GRANULAR

DATA MANAGEMENT PLAN (2)

D 1.4

JANUARY 2025

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).

D1.4 DATA MANAGEMENT PLAN (2)

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	Final version
Dissemination level	Public
Authors	Kirsch A., Dupont A., Berchoux T., Jain P. Damari Y., Chahid N. (IAMM), Ysebaert R. (CNRS), Sorichetta A. (UMIL), Steele J., Voepel H. (UoS), McCallum I. (IIASA), Tapia C., Slätmo E., Stjernberg M. (NOR), Miller D., Hopkins J., Wang C. (HUT), Stortini F. (UNIPI), Ntabuhashe M. (AEIDL), Chasset L. (PPM), Bahta M. (VAS), Doval M., Martins B. (UVIGO), Ricci M. (VdC), Kurdys-Kujawska A. (KUT)
Work Package Leader	Mediterranean Agronomic Institute of Montpellier (IAMM)
Project coordinator	Mediterranean Agronomic Institute of Montpellier (IAMM)

Funded by the European Union

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

TABLE OF CONTENTS

GLOSSARY

Action	GRANULAR project
Beneficiaries	The signatories of the Grant Agreement benefiting from the EU grant awarded for GRANULAR project
CA Consortium Agreement	The agreement to specify with respect to the Project the relationship among the Parties, in particular concerning the organisation of the work between the Parties, the management of the Project and the rights and obligations of the Parties
Coordinator	The CIHEAM-IAMM as the intermediary between the Parties and the Granting Authority
Deliverable	A report that is sent to the granting authority, providing information to ensure effective monitoring of the project. There are different types of deliverables (e.g. a report on specific activities or results, data management plans, ethics or security requirements)
FAIR principles	Data which meet principles of findability, accessibility, interoperability, and reusability as FAIR principles for scientific data management
Grant	The EU grant awarded for GRANULAR project
GA Grant Agreement	The contract setting out the terms and conditions of the relationship between the Granting Authority, the Coordinator and the Beneficiaries
Granting Authority	The body awarding the grant for the Project
Research output	The term 'research output' covers any type of research data. Examples include the contents of a database, interview recordings, images. It also includes source codes or software produced and used in scientific research, models, protocols, etc. It does not include scientific publications or articles produced as part of the project.

Executive summary

This Data Management Plan (DMP) covers the activities and production of datasets, indicators and deliverables required to achieve the aims and objectives of the GRANULAR project. Detailed strategies, recommendations and good practices for data management in line with Open Science and Scientific Integrity principles are provided in the deliverable D1.3 Data Management Plan - version 1, and summarised below.

- The management of research outputs follows the principles of being **Findable, Accessible, Interoperable and Reusable (FAIR)**.
- Datasets are clearly identified using an appropriate **universal identifier such as Digital Object Identifier** (D.O.I.; www.doi.org/) and accompanied by appropriate and sufficient **keywords**, **search terms** and a **GRANULAR metadata standard**, based on Dublin Core standards. Responsibility for their selection is that of the relevant research team and Task and Work Package Leaders.
- Data will be made **freely available** through the **Zenodo online repository** and its GRANULAR community for ensuring long term archive in line with the obligations in the Grant Agreement, **licenced** (under the latest version of the Commons Attribution International Public Licence CC BY-NC Attribution-Non-Commercial) and **comprehensible to external users** (e.g. the European Commission, Living Labs and Replication Labs, other researchers) by providing detailed information on data processing through metadata, which allows to users to understand the content of the data files, in an open format.
- Generated datasets are uploaded as an archive (.zip) containing a description of data processing, associated documentation and a 'Data' folder (inputs and outputs) with the corresponding Metadata file. All metadata will be openly available and licenced (CC BY-NC Attribution-Non Commercial).
- Existing datasets reused in GRANULAR are appropriately acknowledged in reports, papers and other outputs. All data sources used in the research in GRANULAR are properly referenced, with any restrictions placed on the use of third-party data identified and adhered to, which is primarily the responsibility of the relevant teams of researchers.
- All deliverables, policy briefs and practice abstracts, presentations, are **open access** and **licenced under the latest version of the Commons Attribution International Public Licence** (CC BY-NC Attribution-Non Commercial).
- Open access is ensured to **peer-reviewed scientific publications** in full compliance with the conditions set in the Grant Agreement. All scientific publications will be granted immediate access and will be made available through trusted repositories for ensuring long term archive, such as the **Zenodo online repository** and the associated GRANULAR community. Authors retain the needed intellectual property rights to comply with Open Access requirements.

This document, which should be seen as a second version complementary to D1.3, sets out the key recommendations and practices agreed at project level for ensuring FAIR data, with a description of the data management lifecycle for each project research output.

Introduction

Departing from an updated conceptualisation of rurality based on the multi-dimensional nature of contemporary rural-urban interrelations and interdependencies, GRANULAR generates new insights for characterising rural diversity based on a multi-actor and interdisciplinary approach. Through its Multi-Actor Labs and research activities, GRANULAR relies on 16 rural territories to advance scientific research and test innovations. The work is organised through two types of Multi-Actor Labs:

- Living Labs (LL), which co-construct and co-test innovative data collection methods and indicators to support current local policy decisions;
- and Replication Labs (RL), which test the replicability and validate data collection methods and indicators.

Based on the insights from these territories, the project develops conceptual frameworks and generates novel datasets (at the local level but also EU-wide) using a wide range of methods and primary data, such as remote sensing, crowd- sourced data, social media data and web scraping. Data are then used to derive indicators relevant to rural communities for the implementation of the EU Long-Term Vision for Rural Areas (LTVRA), characterising concepts such as resilience, well-being, quality of life and attractiveness. After ensuring their out- and up-scalability, datasets, data visualisation and other tools are directly made available on a dedicated Digital Platform co-designed with rural actors.

By providing and disseminating novel frameworks to conceptualise rural diversity and to rural-proof policies, as well as data collection methods and indicators, the core purpose of GRANULAR is the **provision of citizen-informed and data-driven tools to support decision-making.**

This Data Management Plan (DMP) covers the activities and production of datasets, indicators and deliverables required to achieve the aims and objectives of the GRANULAR project. To guarantee the quality, efficiency and impact of its research, GRANULAR ensures that data generated by the project are as open and reusable as possible, as soon as possible. This includes ensuring the availability of data and outputs as a legacy of the project; using standards keywords, search terms and detailed metadata; ensuring open data licencing and data security.

All chapters of this Data Management Plan are divided into two parts:

- a first generic part, which reiterates the **core principles of the project's data management strategy**, guiding the use of data and its management throughout the project, and in the creation of its legacy (provided in deliverable D1.3 Data Management Plan - version 1);

- a second, more operational part, which describes the **data management life cycle and specific associated procedures with each research output**, for the data to be collected, processed, documented, stored, archived, preserved and shared throughout the project**.**

The DMP OPIDoR[1] tool (Data Management Plan for Optimized Sharing and Interoperability of Research Data) was used to collectively develop this data management plan - version 2 with all the research teams involved in the GRANULAR project. The structure chosen, as well as the guiding questions that help in drawing up the plan, are those provided in the official Horizon Europe Data Management Plan template.

[1] Online tool hosted and managed by Inist-CNRS and based on the DMPRoadmap open source code.

1. Data Summary

1.1. Classification of research outputs

Categories of data that are either re-used or generated during the project are broken down by **research output**. Each research output, including data capture, existing datasets reused and database outputs in GRANULAR, is detailed according to the following classification:

"Data type – thematic *(Task and partner responsible for research data)"*

The list below presents the 15 GRANULAR research outputs.

1. **Crowd-sourced data – accessibility** *Subtask 3.3.1. (CNRS)*
2. **Transaction data – mobility data** *Subtask 3.3.2. Mobility data (UoS/UMIL)*
3. **Nowcasting – social media and housing transaction** *Subtask 3.3.3. Web scraping and nowcasting (NOR)*
4. **Remote sensing data (Earth Observation) - land cover/land use** *Subtask 3.3.4. Earth Observation (IAMM)*
5. **Citizen science data – rural sustainability** *Subtask 3.3.5. Citizen science (IIASA)*
6. **Climate and environmental indicators** *Task 4.1 Climate Neutrality (IAMM)*
7. **Socio-economic resilience, social cohesion and vulnerability indicators** *Task 4.2 (NOR)*
8. **Food systems multidimensional indicators** *Task 4.3 Food Systems and Land Tenure models (IAMM)*
9. **Well-being data** *Task 4.4 (TI)*
10. **Rural Attractiveness indicators** *Task 4.5 Rural attractiveness and perceptions (NOR)*
11. **Typologies of rural areas at local/grid level** *Task 4.6 Characterising rural diversity (NOR)*
12. **Rural Proofing exercise** *WP5 (Ecorys/HUT)*
13. **Geographic information system (GIS) & Storymaps** *Task 5.4 (HUT)*
14. **Multi-Actor Labs participatory collected data** *WP6 (Living Labs and Replication Labs)*
15. **Communication, Outreach, Dissemination and Exploitation (CODE) data** *Tasks 7.1, 7.2, 7.5 (AEIDL)*

1.2. Crowd-sourced data – accessibility

The aim of this dataset is to create open-source accessibility indicators to points of interest in Europe, using several transport modes.

Generated data #1: Accessibility Indicators to services at EU scale - 1km grid indicators

- Dataset type: A zip archive including notebooks, origin-destination matrixes and outputs (accessibility indicators in csv format). Each dataset is associated to a metadata file explaining the data content and a readme file in the root of the archive explaining the organisation of the data delivery.
- Formats : csv, geopackages, geojson
- Size: 2.4 GB, covering all Europe. The next deliveries will be lighter
- Collection methods / protocols: the data has been processed through notebooks explaining the methodology and showing the R code used to produce the resulting indicators. Input data is free and open source. Saying that it is possible for any person with technical ability to reproduce the results.
- Data utility outside the project:
    - Scientists interested in the methodology used to build such indicators (update, technical improvements, etc.)
    - The OSM community: use of routing engines in a large scale, discussion on OSM quality.
    - Policy makers interested in qualifying the accessibility of their respective territories in several context (local, national, EU).
- Reused data for that dataset:
    - Origins: Eurostat, GISCO geodata (© European Union, 1995 – 2024)
    - Destinations : EU towns and cities (© European Commission © REGIOgis)
    - Routing engines: OSRM - car profile (© Project OSRM contributors 2024)
- Pre-existing rights of use: All this data is publicly available with an open-source license

Generated data #2: Accessibility Indicators to services at EU scale - aggregated indicators

- Dataset type: A zip archive including a script showing how accessibility indicators have been aggregated at upper spatial/territorial resolution (NUTS, grid).
- Formats : csv
- Size : 276 MB
- Collection methods / protocols: based on (1) Accessibility indicators, a R script aggregates results obtained so far at upper resolution
- Data utility outside the project:
    - A web-viewer tool that will use these inputs to display interactively the results.
    - Scientists interested in the methodology used to build such indicators (update, technical improvements, etc.)
    - Policy makers interested in qualifying the accessibility of their respective territories in several context (local, national, EU).
- Reused data for that dataset:
    - Based on Generated data #1 and previously.
- Pre-existing rights of use: All this data is publicly available with an open-source license

Generated data #3: Accessibility to education services

- Dataset type: same than generated data #1
- Formats : csv, geopackages, geojson
- Size : ≈ several hundreds of MB
- Collection methods / protocols: same than generated data #1. The analysis will be processed on a set of case studies with consistent data (depending on the destination layer)
- Data utility outside the project:
    - Scientists interested in the methodology used to build such indicators (update, technical improvements, etc.)
    - The OSM community: use of routing engines in a large scale, discussion on OSM quality.
    - Policy makers interested in qualifying the accessibility of their respective territories in several context (local, national, EU).
- Reused data for that dataset:
    - Origins: Eurostat, GISCO Census grid - population aged under 15 (© European Union, 1995 – 2024)

    - Destinations : education services - ISCED levels 2-3 (© European Union, 1995 – 2024)
    - Routing engines: OSRM - car and bike profile (© Project OSRM contributors 2024)
- Pre-existing rights of use: All this data is publicly available with an open-source license

Generated data #4: Train stations and intermodality

- Dataset type: same than generated data #1
- Formats : csv, geopackages, geojson
- Size : ≈ several hundreds of MB
- Collection methods / protocols: creating a hierarchy in train stations based on GTFS tables (stops frequency, number of destinations covered). Explore intermodal transport (car + public transport) to reach destinations.
- Data utility outside the project:
    - Scientists interested in data sources and methods to use GTFS table
    - European Commission / JRC: identifying possible methods for further analysis in a EU context; identifying potential new consistent data sources.
    - Policy makers interested in qualifying the accessibility of their respective territories in several context (local, national, EU).
- Reused data for that dataset:
    - Origins: Eurostat, GISCO Census grid - population aged under 15 (© European Union, 1995 – 2024)
    - Destinations : train stations
    - Routing engines: OSRM - car profile (© Project OSRM contributors 2024) and open source GTFS tables based on operators
- Pre-existing rights of use: All this data is publicly available with an open-source license

1.3. Transaction data – mobility data

The purpose of this research output is a better understanding of rural area mobility (including its characteristics) for producing new knowledge/actionable information to support interventions and applications, as well as evidence-based decision- and policy-making, to improve rural area accessibility and life quality.

Existing data are combined together and used as inputs for modelling human mobility related to rural areas - the used of Twitter (now X) geolocated data was considered but then discarded due to the excessive cost of data.

Generated data: *final name to be decided*

- Dataset type: Facebook/Meta-based mobility metrics at the admin unit level/Digital Data
- Formats: This decision is still pending - data will be at least provided in tabular form, including some spatial components, and as maps for ease of visualization (depending project stakeholders/end-user requests data will be also provided in vector format and potentially raster format).
- Size: Still unknown as modelled outputs are still being developed/generated - size will also be a function of the final format.
- Collection methods / protocols: the modelling approach is still being improved/refined and need to be finalized.
- Data utility: The generated data might be useful for project stakeholders (including the Living Lab Stakeholders), policy and decision makers, as well as the wider research community interested in rural areas and, more in general, in human mobility and connectivity.

Reused data #1:

- Dataset name and type: Facebook/Meta Movement Maps/Digital Data
- Formats: tabular data presenting statistics about aggregate people movements between pairs of locations (representing either admin units or grid-cells) in subsequent daily time intervals (00.00-8.00; 8.00-16.00; 16.00-24.00)
- Size: varies depending on country (in terms of size, number of the considered admin units and/or and grid-cell resolution)
- Origin & provenance: Facebook/Meta (through WorldPop, University of Southampton)
- Pre-existing rights of use (restrictions if any): the permission to use the raw Movement Maps secondary data is granted by Facebook/Meta to WorldPop, University of Southampton.

Reused data #2:

- Dataset name and type: GHS-DUC (v1, 2023) Global Degree of Urbanisation Classification of administrative units according to the application of stage II of the Degree of Urbanisation/Digital Data
- Formats: CSV files; one per GADM v4.1 level (from 0 to 5) and epoch (1975-2030; 5 years timestep), with all administrative units classified according to the Degree of Urbanisation
- Size: Global dataset spanning all years (˜1GB)
- Origin & provenance : EC-JRC (https://human-settlement.emergency.copernicus.eu/ghs_duc2023.php)
- Pre-existing rights of use (restrictions if any): Full Open Access

Reused data #3:

- Dataset name and type: GADM (v4.1); world's administrative areas (boundaries) at multiple levels of sub-division for all countries, territories and dependencies
- Formats: Geopackage, Shapefile, GeoJSON, KMZ; administrative boundaries (over various admin unit levels) can be downloaded by country or for the entire world
- Size: varies depending on the country (areal size) and admin unit level
- Origin & provenance : GADM (https://gadm.org/index.html)
- Pre-existing rights of use (restrictions if any): The data are freely available for academic use and other non-commercial use. Redistribution or commercial use is not allowed without prior permission - data for Austria is covered by a a different license: Creative Commons Attribution-ShareAlike 2.0 (source: Government of Austria)

Reused data #4:

- Dataset name and type: GHS-SMOD (v1, 2023); Global Degree of Urbanisation Classification of gridcells according to the application of stage I methodology recommended by UN Statistical Commission/Digital Data
- Formats: GeoTIFF; 1000x1000km raster tiles with 1km grid cells classified according to the Degree of Urbanisation based on the global population grid generated by the JRC in the epochs 1975-2030 (5 years timestep).
- Size: Varies depending on the specific tile(s)
- Origin & provenance: EC-JRC (https://human-settlement.emergency.copernicus.eu/ghs_smod2023.php)
- Pre-existing rights of use (restrictions if any): Full Open Access

Reused data #5:

- Dataset name and type: 2022 Census Settlement boundaries (2022 Census Geography Products)/Digital Data
- Formats: Esri shapefiles; built-up areas in Scotland which round to 500 people or more. Size: 3.5 MB

- Origin & provenance: Scottish Government (SpatialData.gov.scot; https://www.nrscotland.gov.uk/publications/2022-census-geography-products/)
- Pre-existing rights of use (restrictions if any): Open Government Licence - the following attribution statement must be used to acknowledge the source of the information: Copyright Scottish Government, contains Ordnance Survey data © Crown copyright and database right (insert year).

Reused data #6:

- Dataset name and type: Scottish Government Urban Rural Classification 2020/Digital Data
- Formats: Esri shapefiles; consistent definitions (ie, two-, three-, six- and eight-fold form) of urban and rural areas across Scotland based on population (National Records of Scotland; NRS) and drive-time accessibility.
- Size: 32.5 MB
- Origin & provenance: Scottish Government (SpatialData.gov.scot; https://www.data.gov.uk/dataset/f00387c5-7858-4d75-977b-bfdb35300e7f/urban-rural-classification-scotland)
- Pre-existing rights of use (restrictions if any): Open Government Licence - the following attribution statement must be used to acknowledge the source of the information: Copyright Scottish Government, contains Ordnance Survey data © Crown copyright and database right (insert year).

Reused data #7:

- Dataset name and type: High Resolution Population Density Maps + Demographic Estimates/Digital Data
- Formats: CSV, GeoTIFF, JSON; spatial distribution (at 30m resolution) of overall population density (2) Women (3) Men (4) Children (ages 0-5) (5) Youth (ages 15-24) (6) Elderly (ages 60+) (7) Women of reproductive age (ages 15-49) for about 160 countries
- Size: varies depending on country areal size
- Origin & provenance: Facebook/Meta/Data for Good (https://dataforgood.facebook.com/dfg/docs/methodology-high-resolution-population-density-maps; https://data.humdata.org/organization/meta?q=population%20density&sort=if(gt(last_modified%2Creview_date)%2Clast_modified%2Creview_date)
- Pre-existing rights of use (restrictions if any): Full 0pen Access (Creative Commons Attribution International)

Reused data #8:

- Dataset name and type: ARDECO Population Series, Annual Regional Database of the European Commission (SNPTD)
- Formats: CSV; NUTS/LAU annual average population based on Regional Accounts for EU27, Albania, Iceland, Liechtenstein, Montenegro, North Macedonia, Norway, Serbia, Switzerland, Turkey, United Kingdom (annual 1960 to current + 2 year forecast)
- Size: tabular (1.5MB)
- Origin & provenance: EC-JRC/DG REGIO (https://urban.jrc.ec.europa.eu/ardeco/explorer?lng=en)
- Pre-existing rights of use (restrictions if any): Full Open Access

Reused data #9:

- Dataset name and type: GHS-POP (2023) Global Human Settlement Population Distribution
- Formats: GeoTIFF; global 100m, 1km, 3 arcsec, and 30 arcsec rasters depicting the distribution of residential population, expressed as the number of people per gridcell, between 1975 and 2020 in 5-year intervals and projections to 2025 and 2030
- Size: Varies depending on the specific spatial resolution and tile (for 100m resolution, the size of the global dataset is 5.4GB zipped/8.9GB unzipped)

- Origin & provenance: EC-JRC (https://human-settlement.emergency.copernicus.eu/ghs_pop2023.php)
- Pre-existing rights of use (restrictions if any): Full Open Access

Reused data #10:

- Dataset name and type: WorldPop Age and Sex Structures (Unconstrained global mosaics 2000-2020 & Unconstrained individual countries 2000-2020)
- Formats: GeoTIFF; global 100m (per country) and 1km (global mosaic) rasters depicting the annual distribution of population per gridcell, broken down by sex and age groupings (including 0-1 and by 5-year up to 80+), between 2000 and 2020,
- Size: The size of the country raster, having a resolution of 100m, is a function of the areal size of each country; the size of each global mosaic, having a resolution of 1km, is 3.05GB.
- Origin & provenance: WorldPop data repository
- Pre-existing rights of use (restrictions if any): Full Open Access (Creative Commons Attribution 4.0)

1.4. Nowcasting – social media and housing transaction

The purpose of this research output is to calculate nearly real-time indicators for EU rural areas, showcasing methodologies to create statistical indices from non-conventional sources

Generated data #1: Transport Poverty Risk Index for Europe's rural areas

- Dataset type: A zip archive including spatially-explicit risk transport risk indices and its components (a dimensional-scores).
- Formats: csv, geopackages, geojson
- Size: Approx 2.5 GB, covering all Europe.
- Collection methods / protocols: The collection process will be documented through notebooks explaining the methodology and showing the R/python code used to produce the resulting indicators. Any person with coding skills should be able to reproduce the results.
- Data utility::
    - Policy makers interested in tracking transport poverty in rural areas.
    - Local practitioners interested in replicating our data collection and exploitation methods for the calculation of risk indices in the local area.
    - Scientists interested in transport poverty and risk analysis.

Corresponding Reused data :

- Dataset name and type: The input data will be retrieved from official repositories and platforms providing updated information on (1) living conditions and (2) transport patterns, as well as (3) real-time energy prices. .
- Formats: All the datasets will be alphanumeric, retrieved in different formats (alphanumeric and spatial formats). The data files will be collected via direct downloads or accessed programmatically via APIs. Exceptionally, web scraping could be used to access real time information about energy costs. This will be done in accordance to the data provider's use policy and current legislation. No sensitive information, e.g. including personal information or allowing indirect identification of individuals will be used.
- Size: Unknown.
- Origin & provenance: Items (1) and (3) are retrieved from publicly available, free of charge, sources: including the IEA, EC, Eurostat, and National Statistical Institutes. Additionally, iem (3) includes accessibility information produced in other tasks in GRANULAR.
- Pre-existing rights of use (restrictions if any): All the data used will be retrieved from publicly available sources. Access policies might differ between data providers but, in general, all the data used is expected to be available for scientific use.

Generated data #2: Evolution of Rural Property Prices in the North of Sweden

- Dataset type: A zip archive including a spatially-explicit file with an estimation of the variation on housing prices in Northern Sweden (Swedish living lab) between 2014 and 2024.
- Formats: csv, geopackages, geojson. To avoid potential data confidentiality issues, only modelled variations of relative housing prices (percentage of change) on a continuous 1x1 km grid will be shared externally.
- Size: Approx 0.5 GB, covering the Norrbotten region in Sweden.
- Collection methods / protocols: The collection process is documented through a notebook explaining the methodology and showing the R/python code used to produce the resulting indicators. A preliminary version of the notebook can be accessed here: https://carltap.github.io/housing-prices-demo/. Any person with coding skills should be able to reproduce the results.
- Data utility :
    - Policy makers interested in tracking housing prices in rural areas.
    - Local practitioners interested in replicating our data collection and exploitation method for the calculation of risk indices in the local area.
    - Scientists interested in modelling housing price information.

Corresponding reused data :

- Dataset name and type: Information on property characteristics and transaction prices.
- Formats: The data are collected in the form of descriptive texts informing on the characteristics of the properties in the housing market of the North of Sweden (digital, alphanumeric).
- Size : ~150 MB.
- Origin & provenance: The input data (text describing the properties and price information) are retrieved from Hemnet (https://hemnet.se), the largest digital marketplace for housing in Sweden. Every year, almost 200 000 properties for sale are published on Hemnet, which corresponds to about 90% of the properties sold annually in Sweden. The data are collected via direct web scraping from Hemnet. The scraping is done according to the data provider's terms of use, the current legislation and adhering to our Ethical Guidelines. No sensitive information, e.g. including personal information on individual property costs or information allowing identification of individuals will be disclosed.
- Pre-existing rights of use (restrictions if any): The raw data retrieved from the Hemnet platform will not be disseminated since the data is publicly available from Hemnet.

1.5. Remote sensing data (Earth Observation) – land cover/land use

The dataset has been created to support the development of a deep learning framework for analysing land cover and land use patterns, integrating language modelling and textual queries, as well as enabling image-to-image searches between ground-level and satellite imagery in both directions.

Generated data :

- Dataset name and type: The Sen2LUCAS datasets created consist of sentinel-2 images with RGB bands of 10 m resolution and 100x100 pixels
- Formats:.tif
- Size : 31 GB
- Collection methods / protocols: Geo-tagged Sentinel-2 images are collected based on LUCAS locations using third party API (Planetary computers)
- Data utility: policy advisors, relevant practitioners and researchers. The dataset collected will be used in a Deep Learning framework to derive novel indicators based on text queries.

Reused data #1 :

- Dataset name and type: LUCAS 2018, ground level Image, Geolocated images
- Formats: In-field PNG images with EXIF metadata.

- Size:
  - Spatial extent: European-wide
  - Resolution: between 10 and 100m resolution
  - Compiled datasets around 20Tb
- Origin & provenance : LUCAS
- Pre-existing rights of use: Full Open Access

Reused data #2 :

- Dataset name and type: Sen2LUCAS, Sentinel-2 satellite image datasets
- Formats : tiff.
- Size:
  - Spatial extent: European-wide
  - Resolution: between 10 and 60m resolution
  - Compiled datasets around 31 GB and 200K data points
- Origin & provenance : Copernicus (EU space programme)
  1. Publicly available LUCAS 2018 survey.
  2. Geo-tagged Sentinel-2 images are collected based on LUCAS locations using third party API (Planetary computers)

Reused data #3 :

- Dataset name and type: Sen4Map, Sentinel-2 satellite image datasets
- Formats : hdf5.
- Size:
  - Spatial extent: European-wide
  - Resolution: between 10 and 60m resolution
  - Compiled datasets around 2TB
- Origin & provenance : Copernicus (EU space programme)
  1. Publicly available LUCAS 2018 survey.
  2. Geo-tagged Sentinel-2 images are collected based on LUCAS locations using third party API (GEE).
- The publicly available Sen4Map data is also geo-tagged with LUCAS and includes annual composite time-series along with multi-spectral data.

1.6. Citizen science data – rural sustainability

The purpose of this research output is to generate new and novel indicators of rural sustainability by applying selected citizen science methods (e.g. mobile app data collection, use of crowd-sourced data, use of social media data) via various stakeholder groups to address data gaps, in coordination with the Living Labs.

More efforts will be focused in the coming year on the analysis of social media data (e.g. using TripAdvisor API) to understand tourism drivers, and explore how various citizen science methods can support the different Living Labs, based on their interests.

Generated data:

- Dataset name and type: new and novel indicators of rural sustainability
- Formats: Various types and formats, both vector and raster data formats. Possibility of shapefile formats, geojson, csv, .tif, etc.
- Size: small in terms of file storage, the exception being the collection of photos. Raw data will likely be much larger but this may not be necessarily provided as GRANULAR output.
- Collection methods / protocols: data may be generated via citizen science, crowdsourcing methods (typically mobile applications to collect the data).
- Data useful for: wider academic community. In addition, the JRC may be interested in the results, along with the European Commission.

Reused data:

- Dataset name and type: existing raw data or derived products (Copernicus satellite data, Facebook data, TripAdvisor API etc.)
- Formats: Various types and formats. It is expected that data to be generated or re-used in both vector and raster data formats. We may work with shapefile formats, geojson, csv, .tif, etc.
- Origin & provenance: variety of sources, from proprietary sources to Eurostat, and various public data sources. Particularly Open Street Map data, Eurostat and other EU sources

1.7. Climate and environmental indicators

The objective is to analyse and calculate a climate neutrality index for rural communities with indicators from various domains. The index will be designed to be relevant and hold significance for both policy makers and the scientific community.

Generated data #1: Climate Neutrality Index

- Dataset type: Quantitative index dataset to carry out a comparative analysis of climate neutrality performance across rural areas
- Formats: .xlsx
- Size: 1 MB
- Collection methods / protocols: The data is collected from online open sources (e.g., Eurostat, World bank)

Generated data #2: Downscaling environmental and socio-economic data in rural areas

- Dataset type: Spatially disaggregated intensities dataset to improve the granularity of environmental and socio-economic data in rural areas
- Formats: .xlsx
- Size: 1 MB
- Collection methods / protocols: Calculated through IPAT downscaling method using auxiliary variables (e.g., income data, population size).

Generated data #3: Consolidated list of policy instruments

- Dataset type: Qualitative and structured policy dataset for i) mapping and analysing policy approaches to climate neutrality in small settlement; and ii) identifying common strategies and gaps in policy coverage across Europe
- Formats: .xlsx
- Size: 160kb
- Formats: csv, geopackages, geojson
- Size: Approx 2.5 GB, covering all Europe.
- Collection methods / protocols: Compilation of policy measures from existing European policy databases (e.g., Covenant of Mayors, EEA)

Reused data #1: National-level climate-related indicators

- Formats : CSV, Excel files
- Size : Up to 50 MB
- Origin & provenance : Mainly Eurostat
- Pre-existing rights of use (restrictions if any): Open access, subject to Eurostat's terms of use.

Reused data #2: Geospatial data derived from OpenStreetMap

- Formats : GeoJSON
- Size: Moderate, depending on the geographic area and feature density.
- Origin & provenance: Directly sourced from OpenStreetMap
- Pre-existing rights of use (restrictions if any): Open access under the Open Database License.

Reused data #3: SoilGrids geospatial data

- Formats : tabular format
- Size : Up to 50 MB
- Origin & provenance: Sourced from SoilGrids, developed and maintained by ISRIC - World Soil Information.
- Pre-existing rights of use (restrictions if any): Open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Reused data #4: Socioeconomic and environmental indicators from the World Bank

- Formats : CSV, Excel
- Size : Up to 50 MB
- Origin & provenance: Sourced from the World Bank's open data platform
- Pre-existing rights of use (restrictions if any): Open access under the World Bank Terms of Use

1.8. Socio-economic resilience, social cohesion and vulnerability indicators

The purpose of this research output is to explore the concept of rural resilience from a gender perspective, focusing on the role of labour market attachment

Generated data :

- Dataset name and type: Report on socio-economic resilience from a gender perspective, focusing on the the role of labour market attachment. Alongside the report (text), a dataset (table) including the predicted probabilities of employment gaps for six categories of interaction between gender (female and male) and degree of urbanization (urban rural and intermediate). The interactions will also be visualised using charts and other visuals (included in the report).
- Formats: Report (text and visuals) and supplementary materials (tabular).
- Size: ~100 MB including all materials.
- Collection methods / protocols: The data is collected in tabular format from the Eurostat microdata platform (scientific files)
- Data utility :
    - Policy makers responsible for rural and gender policies
    - Researchers engaged in the analysis of gender studies or rural research
    - Social actors interested in labour market integration

Reused data:

- Dataset name and type: EU-LFS microdata for scientific purposes.
- Formats : tabular
- Size : ~0.5 GB
- Origin & provenance : Eurostat, via the microdata access portal.
- Pre-existing rights of use (restrictions if any): The EU-LFS microdata for scientific purposes is subject to two data protection frameworks:
    - Statistical confidentiality: data on individuals or business entities may only be used for statistical purposes and that rules and measures must be taken to prevent disclosure;
    - General data protection framework (GDPR and EUDPR): applicable whenever information about natural persons is collected.

1.9. Food systems multidimensional indicators

The aim of this research output is to produce indicators in a comprehensive framework at fine spatiotemporal scales (territorial, local) to better monitor food system performance and help policy makers in designing tailored policies for a more sustainable, resilient and local food systems in the EU.

Generated data #1 – Food systems

- Dataset type: Qualitative data from local stakeholders, used in validating the conceptual framework tailored for the characterisation of food systems
- Formats: Microsoft office word documents
- Size: up to 50 MB
- Collection methods / protocols: field visits, interviews with experts

Generated data #2 – Food systems

- Dataset type: Maps representing food system typologies across EU regions/territories
- Formats: GeoJSON
- Size: Moderate, depending on the geographic area and feature densit.
- Collection methods / protocols: use of pre-existing layers collected from geospatial databases

Generated data #3 – Land Tenure

- Dataset type: Categorical - Farmland Ownership Dynamic of Change data and maps; Economic Dynamic of Land Ownership Type data and maps
- Formats: shapefiles and tables
- Size: 70 MB (approximately)
- Collection methods / protocols: 1. Manual data collection (tables from Eurostat); 2. Standardization and cleaning processes; 3. Descriptive Statistics; 4. Time-series Analysis

Generated data #4 – Land Tenure

- Dataset type: Numerical - Farmland Ownership Distribution data and maps; Land Ownership Index (LOI) in Europe - Farmland Areas data and maps
- Formats: shapefiles and tables
- Size: 60 MB (approximately)
- Collection methods / protocols: 1. Manual data collection (tables from Eurostat); 2. Standardization and cleaning processes; 3. Descriptive Statistics; 4. Time-series Analysis

Reused data #1: Datasets in the topic of food system (Food production, food value chains, food consumption)

- Dataset type: Quantitative data from various sources (farm level, municipality, department, national),
- Formats : CSV, Excel files
- Size : Up to 50 MB
- Origin & provenance: Scientific literature data, EU databases (Eurostat, FAOData, WDI databank...), government websites and online sources, data collected through interviews and surveys. Data provenance will be documented in README files.
- Pre-existing rights of use: following the rights of use of the data provider

Reused data #2: Geospatial data derived from OpenStreetMap

- Formats : GeoJSON
- Size: Moderate, depending on the geographic area and feature density.
- Origin & provenance: Directly sourced from OpenStreetMap
- Pre-existing rights of use (restrictions if any): Open access under the Open Database License.

Reused data #3: SoilGrids geospatial data

- Formats : tabular format
- Size : Up to 50 MB
- Origin & provenance: Sourced from SoilGrids, developed and maintained by ISRIC - World Soil Information.
- Pre-existing rights of use (restrictions if any): Open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

1.10. Well-being data

The purpose of this research output is the analysis of large-scale social surveys (longitudinal panel survey and microdata coming from various national and European surveys) to identify correlates of subjective wellbeing in rural areas.

Generated data:

- Dataset name and type: Analysis of large-scale social surveys
- Formats: Report (text and visuals) and supplementary materials (tabular).
- Size: small
- Collection methods / protocols: It makes use of the European Social Survey (ESS), the German Socio-economic Panel (SOEP), and the UK Household Longitudinal Study (UKHLS). The ESS will be used as a stand-alone; SOEP and UKHLS will be augmented with small-scale area characteristics at various scales to explore context effects. Multivariate regression models will be used and area characteristics will be employed as key independent variables (e.g., to differentiate between different types of rural area).

Reused data #1: European Social Survey ESS (2012/13)

- Formats: .sps
- Origin & provenance: European Social Survey European Research Infrastructure (ESS ERIC) (2023) ESS6 - integrated file, edition 2.6 [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess6e02_6.
- Pre-existing rights of use: European Social Survey data is licensed under CC BY-NC-SA 4.0. The data are available without restrictions, for not-for-profit purposes.

Reused data #2: **German Socio-economic Panel SOEP (v39)**

- Formats: .dta
- Origin & provenance : Citation: Socio-Economic Panel (SOEP), data from 1984-2022, (SOEP-Core, v39, EU Edition), https://doi.org/10.5684/soep.core.v39eu;Collection period: 1984-2022, Publication date: 2024-10-17 https://www.diw.de/en/diw_01.c.918103.en/edition/soep-core_v39eu__data_1984-2022__eu-edition.html
- Pre-existing rights of use (restrictions if any): signed Agreement with DIW Berlin required – the data access process at the SOEP Research Data Center (RDC SOEP) is detailed here.

Reused data #3: **Understanding Society: The UK Household Longitudinal Study UKHLS (v20)**

- Formats: .dta
- Origin & provenance: University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Waves 1-14, 2009-2023 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. *19th Edition.* UK Data Service. SN: 6614, DOI: http://doi.org/10.5255/UKDA-SN-6614-20 https://www.understandingsociety.ac.uk/documentation/access-data/
- Restrictions: UK Data Service Special Licence User Agreements obtained by the Thünen Institute and James Hutton Institute (on behalf of the data custodians of the Understanding Society study), to use sensible individual level data with geo-references or georeferenced data to analyse the subjective wellbeing and quality of life of individuals and families in a sub-national context

1.11. Rural Attractiveness indicators

The purpose of this research output is to collect and analyse indicators covering manifold expressions of rural attractiveness, including gender, age and ethnic differentials.

Generated data:

- Dataset name and type: Research report including research results (text, tables, visuals), detailed results (coefficients and standard errors for common confidence intervals and performance metrics), and methodology detailing model specification, assumptions and diagnosis. The data used to fit the model will also be included, alongside a codebook ensuring reproducibility.
- Formats: Report (text and visuals), and supplementary materials (tabular), STATA/R scripts.
- Size: ~ 2.5 GB
- Collection methods / protocols: direct collection (data download) of statistical layers from published sources, production of own datasets for selected variables (e.g. attractiveness to amenities and aesthetic quality), data pre-processing (calculation of point estimates for survey data; disaggregation, statistical downscaling for lattice data), and analysis using multivariate statistics and regression.
- Data useful for:
    - Policy makers responsible for implementing rural policies.
    - Scientists involved in demographic research, territorial development and rural studies.

Reused data:

- Dataset name and type: The input data will include information about: (1) economy and labour markets, (2) accessibility to essential services, (3) living conditions, (4) territorial capitals and (5) place attachment.
- Formats: All the datasets will be alphanumeric, retrieved in different formats (alphanumeric and spatial). The data files will be collected via direct downloads or accessed programmatically via APIs. No sensitive information, e.g. including personal information or allowing indirect identification of individuals will be used.
- Size: Unknown.
- Origin & provenance: Items (1), (3) and (4) are retrieved from publicly available, free of charge, sources: including the IEA, EC, Eurostat, and National Statistical Institutes. Additionally, item (2) includes accessibility information produced in other tasks in GRANULAR and item (5) is collected from published surveys (European Social Survey).
- Pre-existing rights of use (restrictions if any): All the data used will be retrieved from publicly available sources. Access policies might differ between data providers but, in general, all the data used is expected to be available for scientific use.

1.12. Typologies of rural areas at local/grid level

This output's purpose is to develop a territorial typology(ies), allowing to characterise rural areas and shed light on rural diversity across Europe.

Generated data: Main typology and Add-on typology(ies)

- Type of data: Essentially spatial layers of the typology classes.
- Formats: spatial data (likely at grid-level) (exact format yet to be determined), and possible a version aggregated to administrative units (LAU)
- Size: to be determined
- Collection methods / protocols: to be determined
- Data utility outside the project: mainly intended as an analytical tool for research purposes (at the European level), and specifically those interested in rural territories, but also useful for supporting policy and planning

Reused data:

At this stage, it has not yet been determined which exact datasets will be used for constructing the typology. While this cannot yet be fully determined, key datasets will most likely data on population at grid level (1 km), land use, accessibility, and possibly the built environment. It is possible that data developed in other WPs (e.g. regarding land use and accessibility) will be used for the typology development.

The existing data that will be used will most likely come from public and open data sources, probably from Eurostat or JRC.

1.13. Rural Proofing exercise

The origins of the new data created for rural proofing will be the GRANULAR Living and Replication Labs. It will be collated and interpreted by GRANULAR partners Ecorys and James Hutton Institute.

The data will provide assessments of the potential impacts of policies and initiatives on different types of rural areas, negative or positive, and evidence of characteristics required of policies such that they create positive effects on rural areas. The former will be reported in 2025 in GRANULAR Deliverable D5.2, and the latter will be reported in in 2027 in Deliverable D5.3 in a presentation of the tailoring of rural policies.

Generated data: **GRANULAR rural proofing assessments**

- Formats: MS Excel.
- Size: c. 20Mb
- Collection methods / protocols: Details of data collection protocols set out in GRANULAR Deliverable D5.1 (Guidelines for the Rural Proofing of Transition Policies in Europe)
- Data will be used in interpreting rural proofing assessments of selected policies by Living and Replication Labs.
- Data utility outside the project: The data is expected to be of benefit and used by stakeholders in the Living and Replication Labs, and by parties outwith GRANULAR which are interested in the topic of rural proofing (e.g. countries which have commitments to subjecting prospective policies to rural proofing, or equivalent).

1.14. GIS & Storymaps

The origins of the new data on rural proofing will be the GRANULAR Living and Replication Labs (D5.2). It will be collated and interpreted by GRANULAR partners Ecorys and James Hutton Institute. The narratives in the Storymap(s) will be of characteristics of policies which are tailored such that they create positive effects on rural areas (Deliverable D5.3).
The Storymap(s) will be means of presenting these characteristics in ways that are meaningful to interested parties in policy, research and consultancies.

Generated data: Story maps

- Type of data: Components expected to include examples of digital data, text, images, audio or video.
- Formats: Online Storymap (ESRI ArcGIS); *.pdf or equivalent for exported versions to be stored in Zenodo repository.
- Size: c. 10Mb to 100Mb.
- Collection methods / protocols: Details of the data collected will include: i) materials generated for the rural proofing in GRANULAR Deliverable D5.1 (Guidelines for the Rural Proofing of Transition Policies in Europe); ii) supporting images of datasets generated in other Work Packages (3 and 4), and outwith the GRANULAR project (e.g. consultations of the Labs, graphics from report informing the Labs).

- Data utility outside the project: Data will be used in presenting the findings of characteristics of tailored rural policies, derived in Work Package 5. The data is expected to be of benefit and used by parties outwith GRANULAR who are interested in the processes and findings of the rural proofing and the tailoring of rural policies.

Re-used data:

The rural proofing data will be newly generated (D5.2), informed by data about the area of the Living and Replication Labs if available and appropriate. Data that accompanies the Storymaps are expected to include those from Work Packages 3 and 4, and existing data consulted by the Living and Replication Labs. Use of such data will be illustrative (e.g. pictorial) with no new analysis.

1.15. Multi-Actor Labs participatory collected data (Living Labs and Replication Labs)

GRANULAR relies on sixteen rural territories to advance scientific research and test innovations on a representative set of geographies and socio-political contexts across the European Union and Associated Countries. The work is organised through two types of Multi-Actor Labs: (i) **Living Labs**, which have to co-construct and co-test innovative data collection methods and indicators to support current local policy decisions; and (ii) **Replication Labs**, which have to test the replicability and validate data collection methods and indicators.

To date, only Living Labs have undertaken activities. **The overview of primary research data or database outputs planned in GRANULAR Living Labs is available in Annex 2.**

In addition, the Work Package leader UNIPI promotes empowering activities to contribute to experiences and knowledge sharing among rural communities across Europe, and to scaling up dissemination.

Generated data at WP6 level: Framework for the support of Living Labs in their activities

- Type of data: guidelines and paper
- Formats : text file . txt
- Size : very small
- Collection methods / protocols: minutes, notes carried out by UNIPI during meetings, events in presence, review of monitoring documents, etc.
- Data utility outside the project: relevant practitioners and researchers

Reused data #1 :

- Type of data : Text files gathering LLs composition, activities and participants in terms of number and type of actors, risks and mitigation
- Formats: files .txt and excel sheets
- Size : very small
- Origin & provenance: Monitoring and Evaluation tools and Action Plan monitoring, minutes from WP6 internal meetings and Action Plan presentations

Reused data #2 :

- Type of data: audio recordings, text, images - reflection on the limits and possible solutions to better implement the LLs actions
- Formats : .mp3, jpg, .txt
- Size : very small
- Origin & provenance: Reflexive workshops, Annual Knowledge Exchange Meetings, regular internal meetings

1.16. Communication, Outreach, Dissemination and Exploitation (CODE) data

This research outputs concerns partners and stakeholders feedback data for communication, dissemination and knowledge transfer. The Task and Work Package leader AEIDL will primarily generate data for communication and dissemination purposes, including knowledge transfer and valorisation via the Knowledge Transfer Accelerator.

Generated data:

- Type of data: Qualitative data, such as 1) text (eg, interviews with project partners, survey responses, reports); 2) audiovisual (eg, images, sound recordings, video). PowerPoint presentations and videos of external partners interventions as well as recordings from events including audiovisual and event summaries with quotes.
- Formats : PDF, PPT, MP4, MP3, XLSX, DOCX
- Size : more than 100 MB
- Collection methods / protocols: generated data from partners and stakeholders feedback (e.g. through survey, sharing data) in line with GDPR.
- Data utility outside the project: external stakeholders interested about the progress as well as outputs of the project (e.g. other Horizon EU projects, scientific communities, policy and decision makers, rural general public). Preparing the CODE Strategy (D7.1) and for communication and dissemination purposes (e.g. information on their communication channels, websites, social media).

Reused data:

- Type of data: Email, first name, last name, country, Gender, Organisation name, Stakeholder type Event data
- Formats : XLSX
- Size : less than 100 MB
- Origin & provenance: Via subscription to GRANULAR newsletter and KTA events

2. FAIR Data

As described in the Introduction, all following sections are divided into two parts:

- a first generic part, which reiterates the **core principles of the project's data management strategy**, guiding the use of data and its management throughout the project, and in the creation of its legacy (provided in deliverable D1.3 Data Management Plan - version 1);

- a second, more operational part, which describes the **data management life cycle and specific associated procedures with each research output.**

2.1. Making data findable, including provisions for metadata

2.1.1. Strategy and key principles at project level

Core principles, recommendations and good practices for **making data findable**, are provided in the deliverable D1.3 Data Management Plan - version 1, and summarised below:

- The planned levels of dissemination of GRANULAR deliverables (reports and other materials) are those specified in the Grant Agreement, Table p.91-92 in Annex 1, Part A.
- Data will be made available through internationally recognised domain-specific repositories. Please refer to Section 2.2 (Making Data Accessible).
- Data are clearly identified using an **appropriate universal identifier such as Digital Object Identifier** (D.O.I.; www.doi.org/) and accompanied by **appropriate and sufficient keywords** and search terms. A **GRANULAR metadata format** (detailed in Annex 1), based on Dublin Core standards, is designed to ensure that project materials are located by prospective users, usable by project partners in the future and by third parties external to the project.
- The approach to identifying and selecting keywords will use the US Library of Congress Linked Data Service. This is described in more detail in Section 2.3 (Making Data Interoperable).
- The descriptions and metadata will enable their identification by search engines (e.g. Google, Bing) and research tools (e.g. ResearchGate, Web of Knowledge).
- The responsibility for the selection of relevant keywords and metadata will be that of the relevant research team and Task and Work Package Leaders.
- Once final versions of the relevant deliverables, identified for public dissemination, are uploaded to the EC Grant Management Service, they will be made available on a dedicated page on the GRANULAR website. They will be hyperlinked to the relevant web address or other appropriate location (e.g. Zenodo record), with the appropriate citation details, and a point of contact for further information.

In line with Open Science concepts to guide the use of research data and its management during the project, an operational solution adapted to the project's specificities was developed to upload datasets in the GRANULAR Zenodo community as an archive (.zip), following the **GRANULAR Data Delivery Guidelines (Annex 1).**

The archive is structured as follows:

- A 'Readme' file at the root of the archive that explains the data processing: objectives, input data (what we have used for data processing), data processing and associated documentation, output data (proposal of metadata for each dataset);

- A 'Data' folder (inputs and outputs): contains all processed data in an open format (csv, geopackage, tiff), not xls or shapefiles. For each dataset, the corresponding Metadata file has to be filled in an Excel file, presenting all the information useful to understand the context of production and the meaning of dataset and information related to indicators (indicator code, label, unit of measure…).
- A 'Docs' folder for the relevant documentation (notebook, technical documentation) to reproduce the data processing, explaining the methodology, steps of data processing and software used to compute the results.

GRANULAR Metadata format

Metadata are designed to ensure that project materials are located by prospective users, usable by project partners in the future and by third parties external to the project. Metadata must be associated to each dataset / group of coherent datasets. It is an xls documents structured in two sheets. The content of metadata allows to external to users to understand the content of the data files. The name of the metadata file is "metadata_name of the dataset it describes.xls". Please refer to Annex 1 for further details on the GRANULAR Metadata standards.

2.1.2. Overview per research output

Crowd-sourced data – accessibility

All datasets will be identified with a persistent identifier. So far, two datasets were lodged in the Zenodo online repository and allocated a DOI:

- Dataset #1 *Accessibility Indicators to services at EU scale - 1km grid indicators*: https://doi.org/10.5281/zenodo.14160729
- Dataset #2 *Accessibility Indicators to services at EU scale - aggregated indicators*: https://doi.org/10.5281/zenodo.14440274

The responsibility for the selection of relevant keywords and metadata is that of the Task leader CNRS.

Metadata was developed in accordance with the GRANULAR Data Delivery Guidelines and made available in the Zenodo online repository. Keywords were selected using EU specific glossaries, such as the Glossary for transport statistics (https://ec.europa.eu/eurostat/documents/3859598/10013293/KS-GQ-19-004-EN-N.pdf/b89e58d3-72ca-49e0-a353-b4ea0dc8988f)

Transaction data – mobility data

All data will be identified by a persistent identifier (DOI) and rich metadata will be developed in accordance with the GRANULAR Data Delivery Guidelines. Search keywords will be selected using EU specific glossaries and provided in the Metadata.

The responsibility for the selection of relevant keywords and metadata is that of the Sub Task leader University of Southampton, with the support from UMIL.

Metadata will be potentially developed/prepared according to the SpatioTemporal Asset Catalogs (STAC) standard, representing a language for describing geospatial data so that it can be easier to indexed, discovered, and harvest them.

Nowcasting – social media and housing transaction

Derived datasets and algorithms will be uploaded in the Zenodo online repository and allocated a DOI.

The responsibility for the selection of relevant keywords and metadata is that of the Task and Work Package leader NOR.

Metadata will be developed for new datasets. This will follow recommendations developed in the GRANULAR Data Delivery Guidelines and made available in the Zenodo online repository. Metadata will include keywords to enable it to be located from searches. Metadata will be prepared for each dataset derived.

Remote sensing data (Earth Observation) - land cover/land use

All data will be identified by a persistent identifier (DOI) and rich metadata will be developed in accordance with the GRANULAR Data Delivery Guidelines. So far, only one dataset was lodged in the Zenodo online repository and allocated a DOI:

- Dataset #1 *Sen2LUCAS Dataset*: https://doi.org/10.5281/zenodo.14259723

The responsibility for the selection of relevant keywords and metadata is that of the Task and Work Package leader IAMM.

Citizen science data – rural sustainability

The datasets will be identified with a persistent identifier and metadata will be developed in accordance with the GRANULAR Data Delivery Guidelines and made available in the Zenodo online repository. Metadata will include keywords to enable it to be located from searches.

The responsibility for the selection of relevant keywords and metadata is that of the Work Package leader IIASA.

Climate and environmental indicators

All data will be lodged in the Zenodo online repository, identified by a persistent identifier (DOI) and rich metadata will be developed in accordance with the GRANULAR Data Delivery Guidelines.

The responsibility for the selection of relevant keywords and metadata is that of the Task and Work Package leader IAMM.

Socio-economic resilience, social cohesion and vulnerability indicators

All data (report and supplementary materials) will have DOIs. Simple metadata descriptors will be attached to the supplementary materials. Search keywords will be included.

The responsibility for the selection of relevant keywords and metadata is that of the Task and Work Package leader NOR.

Food systems multidimensional indicators

All data will be lodged in the Zenodo online repository, identified by a persistent identifier (DOI) and rich metadata will be developed in accordance with the GRANULAR Data Delivery Guidelines.

The responsibility for the selection of relevant keywords and metadata is that of the Task and Work Package leader IAMM.

Well-being data

No data will be shared, with the exception of the analysis results of large-scale social surveys (no results will be published for specific areas, only report effect sizes / proportions), which will be presented in Deliverable D4.4 under the responsibility of the Task leader Thünen Institute.

Rural Attractiveness indicators

All data (report and supplementary materials) will have DOIs. Simple metadata descriptors will be attached to the supplementary materials. Search keywords will be included.

The responsibility for the selection of relevant keywords and metadata is that of the Task and Work Package leader NOR.

Typologies of rural areas at local/grid level

The datasets will be identified with a persistent identifier and metadata will be developed in accordance with the GRANULAR Data Delivery Guidelines and made available in the Zenodo online repository.

The responsibility for the selection of relevant keywords and metadata is that of the Task and Work Package leader Nordregio.

All the spatial datasets will conform to Article 5(1) of INSPIRE Directive 2007/2/EC. The methodology used for developing the datasets and performing the analyses will be described in the specific deliverables. Keywords will be selected using EU specific glossaries.

Rural Proofing exercise

Metadata will be designed to ensure that project materials are located by prospective users, usable by project partners in the future and by third parties external to the project. The background to the design and use of metadata is set out in Section 2.2 (Making data accessible). The responsibility for the selection of relevant keywords and metadata will be that of the Task and Work Package leaders Ecorys and James Hutton Institute.

The approach to identifying and selecting keywords will use the US Library of Congress Linked Data Service. This described in more detail in Section 2.3 (Making Data Interoperable).

GIS & Storymaps

The Storymap(s) will be lodged in the Zenodo online repository and allocated a DOI.

Metadata will be designed to ensure that project materials are located by prospective users, usable by project partners in the future and by third parties external to the project. This will follow processes developed for research outputs in the Scottish Government Strategic Research Programmes. The background to the design and use of metadata is set out in the Section 2.2 (Making data accessible). The responsibility for the selection of relevant keywords and metadata will be that of the Task and Work Package leaders Ecorys and James Hutton Institute.

As the James Hutton Institute is responsible for the mapping of soils in Scotland, and related derived datasets, the provision of metadata will follow the standards required by relevant regulations or good practices. Examples of datasets which may be used, and for which metadata have been prepared are at the Natural Asset Register of Scotland (https://openscience.hutton.ac.uk/). This model will be followed for the datasets developed for Scotland, in GRANULAR, such that individual datasets can be identified, and information provided relating to coverage (e.g. all or partial coverage of Scotland), cartographic scale, constraints on use (e.g. legal), and technical information such as source scale, data structure type (i.e. vector, raster), coordinate reference system, licence type, and information about frequency of updates.

Reference:
Joint Research Centre (2013). INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, pp 99. http://inspire.jrc.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf

Multi-Actor Labs participatory collected data (Living Labs and Replication Labs)

The Multi-Actor Labs and the Work Package leader UNIPI are not expected to publish aggregated databases, resulting from the combination of various official statistical databases or from monitoring and reflection activities. Data from Multi-Actor Labs and UNIPI data on the monitoring of Living Labs will therefore not be identified by a permanent identifier.

However, if the decision to publish the data is taken later over the course of the project, the datasets will be identified with a persistent identifier and metadata will be developed in accordance with the GRANULAR Data Delivery Guidelines (internal process, metadata format and standard) and made available in the Zenodo online repository. Metadata will include keywords to enable it to be located from searches, and be prepared for each dataset derived.

Communication, Outreach, Dissemination and Exploitation (CODE) data

CODE Data (e..g. conference proceedings) will be lodged in the Zenodo repository (https://zenodo.org/communities/granular/). Two Highlights' Reports on Knowledge Transfer Accelerator have been deposited in Zenodo with a persistent identifier so far:

- Lostrangio, C., Ntabuhashe, M., & Pazos-Vidal, S. (2023, novembre). GRANULAR Knowledge Transfer Accelerator. Living Labs in rural areas: HOW TO? Highlights' report. https://doi.org/10.5281/zenodo.10843083

- Lostrangio, C., Ntabuhashe, M., & Pazos-Vidal, S. (2024, février). GRANULAR Knowledge Transfer Accelerator. One, None, Thousands Ruralities: A New Vision Based on Functionalities. Highlights' report. https://doi.org/10.5281/zenodo.10843217

Metadata will be provided by the Task and Work Package leader AEIDL to characterise data generated and/or collected, such as for text/audio-visual material (e.g. title, subtitle, description, video duration), event attendees' list (e.g. name, surname, email).

Instead of keywords, tags will be used to optimize data access in the project's website.

2.2. Making data accessible

2.2.1. Strategy and key principles at project level

Core principles, recommendations and good practices for **making data accessible**, are provided in the deliverable D1.3 Data Management Plan - version 1, and summarised below:

Repository choice

- The **Zenodo repository** (https://zenodo.org/communities/granular/) is used as the main online repository to store GRANULAR research outputs and make them accessible to external users within the limits of Intellectual Property Rights. The use of Zenodo is free, with no additional costs expected for its use. DOI will be minted by the repository system and attributed to the data.
- However, if a **Disciplinary Repository** exists and is considered more appropriate for the prospective user audience, it will be strongly recommended to the relevant research team and Task and Work Package Leaders to use it to store data in order to match the most relevant and richer metadata specialised standards.

Data and Metadata availability

- Licences for datasets will be permissive (as open as possible), consistent with restrictions on data release and use (as closed as necessary), such as **GPLv3 (for software) or CC-BY-4.0 (for data and publications)**, and will be made available through the GRANULAR Zenodo community.
- Restrictions due to confidentiality, intellectual property rights or limited accuracy of component datasets would lead to access being restricted. In each case, the reason for a restriction would be assessed against the requirements of the Grant Agreement and obligations to the funders.
- Each partner has to comply with the relevant ethics requirements and guidelines. No data which may have personal or sensitive in nature will be made publicly available. It may be shared within the consortium subject to the specific nature of the data and the data protection measures of the parties involved.

- In accordance with the GRANULAR Data Delivery Guidelines, all metadata will be openly available and licenced under the latest version of the Commons Attribution International Public Licence (**CC BY-NC Attribution-Non Commercial**). This licence allows users to distribute, remix, adapt, and build upon the material in any medium or format for non-commercial purposes only, and only so long as attribution is given to the creator;
- The accessibility period of the metadata and associated dataset will correspond to the trusted repository certificate details, with a **minimum period of 10 years**. The metadata will be timestamped. Beyond this period, it will also be possible to use the data responsible person and/or institution contact to request access to the data.

2.2.2. Overview per research output

Crowd-sourced data – accessibility

The Accessibility indicators produced and their associated methodology (literate programming) are lodged in Zenodo with a persistent identifier (https://zenodo.org/records/14160729 and https://zenodo.org/records/14440274).

The Zenodo solution is appropriate to that data: there is no copyright issues on the data used (open licence). Maximum size for a dataset on Zenodo is 50GB (and it is possible to have multiple datasets), which feets with the dataset's specificities.

Datasets are made openly available (with an open-source licence). There is no restricted access, embargo or restrictions on use. Metadata are also openly available and licenced under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial).

The accessibility period of the data corresponds to the trusted repository certificate details (Zenodo) i.e. at least the next twenty years. The metadata is associated to the data in the data repository, as a tabular file.

Metadata was developed in accordance with the GRANULAR Data Delivery Guidelines. Therefore, the data delivery includes:

- Input and data processing description, using literate programming (Quarto Documents, .HTML format)
- Data outputs (.csv file)
- Associated metadata to data outputs (.csv or .ods file).

Transaction data – mobility data

The data will be deposited in the WorldPop Storage, hosted by the University of Southampton, which is a trusted and well-recognized repository for long-term preservation and curation of spatial demographic datasets. WorldPop Researchers are involved in the GRANULAR Project and co-leading the production of the mobility data.

DOIs for all data stored in the WorldPop Storage are assigned and registered by the Southampton Data Library (through DataCite), which ensures the stability of the persistent URL links associated to them.

All modelled outputs will be made openly available (but not likely for the raw mobility data because of legal and contractual restrictions). No embargo is expected for the modelled outputs.

Data will be accessible as HTTPS and perhaps API. No restrictions on use are expected at this time.

Persons accessing the data will remain anonymous.

As per Grant Agreement, metadata will be openly available and licenced under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial). The metadata will contain at least the data DOIs expressed as links to the corresponding data landing pages.

Data will be stored in the WorldPop storage where, as part of the agreement between the WorldPop Research Group and the University of Southampton Library, they will remain available for at least ten years.

Relevant open-source scripts/codes and software (e.g. R, Python, etc.) will be included in the documentation provided along with the data.

Nowcasting – social media and housing transaction

The datasets produced and their associated methodology will be lodged in the Zenodo repository (https://zenodo.org/communities/granular/). It is expected that the final versions of each derived dataset will be shared. A DOI will be generated and linked to the relevant data and outputs.

Datasets are made openly available. No embargo expected. No restrictions on use. Metadata will be openly available and licenced under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial).

No data access committee is envisaged (no personal/sensitive data are used).

The accessibility period of the data corresponds to the trusted repository certificate details (Zenodo) i.e. at least the next twenty years. The metadata are associated to the data in the data repository, as a tabular file. Metadata will be timestamped. Versioning of softwares used will be provided when relevant.

Remote sensing data (Earth Observation) - land cover/land use

The Sen2LUCAS datasets are stored in the Zenodo repository with a persistent identifier (https://zenodo.org/records/14259723).

Datasets are made openly available (with an open-source licence). It is not expected to have any restrictions except in the case of limited accuracy, which would lead to a restricted access. All finalised, quality controlled and non-confidential datasets will be made available to access. Metadata are also openly available and licenced under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial).

The accessibility period of the data corresponds to the trusted repository certificate details (Zenodo) i.e. at least the next twenty years. The metadata are associated to the data in the data repository, as a tabular file.

Citizen science data – rural sustainability

Derived datasets will be lodged in the Zenodo repository (https://zenodo.org/communities/granular/). It is expected that the final versions of each derived dataset will be shared. A DOI is generated and linked to the relevant data and outputs.

Once the work is complete the resulting data will be made openly available. Data will be made accessible through standardized access protocol (Jupyter notebooks).

Metadata will be openly available and licenced under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial).

The accessibility period of the data corresponds to the trusted repository certificate details (Zenodo) i.e. at least the next twenty years. The metadata are associated to the data in the data repository, as a tabular file. Metadata will be timestamped. Versioning of softwares used will be provided when relevant.

Climate and environmental indicators

The datasets will be deposited in Zenodo and allocated a DOI.

All data generated or collected will be made fully and openly available without any restrictions. There are no legal, contractual, or intentional reasons to limit access, and no embargo periods will be applied.

The data will be accessible through Zenodo, which provides a free and standardized access protocol, ensuring seamless availability to all users.

Since there are no restrictions on use, access will not require additional measures either during or after the end of the project. The data will be openly accessible without the need to ascertain the identity of users. Consequently, there is no need for a data access committee or other mechanisms to evaluate or approve access requests.

All metadata will be made openly available and licensed under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial) as per the Grant Agreement. The metadata will include all necessary information to enable users to access the corresponding data, such as descriptions, formats, keywords, and links to the datasets.

The data and metadata will remain available and findable during the period corresponding to the trusted repository certificate details (Zenodo) i.e. at least the next twenty years. Even if the data were to become unavailable in the future, metadata will remain accessible to maintain a record of the datasets and their associated details.

Documentation about the software needed to access or process the data will be provided, including references to Python scripts and Excel templates. Whenever applicable, open-source Python code will be included in the repository to ensure transparency and ease of use for users accessing and analyzing the data.

Socio-economic resilience, social cohesion and vulnerability indicators

All the data generated in the research, including research report, detailed methodology, coding scripts and descriptive tables will be deposited in a trusted repository (Zenodo) as soon as possible. The repository ensures that the data is assigned a DOI.

All the data generated in this task, including scientific publications and supporting materials (tables and scripts), will be deposited and made available free of charge (open access). The data will be made accessible using the access protocols available in Zenodo.

The underlying data (microdata files) will NOT be deposited due to the limitations established in the Terms of Use defined by Eurostat.

Metadata will be made openly available under a public domain and contain information to enable the user to access the data (CC BY-NC Attribution-Non-Commercial). The information will be generated using the metadata generation tools in the trusted repository (Zenodo). Metadata will be timestamped.

The data will remain available and findable for twenty years, according to the standard policy established by the trusted repository (Zenodo). No specific software will be necessary to access the data.

Food systems multidimensional indicators

The data will be deposited in Zenodo, which provides a free and standardized access protocol, ensuring seamless availability to all users.

Since there are no restrictions on use, access will not require additional measures either during or after the end of the project. The data will be openly accessible without the need to ascertain the identity of users. Consequently, there is no need for a data access committee or other mechanisms to evaluate or approve access requests.

All metadata will be made openly available and licensed under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial) as per the Grant Agreement. The metadata will include all necessary information to enable users to access the corresponding data, such as descriptions, formats, keywords, and links to the datasets.

The data and metadata will remain available and findable, with long-term preservation ensured by the Zenodo repository. Even if the data were to become unavailable in the future, metadata will remain accessible to maintain a record of the datasets and their associated details.

Documentation about the software needed to access or process the data will be provided, including references to Python scripts and Excel templates. Whenever applicable, open-source Python code will be included in the repository to ensure transparency and ease of use for users accessing and analyzing the data.

Well-being data

The SOEP and UKHLS data are not openly available and can only be accessed under special or secure licensing conditions. Access needs to be arranged through individual research teams. Onward sharing of the micro-data is not permitted. Therefore, results for specific areas will not be published, only report effect sizes / proportions.

The analysis of large-scale social surveys will be presented in Deliverable D4.4, which will be deposited in Zenodo. Metadata will be made openly available under a public domain and contain information to enable the user to access the data (CC BY-NC Attribution-Non Commercial). The information will be generated using the metadata generation tools in the trusted repository (Zenodo). Metadata will be timestamped.

Rural Attractiveness indicators

All the data generated in the research, including research report, detailed methodology, coding scripts and descriptive tables will be deposited in a trusted repository (Zenodo) as soon as possible and made available free of charge (open access).

The data will be made accessible using the access protocols available in Zenodo.

Metadata will be made openly available under a public domain and contain information to enable the user to access the data. The information will be generated using the metadata generation tools in the trusted repository (Zenodo). The data will remain available and findable for twenty years, according to the standard policy established by the trusted repository (Zenodo).

No specific software will be necessary to access the data.

Typologies of rural areas at local/grid level

The datasets produced and their associated methodology will be lodged in the Zenodo repository (https://zenodo.org/communities/granular/). It is expected that the final versions of each derived dataset will be shared. A DOI will be generated and linked to the relevant data and outputs.

Add-on typologies will be described in deliverable D4.7. The spatial layers of the typology(ies) that will be developed will be made accessible through a free and standardized protocol (exact format to be determined)

No embargo expected. No restrictions on use. No data access committee is envisaged (no personal/sensitive data are used).

Metadata will be openly available and licenced under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial).

The accessibility period of the data will correspond to the trusted repository certificate details (Zenodo) i.e. at least the next twenty years. Metadata are associated to the data in the data repository, as a tabular file.

Versioning of softwares used will be provided when relevant.

Rural Proofing exercise - data collected from the Multi-Actors Labs

The rural proofing data will be stored in the Zenodo data repository. A DOI will be minted by the repository system and attributed to the rural proofing dataset.

The collated rural proofing dataset will be openly available. The process of cleaning the data will identify any points that may be confidential, or conflict with obligations for data security or General Data Protection Regulation (GDPR).

Once data is made available for re-use, no embargo will be in place. By this stage, partners will have discharged their responsibilities to produce the scientific deliverables using the data, and any academic papers which draw directly on the data gathered.

Data are planned to be free to use and available for download from the Zenodo repository and project website. Initially, a request for those accessing the data to record their contact details is proposed. Recording the identity of those accessing or using the data will be kept under review.

No data access committee is envisaged.

Metadata will be made openly available under a public domain license. It will provide a link to the dataset on the Zenodo repository and the project webpage. The metadata and associated dataset will be accessible on the repository in perpetuity.

The background to the dataset and template is described in deliverable D5.1. The dataset will be described in relevant deliverables (D5.2 and D5.3). All deliverables will be in the public domain. Software required to read the spreadsheet will be MS Excel, or another package that can read common file formats for spreadsheets or text files (e.g. ascii, txt, csv).

GIS & Storymaps - secondary data analysis

Exported versions of Storymap materials will be saved as .pdf in the Zenodo data repository (https://zenodo.org/communities/granular/) or an equivalent considered more appropriate for the prospective user audience. A DOI will be minted by the repository system and attributed to the Storymap(s).

It is expected that component or underlying data will be shared subject to issues of confidentiality or licencing. Component data may be embargoed until academic publication.

It is expected that the final versions of each derived dataset will be shared. This will depend upon whether any issues of confidentiality or licencing arise regarding the underlying data. The process of cleaning the data will identify any points that may be confidential, or conflict with obligations for data security or General Data Protection Regulation (GDPR).

Once data are used in a Storymap it is expected to be available for re-use with no embargo in place. By this stage, partners will have discharged their responsibilities to produce the scientific deliverables using the data, and any academic papers which draw directly on the data gathered.

The Storymap(s) will be accessible as public domain materials from the ESRI ArcGIS server. Access to the *.pdf version in the Zenodo repository will be free with the default of a direct download.

Restrictions due to confidentiality, intellectual property rights or limited accuracy of component datasets would lead to access being restricted. In each case the reason for a restriction would be assessed against the requirements of the Grant Agreement and obligations to the funders.

At the end of the project the status of each dataset will be assessed and a decision made on any legacy restriction, or withdrawl of a dataset.

The expected status is that all finalised, quality controlled and non-confidential datasets will be made available as free-to-access.

No information will be sought on the identity of an individual or organisation accessing the Storymaps

However, access and uses of component data would be subject to certain agreements (e.g. legality, liability and appropriateness over uses), which may require some form of exchange of details. Information on arrangements would be provided under each specific dataset.
No specific committee is required for assessing and approving access to the Storymap(s).

Each partner has to comply with the relevant ethics requirements and Intellectual Property guidelines. No data which may have personal or sensitive in nature will be made publicly available. It may be shared within the consortium subject to the specific nature of the data and the data protection measures of the parties involved.

Metadata will be openly available and licenced under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial). It will provide a link to the dataset on the Zenodo repository and the project webpage. The data and metadata will remain available and findable during the period corresponding to the trusted repository certificate details (Zenodo) i.e. at least the next twenty years. The metadata will be timestamped.

The product will be described in the relevant deliverable (D5.3). Any limitations on the use of the data (e.g. sensitivity to changes in input data) will be highlighted to provide prospective users with information to inform their decision over the uses of the data.

The Storymaps and multi-media products will be readable in a free-to-use web-based interface (ESRI ArcGIS storymap).

Multi-Actor Labs participatory collected data (Living Labs and Replication Labs)

Multi-Actor Labs are not expected to publish datasets. However, if this is the case, datasets will be lodged in the Zenodo repository (https://zenodo.org/communities/granular/) and be given a DOI.

Data collected in the GRANULAR Multi-Actor Labs (see Annex 2) are expected to include interviews and post-event surveys. These means of collecting data are all subject to relevant legal requirements and appropriate best practices to protect personal data (i.e. comply with data protection obligations), and ethical considerations (e.g. rights of participants to know the purposes to which information will be used, and by whom) as set out in Article 14 (Ethics and Values), and GRANULAR Deliverable D1.1 Ethics Guidelines.

At Work Package level (responsibility of the Task and Work Package leader UNIPI), non-sensitive data will be accessible upon GRANULAR open access principles. Once data are made available for re-use, no embargo will be in place. At this stage, partners will have discharged their responsibilities to produce the scientific deliverables from GRANULAR uploaded in the Zenodo repository (https://zenodo.org/communities/granular/), and academic papers which draw directly on the data gathered.

Specific Multi-Actor Labs data will be accessible only to authorized users, mainly to data uploaders, Living Labs members and the person in charge of each Living Lab in the project. Much of the information will be of a personal nature, and will need to be taken into account to track the Living Labs activities, but will not be publicly available.

Video and audio files and their transcripts are used by UNIPI to make the Task reports according to the Grant Agreement at Confidential or Public level. Non-digital data (handwritten laboratory notebooks, paintings...) are converted to a digital source. If a real-world object is not easy to scan, a digital photograph(s) at the highest resolution possible will be taken.

Data are stored in the UNIPI institutional repositories (certified/Trustworthy Repository) and may be shared through the project collaboration platform, the GRANULAR Nextcloud, an open-source self-hosted private cloud solution based on IAMM's servers (see Section 5. Data Security).

The deliverables, if public, all data and files will be uploaded in the Zenodo repository (https://zenodo.org/communities/granular/). Metadata will be openly available and licenced under the latest version of the Commons Attribution International Public Licence (CC BY-NC Attribution-Non Commercial). It will provide a link to the dataset on the Zenodo repository and the project webpage. The data and metadata will remain available and findable during the period corresponding to the trusted repository certificate details (Zenodo) i.e. at least the next twenty years. The metadata will be timestamped.

CODE data

CODE Data (e..g. conference proceedings) will be lodged in the Zenodo repository (https://zenodo.org/communities/granular/).

Not all datasets will be publicly available to comply with GDPR rules (e.g. participant lists to events, newsletter subscribers etc). Other data will be publicly available and used for communication and dissemination purposes (e.g. videos, social media posts). No embargo expected.

According to data, different privacy policies and data management policies apply.

Metadata is guaranteed to remain available after data is no longer available. No dedicated software required for accessing or reading the CODE Data.

2.3. Making data interoperable

2.3.1. Strategy and key principles at project level

Core principles, recommendations and good practices for **making data interoperable**, are provided in the deliverable D1.3 Data Management Plan - version 1, and summarised below:

- The vocabularies used for describing the datasets produced in GRANULAR will be those recognised as standard for use by libraries, in particular standards and vocabularies such as those promulgated by the US Library of Congress (Library of Congress, 2017) via the Linked Data Service, and terminology developed by relevant peer groups.
- Specific identifiers will be embedded in the metadata for datasets and public deliverables leading to them being more effectively located by internet search engines, and documented in international libraries.
- Data generated in the framework of GRANULAR will include qualified references to re-used data and/or other data from previous work.
- The responsibility for selecting the most appropriate keywords and terms will lie with the authors of each report and associated datasets. Overall checking of the presence of keywords, terms and metadata will be part of the quality control process at Work Package level.
- For specific forms of data, relevant thematic or technical guidelines will be used. For specific forms of data, relevant thematic or technical guidelines will be used. For example, spatial data will use terms from the EU INSPIRE Directive, European Environment Agency glossary of definitions, ESPON Database Dictionary of Spatial Units, or widely used ESRI GIS reference dictionary. The glossary of the UK Digital Curation Centre will also be used for technical terms relating to data preservation.

2.3.2. Overview per research output

No specific actions are required to ensure data interoperability as all datasets will use widely accepted standards, formats, and methodologies, such as CSV, GeoJSON, and Excel, which are inherently interoperable and allow for data exchange and reuse across disciplines.

Metadata will also follow standard practices, ensuring compatibility without the need for project-specific ontologies or vocabularies.

However, it should be noted that some measures are taken in particular for the following research outputs:

- Socio-economic resilience, social cohesion and vulnerability indicators (see Section 1.8): Standardized vocabularies and ontologies to used are those provided by the Social Science Research Network (SSRN). The following disclaimer regarding microdata use will be added to all published materials: "This study/report/paper is based on data from Eurostat, LFS, 2025. The responsibility for all conclusions drawn from the data lies entirely with the author(s)".
- Rural Attractiveness indicators (see Section 1.11): Standardized vocabularies and ontologies to used are those provided by the Social Science Research Network (SSRN).

2.4. Increase data re-use

2.4.1. Strategy and key principles at project level

Core principles, recommendations and good practices for **increasing data re-use**, are provided in the deliverable D1.3 Data Management Plan - version 1, and summarised below:

- Data will be made **freely available, accessible** from a repository and project website in line with the obligations in the Grant Agreement, **licenced** (CC BY-NC Attribution-Non-Commercial) and **comprehensible to external users** (e.g. the European Commission, Living Labs and Replication Labs, other researchers) by providing detailed information on data processing through metadata, that allows to external to users to understand the content of the data files, in an open format.
- All the relevant **documentation** needed to validate data analysis and facilitate data re-use is provided in the metadata and data processing notebooks, in accordance with the GRANULAR Data Delivery Guidelines. The provenance of the data is also documented in metadata and in the data processing documentation.
- At project-level, roles and responsibilities in data management, including the **data quality assurance process**, are detailed in Annex 3.
- The details of existing datasets used in the research, or production of new data, will be reported in the .zip archive for metadata or the relevant Deliverable. These will identify the source(s) of the data used, refer to any relevant limitations on its use with respect to the purpose in GRANULAR, clearance of any ethical considerations associated with the data, and ensure that it is cited as guided by the data provider or at least to appropriate standards.

The project outputs and activities, and associated benefits from project outcomes, will be targeted at the most relevant stakeholder types (defined in the Communication, Outreach, Dissemination and Exploitation (CODE) Strategy, D7.1) to seek impact in relation to creating avenues for long-lasting mechanisms improving interfaces between society, science and policy makers on rural development, as set out in the Grant Agreement.

2.4.2. Overview per research output

Crowd-sourced data – accessibility

All the relevant documentation needed to validate data analysis and facilitate data re-use is provided in the metadata and data processing notebooks, in accordance with the GRANULAR Data Delivery Guidelines. The provenance of the data is also documented in metadata and in the data processing documentation.

Data are made freely available with the CC-BY-SA license.

Data have been described in official EU territorial divisions (1km grid, NUTS). Consequently, it can easily be used for further studies or data visualisation tools.

Several aspects in the data processing is systematically considered:

- Completeness of Point of Interest POIs (comparisons to reference institutional data);
- Geographic accuracy of POIs;
- Comparing routing engines outputs to other reference database (such as carpooling data https://www.data.gouv.fr/fr/datasets/trajets-realises-en-covoiturage-registre-de-preuve-de-covoiturage/);
- Reproducibility of the methodology proposed;
- Feedbacks from Living Labs on indicators accuracy;
- Exchanges with the EU Joint Research Centre (JRC) on data quality, EU reference datasets and indicators accuracy.

Transaction data – mobility data

All the relevant documentation needed to validate data analysis and facilitate data re-use will be provided in the metadata and data processing notebooks (readme file with information on methods and code development), in accordance with the GRANULAR Data Delivery Guidelines. The metadata associated to the modelled outputs will document both the provenance of the data and the sources of the input data used to produce them.

As per Grant Agreement, the modelled outputs will be made freely available with the CC-BY-SA license (as the raw data are usually restricted by the proprietor).

The modelled outputs will be useable by third parties both during and after the end of the project.

Input data will be screened and cleaned prior to modelling - model validation metrics and quality check results will then be provided in the documentation associated to the generated data.

All research outputs, including data and findings, will be made freely available online with a DOI in digital format and published on open access journals (to maximize FAIR principles) using the allocated project budget or, where possible, relying on in-place bilateral agreements between publishers and universities enabling staff to publish their research without having to pay APCs. All input data will be stored in a password secured database, servers along with readme files and metadata, and will only be made available to project staff. Before using the input data, ethical clearance was sought and obtained through the University of Southampton.

Nowcasting – social media and housing transaction

All the relevant documentation needed to validate data analysis and facilitate data re-use is provided in the metadata and data processing notebooks, in accordance with the GRANULAR Data Delivery Guidelines. The provenance of the data is also documented in metadata and in the data processing documentation.

Data will be made freely available with the CC-BY-SA license.

It is expected that the data produced will be usable by third parties once it is made available publicly.

The provenance of each of the input data will be clearly stated, and of the organisation creating a derived dataset, or publishing a dataset from new measurements.

Several aspects in the data processing is systematically considered for quality assurance:

- Conceptual consistency, according to state-of-the-art research;
- Reproducibility of the methodology proposed;
- Validation of results with existing assessments of transport poverty for specific countries (Germany, Spain) and official statistics on housing prices in Sweden;
- Feedbacks from Living labs on indicators accuracy;
- Exchanges with the EU commission (JRC) on data quality, EU reference datasets and indicators accuracy.

Remote sensing data (Earth Observation) - land cover/land use

All the relevant documentation needed to validate data analysis and facilitate data re-use is provided in the metadata and data processing notebooks, in accordance with the GRANULAR Data Delivery Guidelines.

The provenance of the data is also documented in metadata and in the data processing documentation.

Data will be made freely available with the CC-BY-SA license. It is expected that the data produced will be usable by third parties once it is made available publicly.

The provenance of each of the input data will be clearly stated, and of the organisation creating a derived dataset, or publishing a dataset from new measurements.

The dataset was downloaded using the Planetary Computer API, which provides pre-processed satellite imagery. To ensure data quality, the following steps were undertaken:

- Cloud Coverage Filtering: Images were processed to minimize cloud coverage, ensuring clear and usable data.
- Pixel-Level Verification: Each image was examined to check for missing portions caused by tile selection errors, ensuring completeness.
- Temporal Filtering: A month-based filtering process was applied. For instances where images for the target month were unavailable, the closest available month with cloud-free images was selected to maintain consistency across the dataset.

Citizen science data – rural sustainability

All the relevant documentation needed to validate data analysis and facilitate data re-use is provided in the metadata and data processing notebooks (code where possible, using in some cases R markdown or similar code based descriptions) in accordance with the GRANULAR Data Delivery Guidelines. The provenance of the data is also documented in metadata and in the data processing documentation.

Data will be made freely available with the CC-BY-SA license. It is expected that the data produced will be usable by third parties once it is made available publicly.

The provenance of each of the input data will be clearly stated, and of the organisation creating a derived dataset, or publishing a dataset from new measurements.

As part of the quality assurance process, the results will be compared with officially generated datasets wherever possible - i.e. EuroStat.

Climate and environmental indicators

All data will be openly licensed under CC-BY-SA. A README file and detailed instructions will be included with each dataset, ensuring clear guidance for users), in accordance with the GRANULAR Data Delivery Guidelines.

Python scripts will be provided where applicable to support replication and reuse.

Socio-economic resilience, social cohesion and vulnerability indicators

Detailed methodology and scripts will be included as supplementary materials to the main report. These will make the research entirely reproducible.All the output data will be made available in a Zenodo repository, free of charge, making it usable by third parties.

The provenance of the data will be thoroughly documented (referenced) using the appropriate standards.

Data assurance will be ensured through the definition of reliability thresholds (e.g. estimates below reliability limit 'a', weighted, shall not be published; Estimates below reliability limit 'b' , weighted, shall be published with a warning concerning their limited reliability). Additionally, a two-tier peer review process involving internal and external (independent) reviewers.

Particular attention will be posed to avoid disclose personal information: no result or visual (such as table, graph, or map) including geographically explicit information on sensitive information (such as income-distribution indices) shall be disclosed for statistical units below the Functional Rural Area level (population of at least 25,000 inhabitants) level. In general, results below three observations (unweighted sample), shall not be published.

Food systems multidimensional indicators

All data will be openly licensed under CC BY. A README file and detailed instructions will be included with each dataset, ensuring clear guidance for users), in accordance with the GRANULAR Data Delivery Guidelines.

Well-being data

The SOEP and UKHLS data are not openly available and can only be accessed under special or secure licensing conditions. Therefore, due to data protection law, the raw microdata used in the analysis cannot be shared.

The analysis of large-scale social surveys (no results will be published for specific areas, only report effect sizes / proportions) will be presented in Deliverable D4.4.

Additionally, programmes used to clean and analyse the micro-data will be posted on the GRANULAR project Digital Platform to permit third parties to replicate the set-up of the analysis (see Section 3. Other Research Output).

Rural Attractiveness indicators

Detailed methodology and scripts will be included as supplementary materials to the main report. These will make the research entirely reproducible.

All the output data will be made available in a Zenodo repository, free of charge, making it usable by third parties.

The provenance of the data will be thoroughly documented (referenced) using the appropriate standards.

Data assurance will be ensured through a two-tier peer review process involving internal and external (independent) reviewers.

Particular attention will be posed to avoid disclose personal information: no result or visual (such as table, graph, or map) including geographically explicit information on sensitive information (such as income-distribution indices) shall be disclosed for statistical units below the Functional Rural Area level (population of at least 25,000 inhabitants) level. In general, results below three observations (unweighted sample), shall not be published.

Typologies of rural areas at local/grid level

NOR will document all our methods in the specific deliverables and publish the code used to produce our results online (e.g. on GitHub). All the relevant documentation needed to validate data analysis and facilitate data re-use is provided in the metadata and data processing notebooks, in accordance with the GRANULAR Data Delivery Guidelines. The provenance of the data is also documented in metadata and in the data processing documentation.

Data will be made freely available with the CC-BY-SA license. Data will be useable by third parties. However, the data should be available "as-is". No commitment whatsoever to provide guidance on how to use it or further updates.

The provenance of each of the input data will be clearly stated, and of the organisation creating a derived dataset, or publishing a dataset from new measurements.

Regarding data quality assurance, this cannot yet be detailed, as the exact process and specific details will be determined in the coming stages of the work.

Rural Proofing exercise

Documentation will be provided in the associated GRANULAR project deliverables D5.1, 5.2 and 5.3, and note in a tab in the spreadsheet containing the data.

The data will be free to use, accessible from a repository and project website in line with the obligations in the Grant Agreement.

The expectation is that the data will be re-usable by third parties after the end of the project (e.g. in research, or informing policy teams of the process and, or, findings relating to specific policies).

The data origins will be clearly recorded to the level of the Living or Replication Lab and organisation of the Facilitator. The names of individual members of the Lab will not be published unless with their agreement.

Data quality will be reviewed by the Living and Replication Labs to check for clear errors in data entry, transcription or translation. The logic of the data entries will be checked by the Task leaders (Ecorys and James Hutton Institute). Details of wording will be checked by the Task leaders (Ecorys and James Hutton Institute), in communication with the relevant Living or Replication Lab teams as appropriate. A final check of data content will be carried out with the Labs in advance of uploading of data and publication on a repository. Checks will be made on issues of research ethics, confidentiality and resource requirements which may restrict use, and noted in the associated data description.

GIS & Storymaps

Documentation will be provided in the associated GRANULAR project deliverables D5.1, 5.2 and 5.3, in line with the GRANULAR Data Delivery Guidelines. The data used for these outputs will be from existing materials and thus no further data analysis or cleaning will be required.

The data will be free to use, accessible from a repository and project website in line with the obligations in the Grant Agreement, under a CC-BY-SA license. The expectation is that the data will be re-usable by third parties after the end of the project (e.g. in research, or informing policy teams of the process and, or, findings relating to specific policies).

The data origins will be clearly recorded to the level of the project partner, Work Package and Task, or Living or Replication Lab if appropriate. The names of individual participants, including members of Labs, will not be published unless with their agreement.

The specific quality control processes cannot be specified until the nature and contents of the data are known. Data quality will be reviewed by the providers (e.g. Task, deliverable authors, Living and Replication Labs) to check for clear errors in representation of materials. Checks will be made on issues of research ethics, confidentiality and resource requirements which may restrict use, and noted in the associated data description.

Multi-Actor Labs participatory collected data (Living Labs and Replication Labs)

All the relevant documentation needed to validate data analysis and facilitate data re-use is provided in the metadata and data processing notebooks (Readme files with information on methodology, analyses, variable definitions; glossary of terms; electronic notebook), in accordance with the GRANULAR Data Delivery Guidelines.

Provenance will be mainly recorded in a single README text file that describes the data collection and processing methods, but it might also involve a visualisation of the machine-readable representation such as VisTrails.

The data will be checked automatically during the upload process and manually by Multi-Actor Labs' staff and the uploader. This also involves cross-checking relevant information found in different databases and, in general, ensuring that there are no inconsistencies in the information found in a database or set of databases.

The management of other research outputs that are generated/re-used in the project (e.g., protocols, models, draft policies, new educative resources...) will be discussed and, when relevant, their compliance to the FAIR principles will be detailed.

CODE data

No documentation to validate analysis is expected for CODE Data. When possible, data will be made freely available with the CC-BY-SA license (e.g. certain communication outputs). Depending on the data type, data will be useable by third parties.

The quality assurance process for CODE Data does not include data checking, data correction, proofreading. Given the nature of the collective data (audiovisual), this will be the responsibility of the relevant Communications Manager to ensure accuracy and quality of materials collected.

3. Other research outputs

In addition to the management of all the research outputs detailed in this Data Management Plan, other research outputs are generated or re-used throughout the GRANULAR project. Project partners must consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

Other research outputs generated include:

- At project-level: the Rural Diversity Prototype Compass (Deliverable D2.2) as an analytical tool designed to conceptualise diversity to inform EU, national and local actors in the development of evidence-based, place-based, integrated and tailored policies. All GRANULAR novel data and indicators will be aligned with the general framework of the Rural Compass.

This tool provides a refined understanding of functional characteristics of rural areas that aims to provide deeper insights in contemporary rural differentiation drivers, tendencies and outcomes. On the basis of this Rural Compass prototype, the Rural proofing exercise (see Section 1.13) will develop an agile framework for translating the EU ambitions for rural areas into specific targets, which will be tested with a subset of Multi-Actor Labs.

- At project-level: the GRANULAR Digital Platform, central to disseminating project results.

It will serve as an online interface to navigate datasets and indicators produced by WP3, WP4, and WP5, ensuring accessibility and usability for diverse stakeholders, including policymakers, researchers and stakeholders from Multi-Actor Labs and other rural areas.

The platform focuses on optimising navigation through data, and producing meaningful visualizations tailored to specific audiences. The GRANULAR Digital Platform provides:

- a **Repository** that integrates different types of data. Data are either obtained or developed from various sources[2];

- a **Visualisation Interface** divided into two tools: (i) a **'Data Viewer'** which will provides a geospatial interface enabling an intuitive and user-friendly exploration of WP3 and WP4 data; and (ii) a **'Visualisation Toolkit',** which creates interactive visual representations of higher-level, policy-oriented aggregated information from WP4/WP5.

All datasets and indicators displayed in the Digital Platform will be made available with their metadata, previously developed by the relevant research team in accordance with the GRANULAR Data Delivery Guidelines. Only aggregated results will be visualised within the platform; with links to reliable repositories (with DOIs) for accessing full datasets. Datasets displayed on the Digital Platform are made openly available. No restricted access are planned. Keywords and metadata should have been previously developed and selected by the relevant research team (as described in the other sections of this Data Management Plan).

- Crowd-sourced data (see Section 1.2): A **web viewer tool** (under construction) and based on data delivery #1 and #2 will allow to explore easily the content of these data deliveries. It will be based on the Gridviz JS library, developed especially to visualize EU grid cells.

- Remote sensing data (see Section 1.5): The code used for deep learning framework to derive satellite land use images based on text queries is freely available through github (https://github.com/pallavijain-pj/SenCLIP). Code also provide the dataset loading and pre-processing scripts.

[2] The GRANULAR digital platform enables access to a total number of 17 datasets and their associated metadata so far

- Well-being data (see Section 1.10): **programmes (Stata and R-code) used to clean and analyse the micro-data** will be posted on the GRANULAR project Digital Platform to permit third parties to replicate the set-up of the analysis.

4. Allocation of resources

The costs for collecting and storing the data are allocated to relevant partners under the Grant Agreement or other arrangements (e.g. UKRI Guarantee Fund).

The principal costs for making data FAIR are for preparing the suitable documentation and metadata, upload to the repository, and to provide relevant links to project and institutional websites. No specific non-staff costs are anticipated for making such research outputs available.

All costs for the visualisation of data on the GRANULAR online Digital Platform are included in the budgets of relevant partners and Work Packages (mainly Work Package 7). Additional costs could be incurred for publishing data in open access scientific journals which will require to be covered by partner budgets for non-staff costs, or other funds made available by individual partner organisations.

The responsibility for collecting, analysing and managing scientific data (e.g. surveys, workshops, case study interviews, spatial data) is that of the relevant partner. These responsibilities include the quality control of the data, its documentation and safe storage (see Section 5. Data Security). Each partner has designated the person(s) responsible for data management.

Please refer to Annex 3 for further information on responsibilities chart for data management (including GRANULAR project naming convention).

5. Data Security

All partners have **back-up systems** operated by their organisations. All final outputs (datasets, databases, reports) will be stored in the Zenodo repository for **long term preservation**. In addition to Zenodo, one research output will be safely stored in other trusted repositories. This is the case of Mobility data (see section 1.3) which will be stored in the WorldPop Storage, which is a trusted and well-recognized repository for long-term preservation and curation of spatial demographic datasets. Data stored/archived in the WorldPop Storage are backed up at regular intervals on secure, password protected servers managed by University of Southampton's iSolutions.

Best practice will be followed in **cybersecurity and preventing unauthorised access to research, administration and financial data** (e.g. UK Government National Cyber Security Strategy 2016-2021). Among the partners, James Hutton Institute is Cyber Essentials Plus Certified, which is renewed annually. The Institute follows and implements appropriate data protocols, an example of which is the Scottish Government Cyber Security Public Sector Action Plan. The Institute has robust backup procedures in place for data recoverability, auditing, security, and research integrity and data use, which researchers RESAS researchers are required to follow. Training is provided in the relevant data procedures, and mandatory training on cyber security.

Intermediate versions of data will be managed through the storage and back-up arrangements of the relevant partners, complying with good practice on back-up and deletion of data (e.g. 'Data protection and privacy ethical guidelines', (European Commission, 2009) and the General Data Protection Regulation (GDPR) (EU) 2016/679).

Data relating to individuals (e.g. survey respondents, interviewees, etc.) will be managed with adherence to the ethical guidelines set out in the Grant Agreement Article 14, and summarised in Section 7 below. No transfer of sensitive data is envisioned.

GRANULAR internal management tools

GRANULAR is using tools for data and document management and communications within the project team, supervised by IAMM as Coordinator.

IAMM has set-up a **project collaboration platform, the GRANULAR Nextcloud**, an open-source file sync and share software, used by all partner to share, coordinate and collaboratively work on the project activities. The GRANULAR Nextcloud is a self-hosted private cloud solution based on IAMM's servers. Consideration is given to the availability and integrity of data, and to the appropriate levels of authorised access: IAMM remains the Nextcloud administrator and is solely responsible for creating new user accounts.

Access to data stored on the **website** is also restricted, with the same levels of authorised access: IAMM remains the website administrator along with AEIDL. The website backend and mailing list platform are password-protected, with two-factor authentication.

IAMM services are backed up daily (stats) and several times a day for project data (NextCloud and Digital Platform), all in line with IAMM's IT service continuity management policy.

6. Ethics

The ethical considerations for GRANULAR are set out in Article 14 of the Grant Agreement, detailed in the Deliverable D1.1 Ethics Guidelines prepared to manage ethical issues in the activities involving human participants foreseen by the project. D1.1 provides the regulations related to:

- the collection and handling of primary data, including guidelines for obtaining informed consent in accordance with the GDPR with regard to recruitment processes. A **template for obtaining informed consent** from participants in project activities, along with an accompanying **information sheet**, has been provided. Under the coordination of UNIPI, all Living Labs translated the English consent form and associated information sheet in their language (French, Italian, Polish, Swedish, Dutch and Spanish);
- the use and processing of secondary data;
- all relevant EU requirements and best practices in research procedures.

The responsibility for ensuring that the conduct of the research is in line with these principles rests with the relevant research teams. All data collection involving human participants will comply with the relevant project partner's ethics process.

If the partners have their own procedures for obtaining ethics approval, they must provide evidence of approval for individual activities to IAMM.

Appropriate safeguards for protecting personal data will include respecting the principle of **data minimisation, anonymising/pseudonymising data** where required, ensuring that researchers have received appropriate **training or guidance in data protection**, and conducting **data privacy risk assessments**. All applications to re-use data or information for a purpose other than that for which it was collected will be treated in an open, transparent, fair and non-discriminatory way.

Processing and handling of personal data is covered by Article 15 of the Grant Agreement, 15.1 for handling by the European Commission and 15.2 for handling by beneficiaries. No circumstances are foreseen that personal data will be transferred from project beneficiaries to the European Commission. If any such circumstance does arise then a service specific privacy statement (SSPS) will be downloaded from the European Commission site and the relevant requirements followed.

Annex 1: GRANULAR Data Delivery Guidelines

In line with Open Science concepts to guide the use of research data and its management during the project, an operational solution adapted to the project's specificities was developed to upload datasets in the GRANULAR Zenodo community as an **archive (.zip).**

1. A readme document at the root of the archive

The **readme** document is a technical report l**ocated at the root** of the archive. It aims at making understandable your data delivery and processing to external users.

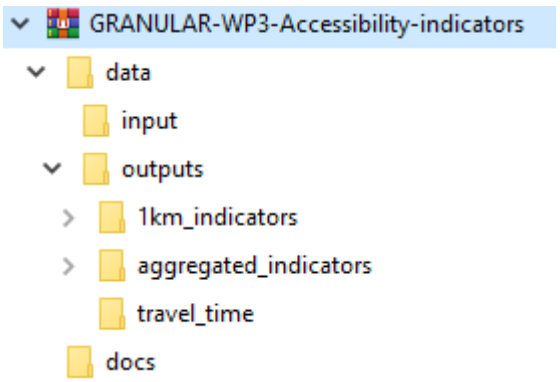
It explains the data processing: objectives, input data (what we have used for data processing), data processing and associated documentation, output data (proposal of metadata for each dataset).

This document can be a .pdf, using this template or a .html file (ouput of a notebook for instance). It is short a can refer to additional documentation (cf below).

2. How the archive is organised?

- Data and metadata are included the folder **data**.
  - Input data is included in a sub-folder **input**. Making available input data depends on the interest for external users (a specific interest to disseminate this data for other analytical purposes, making reproducible a workflow, etc.) and data specificities (the licence allows to disseminate the data, risk that the resource will not be available in the future, importance to make available the specific version of the dataset that has been used, etc.)
  - Outputs from the data processing has to be included in a sub-folder **ouputs**. It can be divided in several sub-folders if relevant. You are free to choose the format the most adapted to your data : it can be tabular files (.csv, .xls), spatial format (.tiff, .geojson, .gpkg, etc.). We only recommend to use as far as possible open format (.csv instead of .xls, .geojson instead of .shapefiles). Each dataset must be associated to an associated metadata file (cf below).

- **Docs** folder is optional and can be used to make available additional technical documents / notebooks explaining more in depth the data processing.



3. Metadata format

3.1 Dataset and Indicator

Metadata must be associated to each dataset / group of coherent dataset. It is an xls documents structured in two sheets. The content of metadata allows to external to users to understand the content of the data files. The name of the metadata file is “metadata_name of the dataset it describes.xls”.

Dataset sheet

Dataset information	Description
File Name	accessibility_EU, also available by country (accessibility_country folder)
File format	csv
Ressource name	Accessibility indicators
Provider name	UAR RIATE (CNRS - Université Paris Cité)
Spatial Extent	European Union + CH + LI + NO + UK)
Coordinate Reference System	Not adapted
Resolution or territorial level	EU reference 1km grid cells (version 1.3)
Copyright	© UAR RIATE © OpenStreetMap and contributors, 2024 © European Union, 1995 – 2024 © European Commission © REGIOgis
URL	
Use constraint	CC BY-SA 4.0
Downloaded or created	Created
Download or creation date	2024-05
Project	Horizon Europe GRANULAR
Metadata date (GRANULAR)	2024-06
Point of Contact (GRANULAR)	Ronan Ysebaert
Point of Contact (GRANULAR organizatio	UAR RIATE (CNRS - Université Paris Cité)
Point of Contact (GRANULAR email)	ronan.ysebaert@cnrs.fr
Abstract	Accessibility indicators based on OpenStreetMap network (OSRM routing engine, car profile) between 1km population grid and towns and city layer (European Commission). Include time-based indicators (travel time to the n nearest towns and cities) and cumulative opportunities (number of towns and cities or population reachable in x minutes).

Table 1: Dataset sheet, example derived from accessibility indicators

- **File Name:** Name of the data file. **[mandatory]**
- **File format** : File format of the dataset. **[mandatory]**
- **Ressource name :** a short name summarizing all the data delivery (all the datasets included in the .zip archive) **[mandatory]**
- **Provider name** : Name of the data provided (if input data) ort name of your organization (if output data). **[mandatory]**
- **Spatial Extent** : If spatial dataset bounding box (Xmin, Xmax, Ymin, Ymax). If tabular file a description of the spatial extent of your dataset. **[mandatory]**
- **Coordinate Reference System** : If spatial dataset the EPSG. *Example : EPSG:3035 (ETRS89-extended / LAEA Europe)* **[depends on the data file]**
- **Resolution or territorial level** : If raster data the granularity of the pixel. If vector data the name of the nomenclature and its version. *Example : NUTS2, version 2016* **[mandatory]**
- **How to cite**: Copyright of your organization and/or input data sources used for creating the indicators **[mandatory]**
- **URL**: Relevant for input data description. The URL to be considered to gather the input data you used. **[optional]**
- **Use constraint** : CC BY-SA 4.0 by default. Has to be checked carefully for input data. **[mandatory]**

- **Downloaded or created**: downloaded for input data, created for output data. **[mandatory]**
- **Download or creation date:** Date the dataset has been gathered / created (format : YYYY-MM) **[mandatory]**
- **Project :** Horizon Europe GRANULAR by default. **[mandatory]**
- **Metadata date (GRANULAR):** Date the metadata has been filled (format : YYYY-MM) **[mandatory]**
- **Point of contact (GRANULAR):** Surname and name of the point of contact. You can put several point of contacts if necessary. **[mandatory]**
- **Point of contact (GRANULAR organization)**: Name of your organization. **[mandatory]**
- **Point of contact (GRANULAR email)**: email of the point of contact. **[mandatory]**
- **Abstract:** Free text making possible to explain more in detail the content of the dataset. **[mandatory]**

Indicator sheet

Attribute	Description	Temporal validit	Datatype	Unit of measure	Comment
GRD_ID	The grid cell identifier according to INSPIRE specification	2023	String	/	
TIME_TOWN_5K_1	Travel time to the nearest little town and city (more than 5000 inh	2024	Double	Minutes by car	
TIME_TOWN_10K_1	Travel time to the nearest medium town and city (more than 1000	2024	Double	Minutes by car	
TIME_TOWN_50K_1	Travel time to the nearest city (more than 50000 inhabitants)	2024	Double	Minutes by car	
TIME_TOWN_5K_2	Travel time to the second nearest little town and city (more than 5	2024	Double	Minutes by car	
TIME_TOWN_10K_2	Travel time to the second nearest medium town and city (more th	2024	Double	Minutes by car	
TIME_TOWN_50K_2	Travel time to the second nearest city (more than 50000 inhabita	2024	Double	Minutes by car	
TIME_TOWN_5K_3	Travel time to the third nearest little town and city (more than 500	2024	Double	Minutes by car	
TIME_TOWN_10K_3	Travel time to the third nearest medium town and city (more than	2024	Double	Minutes by car	
TIME_TOWN_50K_3	Travel time to the third nearest city (more than 50000 inhabitants	2024	Double	Minutes by car	
N_TOWN_5K_15MN	Number of towns and cities (more than 5000 inhabitants) that it is	2024	Integer	Number of towns	and cities
N_TOWN_5K_30MN	Number of towns and cities (more than 5000 inhabitants) that it is	2024	Integer	Number of towns	and cities
N_TOWN_5K_60MN	Number of towns and cities (more than 5000 inhabitants) that it is	2024	Integer	Number of towns	and cities
N_TOWN_5K_120M	Number of towns and cities (more than 5000 inhabitants) that it is	2024	Integer	Number of towns	and cities
N_TOWN_10K_15M	Number of medium / large towns and cities (more than 10000 inh	2024	Integer	Number of towns	and cities
N_TOWN_10K_30M	Number of medium / large towns and cities (more than 10000 inh	2024	Integer	Number of towns	and cities
N_TOWN_10K_60M	Number of medium / large towns and cities (more than 10000 inh	2024	Integer	Number of towns	and cities
N_TOWN_10K_120M	Number of medium / large towns and cities (more than 10000 inh	2024	Integer	Number of towns	and cities
N_TOWN_50K_15M	Number of cities (more than 50000 inhabitants) that it is possible	2024	Integer	Number of towns	and cities
N_TOWN_50K_30M	Number of cities (more than 50000 inhabitants) that it is possible	2024	Integer	Number of towns	and cities
N_TOWN_50K_60M	Number of cities (more than 50000 inhabitants) that it is possible	2024	Integer	Number of towns	and cities
N_TOWN_50K_120M	Number of cities (more than 50000 inhabitants) that it is possible	2024	Integer	Number of towns	and cities

Table 2: Indicator sheet, example derived from accessibility indicators

- **Attribute**: Indicator code included in the dataset. **[mandatory]**
- **Description**: Label of the indicator. **[mandatory]**
- **Temporal validity :** Temporal reference of the indicator (Format : YYYY or YYYY-MM or YYYY-MM-DD). **[mandatory]**
- **Data type** : String, Double or Integer or Boolean (you can add another modality if relevant). **[mandatory]**
- **Unit of measure**: What is the measure unit of the indicator. **[if relevant]**
- **Comment:** If you want to add something to describe your indicator. **[optional]**

3.2 GRANULAR Indicators Library

To bring together results and facilitate the good flow of information among project partners, IAMM has set up an **Indicators Library**, which aims to (i) list all datasets generated during the project; and (ii) highlight the links between data/indicators (e.g. to identify synergies/redundancies between Work Packages).

This Indicators Library, in the form of an Excel Table, is shared at project level with all the partners and structured according to the minimal GRANULAR Metadata format (as detailed above), to which project-specific information is added. The responsibility for feeding the Indicators Library with the enriched Metadata is that of the Coordinator IAMM.

All the fields, as presented, below are divided into six categories and are mandatory.

Definition

- **ID: [internal use only – generated by IAMM]**
- **CODE: [internal use only – generated by IAMM]**
- **Title_short:** Short name of the dataset.
- **Title_full**: Long name of the dataset.
- **Subject:** Broad scientific field relevant to the dataset.
- **Description:** Summary describing the purpose, nature and the scope of the dataset.

GRANULAR-specific fields

- **Functionality:** Principal functionality of the Rural Diversity Compass associated with this dataset.
- **Component:** Principal component of the Rural Diversity Compass associated with this dataset.

Status

- **Access_rights:** Indicates use and access conditions to the dataset.
- **Licence:** Licence/Data Use Agreement
- **dmpID:** ID of the data management plan or of the data (PID/DOI…)
- **Date_start:** First date covered by the dataset.
- **Date_end:** Last date covered by the dataset.
- **Date_issued:** Date that the dataset was created.

Management

- **Contacts:** The entity, e.g. a person or organization, that users of the dataset can contact with questions.
- **Creator:** The entity, e.g. a person or organization, primarily responsible for making the resource.
- **Authors**: Entities (person or organization), that participated to the creation of the dataset.
- **Depositor**: The entity, e.g. a person or organization, that deposited the dataset in the repository.
- **Producer**: The entity (organization) that serves to produce the dataset.

Descriptors

- **Data_Format**: Describes the physical format of the data documented.
- **Data_Origin:** Primary origin of the data used (i.e. observational data, experimental data, survey data, analysis data, text corpus, simulation data, aggregate data, audiovisual corpus, computer code, other)
- **Data_Provenance:** Complementary information on origin of the data specific to GRANULAR methods (i.e. reused/external - OSM, reused/external - social media, reused/external - citizen, reused/external - heterogeneous, reused/external - public social and healthcare data, reused/external - government dataset, GRANULAR in-situ/remote-sensing observational data, GRANULAR experimental data GRANULAR survey data, GRANULAR analysis data, GRANULAR simulation data, GRANULAR aggregated data, other)

Spatial Descriptors

- **Spatial_Extent:** Geographic coverage of the data.
- **Spatial_Representation_Type:** Method used to represent geographic information in the dataset.
- **Spatial_Resolution**: Granularity of the dataset: NUTS1/2/3, LAU, gridded (resolution in m or km)
- **EPSG_Code**: Unique codes to reference to geographic objects like ellipsoids, datums, spheroids, and measurement units
- **Spatial_Keyword:** Specific keywords for spatial datasets

4. Uploading your data file in GRANULAR Zenodo repository

Do not hesitate to add additional metadata for a better visibility of your dataset on the Web.

Annex 2: Overview of primary research data or database outputs planned in GRANULAR Living Labs

Living Lab	Purpose of data	Data description	Secondary data	Primary data	Data reuse outside the project
France – Pays Pyrénées Méditerranée (PPM)	**Mobility.** What are the main mobility flows and demand all along the year, both during "everyday" periods and in the summer months? Specify the spatial distribution of travel flows and their volume for all reasons for travel, to anticipate and size alternative transport offers more accurately.	Content: location of the trip-generating centre/mobility flows origin & destination, number of travellers throughout the day/year, means of transport and factors of choice. Data type: commuting datasets, data derived from GPS geolocation of GSM phones, interviews Format: .xls, PDF, .txt	•Annual Home: work daily travels dataset [INSEE]: www.insee.fr/fr/information/2383337 •Car and bicycle counters on main local roads [CD66]: www.ledepartement66.fr/; www.data.gouv.fr/fr/datasets/ •Trains and buses attendance dataset [Région OPM] •Diagnosis of the cycling master plan [Pays PM] All this data is publicly available with an open-source licence. No private sector data are used. •Geolocation of mobile phones (data sets acquired by the communities of communes over 2024 from Orange Flux Vision)	Carrying out a quantitative survey of tourists staying in Argelès-sur-mer during the summer of 2025. Gathering information on journeys made (origin/destination, mode of transport) and opinions on travel solutions. Protocol and quality control are to be defined with scientifi and local partners. It is also proposed to combine the tracking of people's movements using trackers throughout the experiment (using the TraceMob application) and several interviews with these people to understand their choices in terms of mobility. Particular attention will be paid to RGPD requirements when collecting this voluntary data. This data collection method is qualitative and the number of people to be involved in the process has yet to be determined.	The French Living Lab is not expected to publish aggregated databases, resulting from the combination of various official statistical databases or from monitoring and reflection activities.
	Uses of natural areas. When and where do visitors and local people go to natural areas? Locate and characterise natural areas where the use of these areas by outdoor enthusiasts has an impact on biodiversity.	Content: characteristics of the movements of outdoor practicioners who use apps and volunteer to share their tracks, fauna & flora inventories Data type: GPS data (Outdoor vision and Strava), counters, maps and fauna & flora inventories	•Several visitors' counters (implemented since 1990 for some of them); •Several one-shot observation campaigns implemented by natural reserves; •Observation means' coordination for summer 2023: visitors' counters, satellite data and phones tracking [Syndicat Mixte du Canigou]; •National OutDoorVision project, collecting GPS data from sport applications	/	
	Impacts drought on biodiversity. Focus on the chestnut tree to better understand the effects of climate change on the biodiversity supported by the local chestnut grove, to anticipate and adapt forest management, and to better prevent fires.	Content: Ground level Image and satellite image datasets from tools available to assess forest dieback, particularly in the Vallespir chestnut grove. The format and type will be precised in the coming year.	Diagnosis of existing tools and associated datasets to assess forest dieback, particularly in the Vallespir chestnut grove	The work will include interviews with local actors involved in the Territorial Forest Charter. The results will also be compared with the reality on the ground where data exists (monitored plots), or even by carrying out a targeted dieback diagnosis exercise.	

Italy – Distretto Rurale Val di Cecina (VdC)	**Agri-food supply chains.** Assess the need for construction of infrastructure for the logistics of agri-food products	Content No.1: Interview promoted by UNIPI **Qualitative interviews** on members perceptions related to the District governance and processes Data type: audiorecordings and textual transcriptions Format: .docx, .mp4, .m4a Expected size: small size (5 GB) Content No.2: **Agricultural products** - data on production, processing and waste Data on applied business models. Data type: digital data, text	**Data from Tuscany Region:** annual reports from Rural Districts Origin: downloaded PDF Provenance: official website of Tuscany Region Pre-existing rights of use: no restriction	**Interviews** Context: Online or in-person interviews to members of the Living Lab and related actors through the snowballing procedure. Collection methods: semi-structured interviews. Protocol: the interviews are personal but for the project are made anonymous and analysed through the coding method. Quality control: triangulation and workshop organization with the Living Labs participants **Questionnaires** Context: online administration to members of the Living Lab Collection methods: open-ended and multiple-choice questions Protocol: anonymous administration	**Interviews** Data utility: practitioners, researchers and policy makers Restrictions: data are not sensitive
	Water resources. Improve the management of water resources for environmental objective	Content: **Quantitative data on water consumption** (at farmer level) and soil moisture retention level (irrigation needs). Data type: digital data, text	**Data from public agencies** Pre-existing rights of use: no restriction	**Questionnaire** Context: online questionnaire to members of the Living Lab about water needs and use destination (irrigation, farmhouses, livestock) Collection methods: open-ended and multiple choice questions, online administration Protocol: anonymous administration	*/*
	Water resources. Support the implementation of integrated territorial planning to cope with climate change especially concerning the availability of water.	Content No.1: Interview promoted by UNIPI **Qualitative interviews** on farmers perceptions [water, farm practices, climate change etc] Data type: audio recordings and textual transcriptions Format: .docx, .mp4, .m4a Expected size: small size (5 GB) Content No.2: **Questionnaires** promoted by the District	*/*	**Interviews** Context: Online or in-person interviews to members of the Living Lab and related actors through the snowballing procedure. Collection methods: semi structured interviews. Protocol: the interviews are personal but for the project are made anonymous and analysed through the coding method. Quality control: triangulation and workshop organization with the Living Labs participants **Questionnaires** Context: online administration to members of the Living Lab about their water needs and use destination (livestock, irrigation, farmhouses, etc.) Collection methods: open-ended and multiple choice questions, online administration Protocol: anonymous administration	**Interviews** Data utility: practitioners, researchers and policy makers Restrictions: data are not sensitive

	Tourism. Support the implementation of sustainable management of the touristic flows	Content No.1: **Tourism activities** developed from the farmhouses Content No.2: **Data on tourist flows**, especially linked to the accommodation capacity of agricultural companies Data type: digital data, text Format: .xlsx Expected size: small size	**Data from Municipalities** Origin: data sharing/exchange (tourism tax) Pre-existing rights of use: no restrictions **Data from public agencies** Origin: data sharing/exchange (Ambiti Turistici) Pre-existing rights of use: no restrictions	**Questionnaire** Context: online questionnaire to members of the Living Lab about water needs and use destination (irrigation, farmhouses, livestock) Collection methods: open-ended and multiple choice questions, online administration Protocol: anonymous administration	/
Spain – Ourense (RIO)	**Demographic change (hyperlongevos).** General characteristics of this sector of the population, their needs, and the situation of the services associated with them	Data type: Cartographic materials with distribution of hyperlongevos people about services offered. Combine statistical data related to the socio-economic and territorial data of the province with information on the location of hyper-aged people. Format: the format will be decided during the course of the project	Qualitative information collected from **unpublished semi-structured interviews** with people over 85 and 100 years old Origin: a private company has been interviewing respondents for years in our territory. We would like to scientifically analyse all these data and compare them with our own interviews. This company has signed the data transfer documents to do these interviews. Format and size: Excel format and approximately 15000 words analysed and ranked.	Qualitative data from **interviews or focus groups**. Methodology: word cloud methodology used to validate respondents' own answers. Qualitative information from third parties is obtained with the consent of our LL members in compliance with the data requirements of the programme as well as national regulations. Both the implementing entity and the participants are informed in advance of the ethical and data standards. The specific data security process for the Task Leader UVIGO is detailed below[1]. All information received is previously anonymised.	Data will be made available inside the GRANULAR consortium - only to authorized users, mainly to data uploaders, the Facilitator and the person in charge of each Living Lab in the project, through GRANULAR Cloud-based storage and collaboration service and the UVIGO institutional repositories (certified). A replication is proposed to all interested partners and territories. The specific data publication process for the Task Leader UVIGO is detailed below[1]. The data will be useful especially for decision makers and researchers who can make decisions on new rural policies. Ownership rights will be stipulated by the results publication system.
	E-governance (hyperlongevos). Data analysis for an e-governance policy for older people in rural areas	Reused data: data derived from the elderly-oriented telecare programme (the most extensive digital process in the province). Generated data: territorial analysis of users + pilot of the emergencies of elderly people using the telecare system Format: Data in Shape and other formats for data analysis in Geographic Information Systems programmes. Also, data in excel format and graphics	**Quantitative data derived from the telecare programmes** aimed at elderly people in rural areas Quantitative data are anonymised. Each person's name and any identifying information (including overly precise geolocation) is replaced or removed and replaced by a code. **Data derived from qualitative interviews** with pre-established usage rights.	/	
	Climate change exposure (Mega fire). Analyse the relationships between forest fires and the other priorities of the Living Lab (ageing and e-governance)	Content: Interviews and qualitative analyses targeting pupils and older people to relate environmental (fires) and social problems such as education, ageing. Also to analyse e-governance policies linking these three concepts Format: .doc format and texts with specific analyses	/	Data collected during **focus groups** (10 carried out throughout the territory in 2024), The same protocol (as described above) applies.	

Poland – West Pomerania (KUT)	**Agri-food supply chains.** Analyse the activities of food producers, and in particular identify what strengthens the short supply chains and organic food production and sale	Content: Data on local processing infrastructure ; Data on resources of food producers ; Data on forms of sale of manufactured products ; Data on activities of pro-ecological food producers ; Data on cooperation and competition; Data on online sales Data type: Text data, numerical data, questionnaires, survey results Format (e.g ".txt file"): PDF, MS Word (.doc/.docx), MS Excel (.xls/.xlsx)/XML marked-up text (.xml) Expected size: Estimated data volume 1TB	**Data from public collections and public statistics** in open access. Secondary data are in an aggregated from, which does not allow for the identification or individual participants.	Context: Collecting primary data to identify the factors that strengthen short supply chains and support food production and sales for a better understanding of production processes, distribution methods and market strategies. Collection methods: Data obtained by direct survey using a survey questionnaire (participation in the survey voluntary and anonymous). The questionnaire contains open-ended and multiple-choice questions. Some questions use a Likert scale. A survey questionnaire addressed to farmers and food producers in rural areas. Protocol: anonymous administration Quality control: The source data and documents will be organized by the Koszalin University of Technology (KUT) according to the principles detailed below[2].	All finalized, quality-controlled and non-confidential data sets will be made available for review within the GRANULAR consortium. The raw data will not be published. Research outputs will be published in scientific articles with open access, and presented at scientific conferences and regional agricultural organisations. The specific data publication process for KUT is detailed below[2]. The collected data on the activities of food producers, short supply chains and food sales can be valuable to: food producers (farmers, processors, food entrepreneurs), consumers and consumer organisations, government institutions and agricultural and non-governmental organisations (NGOs), researchers.
The Netherlands – P10	**Accessibility.** Gain a understanding of the accessibility of basic services in the most typical Dutch rural municipalities.	Content: (i) Secondary data on the accessibility of basic services at national, regional, municipal neighbourhood and district level, with specific attention to P10 municipalities Berkelland, Hof van Twente and Bronckhorst (ii) Data to be collected: Qualitative data - interviews on perceptions about the accessibility and availability of essential services	**Data from Municipalities and public agencies** Quantitative data, Literature analysis: report on the accessibility of basic services in the typical Dutch rural municipalities	Collection methods: interviews with individuals on perceptions about the accessibility and availability of essential services. Data collected through one national P10 workshop with more than 30 members of the P10. Additional data collection through interviews with civil servants of three P10 municipalities and also village representatives of these municipalities. The interviews were personal but for the project it has been made general / anonymous.	At this stage, no information has been shared or published.
United Kingdom – Rural Scotland (HUT)	**Rural Proofing.** Contribute to the testing and operation of a process of rural proofing of policies, public and private, as they affect rural areas	The datasets most likely to be re-used are for the provision of: i) context (e.g. mapping of characteristics of rural areas) in reports, workshops and analysis; ii) deriving indicators (e.g. for characterising rural Scotland);	•Datasets from the national mapping agency (Ordnance Survey) (e.g. biophysical data such as water bodies); •Datasets from public agencies and research organisations (e.g. natural heritage designations; soils, Land Capability for Agriculture;	/	The specific data publication and quality assurance process for the Task and Work Package Leader James Hutton Institute (HUT) is detailed below[3].

		iii) analysis of characteristics of rural Scotland (e.g. under specific policies or scenarios of alternative futures). Types and Formats: Spatial datasets: Vector and polygon data, in formats of ESRI Shapefiles or geodatabases; Derived raster data at various spatial resolutions (ESRI GRID format, ERDAS IMAGINE format). Timeseries: In-field observations of characteristics of soils (e.g. greenhouse gas emissions, soil temperature, soil moisture) in bespoke formats (e.g. MS Excel, MS Access).	•Datasets from NGOs and voluntary bodies (e.g. National Biodiversity Atlas, including data contributed by citizens; •Classifications of rural areas such as that of the Sparsely Populated Areas of Scotland; Rural socio-economic performance); •Datasets with indicators (e.g. Scottish Index of Multiple Deprivation, population and demographic change; •G2 experimental datasets of in-field observations (e.g. GHG emissions). The format and type will be detailed in the coming year.		Almost all the datasets under consideration are used by policy advisors, relevant practitioners and researchers. Few of these datasets are used by members of the public, except those involving citizen science.
Sweden – Regions of North Sweden (SWE)	**Green Transition**. The objective is to gain a better understanding of rural areas' diversity, i.e., how the green transition impacts rural areas, and how to create a policy that can tackle the combination of depopulation, significant green investments, and regional hubs of growth within the region	Content #1: Data on transformation readiness i.e., the capacity for change. Data will be reused to provide another level of socioeconomic analysis that understands sparsely populated rural areas. The Swedish Living Lab needs better tools to measure innovativeness in rural areas. Content #2: Data on basic services availability [travel time, accessibility and robustness] and related costs to investigate the accessibility of public services within regions	**Data from local and regional administration and other associations** Analysis of the development traps presented in the 9th Cohesion Report at the local level	The Living Lab will organise stakeholder dialogues, interviews and surveys to collect better data at NUTS3 and local levels to measure rural development. The data collection activities will target key actors in the municipalities related to resilience to the green transformation.	At this stage, no information has been shared or published.

[1] *UVIGO's specific data security, publication and quality process*:

- For personal /sensitive data fully informed consent will be given for collecting, processing and storing data. UVIGO will use GRANULAR Cloud-based storage and UVIGO institutional repositories (certified/Trustworthy Repository). Anonymised data will be separated from informed consent forms, by keeping the forms in a locked cabinet in the institutional repository. In cases of public emergency, if requested by the legal entities, immediate open access will be provided to the legal entities that need the data to address the public emergency.
- Provenance will be mainly recorded in a single README text file that describes the data collection and processing methods, but it might also involve a visualisation of the machine-readable representation such as VisTrails. The data will be checked automatically during the upload process and manually by UVIGO staff and the uploader. Also by cross-referencing relevant information found in different databases, and in general making sure there are no inconsistencies with the information found within a database or a set of databases.
- All validated data will be publicly available and deposited in Zenodo. Data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement. However, some data embargos are also possible (access to the data is delayed to allow us to complete our analyses or publications); if needed, a grace period of up to three years can be requested for particular data to allow researchers to protect our intellectual property rights. However, assay description and metadata will be accessible by default. There will be access restrictions: we can request the access to the data are restricted because of their sensitive nature. This means people would need to request access and that we can decide whether, and how, the data will be accessed.

[2] *KUT's specific quality, data security and publication process*:

- The source data and documents will be organized according to the following principles. First, ensuring high data quality: a) clear and precise wording of questions and use of appropriate question types and response scales in the survey questionnaire. Additionally, conducting pilot tests that will allow detecting any problems related to the questionnaire before starting the actual study; b) validation of responses after data collection, before analysis, which allows for eliminating false data and increasing the quality of the study results; c) application of the Single IP Access method in the case of the CAWI study, which consists in limiting access to the survey from one IP address. The person responsible for substantive verification is the person managing the project together with the research team. Second, data organization. Each data set will be provided with a metric defining the currency of the data, the last change, the source of the data, etc. The person managing the project together with the research team is responsible for data quality control. The data will be organized in appropriate folders assigned to the individual stages of the study. The activities related to updating and processing of research data are subject to control. It is carried out in order to ensure the continuity of the organization and description of data.
- Dissemination of research results will not violate property rights.
- Time embargo: data sets will be embargoed until a scientific publication is the result of the work of the entire project partner team. It is currently not possible to precisely estimate the embargo period. Factors that could affect such an embargo period are: time of data collection, time of data processing, time of manuscript preparation, time of manuscript publication and any other publications that are the result of the work of the entire project partner team that will be associated with this data. The project manager is responsible for managing data and resources in the project.
- The Polish Living Lab does not plan to collect personal data. Personal data (if collected) will be collected in compliance with the principles of: legality, purpose limitation, data minimization, accuracy, processing limitation, integrity and confidentiality. Participants in the study will be asked to complete a consent form and consent to the processing of personal data. Data will be collected in a manner that is adequate to the purpose, while maintaining its confidentiality and integrity, and will be stored in a form that allows identification of the data subject for no longer than is necessary for the purposes of the study. Only the minimum number of people necessary for conducting scientific research, who have appropriate authorizations to process personal data, will have access to personal data.

[3] *HUT's specific data publication and quality process*:

- It is expected that the final versions of each derived dataset will be shared. This will depend upon the underlying data and any issues of confidentiality or licencing arise. As each dataset is derived an assessment will be carried out of any constraints on the sharing of the data.
- Depending upon the intellectual capital datasets may be embargoed until academic publication. At present it is not possible to provide an accurate assessment of an embargo time. Factors that would influence such an embargo time are the time to prepare a manuscript, time to publication of a manuscript, and any subsequent publication that would be associated with these data.
- Restrictions due to confidentiality, intellectual property rights or limited accuracy would lead to access being restricted. During the project, in each case the reason would be assessed against the requirements of the Grant Agreement and obligations to the funders.
- At the end of the project that status of each dataset will be assessed and a decision made on any restriction, or withdrawl of a dataset. The expected status is that all finalised, quality controlled and non-confidential datasets will be made available to access.

Annex 3: GRANULAR Data Management responsibilities and quality process

This document sets out all the steps and processes for ensuring data quality management. The procedures described here apply not only to deliverables, datasets, databases and reports, but also to all project results, and therefore to all consortium partners.

The following sections are divided according to the different roles in the production of outputs, starting with the partners who are producing outputs (referred to hereinafter individually as the "lead researcher"), then at the level of the Coordinator, IAMM.

A – Relevant Partner (Lead Researcher /Author)

1. Naming and versioning

Lead authors working in relevant Work Packages determine the significance of the changes made in the report/dataset, and then allocate major and minor version numbering.

- Project Deliverable: the naming convention for final copies of project Deliverables is: GRANULAR D<number><Deliverable title>

- Datasets to be stored in trusted repositories: the naming convention will follow the same approach, with an identifier representing a sequence of such data coming from GRANULAR and a title reflecting the dataset content: GRANULAR Dataset<number><Dataset title>

A multi-level version management labelling will be used for reports, datasets and other project outputs. The labelling uses a syntax comprising major and minor version properties. *e.g. 3.1 where first digit (3) represents the major version, and the second digit (1) represents the minor version.*

The components are as follows:

1. new documents or datasets with no existing major version, numbered 0.0.
2. major versions with significant changes in status, numbered 1.0, 2.0, 3.0, …
3. minor versions based working edits and changes, numbered 1.1, 1.2, 1.3, …

A unique version number is assigned to each version of a document whether or not the content is a major or a minor version. A final check will be performed by IAMM as Coordinator, responsible for renaming the final versions of all deliverables and reports.

2. Data storage

Each project partner is responsible for:

- the storage of its own datasets, collected or produced;
- following best practices in **cybersecurity** and **preventing unauthorised access to data**
- managing intermediate versions of data: please refer to the organisation's storage and back-up arrangements, complying with good practice on back-up and deletion of data.

Concurrently, all final outputs (datasets, databases, reports) will be stored in the GRANULAR project data archive stored by IAMM, which complies with the EU requirements.

3. Quality monitoring

Quality assurance formalising all quality methods and processes at project-level has been put in place and integrated into Deliverable D1.2 Quality assessment guidelines. In particular, quality assurance foresees a two-step process, for project outcomes where scientific outcomes of the project are presented – i.e., deliverables and milestones: (i) the WP leaders are in charge of approving the deliverables and milestones within their WP; then (ii) IAMM as Coordinator validates and approves definitively the deliverables and milestones and submits all project deliverables to the European Commission.

Responsibilities for quality of data at the level of its capture and processing (plus the associated documentation) is that of the lead partners.

⚠ **Focus for the research teams dealing with data relating to individuals** (e.g. survey respondents, interviewees, etc.) have to be managed with adherence to the Deliverable D1.1 **Ethics Guidelines.**

4. Metadata format, keywords and data documentation

The choice of detailed metadata entries will be the responsibility of the relevant lead researchers who are producing Deliverable, or creating datasets to be made publicly available in accordance with the GRANULAR Data Delivery Guidelines (see Annex 1 of the Data Management Plan).

All GRANULAR project datasets will be stored in accordance with a uniform folder structure.

In order to achieve this objective, all partners who are producing datasets shall use the template “archive (.zip)”, which includes:

- **A “readme” document, to be written and added** at the root of the archive. It is a technical report located, aiming at making understandable your data delivery and processing to external users. This document can be a .pdf or a .html file.
- A folder “DATA” with two subfolders:
  - Input - including for each dataset the corresponding metadata file
  - Outputs - including for each dataset the corresponding metadata file
- A “Docs” folder (optional) and can be used to make available additional technical documents or notebooks explaining more in depth the data processing

⚠ Even if some data have dissemination restrictions and cannot be added to the deposited folder, the partner must add the corresponding metadata file to the folder, indicating the contact details of the person and institution responsible for the data, so that a hypothetical future user of these data can request them.

The responsibility for selecting the most appropriate keywords and terms will lie with the authors of each report and associated datasets. **Please refer to the US Library of Congress Linked Data Service to identify and select keywords.**

5. Licence choice

Regarding datasets, licences will be permissive (as open as possible), consistent with restrictions on data release and use (as closed as necessary), such as GPLv3 (for software) or CC-BY-4.0 (for data and publications).

For data for which dissemination restrictions or constraints are needed, the contact of the responsible person and institution of the data will be given in order to allow a hypothetical future user of these data to request them. This request will then be submitted to the responsible parties who will decide if they can share data or not and the protocol and connection details specially tailored to answer to the request.

6. Repository choice

The choice of relevant thematic repositories is the responsibility of each relevant partner.

Domain-specific Repositories linked to thematic metadata models and standards (e.g. list on Digital Curation Centre website) are strongly preferred for storing and to obtain permanent identifiers such as DOI. If no thematic repository is chosen, the default repository for datasets is Zenodo repository - GRANULAR community.

Verification that quality control procedures have been completed (final approval by IAMM, regarding metadata) will be ensured by each **research team quality contact**.

- *If a domain specific repository is chosen:* The relevant partner must inform the IAMM team of the deposit, and send them the Digital Object Identifier (D.O.I.; www.doi.org/) and Digital platform metadata file.

- *If the Zenodo repository is chosen:* Zenodo requires limited metadata as a mandatory requirement for storage of the data; this leads to a lack of metadata quality control and to a potential lack of reliable metadata. Please note that an automatic revision request will be sent (by Zenodo) to IAMM, who will make a final check and accept the publication.

B – Coordinator (IAMM)

1. Naming and versioning

FOR DELIVERABLES, REPORTS AND SCIENTIFIC ARTICLES
When received by the partner, WP1 (IAMM) will check the naming (and versioning) of documents to ensure consistency, spelling and formatting as part of the quality control process. **IAMM designates the final version** by assigning a major version number to a document, and renaming the final version for uploading and public dissemination.

2. Quality control of deliverables and associated metadata

As detailed in Deliverable D1.2 Quality assessment guidelines, IAMM has the overall responsibility for Quality Control of Deliverables, and of Metadata created.

3. Data storage

All final outputs (datasets, databases, reports) will be stored in the GRANULAR project data archive stored by IAMM, which complies with the EU requirements.

4. Deposit on repository

FOR DELIVERABLES, REPORTS AND SCIENTIFIC ARTICLES
Depositing those documents in thematic repositories proposed and approved by the relevant partner or in Zenodo is the responsibility of IAMM.

5. Visualisation on the Digital Platform

Only aggregated results will be visualised within the platform; with links to reliable repositories (with DOIs) for accessing full datasets.

Once a deposit is completed, IAMM will collect the Digital Object Identifier (DOI) and the metadata file sent by the relevant partner, and complete the GRANULAR Indicators Library accordingly.