primary data collection exercise, as this can save time and resources. In most planning activities, a combination of primary and secondary data is used, with secondary data often serving as a starting point.

2.3. Inventory of secondary data

Secondary data, which is data collected by someone else for different purposes but used for a new analysis, plays a critical role in rural development, in part due to data scarcity in rural regions. This data can be publicly available or proprietary - whereby some propriety data is accessible via data agreements, web scraping, etc. The quality and usefulness of secondary data can vary significantly, but in many EU countries, local authorities (NUTS2/3) are encouraged to maintain a central repository of such data. This repository, either kept in a planning office or online, should be regularly updated and easily accessible for research and planning purposes. Unfortunately, such organized platforms do not always exist, requiring local actors to gather secondary data anew for each project. Those regularly involved in decision-making should familiarize themselves with available secondary data sources to expedite the process when needed.

Within GRANULAR we have made an initial effort to inventory a wide variety of available datasets for rural development here. Furthermore, we performed a benchmark of the performance and costs of a selection of datasets available here.

After giving an overview of relevant open data banks where a wide range of data can be found, this section will explore five key types of secondary data that are particularly relevant to the social aspects of rural area planning and are typically accessible at the local level: demographic data, sample agricultural household surveys, official government records, maps, and general knowledge. These categories are not exhaustive but provide a sense of the range of secondary data available and the challenges associated with its use.

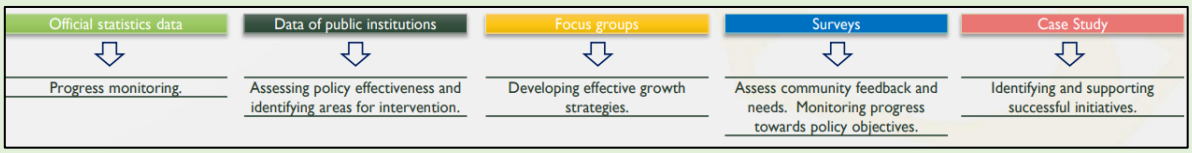
IN PRACTICE

Secondary data: Using different data sources

The region of **West Pomerania Voivodeship (Poland)** is characterized by large farms, a high degree of mechanization, and favourable conditions for high commodity production. However, the territory has **alternative food networks** that need to be supported to increase their competitiveness.

The Living Lab identified needs for new data that can help strengthening and consolidating short food supply chains. First, it identified available data and missing information using **different tools and methods**, such as desk research and surveys, and analysed data from public statistics that were then **integrated with various strategic documents** at the regional level.

The Living Lab members used regional/national datasets (stat.gov.pl) and then explored **alternative data sources** from documents developed at the regional level. The data collected included information on: retail and wholesale sales, number and structure of marketplaces, number of local and organizational marginal entities, number of agricultural retail operators, number of direct sales entities, land use structure, and number of organic producers. The verification of non-public data and the determination of data collection methods was done through meetings with researchers, experts and local actors (farmers).



The different data sources and the beneficial outcomes for practitioners (sources: Granular KTA 18/03/2025)

The diagram above, available in its full version here, summarizes the main results of these activities and recommends the use of the different data sources and approaches at certain stages, from progress monitoring – through official statistics data and surveys – to the development of effective strategies with focus groups, to the strategies development with focus groups.

2.3.1. Data Banks

In the rapidly evolving field of rural development, access to comprehensive and reliable data is crucial for informed decision-making and effective planning. To address this need, we have compiled a curated list of authoritative databanks that offer a wide range of relevant information. The following table presents an overview of 13 key data sources, each specializing in different aspects of rural and global development. These resources span from European statistics and global development indicators to environmental data, geospatial information, and agricultural statistics. Collectively, they provide a robust foundation for researchers, policymakers, and practitioners working in rural development.

The table below includes the name of each databank, a concise description of its primary focus, and the corresponding URL for easy access. This compilation aims to streamline the data discovery process, enabling more efficient and data-driven approaches to rural development challenges.

Table 1: Compilation of main databanks relevant to rural data

Databank	Description	URL
Eurostat	European statistics	https://ec.europa.eu/eurostat/fr/
World Bank Data	Global development indicators	https://data.worldbank.org/
Copernicus	Earth observation data	https://land.copernicus.eu/en
ArcGIS Open Data	Geospatial information	http://opendata.arcgis.com/
SEDAC	Socioeconomic and environmental data	https://sedac.ciesin.columbia.edu/
Zenodo	Multidisciplinary research outputs	https://zenodo.org/
Natural Earth	Public domain map data	https://www.naturalearthdata.com/downloads/
UNEP	Environmental publications and data	https://www.unep.org/publications-data
IPUMS Terra	Integrated population and environmental data	https://terra.ipums.org/
EcoDataCube	Ecological datasets	https://stac.ecodatacube.eu/
OpenStreetMap (via Geofabrik)	Crowdsourced geospatial data	https://download.geofabrik.de/
Free GIS Data	Curated list of geographical information sources	https://freegisdata.rtwilson.com/
FAO Catalogue	Agriculture-related data	https://data.apps.fao.org/map/catalog/srv/eng/catalog.search

Table 1: Data Banks list

2.3.2. Demographic data

Demographic data provides information on the size and structure of a population, including total population, household size, age and sex distribution, growth rates, fertility, mortality, migration, and population density. This type of data is crucial for various local level planning activities, whether the need is for aggregated local data, data disaggregated by administrative subdivision or social group, or data specific to a particular area or community.

Obtaining accurate demographic data without a comprehensive house-to-house survey is challenging and resource-intensive. Therefore, secondary data, particularly from national population censuses, is often used. Censuses, typically conducted every ten years, provide a wealth of information, sometimes supplemented by inter-census sample surveys. These censuses often include data on occupation, education, and access to services.

However, there are two primary challenges with using census data at the local level. First, the data may not be disaggregated in a way that aligns with local planning needs; census enumeration areas might not correspond to local administrative units, and the data might not be broken down by social criteria such as gender or other relevant socio-economic characteristics. Secondly, the data can be outdated by the time it becomes publicly available, creating difficulties in communities with rapidly growing or unevenly distributed populations, where accurate projections are essential.

Examples of openly available demographic data include GISCO, the Global Human Settlement Layer and WorldPop.

2.3.3. Agricultural surveys

Sample agricultural household surveys are regularly conducted by agricultural ministries or national statistical agencies to gather basic data on agricultural production volumes and methods.

These surveys usually sample households from different agro-economic zones and collect data over a full agricultural year. Information gathered typically includes household size, cultivated area by crop, inputs (including labor), yields, crop sales, income, and livestock numbers and sales.

This data is vital for policy planning, especially in the agricultural sector. However, its utility at the local level depends on the sample size, selection method, and data disaggregation. Often designed for national planning, these surveys may have relatively small samples and use nationwide agro-economic zones as the basis for selection and disaggregation.

Examples of available data include harmonised datasets from payments from the European Union's Common Agricultural Policy (Nicholas et al., 2021), harmonised national agricultural census datasets (Villani et al., 2019), or geo-referenced polygons of agricultural croplands (Schneider et al., 2023).

2.3.4. Local official records

Official government records, maintained by local government agencies and submitted periodically to national headquarters, are another valuable source of secondary data for policy planning. These records typically include reports on activities and problems and statistics, primarily intended for national planning and monitoring. These records contain a wealth of information that can be repurposed for local policy making. These records can be adapted to collect additional information while still meeting national reporting requirements.

Examples of this data can be found at the repository of Eurostat.

2.3.5. Other spatial data

Spatial data are especially useful for illustrating geographical variation within a territory and showing the relationship between different functionalities in a specific area, or to highlight socio-economic structures relevant for policy elaboration.

Examples of openly available EU-wide maps include: Open Street Maps for infrastructure and services, and Copernicus Services for climatic data, land cover and more.

2.4. Primary data collection

Primary data, as previously mentioned, refers to data collected specifically for a particular purpose. In most types of policy planning activities, collecting some primary data is necessary because secondary data rarely provides all the needed information. This is especially true for socio-economic policies, where secondary data often lacks the detailed insights into people's needs, problems, and attitudes that such planning requires. There are several ways to collect primary data, which depend not only on the type of information needed but also on the resources available for data collection. While collecting primary data generally demands more time, money, and human resources compared to using secondary data, some methods are relatively cost-effective. Primary data may also involve the collection and processing of personal data, which must be protected: appropriate processes must therefore be implemented to ensure the protection of all personal data of research participants, in compliance with the EU General Data Protection Guidelines (GDPR) or equivalent that pertain in each country (see Annex 1).

2.4.1. Surveys

Survey methods in rural areas encompass a range of techniques tailored to capture both socio-economic and environmental data, each with its own strengths and challenges. Socio-economic surveys often employ household questionnaires, focus group discussions, and key informant interviews to gather information on demographics, livelihoods, and community dynamics. These methods allow for in-depth understanding of rural social structures and economic activities. Environmental surveys, on the other hand, frequently utilise field sampling, remote sensing, and GIS mapping to assess land use, biodiversity, and natural resource management. While socio-economic surveys typically require direct interaction with local populations, environmental surveys may involve more technical data collection methods. Both types of surveys benefit from participatory approaches, such as community mapping or citizen science initiatives, which can enhance data accuracy and local engagement. However, as rural areas often present unique challenges, including limited accessibility, sparse populations, and seasonal variations, surveys may necessitate adaptive sampling strategies.

The choice of survey method depends on the specific research objectives, available resources, and local context, with many rural development projects adopting mixed-method approaches to capture the complex interplay between socio-economic and environmental factors. With regard to quantitative data, the most commonly used tool is the questionnaire. For qualitative data collection, the main methods are interviews and focus groups.

Sampling

When conducting surveys, different sampling techniques are used to select respondents based on the study's goals, population size, and logistical constraints. Most common sampling types include census, sampled surveys, key informant, snowball sampling (Table 2).

Table 2: Different sampling methods used in surveys

Sampling Type	Definition	How it works	When to use	Pros	Cons
Census	Collecting data from every individual in the population.	Survey the entire population.	When full data coverage is needed or when population is small.	Comprehensive and accurate. No sampling bias.	Time-consuming and expensive Logistically difficult for large populations.

Sample	Collecting data from a subset of the population.	Select a representative subset using random or stratified sampling methods.	When it's impractical to survey the entire population.	Cost-effective and faster than a census. Results can be generalized to the population.	Potential sampling error. Bias may occur if the sample isn't representative.
Key Informant Sampling	Selecting individuals with specialized knowledge or insights.	Interview key individuals, such as community leaders or experts.	When in-depth, context-rich data is needed from knowledgeable individuals.	Provides detailed insights. Requires fewer participants, making it less resource-intensive.	Not representative of the whole population. Results cannot be generalized.
Snowball Sampling	A method where existing participants recruit future subjects from their networks.	Initial participants suggest others to be surveyed, creating a "snowball" effect.	For hard-to-reach or niche populations.	Effective for accessing hidden populations. Cost-effective, as participants help recruit others.	High risk of sampling bias. Limited diversity, as participants are often from the same network.

The term 'census' generally refers to a survey that includes all households in a specific area to gather basic data on household characteristics. When planning a project in a specific community, where basic information on all households is needed to design project details or provide baseline data for evaluating the project's impact later, it is possible to conduct a census focusing on a very small area.

Sample surveys are useful methods for obtaining statistically significant information at a unit level. Because these surveys involve only a sample of the total population, they require fewer resources than a traditional census. and are, therefore, more feasible for gathering information for policy planning. However, they still demand a significant amount of time, money, and human resources. As a result, they are typically used for specific purposes within a limited geographical area rather than to provide basic data for an entire rural area.

The sample for a survey can be selected in three main ways, depending on the type of information needed:

- **A random sample** is selected completely at random, without any prior knowledge or consideration of specific characteristics (e.g., by choosing every n^{th} household from an alphabetical list or as they appear on the ground).
- **A stratified random sample** is chosen by first dividing the population into categories based on certain predetermined characteristics and then selecting a random sample from each category.
- **A purposive sample** is selected based on one or more predetermined characteristics, with the aim of obtaining information specifically about those members of the population who exhibit these characteristics.

In all cases, the sample size is crucial because it must be large enough to provide data that is statistically representative of the entire population. The minimum sample size depends on the population size and the type of sample being used. For instance, the proportion of the population included in the sample should be larger for smaller populations than for larger ones, and random samples should generally be

larger than stratified random or purposive samples. However, the sample size should typically not be less than about 10% of the population[1].

Questionnaire

A questionnaire is generally considered the most important quantitative data collection tool. It involves asking a set of structured, semi structured, and/or open questions to a representative sample of people to gather information. It is generally used for producing quantified and statistically significant data representing a phenomenon. Questions can be about different aspects such as living conditions, access to services, agricultural activities, and community needs. A questionnaire can be administered by surveyors (in person, over the phone, or online) or self-administered (on paper or digital).

A questionnaire can be used to apprehend the challenges faced by rural communities, assess their needs, gather data on various issues, and monitor changes over time. This information can help policymakers and organizations to develop targeted programs and policies to improve living conditions and address specific issues in rural areas.

Example: a territory wants to understand the living conditions in its rural areas. They conduct a survey, asking people questions about their income, education, and access to healthcare. They also gather administrative data from government agencies about agricultural production and infrastructure. By analyzing this information, the government can make informed decisions about improving rural development and providing better services to its rural population.

Interviews

Interviews are a data collection method that involves an individual conversation between a researcher and the participant. The researcher asks questions to understand people's thoughts, experiences, and feelings about specific topics. Interviews may differ in their degree of standardization, that is, the different degree of freedom or constraint granted to the two actors: they can be unstructured, semi-structured and structured (structured interviews can be considered equivalent to a questionnaire).

Unstructured or semi-structured interviews are considered among the most important qualitative data collection methods. They can be conducted face-to-face, by telephone or online. Their main objective is an in-depth understanding of facts, thoughts, and feelings from the point of view of the interviewee. The interviewer guides the conversation with a flexible and non-standardized scheme that allows to access the perspective of the interviewee. The views of social actors are captured on two related levels of their social experience: the level of objective practices and facts, and the level of subjective representations, judgements, perceptions, or worldviews.

Focus group

A focus group is a qualitative data collection method based on bringing together a small group of people to discuss a specific topic of interest. It allows to create new knowledge and skills through the exchange of ideas and experiences between the participants. A moderator leads the discussion, asking open-ended questions and encouraging participants to share their opinions and experiences. While interviews are effective for collecting data on personal histories and perspectives, focus groups are optimal in eliciting data on social and cultural norms, issues of concern, and opinions of a group or different subgroups of stakeholders. They may be conducted in person or online.

[1] It is possible to calculate the sample size to collect the correct number of responses by using online tools (e.g. https://www.surveymonkey.com/mp/sample-size-calculator/)

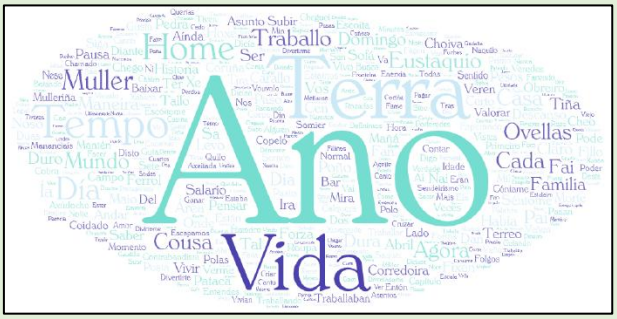
Focus groups provide a platform for community members to interact, share diverse perspectives, and collectively identify key issues and solutions. Involving different participants is relevant: they influence each other through their attitude, words, and reactions to what others say. Not everyone will have the same views and experiences – due to differences in age, gender, education, resources, and other factors – and many different points will likely be expressed. The involvement of stakeholders in problem-solving makes decision-making more appreciated: everyone can contribute to plan a shared path. Also, an added value of this technique may be the possibility of finding unexpected results. The richness and depth of focus group data emerges from the group dynamic and depends on a careful management of the dynamic of the discussion.

Example: a non-profit organisation is working on a project to improve access to healthcare services in a remote rural community. The organization arranges a focus group with local community members, including farmers, teachers, healthcare workers, and elders, asking questions such as "What challenges do you face in accessing healthcare?". The participants share their experiences and perspectives on issues such as distance to medical facilities, transportation difficulties, and the affordability of healthcare services. The organisation learns about the lack of nearby healthcare facilities and the financial constraints that hinder access to medical services, which helps the organisation to develop targeted solutions, such as mobile medical clinics, health awareness campaigns, or partnerships with healthcare providers to bring essential services closer to the community.

IN PRACTICE

Qualitative data: Unconventional data sources and mixed methods

The Ourense province is predominantly rural with a low population density. The province is affected by the decreasing of number of inhabitants, while it has one of the **highest life expectancies** in Spain and Europe. Building on the availability of traditional demographic data, the Ourense Living Lab aims to explore the different factors that contribute to the **ageing and hyperlongevity** of its population and identify **existing and needed services**. To do this, the Living Lab used a mixed methods approach for its data collection activity, based on **cultural content**. It then provided an **analysis of unpublished interviews** in collaboration with **audiovisual companies and social organisations**.



Data collected by Ourense LL based on cultural content, the results is the count of most repeated words

Qualitative data from interviews can reveal important details about daily life, crucial for understanding the factors that contribute to long and healthy lives in rural areas. In addition, the use of **word cloud** (image above)can capture what needs and social aspects to be considered. The final aim of this data collection was to transform pre-existing and unexploited information into new forms of data integrated with other forms of traditional methods such as **questionnaires**.

2.4.2. Rapid Rural Appraisals

Rapid rural appraisal techniques, combining elements of both socio-economic and environmental assessments, have gained popularity for their ability to quickly gather comprehensive data in resource-constrained settings. They emerged as an alternative to conventional social surveys for gathering information in resource-constrained rural development planning. The concept has expanded beyond primary data collection, becoming associated with participatory planning due to its emphasis on local involvement. However, its core purpose remains collecting adequate primary data with limited resources.

The goal of rapid rural appraisal is to obtain sufficient information for specific planning objectives within constraints of time, money, and personnel. While the data may be less comprehensive than traditional surveys, it is often more qualitative and subjective. A well-conducted rapid appraisal can yield more accurate results than a poorly executed conventional survey. Additionally, the participatory nature of these techniques can enhance project sustainability.

Rapid rural appraisal encompasses various approaches, often combining multiple techniques to cross-check information and mitigate inaccuracies or biases. The versatility of rapid appraisal techniques makes it impossible to exhaustively list or describe them all. Examples include modified village surveys to assess crop production or using diverse methods to gather information about local farming systems. These techniques can supplement limited data from national surveys or serve as alternatives to full censuses.

2.4.3. Sensors and observation

Earth Observation

Earth observation involves using satellites and other remote sensing technologies to collect data about the Earth's surface and atmosphere from space. Satellites equipped with various sensors capture images and measurements, providing valuable information about land cover, vegetation, weather patterns, and environmental changes in rural areas. This data can be used to monitor agricultural practices, assess natural resources, track changes in land use, and detect potential environmental hazards like deforestation or droughts. Earth observation plays a crucial role in rural areas, assisting farmers in making informed decisions, supporting disaster response and planning, and contributing to sustainable development initiatives for rural communities.

Example: a regional government is developing a spatial plan to guide sustainable land use and development in a rural area. The regional government utilises Earth observation data to create accurate land-use maps, identify areas of ecological significance (e.g., wetlands, forests), and assess the impact of human activities on the environment. This information helps in formulating spatial planning policies and zoning regulations that promote responsible development, protect ecologically sensitive areas, and support balanced urban-rural growth. Earth observation data enhances the spatial planning process by providing objective and up-to-date information, ensuring that decisions are based on a comprehensive understanding of the rural area's current status and potential for sustainable development.

IN GRANULAR

We explore the use of Deep Learning algorithms to detect and map landscape features using Earth observation data connected with geo-localized photos and text prompts. The final goal is to produce probability maps showing the distribution of specific concepts like rural structures, forest types, and human activities, expressed in natural language.

Transaction data

Transactional data refers to information collected from everyday transactions, such as purchases, phone calls, or internet activities. In the context of rural areas, transactional data can be obtained from mobile phone network operators to understand people's travel patterns, movement, and communication. By analyzing this data, researchers and policymakers can gain insights into rural mobility, commuting habits, and social interactions. This information helps in planning transportation infrastructure, improving connectivity, and making informed decisions about public services in rural communities.

Example: a transportation authority wants to improve public transport in rural areas. They use data from mobile phone networks to analyse people's commuting patterns. By understanding where people travel most and when, the authority can optimize bus routes and schedules, making it easier for rural residents to access essential services and jobs.

IN GRANULAR

We explore the use of location-based information available through Facebook's platform to characterise mobility patterns. By examining check-ins, location tags, and other location-sharing features, we gain insights into people's movements and travel patterns. This data provides valuable information about where individuals live, work, and travel, helping to understand commuting habits and mobility between different areas. By aggregating and anonymizing this data, we can study broader trends in mobility within and between regions, aiding in spatial planning, transportation optimization, and addressing mobility challenges in rural areas.

2.4.4. Citizen science

Crowd-sourcing

Crowd-sourced data is information collected from a large group of people, often through mobile apps or online platforms where individuals voluntarily contribute data. In the context of rural areas, people can share their experiences, opinions, and observations about various aspects like road conditions, access to services, or environmental changes. This data can be used to create maps and analyze trends, providing valuable insights for rural development.

Example: individuals, community groups, and organizations can participate in crowd-sourcing data for OpenStreetMap by adding and editing geographic information such as roads, buildings, landmarks, and points of interest. They use satellite imagery, GPS data, and their local knowledge to make contributions to the map. For instance, if someone notices that a new road has been built in their neighbourhood, they can use the OSM platform to add the road to the map. Similarly, they can add a new restaurant, update the name of a street, or make any other relevant changes.

IN GRANULAR

We use OpenStreetMap data to characterise accessibility to services in rural areas, as it provides detailed data on road networks, footpaths, public facilities, and other amenities in rural communities. By studying this data, we can assess the connectivity between different locations, identify areas with limited infrastructure, and characterise travel times to main services for residents. This information enables them to create accessibility maps, pinpoint regions with potential access challenges, and design targeted interventions to improve mobility and resource distribution in rural areas.

Citizen sensing

Citizen sensing involves individuals using their smartphones or other devices to collect data about their environment or community. In the context of rural areas, people can use their phones to measure air quality, report changes in land use, or document wildlife sightings. This data is then shared through apps or online platforms, creating a collective dataset that helps monitor environmental conditions and community activities. Citizen sensing enables rural residents to actively participate in data collection, giving them a voice in local decision-making and facilitating community-driven initiatives. It can be used to address rural challenges like environmental monitoring, natural resource management, and community engagement, promoting sustainable development and enhancing the well-being of rural communities.

Example: citizens can use personal air quality sensors to measure air pollution levels in their local areas. The sensors detect various pollutants such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), ozone (O3), and other harmful substances present in the air. The data collected by these citizen sensors are then aggregated and displayed on interactive maps or mobile apps, allowing the public to access real-time information about air quality in different locations.

2.4.5. Scraping

Web-scraping

Web scraping is a method of collecting data from websites automatically. It involves using computer programs to extract information from web pages, such as text, images, or tables. In the context of rural areas, web scraping can be used to gather relevant data about agriculture, weather, local events, or community services from various websites. This data can be analyzed to understand rural trends, monitor environmental changes, or assess the availability of resources in different regions. Web scraping helps researchers and policymakers access a wide range of information quickly and efficiently, aiding in evidence-based decision-making and supporting rural development initiatives with up-to-date and comprehensive data.

Example: a territory wants to gather data from different websites, such as housing prices. Instead of manually copying and pasting this information one by one, web scraping automates the task. It works by sending requests to websites (real estate agencies, government land registration), then the web scraper parses the HTML code (the language websites use to display information) and extracts the relevant data, such as the size of the house, its location, price, date of transaction.

IN GRANULAR

We collect and analyse real-time data on property transactions. The data obtained is valuable for studying the performance of local housing markets, particularly in areas where tensions may arise due to external shocks. By examining these property transactions, we gain insights into the dynamics of the housing market and make informed decisions to address emerging challenges in affected regions.

Table 3: Data produced in GRANULAR

Title	Description	Delivery	Coverage	Resolution	URL	Lead
Accessibility	Accessibility from population grid cells to a variety of services	Mid-2024	EU + UK	1km + NUTS	Zenodo Record	CNRS
Mobility	Facebook-derived mobility of people from one region to another	End-2024	Living Labs	NUTS3/LAU		UoS / UniMi
Housing prices	National housing prices, web scraped	End-2025	Sweden, Czech	NUTS3/LAU	Sweden record	NOR
Landscape beauty	One of the input layers to rural attractiveness	End-2024	Germany, EU	1km		TI
Land Use	Land use classification via LLM	Mid-2025	EU + UK	100m		IAMM
Tourism	Use of TripAdvisor data to determine tourism flows, attractions, seasons, and origin	End-2024	French LL (possibly upscaled)	1km + NUTS		IIASA
Climate neutrality	Indicators of climate neutrality	Start-2025	EU	NUTS3/LAU		IAMM
Socio-economic resilience	Economic diversity index	Start-2025	EU	NUTS3/LAU		NOR
Food systems	Summary metrics of food system dynamics	End-2025	EU	NUTS3/LAU		IAMM
Wellbeing	Summary indices of different subjective wellbeing dimensions	End-2025	EU	NUTS3/LAU		TI
Rural attractiveness	Rural attractiveness indicator	Mid-2026	EU	NUTS3/LAU		NOR
Rural typologies	Typology of rural areas	Mid-2026	EU	1km + NUTS		NOR

2.5. Analysis and visualisation

Analysis and visualization can take many forms and may range from simple methods up to very complex applications which require highly specialized skills. Ultimately the chosen approach will depend upon the skills of the analyst, the data availability, financial resources (to for example hire expert help and or purchase software) and the level of analysis required.

Depending upon the data at hand, various methods could be applied. If the data is text based, an initial screening and summary by an expert(s) in verbal or quantitative form (simple graphs, etc.) could suffice. Text can however be analysed with a variety of software and increasing levels of sophistication can be applied, moving from e.g. word clouds (e.g. Wordle) up to and including the use of large language models to synthesize and visualize the data (e.g. ChatGPT).

Example: representing qualitative data

Qualitative data can be represented visuallyand textually to help convey patterns, themes, or insights. In addition to the word cloud – as the case of Ourense Living Lab – there are several ways to represent qualitative data through:

Thematic and mind maps to visually show the relationships between different themes or codes within the data. They help illustrate how key themes are connected or layered.

Tables and matrices, tables can be used to systematically organize and compare qualitative data across categories, themes, or cases. This method is particularly useful for comparing different interviewees, locations, or groups.

Conceptual diagrams depict relationships between different concepts and variables. These are useful when you want to explain how certain factors interact or contribute to broader phenomena.

Photographs and visual data, when you have visual data from interviews (e.g., photographs of the area or community), you can use these visuals alongside quotes or themes to support the analysis.

Narratives and case studies, writing up detailed case studies or narratives is a common way to represent qualitative data. This method helps convey the depth and complexity of individual experiences or cases.

If the data is numeric, again another set of tools are available, typically quantitative, to analyze and visualize the data (e.g. R, MatLab, etc.). This becomes further enhanced if the data has a spatial component (e.g. QGIS, ArcGIS, etc.).

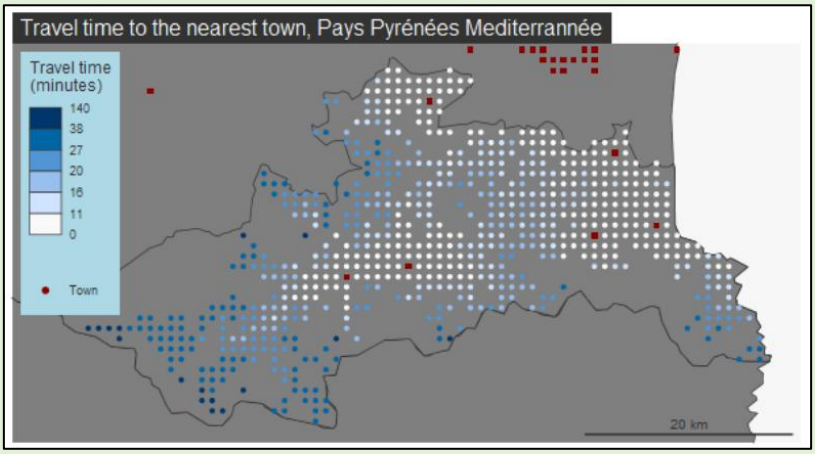
Basic tools for analysis also include Microsoft Excel or Google Sheets, which are relatively easy to use for most computer users, allowing for the creation of a variety of graphical objects. This can be a good starting point for initial exploration. Increasing the sophistication, packages such as R Shiny and Google Earth Engine allow for the creation of dashboards and advanced methods of visualization.

IN PRACTICE

Data capacity building: Facilitating exchanges between data producers and users

One of the objectives of the Pays Pyrénées Méditerranée Living Lab (France) is to **improve skills of local population** empowerment and raise awareness **on data and observation tools**. This Living Lab is located near the French-Spanish border along the Tech River and includes a **great diversity of rural areas** between mountain and coastline areas.

The Living Lab has promoted various meetings with the ambition to introduce local actors to the **data culture**, with a focus on **new forms and uses of data**, through direct relations and collaboration between local data communities and National Institutes.



Example from RIATE-CNRS on OSM use in PPM territory

For instance, through the meeting gathering **Local Open Street Map (OSM) Community members, GIS experts from municipalities, local urbanism association AURCA** and **researchers from RIATE-CNRS, the** Living Lab discusses how the OSM Community and the municipalities could work together to share knowledge including with citizens and use available data to better understanding mobility flows and outdoor activities. In addition, RIATE-CNRS shows several examples of OSM use applied directly to the PPM territory.

Workshop INSEE/IDESCAT

Due to the cross-border nature, the Living Lab organized a specific workshop with two statistical organizations: the **National Institute for Statistics** (INSEE) and the **Institut d'Estadística de Catalunya** (IDESCAT). The institutes shared with local stakeholders (municipalities, economic and job agencies, journalists...) how to access open data and some examples of how to exploit it. Moreover, they involved local actors through a quiz and practical exercises on specific territorial aspects such as population characteristic and tourism. Data users could also discuss their working practices and data wish list with producers.

For more info visit the official website: https://www.payspyreneesmediterranee.org

2.6. Towards policy formulation

To formulate effective policies in rural areas, reliable data is essential, but challenges exist in obtaining it across all sectors.

For example, in nutrition policy, data is needed on both the nutritional status of the community and the factors influencing it. While food intake data can be gathered through household surveys or rapid appraisal methods like group discussions, information on the physical impacts of nutrition—such as body size and weight—provides clearer insights for young children. Additionally, data on food production, storage, and cultural practices around food consumption is crucial but often difficult to collect. Improving nutrition requires addressing the broader issue of poverty, a root cause of malnutrition, and ensuring that subsistence production is improved, cash cropping does not negatively affect nutrition, and nutrition education is provided. Nutrition planning must be a coordinated, inter-departmental effort involving local communities to ensure sustainable and effective interventions.

In education policy, gathering data on both the supply of and demand for education is vital. Supply-side data, such as the number of schools, teacher-to-pupil ratios, and infrastructure, is typically accessible, but demand-side data is more complex. This includes information on the number of children eligible for various levels of schooling, the geographic distribution of the population, and the relevance of education to local labour needs. Mapping school locations and catchment areas can highlight gaps in educational access, while qualitative data on local workforce demands helps ensure curricula align with job opportunities. Education policy must also account for different levels of education, geographical and social disparities in access, the quality of instruction, and its relevance to local development. Though major decisions are made at the national level, local level planning still has some flexibility. Collaboration across government agencies and local community involvement is essential to ensure that education planning meets the diverse needs of the population.

For rural water supply policies, comprehensive data on both traditional and improved water sources is required, along with information on water use for domestic and productive purposes. Improved water source data is generally accessible, though records may not always reflect operational status. Collecting data on traditional sources often requires community visits and rapid appraisal techniques. Mapping these sources can identify areas with limited access, while cultural practices around water use, including rights of access and the role of women, must be understood for effective planning. Water policy must address both domestic and productive needs, access disparities, maintenance provisions, and the potential for cost-sharing. Hygiene education is also critical. Water supply planning requires interdepartmental collaboration and, importantly, community participation to foster a sense of ownership and responsibility for maintenance.

IN PRACTICE

Testing Rural Proofing

The aim of the United Kingdom Rural Scotland Living Lab is to identify and recommend approaches to **providing evidence that can support the monitoring and the evaluation of public policies** with particular significance for rural areas. A specific objective was to **test the role of the prototype process of rural proofing**, delivering to the GRANULAR work on tailoring rural policies.

The GRANULAR approach to rural proofing was tested with reference to **Scotland's Land Use Strategy**. The first strategy was published in 2011, and the second in 2016. The current strategy was published in 2021. This Third Strategy reflects an overarching and holistic approach of sustainable land use in Scotland, and away from a sector-by-sector approach. Its time horizon is towards targets of specific policies in 2032 (relating to biodiversity loss) and 2045 (relating to climate change), rather than its formal duration of 5 years. This represents the closest alignment in Scottish policies with the **EU Long Term Vision for Rural Areas** (LTVRA).

The test provided an opportunity to work through the process and thinking of rural proofing, reflecting upon its application in practice. The process necessitated reading the Strategy in detail, mapping it onto the components of the Rural Diversity Compass.

The Third Land Use Strategy was assessed as having no specific weaknesses with respect to rural areas, as measured by having a negative effect for any of the components (see Figure below). However, overall, it could be criticised for having a lack of specific policies on land use, instead relying on linking to policies and measures developed for sectors or topics.

The process of testing the rural proofing stimulated discussion of the differences within rural Scotland of potential benefits and implications of specific measures. Notable issues identified are those of **woodland expansion** and where there is biophysical and socio-economic scope for certain types of trees to be planted, and the management approaches to incorporating different mixes of tree species within stands at the level of estate planning.

The relative weaknesses of the Strategy were reported as those components of the **Rural Diversity Compass** for which the effect was assessed as being neutral, with a low magnitude of impact. Four such components were identified, listed under the Residential functionality (digital transformation, basic services, commuting and migration, and demographic trends), one under productive (foodscapes), one under environmental (circularity), and one under Recreational (seasonality and multilocality).

Functionality & Components	2.If Yes, please explain why *(provide examples of those effects)*	3. Is this effect direct or indirect? *(please elaborate your answer)*	4. How would you judge the **effect** on this component (-2 = very negative; -1 = negative, 0 = neutral, 2 = positive; 3 = very positive)	5. What is the magnitude **of the effect** on this component (1 = low; 2=medium; 3=significant impact)
Environmental				
Natural capital stocks This component focuses on rural areas as holders of natural resources and ecosystem services. It acknowledges the variety of available resources and associated biophysical flows.	The Strategy sets the context for protecting, managing and investing in natural capital. It recognises natural capital on an equal footing to people, social and economic capital as the four pillars in which investment underpins the economy.	The Strategy has a direct role in setting out the importance and role of natural capital stocks and their management. Natural capital is monitored by the Scottish Natural Capital Asset Index (NCAI) supplemented with information from UK National Ecosystem Assessment. This indicator is being updated and improved through Scottish Government Strategic Research Programmes.	*3*	*2*
Agri-environmental performances This component focuses on measuring agri-environmental performance using indicators for CAP monitoring. The objective is to understand the outcomes, opportunities, and limitations of agricultural activity in relation to societal ecosystem service concerns that go beyond food provision.	The promotion of sustainable practices in agriculture has a significant role in the Strategy. Cross-compliance is a set of rules made up of 'Statutory Management Requirements' (SMRs) and 'Good Agricultural and Environmental Conditions' (GAECs) stipulations. Land management practices are detailed in the Climate Change Plan (2018-2030) and sectoral plans.	Impacts will be indirect, through the measures and regulations which support and encourage agri-environmental actions.	*1*	*2*
Climate change This component aims to differentiate between climate change mitigation and adaptation aspects that can impact the functionality of rural areas.	The Land Use Strategy was introduced by the Scottish Parliament as a part of the Climate Change Act (2009). The roles of climate change mitigation and adaptation are recognised, and their impacts on the functions of rural areas.	The Strategy articulates policies on tackling climate change through the lens of land use. It's measures for implementation are focused on spatial planning and multi-actor development of approaches.	*1*	*1*
Renewable energy This environmental functionality component focuses on rural efforts to reduce fossil fuel dependence through various renewable energy sources like solar, wind, and hydro.	The Strategy provides part of the rationale for the spatial planning of renewable energy developments, and links to land characteristics and uses (e.g. solar farms on agricultural land; wind turbines on carbon rich soils). Most such developments will, be in rural areas.	The effect is indirect. The contribution of land uses to reducing greenhouse gas emissions is that of Land use, land use change and forestry (LULUCF) greenhouse gas emissions, supplemented with information on renewable energy.	*1*	*1*

Extract from the GRANULAR rural proofing template used in the test phase

The experience of the test showed that weaknesses in the Strategy could not be effectively articulated at the level of the four functionalities of the Rural Diversity Compass. For this purpose, those were not at a sufficient level of granularity and too heterogeneous in their categories.

The experience of the test will be valuable in preparing for the process of rural proofing using the final GRANULAR template developed to utilise aspects of the Long-Term Vision for Rural Areas and the Rural Diversity Compass.

3. Practical application: Results from Living Labs

The initial assessments by Living Labs highlighted thematic priorities, including mobility, tourism, biodiversity, agri-food value chains, ecosystem services, and well-being. These priorities guide the project's work packages (WPs), particularly WP3 and WP4, which focus on data collection and the creation of policy-relevant indicators.

Key initiatives and methodologies developed across different Living Labs include:

- **France:** Focus on quality tourism and diversified agriculture, development of travel-time OD matrices for accessibility based on OSM, and exploration of rural attractions using TripAdvisor data.
- **Italy:** Addressing agricultural abandonment and tourism through remote sensing imagery to monitor ecosystem services and land use changes.
- **Netherlands:** Investigating seasonal agricultural migration and enhancing well-being and quality of life in rural areas.
- **Poland:** Enhancing competitiveness of farms and food producers through direct sales and consumer trend analysis via interviews.
- **Spain:** Tackling demographic challenges, forest fire prevention, and digital administration through land use monitoring using earth observation data.
- **Sweden:** Leveraging natural resources for economic growth while addressing vulnerabilities in sparsely populated areas, with innovative web-scraping techniques for real-time indicators.
- **United Kingdom:** Implementing climate change responses in rural Scotland, with mobility data analysis to understand movement patterns and their socio-economic impacts, and data to inform the potential designation of a new National Park.

The project's innovative approaches, such as the use of remote sensing, web-scraping, and various data integration techniques, aim to produce comprehensive datasets and indicators that support rural policy-making and development. The collaboration between Living Labs and work packages ensures the relevance and applicability of the developed tools across diverse rural contexts in Europe.

Based on the initial assessment, Living Labs and multiple GRANULAR WPs have jointly explored data requirements and methods which could address some of the priority themes that were identified (as pointed out in Figure 3). As shown in Table 2, the Living Labs conducted preliminary research on existing databases relevant to their priorities and the data currently used by official statistics. This helped in performing an initial assessment of the data, indicators, and methods needed to identify gaps in the territory. To support this effort, Living Labs, such as Rural Scotland, began collecting quantitative data from official documents, national databases (e.g., national statistical research institutes), and other secondary sources. However, challenges were identified during this process, such as the fragmentation of data, which is scattered across multiple portals, and varying levels of authority competence in providing data. Despite these challenges, opportunities emerged to address key priorities in a more integrated and complementary way.

To enhance their preliminary data overview and strengthen their efforts, Living Labs initiated a series of field activities, including workshops, semi-structured interviews, focus groups, and mixed-method approaches that combine qualitative and quantitative data. This approach has been employed by P10 and Sweden to address basic services (Box 2 and 3, respectively "Selection of indicators for catching the diversity and complexity in Northern Sweden" and "Combination of quantitative and qualitative data: mixed methods for essential services"), and by RIO to focus on hyper-longevity priorities (Box 5 "Qualitative data: Unconventional data sources and mixed methods").

In addition, some Living Labs, like Val di Cecina (as highlighted in Box 1: “In Practice – Preliminary Assessment: Identification of Priorities and Data Gaps”) and West Pomerania, have started their activities by integrating data collection through surveys and questionnaires (Box 4 “Secondary data: Using different data sources).
However, one of the key challenges Living Labs encountered was engaging specific stakeholders and effectively communicating the collected information to citizens. To address this, the PPM Living Lab organized events with public data providers and initiated collaboration with the OpenStreetMap community (as reported in Box 6 “Data capacity building: Facilitating exchanges between data producers and users”).

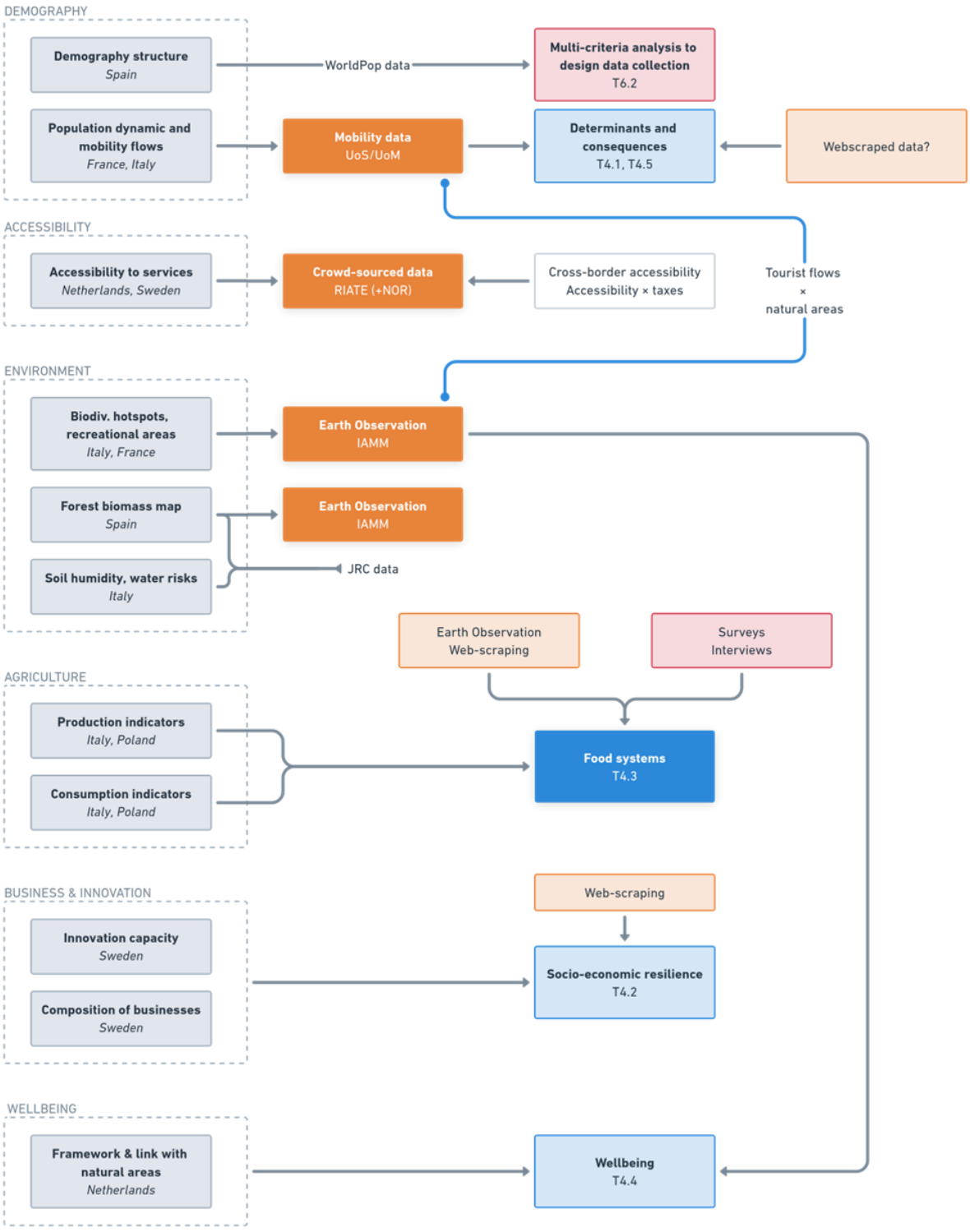


Figure 3: Living Labs thematic priorities and their links with WPs

Table 4: Living Labs and main actions to reach their ambitions

Living Lab	Priorities	Data to be collected – data gap	Methods	Ambitions
France – Pays Pyrénées Méditerranée (PPM)	Demographic change Commuting Recreational flows	Location of the trip generating center/mobility flows origin and destination, number of travelers throughout the day/year, means of transport and factors of choice. Level/intensity of attendance of places in natural areas, natural areas' hotspots and sensitive places, impacts of over-use or misuse of places in natural areas. Level/intensity of the drought in various natural and agricultural areas' sectors, natural areas' hotspots and sensitive places, impacts of over-use or misuse of places in natural areas.	Gathering existing data, such as path meters and mobility flows monitored within the previous priority theme, combined with remote sensing tools to discuss local knowledge of biodiversity hotspots and subject to the impact of over-use and misuse Remote sensing and comparison with local existing data and field experience to contribute to the debate on a current and pressing issue, allowing acting with lace.	Improve mobility flows and demand to develop targeted transport alternatives. Objectify and monitor tourists and local people's uses of natural areas to improve their reception and handle impacts on biodiversity. Evaluate the impact of drought on biodiversity. The Living Lab's objective is also to upgrade the skills of local players in observation tools and, more broadly.
Italy – Distretto Rurale Val di Cecina (VdC)	Recreational flows Natural capital assets Food systemic features	Data on water consumption (at farmer level) and soil moisture retention level. Data on the need for processing infrastructure for farms and applied business models. Data on tourist flows, especially linked to the accommodation capacity of agricultural companies.	Questionnaires for farmers. Request of specific data to entities.	The implementation of integrated territorial planning to cope with climate change especially concerning the availability of water. The construction of infrastructure for the logistics of agri-food products to allow the right support of agriculture and food chains. The planning of sustainable management of the touristic flows to deal with the lack of entrepreneurial culture and the localism and administrative fragmentations.
Netherlands – P10	Basic services Rural wellbeing	Qualitative data on perceptions about the accessibility and availability of essential services.	Inventory available statistical material on essential services (3 Municipalities); Interviews municipal representatives. Additional interviews with village representatives. Broader wellbeing method through blind spot ands and extra indicators for rural areas and	Implementation of the wellbeing monitor. Support the development of rural/regional development policies and the position of rural municipalities in multi-level governance settings, particularly those in more peripheral regions.

			how rural areas contribute to wellbeing in urban areas.	
Poland – West Pomerania (POL)	Food systemic features Land use system	Data on short supply chains with attention on food producers' activities.	Analysis of the activities of food producers through desk research and survey.	
Spain – Ourense (RIO)	Demographic change Public services Climate change	Data related on hyperlongevity (needs of elderly people), e-governance and fires and their nexus.	Analysis of the relation between services and needs. Existing data through an analysis of digital platform and data on telecare programme. Qualitative (semi structured interview) and quantitative data, mixed methods to combine statistical data (socio-economic) and territorial data. Cartographic materials with distribution of hyperlongevos people about services offered. Compering information with new needs.	Identify the relationship between the services location and needs within a rural-urban context. Understand how e-governance is oriented towards the elderly. Identify the nexus between forest fires, education, culture and aging.
Sweden – Regions of North Sweden (SWE)	Public services	Data on transformation readiness.	Interviews and surveys.	Provide better understanding of rural areas in terms of green transition.
United Kingdom – Rural Scotland (RS)	Demographic change Natural capital assets Climate change	Data in support of assessments of rural resilience, social cohesion and vulnerability in rural areas (Task 4.2). Updated data on demographic change based on the 2022 population census.	Review of preliminary findings from new data collection forms. Analysis of data relating to islands and remote areas (National Islands Plan Survey, 2022-2023). Methods integrating datasets for the derivation and maintenance of indicators (e.g. sparsely populated areas).	Contribute to the testing and operation of a process of rural proofing of policies, public and private, as they affect rural areas.

4. Conclusion

In summary, this deliverable serves as a set of operational guidelines to help practitioners address local challenges within their territories. The practical tools and methodologies presented here provide essential information to guide each step of the local data and policy cycle. The seven GRANULAR Living Labs, each with distinct characteristics, needs, and challenges, offer a diverse range of examples for collecting and using data that can be adapted to local contexts at various stages of the planning process and across different geographical scales.

While many relevant datasets are available at the appropriate granularity for rural areas, the success of data-driven initiatives largely depends on how the data and tools are utilized. Local capacity—skills, time, and resources—plays a significant role in the effective use of open data for policy-making. In-depth analysis and visualization often require specialized expertise and financial investment. Therefore, before embarking on data collection, it is critical for practitioners to define minimum needs, ensure a balance between quantitative and qualitative data, establish the level of objectivity required, determine disaggregation levels, set collection timelines, and assess accuracy requirements, all while deciding the importance of primary versus secondary data.

Beyond data collection, collaboration among local stakeholders is key to fostering peer-to-peer learning and promoting the exchange of best practices across rural regions. Encouraging knowledge sharing on open data use and nurturing a broader data culture is vital for effective rural policy-making. With the right methodologies, skills, and resources, data-driven initiatives can become a powerful tool to generate reliable, granular information that enhances local planning and decision-making.

Annex 1: GDPR and ethics considerations in local data collection

Primary data may involve the collection and processing of personal data. These data have to be protected in compliance with the EU Regulation No. 2016/679, which defines personal data as “any information relating to an identified or identifiable person”, and any other relevant national regulations. Appropriate processes must therefore be implemented to ensure the protection of all personal data of participants in the research, in compliance with the EU General Data Protection Guidelines (GDPR) or equivalent that pertain in each country.

Data minimisation

- The collection and handling of personal data must follow the principle of **data minimisation**, as outlined in Article 5(1)(c) of the GDPR and Article 4(1)(c) of Regulation (EU) 2018/1725. This means that the personal data collected will be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed."

Protection of privacy and confidentiality

- Data must be **de-identified, pseudonymized and/or anonymized** before it is released. This means that any personal information that could, potentially, identify an individual will be removed or obscured;
- Study subjects must be required to have given **their consent for the use of their data**. This will be achieved through the use of consent forms (example provided below), along with an accompanying information sheet, or other means of obtaining consent;
- the outcomes of any analysis **should not allow for the re-identification of participants**, and **results should be aggregated** either spatially or categorically when feasible;
- Sensitive data[2] may need extra procedures (depends on the nature of the data).

Safeguarding rights and freedoms of Participants

- **Right to be informed**: an information sheet must be shared with participants, explaining the project, the use of the information collected, and their rights;
- **Right of access**: Participants must be able to request information about the activities they contributed to (e.g. meeting documentation) and the information held about them (e.g. contact details);
- **Right to rectification**: Participants must be able to request that inaccurate personal information be corrected;
- **Right to erasure**: personal data must be erased if the data is no longer necessary for the original purpose or if the participant withdraws their consent for the data to be held;
- **Right to restrict processing**: this may involve removing published data.

Other key requirements

Best practice in the stewardship of data must be followed by the application of reasonable and appropriate technical, physical, ethical and procedural security measures. Consideration must be given to the availability and integrity of data, and to the appropriate levels of authorised access. It is the responsibility of the organisation to have robust backup procedures in place for data recoverability, auditing, security, and research integrity and data use, which staff and researchers are required to follow.

[2] Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs or trade-union membership, as well as the processing of genetic data, biometric data for the purposes of identifying a person, data concerning (medical) health or data concerning the sex life or sexual orientation of a person.

Template for obtaining informed consent

Consent Form

Description of the project

Please tick the boxes

I confirm that I have read, or had read to me, and understand the information sheet dated …./…./…..[3] for the project. I have had the opportunity to ask questions and these have been answered fully and explicitly.	
I understand that my participation is voluntary, and I am free to withdraw at any time without providing any reason and without my legal rights being affected.	
I understand that confidentiality will be maintained at all times, data will be pseudonymised at the earliest possible time and it will not be possible to identify me from any publications/outputs.	
I understand that this form relates to all of the activities to which I am contributing within the project.	
I agree to take part in the project.	
I agree to being contacted at a later date in relation to the project.	
I agree for the information I provide to be recorded (e.g. audio, flipchart notes) and transcribed.	
I acknowledge that I have read and understood the privacy notice (presented p.2 as below).	

Name of the participant Signature Date

[3] Please insert the date of the information sheet

PRIVACY NOTICE

The *[Name of the organization in charge of the collection of data]* will use your personal data (name, age, gender, profession, employer, area of expertise, email) for the purposes of the research undertaken in [name of the project]**.** Our legal basis for processing your data is that it is necessary for the performance of a task carried out in the public interest in relation to research [description].

We are the Data Controller over your personal data. The level of access to the data is determined by a secured computer account linked to the project. No data is accessible without a computer account, except for anonymised public data. We will not share your personal data beyond the project consortium[3], unless required by law and shall only retain it according to good scientific practice for as long as is necessary to fulfil the research undertaken on the project, to deliver project outcomes, and to fulfil the requirements of the funders. In particular, as a requirement of EU-funded project, data will be anonymised and made open access. You have rights in relation to your personal data as defined in [*relevant articles of your national data protection laws / regulations*][4]. For further information, please contact [*add email address of person responsible for data protection*].

Contact details:

[Insert Name of the Organization]

[Address]

[Email address]

[Telephone]

[4] To be completed by the organisation in charge of collecting data with the relevant articles of your national data protection laws / regulations. For a checklist on the GDPR on EU level, please see https://gdpr.eu/checklist/