Mapineq Link: Geospatial Dashboard and Database

Melinda Mills
University of Oxford / University of Groningen

Douglas Leasure
University of Oxford

Mapineq deliverable D1.5

September 2024

Funded by the European Union

Suggested citation:

Mills, M.C. & Leasure, D. (2024). Mapineq Link: Geospatial Dashboard and Database. Turku: INVEST Research Flagship Centre/University of Turku. DOI: 10.5281/zenodo.13864000

Summary history		
Version	Date	Comments
1.0	30.09.2024	Finalised Author Draft
2.0	07.10.2024	Reviewed manuscript for Zenodo

Mapineq – Mapping inequalities through the life course– is a three-year project (2022-2025) that studies the trends and drivers of intergenerational, educational, labour market, and health inequalities over the life course during the last decades. The research is run by a consortium of eight partners: University of Turku, University of Groningen, National Distance Education University, WZB Berlin Social Science Center, Stockholm University, Tallinn University, Max Planck Gesellschaft (Population Europe), and University of Oxford

Website: www.mapineq.eu

The Mapineq project has received funding from the European Union's Horizon Europe research and innovation programme under the grant agreement No. 101061645.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, the European Research Executive Agency, or their affiliated institutions. Neither the European Union nor the granting authority can be held responsible for them.

Acknowledgement:

This document was reviewed by Jani Erola and Mijail Figueroa González (University of Turku) as part of Mapineq quality assurance procedures. The content of the document, including opinions expressed and any remaining errors, is the responsibility of the authors.

Publication information:

This work is licensed under the Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International (CC BY-NC-SA 4.0) license. You are free to share and adapt the material if you include proper attribution (see suggested citation), indicate if changes were made, and do not use or adapt the material in any way that suggests the licensor endorses you or your use. You may not use the material for commercial purposes.

Executive summary

The Mapineq project focuses on the trends and drivers of intergenerational, educational, labour market and health inequalities across the life course. A central aim of economic, social and health policy interventions is to diminish inequalities. To do so, we require an understanding of how inequalities may accumulate and how they are influenced by geographical context across multiple domains.

Mapineq Link is an online interactive geospatial dashboard and database that improves existing resources by offering a **unique and rich multiverse of data for multiple uses**. It will allow policy-makers and journalists to visually explore regional inequalities and case studies to develop new policies and investigative insights. It offers researchers an interactive data dashboard to form hypotheses and visualise data and invites data scientists and programmers to unlock tools for further research via API access to our databases.

Mapineq Link contains **diverse information that is geospatially patterned**. Indicators have been initially gathered under four modules: (1) institutional policies (e.g., education, labour market, housing, immigration policies), (2) environmental (e.g., pollution, temperature, green space, infrastructure, road system), (3) real-world data (e.g., commercial house price listings, google searches, social media aggregated user data, energy use), (4) demographic, socioeconomic, health and lifestyle (e.g., births, marriages, education type/levels, labour market activity, migration, smoking, death rates, population density).

Mapineq Link **has an API** (Application Programming Interface) which allows: (1) researchers to query the database, (2) programmers and data scientists to retrieve data for specific geolocations; (3) external web developers to easily integrate MapIneq data into their own interactive web applications; and (4) other users (e.g. the public) to view data and visualizations of a wide range of indicators. The MapIneq Link team is **dedicated to FAIR principles** to make our innovations Findable, Accessible, Interoperable, and Reusable.

Mapineq Link **improves existing resources** by: (1) harmonising and blending geospatial data from multiple sources, (2) including fine-grained geospatial indicators at multiple levels, (3) expanding coverage to a multiverse of data types, (3) developing an API tool to allow data scientists to access geo-coded data to link to individual data and integrate into their own web applications.

Funded by
the European Union

Abbreviations

API	Application Programming Interface
Eurostat	Statistical Office of the European Union
FAIR	Findable, Accessible, Interoperable and Reusable
GADM	Globally-harmonised administrative boundaries
LAU	Local Administrative Units
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organisation for Economic Cooperation and Development
UN	United Nations

Funded by
the European Union

Content

Figures

Funded by
the European Union

Mapineq Link: Geospatial Dashboard and Database

Mapineq Link is an online interactive geospatial dashboard and database that improves existing resources by offering a unique and rich multiverse of data for multiple uses. It will allow policy-makers and journalists to visually explore regional inequalities and case studies to develop new policies and investigative insights. It offers researchers an interactive data dashboard to form hypotheses and visualise data and invites data scientists and programmers to unlock tools for further research via API (Application Programming Interface) access to our databases.

Mapineq Link is a rich **open-access online interactive dashboard and database** that contains diverse information that is geospatially patterned. Indicators have been initially gathered under four modules: (1) institutional policies, (2) environmental indicators, (3) real-world data, (4) demographic, socioeconomic, health and lifestyle.

Mapineq Link **has both an online interactive dashboard and an API** which allows: (1) researchers to query the database, (2) programmers and data scientists to retrieve data for specific geolocations or bespoke regions; (3) external web developers to easily integrate MapIneq data into their own interactive web applications; and (4) other users (e.g. the public) to view data and visualizations of a wide range of indicators. The MapIneq Link team is **dedicated to FAIR principles.**

Mapineq Link **improves existing resources** by: (1) harmonising and blending geospatial data from multiple sources, (2) including fine-grained geospatial indicators at multiple levels, (3) expanding coverage of to a multiverse of data types, (3) developing an API tool to allow data scientists to access geo-coded data to link to individual data and integrate into their own web applications.

Funded by
the European Union

1. Introduction

The Mapineq project focuses on the trends and drivers of intergenerational, educational, labour market and health inequalities across the life course, a core area of social, economic and policy research. Life course domains include those spanning early childhood, parenting, childcare and family, transition to adulthood, education, labour market, occupational mobility to retirement. These life course events and domains are, however, experienced differently by individuals depending on the context they live or grew up in, which motivates the exploration of life course differences from a spatial perspective. While differences in experiences across the life course can relate to variation in individual and family traits, they are also shaped by local, regional, national and supranational opportunity structures and influenced by local spatial as well as cross-regional and cross-national variations in resources, institutions and norms across. The emergence and shaping of life course inequalities between individuals and groups will always be contingent on social and physical context, which motivates our pursuit of a rich spatial database on life course inequalities.

Mapineq Link is a rich open-access online interactive dashboard and database that contains **diverse information that is geospatially patterned**. Indicators have been initially gathered under four modules: (1) institutional policies (e.g., education, labour market, housing, immigration policies), (2) environmental (e.g., pollution, temperature, green space, infrastructure, road system), (3) real-world data (e.g., commercial house price listings, google searches, social media aggregated user data, energy use), (4) demographic, socioeconomic, health and lifestyle (e.g., births, marriages, education type/levels, labour market activity, migration, smoking, death rates, population density).

Mapineq Link **has both an online interactive dashboard and an API** which allows: (1) researchers to query the database, (2) programmers and data scientists to retrieve data for specific geolocations or bespoke regions; (3) external web developers to easily integrate MapIneq data into their own interactive web applications; and (4) other users (e.g. the public) to view data and visualizations of a wide range of indicators. The Mapineq Link team is **dedicated to FAIR principles** to make our innovations Findable, Accessible, Interoperable, and Reusable.

Mapineq Link is an online interactive geospatial dashboard and database that improves existing resources by offering a **unique and rich multiverse of data for multiple uses**. It will allow policy-makers and journalists to visually explore regional inequalities and case studies to develop new policies and investigative insights. It offers researchers an interactive data dashboard to form hypotheses and visualise data and invites data scientists and programmers to unlock tools for further research via API access to our databases.

This report is the fifth and final report outlining the development of the Mapineq Link dashboard and database. It builds upon the first four reports, which described the different substantive modules, which were the institutional and structural policies (Leasure et al. 2023), environmental measures (Verkroost et al. 2023), real-world data (Mills 2024) and demographic, socioeconomic, health and lifestyle indicators (Scotti-Bentivoglio et al.

2024). The current report is written to read as a standalone document and for that reason does have some repetition from previous reports such as the description of the database modules, geographical units and database types. In this report we first explain why we developed Mapineq Link, highlighting the innovation features and added value. We then briefly review the previous four aforementioned substantive topic models developed throughout the project. This is followed by a review of geographic units and geospatial data types. We then provide a description and visual overview of the Alpha version of the Mapineq Dashboard and Database. We conclude with a description of our Launch event on October 7, 2024 and conclude with future directions.

1.1. Why Mapineq Link?

Although many databases exist, they are often in siloes of specific areas such as social policy or pollution or lack a geospatial component. Mapineq Link improves existing resources by:

(1) **harmonising and blending geospatial data from multiple sources**, moving beyond fragmented databases which focus on measurement of one indicator such as pollution, housing, demographics or social policy alone;

(2) **including fine-grained geospatial indicators** at multiple levels, depending on the indicator from very small local areas, neighbourhoods, regions to countries,

(3) **expanding coverage of to a multiverse of data types,** including not only social policy and institutional data to environment and built landscape, experimental real world data and classic demographic, socioeconomic, health and lifestyle indicators;

(4) **developing an API (Application Programming Interface) tool to allow data scientists to access geo-coded data to link to individual data and integrate into their own web applications**. Users who have access to geolocated individual-level data can use locations (e.g., latitude and longitude coordinates), to retrieve spatially-linked data across multiple data sources and, external web developers to integrate MapIneq Link data to their own web applications.

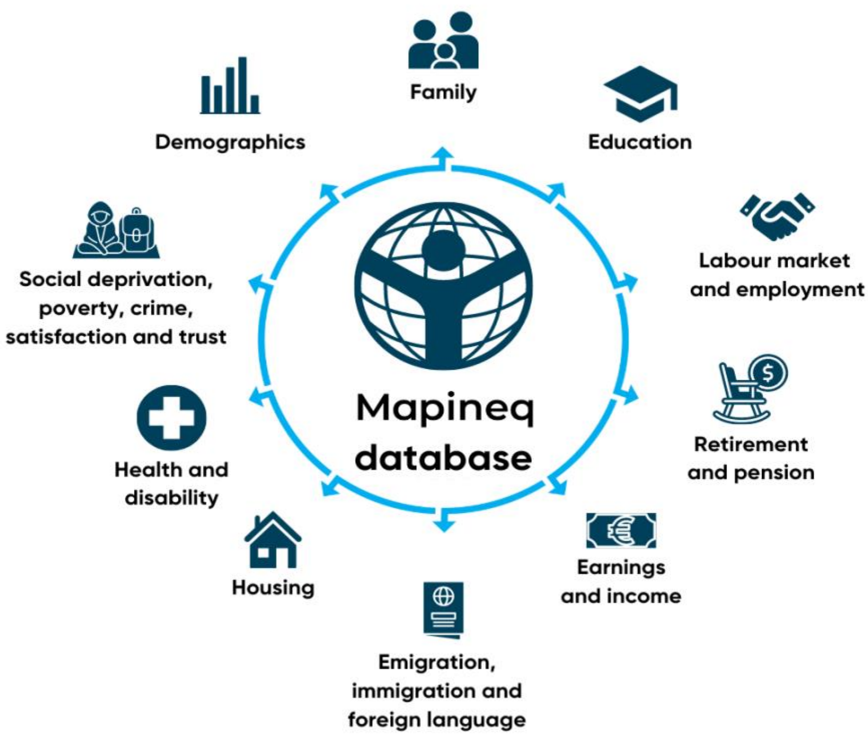
In the first instance our database will focus on the **European context and associated countries**. As data becomes available and the database evolves over time, we may expand to other regions of the world.

2. Mapineq Link Database modules

2.1. Socio-economic policy indicators and institutional measures

Phase 1 of MapIneq Link contains **ten thematic modules** that can be used for standalone analyses or to linked to individual-level data, with data covering structural and policy indicators for: (1) family, (2) education, (3) labour market and employment, (4) retirement and pension, (5) earnings and income, (6) migration and language, (7) housing, (8) health, (9) deprivation, poverty, crime; and, (10) demographics. A conceptual overview is provided in Figure 1, with more detailed information about the specific measures described in Leasure et al. (2023).

Figure 1. An overview of structural and institutional policy measures in Mapineq Link.



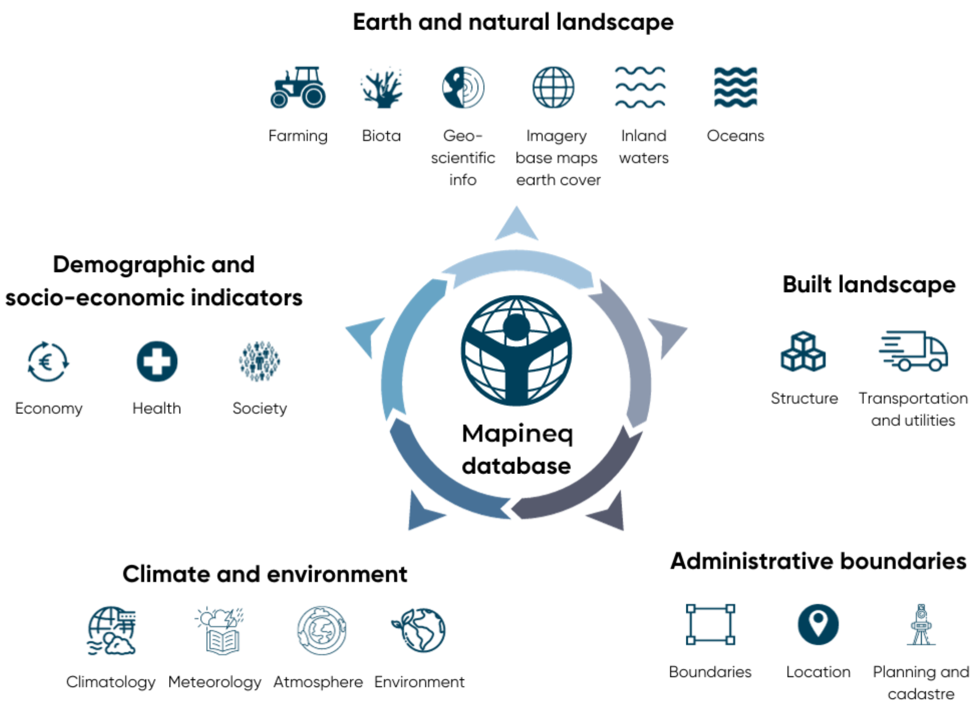
2.2. Environmental measures

The environmental module of the database consists of variables across five substantive themes and four geo-spatial data types, which are described in detail in the report by Verkroost et al (2023).

These five substantive themes aggregate parts of the ISO19115 Topic Category1 (see Figure 1) and include:

1. earth and natural landscape (including farming, biota, elevation, geo-scientific information, imagery base maps earth cover, inland waters and oceans);
2. built landscape (including structure, transportation and utilities);
3. administrative boundaries (including boundaries, location, planning and cadastres);
4. climate and environment (including climatology, meteorology, atmosphere and environment);
5. socio-economic indicators (including economy, health and society).

Figure 2. Visualisation of the five thematic modules of Module 2 environmental indicators.



2.3. Real-world data

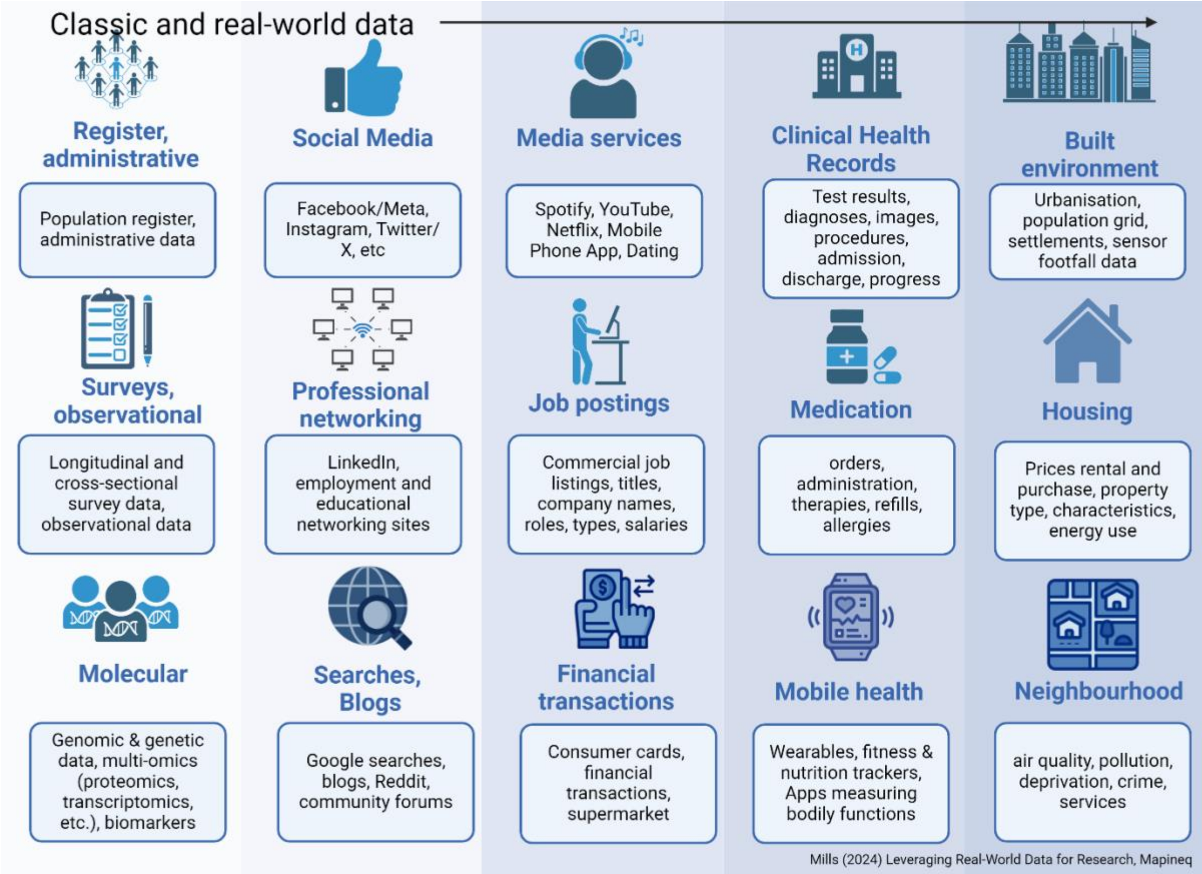
Our third report provided an overview of exemplary types of real-world data which we will continue to explore and integrate over time. These include often aggregated and geolinked data from: (1) social media (e.g., Facebook), (2) professional networking sites and job postings (e.g., LinkedIn), (3) searches, views, blogs and forums (e.g., Google, Reddit), (4) media services (e.g., Spotify, Netflix, YouTube), (5) financial transactions (e.g., credit, loyalty cards), (6) mobile health (e.g., phones, wearables), (7) built environment, pollution and neighbourhood (e.g., prices, Street Maps, air pollution), and, (8) internet speed and

[1] https://apps.usgs.gov/thesaurus/thesaurus-full.php?thcode=15

Funded by the European Union

information communication technology (ICT) (e.g., broadband speed). An overview of the data uses and potential types of data that can be used is discussed in Mills (2024) and shown in Figure 3.

Figure 3. Classic and real-world data types and sources.

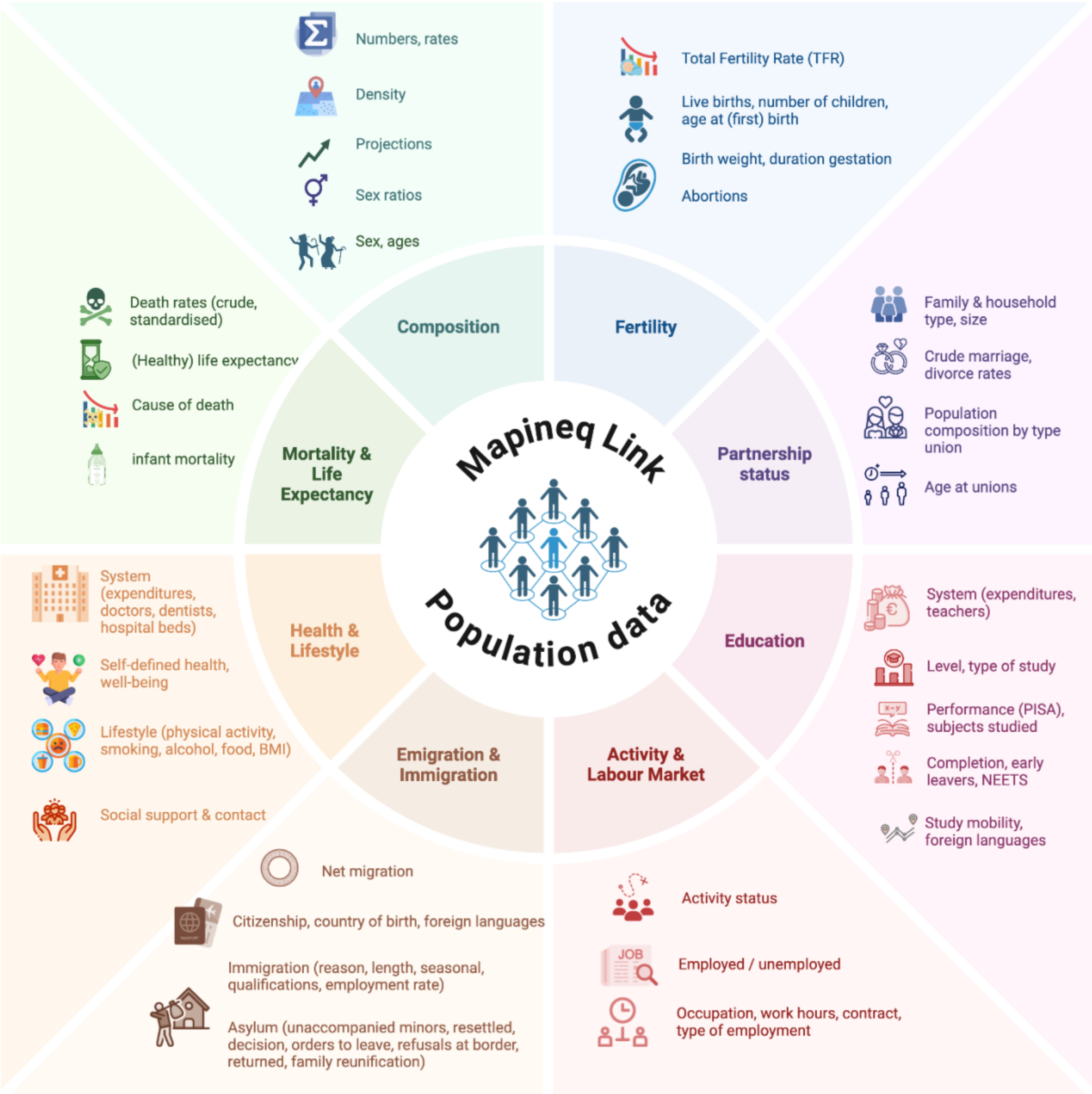


2.4. Demographic, socioeconomic, health and lifestyle indicators

The fourth module of the Mapineq Link database included national and sub-national measures of: (1) population composition; (2) fertility, (3) partnership status, (4) mortality and life expectancy, (5) health, disability and lifestyle, (6) emigration, and immigration; and, (7) socioeconomic indicators (education, activity status). A full description of this module can be found in Scotti-Bentivoglio, Leasure and Mills (2024), with an overview provided in Figure 4.

Funded by
the European Union

Figure 4. Demographic, socioeconomic and health and lifestyle indicators.



3. Geographic Units and Geospatial data types

3.1. Geographic units of analysis

For the Alpha version of the first Mapineq Link dashboard and database, we prioritised indicators from OECD (2022; 2023a-e) and EuroStat (2023a; 2023b) for the initial build of the Mapineq database aiming for direct compatibility with these important data sources and to resolve any data harmonisation challenges, particularly with differences in geo-coding. Both data providers have adopted **SDMX standards** (Statistical Data and Metadata eXchange; sdmx.org) which allowed us to access data in bulk that were relevant to the Mapineq Link geospatial database along with detailed metadata in a machine-readable and standardised format. The Mapineq database reuses these formats for data and metadata with some modifications for harmonisation, documentation and spatial linkage purposes.

When it comes to data on policies, the geographical unit of analysis matters not only in terms of the implementation of those policies but also the data collection regarding their efficacy. In terms of the former, policies can be implemented and relevant at the national, regional or local level. In terms of the latter, data on the policies themselves or on outcomes influenced by them can be collected only at the national but also at regional and fine-grained geo-coded levels. For instance, federal government spending, national educational policies (e.g. required years of schooling), or national labour market policies (e.g. minimum wages), tend to be measured at the national level. If, however, regional or local authorities allocate spending or educational policies are handled by region or province/state, the relevant geographic unit of variables might change. Furthermore, many variables that are influenced by national-level policies, such as the unemployment rate or country of birth of residents may be more relevant to measure at a fine-grained lower geographic unit. The Mapineq Link geospatial database accommodates data queries across these geographic levels and subnational variations. Furthermore, it facilitates various widely used administrative boundaries datasets such as European NUTS (Nomenclature of Territorial Units for Statistics) regions, International Territorial Levels used by OECD and the United Kingdom, and GADM (globally-harmonised administrative boundaries). What is essential is the harmonisation of subnational information not only between different definitions of administrative regions, but also within these definitions (e.g., when it comes to inconsistencies over time). For Phase I, we focused on NUTS regions and OECD International Territorial Levels.

3.1.1. NUTS Divisions

The Nomenclature of Territorial Units for Statistics (NUTS) is a geocode standard, which was adopted in 2003 by the European Union. The Mapineq Link database includes the NUTS polygon geometries and geocodes all EuroStat data to these boundaries. These boundaries can also be used to produce zonal statistics summarising other spatial layers that will be incorporated into the database in later phases (e.g. phase II incorporates environmental spatial layers).

A hierarchy of three NUTS levels (i.e. NUTS1, NUTS2 and NUTS3) was established by Eurostat. The subdivisions do mostly, but not necessarily, correspond to administrative divisions within the countries. Below the three NUTS levels are local administrative units (LAUs). The current NUTS classification, valid from 1 January 2021, lists 92 regions at NUTS 1, 242 regions at NUTS 2, 1166 regions at NUTS 3 level, and 99,151 local administrative units (LAUs). The NUTS framework provides spatial units of analysis that allow for a comparative spatial setup of European countries and that captures existing administrative boundaries. The wide usage of this system in EU, national and regional data collection makes this the most prevalent spatial classification system for data collected thus far.

Funded by
the European Union

3.1.2. OECD International Territorial Levels

The territorial levels used by the OECD roughly correspond to NUTS regions, but the spatial extent includes OECD countries outside of the European Union. There are also differences in Germany, United Kingdom, France, Greece, The Netherlands, Belgium, and Iceland where the hierarchy of TL levels do not map directly to NUTS levels. The polygon geometries for OECD territorial levels are not publicly available, so we have a request pending through project collaborators at OECD to access these boundaries for use in the Mapineq database. Currently, the database maps OECD TL geocodes to spatial geometries of NUTS regions. For non-European countries where NUTS regions are not available, the Mapineq database may need to link TL region codes to GADM boundaries.

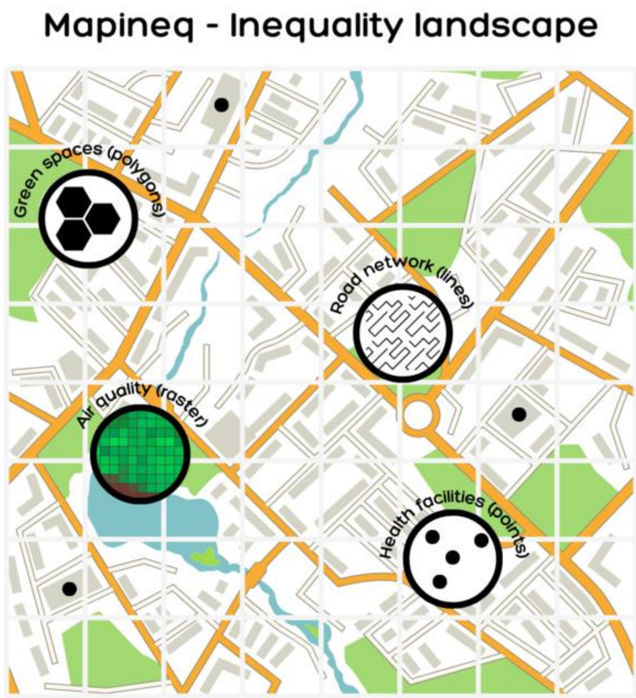
3.1.3. GADM Database of Global Administrative Areas

GADM administrative boundaries (gadm.org) are openly available for free with global coverage and general alignment with NUTS regions in the European Union. The Mapineq database may need to rely on GADM boundaries for any countries included that are outside of the European Union, including some OECD countries. While Mapineq focuses on the EU, it is important to have spatial geometries defining the boundaries of non-European OECD geographies so that these national and subnational indicators can be spatially linked to other datasets in the database and to facilitate cross-country comparisons outside of Europe.

3.2. Geospatial data types

The four geo-spatial data types that are considered in the Mapineq Link resource include: (1) point data; (2) line data; (3) raster data; and (4) polygon data. These are represented visually in Figure 5.

Figure 5. Visualisation of the four geospatial data types of Module 2 environmental indicators.



3.2.1. Point data

For the initial Alpha development phase Mapineq Link, we have focussed on a number of point type of geospatial data. Point data are often found in the second thematic module concerning the built landscape (e.g., when it comes to facilities and buildings) or the fourth thematic module concerning climate data (i.e., when the data come from specific point stations).

Examples of point data are levels of ozone as well as UV radiation, for different periods of time as well as measuring devices and profiles. That is, we include total column ozone as well as vertical ozone profiles (lidar, ozone sonde, rocket sonde and Umkehr-N14), and for UV radiation we include broadband, multiband and spectral data. These data are available from the world Ozone and Ultraviolet Radiation Data Centre (WOUDC).[2] The geospatial points refer to measurement stations worldwide. Other examples of point data that we are considering for future inclusion in the database are locations of schools, hospitals, clinics, and daycare facilities. Accessibility to services depends on their locations relative to the populations they serve, and geographic disparities in access can underpin inequalities in health outcomes. Inclusion of these data depends on our ability to retrieve them with consistent coverage from across Europe, and we are currently assessing the feasibility of this.

3.2.2. Line data

In our initial development of Mapineq Link, we focused on a number of line type of geospatial data. Line data are often found in data contained in the second thematic module, namely the built landscape and particularly infrastructure and transportation (Verkroost et al 2023).

An example of line data currently contained in the database is a detailed network of the road network in Great Britain based on the Ordnance Survey.[3] In the future, these data could be extended to include European countries, if they are openly available. The geospatial lines refer to the network of roads classified by the National or Local Highway authority, and are updated bi-yearly.

Road networks data are a critical component to support calculation of "distance to" measures within the database. For example, a user could provide point locations and ask to retrieve the distances to the nearest health care facilities or other points-of-interest that are in the database. For many research questions, Euclidean distances (i.e., straight line distance) would be less suitable than road distances (i.e., driving distance). As we develop the resource over time, our aim is to support user queries to calculate both Euclidean distances and driving distances from user-provided point locations to various points-of-interest in the Mapineq Link database. Other types of line data we are considering to include in the future are rivers, trails and streets. These data are interesting from a substantive point of view because road proximity and connectivity are related to various life course domains and outcomes as described in Verkroost et al. (2023).

[2] https://woudc.org/data/explore.php?lang=en

[3] https://www.ordnancesurvey.co.uk/products/os-open-roads

Funded by
the European Union

3.2.3. Raster data

To date we have focused on a number of raster types of geospatial data. Raster data are usually linked to the module of climate and environment data (Verkroost et al. 2023), when weather station data are interpolated across a pre-specified grid of smaller areas. Raster data that are being integrated into the database include precipitation, solar radiation, water vapour pressure, wind speed, and temperature (minimum, average and maximum).[4] The geospatial rasters refer to these blocks of earth surface.

Raster data are usually found as remotely sensed and topographic data, including satellite and aerial imagery. Such data are substantively relevant, especially when it comes to mapping (the consequences of) climate change throughout the past decades. As discussed in our previous reports, we can use this type of data to examine migration, climate and environment and how they relate to inequalities in health and well-being such as access to green space, internet speed or infrastructure.

Satellite-based measures of buildings are another important raster dataset that we will incorporate into the database such as the Global Human Settlement layer from the European Commission and the World Settlement Footprint from the German Aerospace Centre (see Verkroost et al. 2023). These datasets quantify the coverage of buildings–and sometimes building heights–for small grid cells globally and their changes through time. These space-based observations of the built environment are fundamental for understanding the inequality landscape in terms of geographic distributions of building densities, urbanization, settlement sizes, proximity to undeveloped areas, and so forth.

A priority set of raster datasets that we will include in the database is gridded population data. WorldPop is one widely-used example that provides annual population estimates (until 2020) disaggregated by age and sex for every 100 m grid square globally. High-resolution gridded population data provide the ability to aggregate 100 m grids to generate population estimates into spatial units appropriate for a given research question. This could range from administrative areas (e.g., NUTS regions) to school or hospital catchment areas that are provided by users. This provides an important source of population denominators for producing per capita inequality indicators for relevant subnational areas.

3.2.4. Polygon data

We also focused on a number of polygon types of geospatial data. Polygon data are usually found in data that refer to bounded areas, such as the natural boundaries (e.g., wetlands, forests) as well as the administrative boundaries (e.g., postal codes, NUTS regions) – which usually also are the geographical units of the socio-economic and demographic data.

This includes data collected from Natural Earth, funded by the North American Cartographic Information Society (NACIS), and contain information about roads, railroads, airports, ports and urban areas at the 1:10M scale.[5] The second set of data are collected from the Global Lakes and Wetlands Database, provided by the WWF and the Center for Environmental Systems Research (University of Kassel, Germany).[6] The first level (GLWD-

[4] https://www.worldclim.org/data/worldclim21.html

[5] https://www.naturalearthdata.com/downloads/

[6] https://www.worldwildlife.org/pages/global-lakes-and-wetlands-database

1) consists of the world's largest lakes and reservoirs, whereas the second level (GLWD-2) consists of permanent open water bodies not included in GLWD-1. There are 3,067 lakes of an area larger than 50 squared kilometres as well as 654 reservoirs of storage capacity over 0.5 cubic kilometres in GLWD-1; GLWD-2 contains open water bodies that are at least 0.1 squared kilometres in surface area. The geospatial polygons refer to the shapes of these lakes and wetlands. Most often, polygon data refer to bounded areas such as the aforementioned lakes and wetlands, but also for example forests and cities. These data could be relevant to explore a variety of questions in the socio-economic inequality domain, such as when it comes to the relationship with urban forestation or street tree abundance.

In the Alpha phase of development, we included Ookla internet speed tests to capture digital inequality. These data provide quarterly aggregated results for millions of internet speed tests conducted by the public. These results are provided within 600 m grids for users connected via mobile networks as well as fixed "landline" networks. Geographic differences in access to high speed broadband and mobile networks is a contributing factor for inequalities in job opportunities, access to information to improve both education and health. This often manifests as an urban/rural divide.

4. Mapineq Dashboard and Database

This section describes the targeted audience and users of Mapineq Link and how they are invited to interact with content on the dashboard and database. We then describe various elements of the web interface, with examples of static screenshots to demonstrate functionality. We note that as of September 2024, Mapineq Link is in the Alpha testing phase and we have established a Minimum Viable Product (MVP). This means that we are testing the functionality and embarking upon user-experience and testing within our multiple user groups to provide feedback for further development. After this phase the web interface will become openly available with a selection of indicators, with data and functionality expanded throughout the duration of the project. We have developed this version together with programmers and developers at Geodienst, located at the University of Groningen in the Netherlands.

4.1. Who are the users and what the uses of Mapineq Link?

Mapineq Link is anticipated to serve multiple user groups, with multiple purposes (Figure 6a, 6b). Policy-makers may want to understand how different types of inequalities or inequalities are patterned in their region, district or country and how that compares to others. Journalists may want to cite and explore one of our existing case studies or use the interactive dashboard to do investigative journalism or produce figures. Researchers can use the data to form hypotheses and visualise data and also join with data scientists to use the API to unlock tools and link the geospatial indicators to their own micro-level individual data for further analysis. As Figure 6b illustrates, on the landing page we have developed bespoke portals for each type of user, tailored to their different needs.

Figure 6a. Users and Usage of Mapineq Link: Policy-makers, journalists, researchers and data scientists.

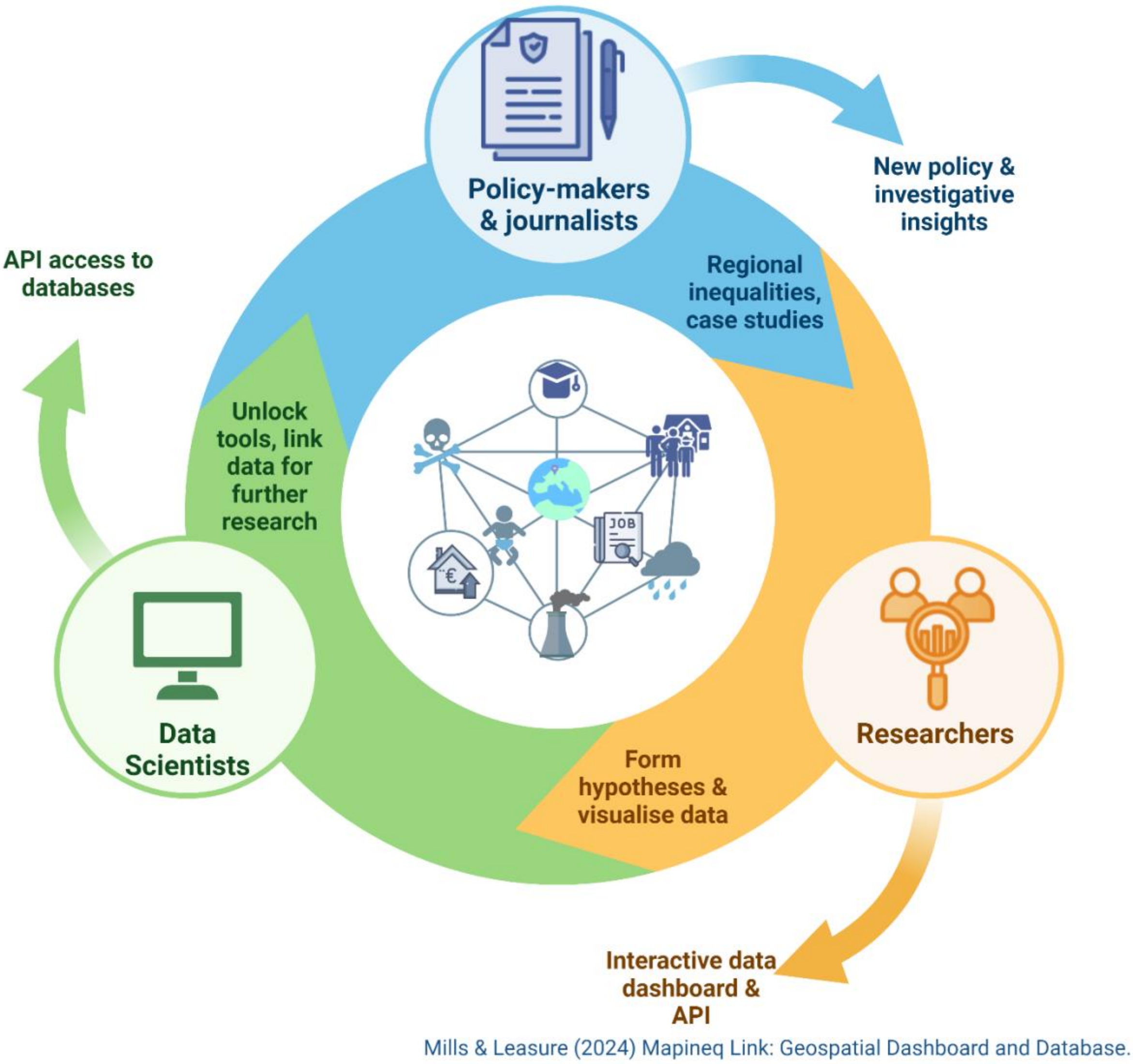
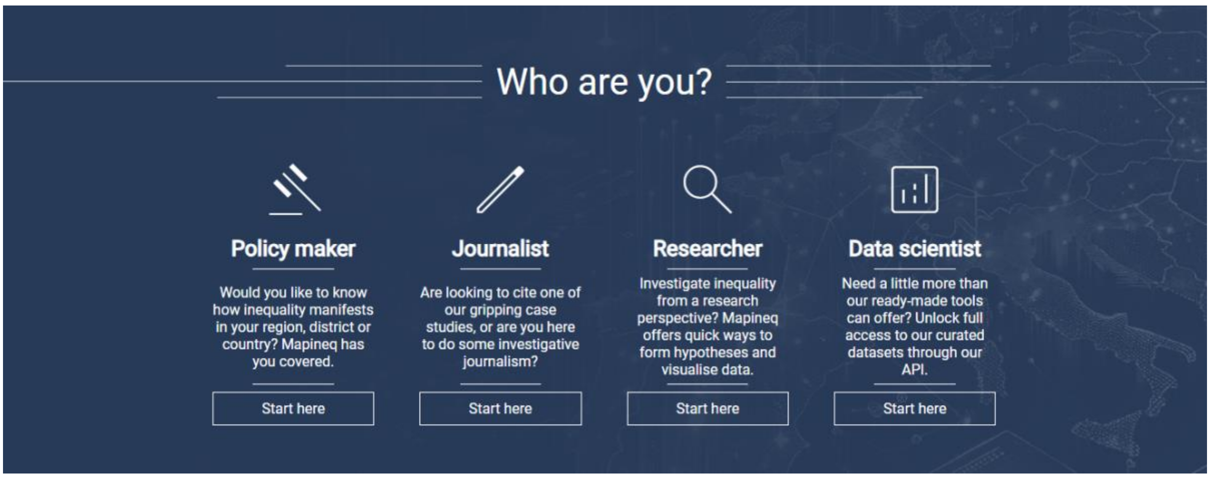


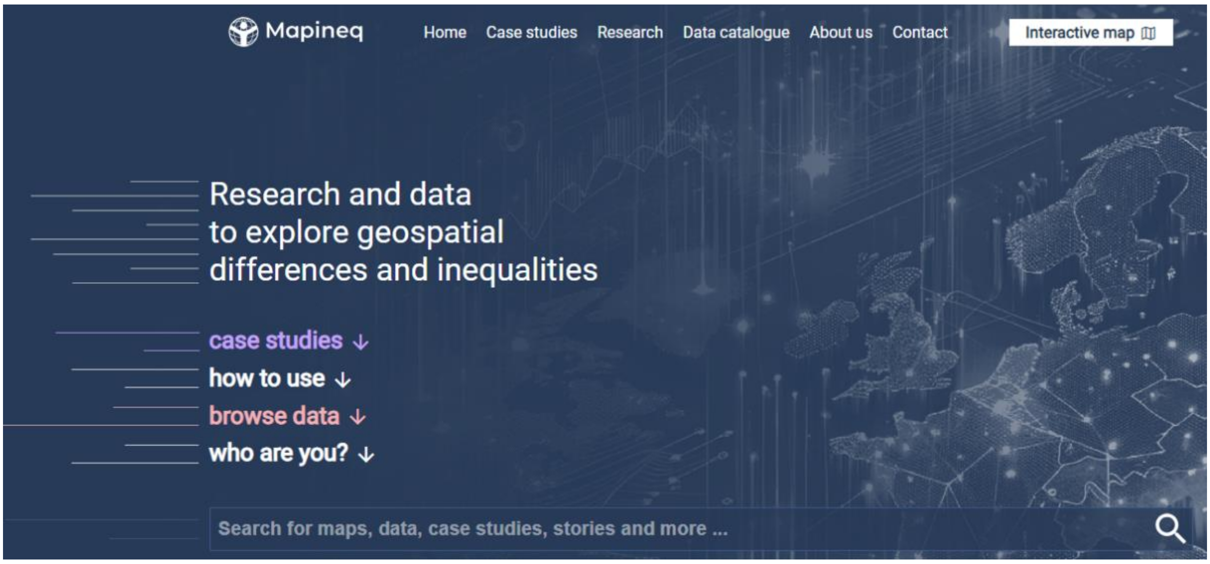
Figure 7. Web interface portal for each type of user, Mapineq Link.



4.2. Web interface

The figures below show the **Alpha MVP version of the Web interface**. As Figure 7 illustrates, the **landing page** offers a clear value proposition and message to explain that the resource contains 'Research and data to explore geospatial differences and inequalities'. We also provide immediate options and calls to action and to address different user needs including: case studies unpacking some interesting examples, instructions and tutorials on how to use the interactive map functions and APIs, a link to the data catalogue to browse data, tailored content for different audiences and a search bar to navigate and search for information across the webpage.

Figure 8. Mapineq Link Web interface landing page.

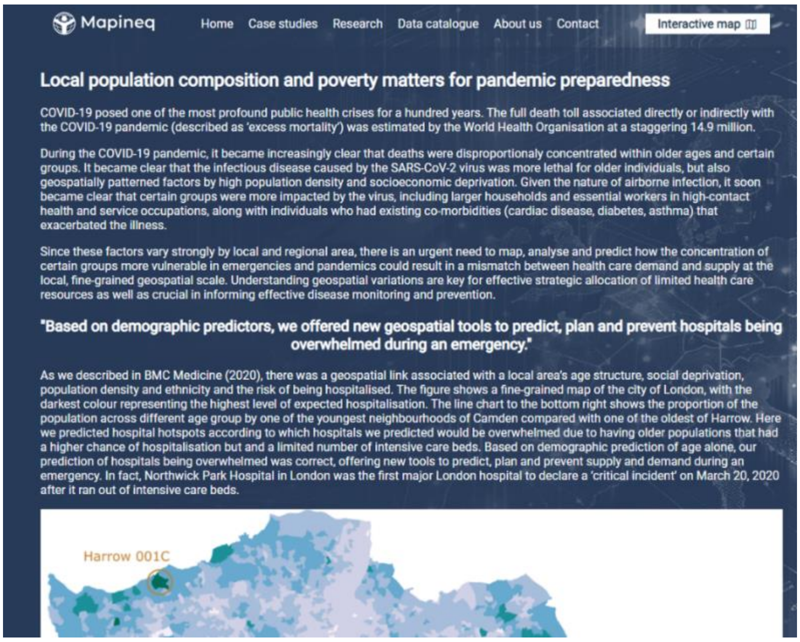


When the user scrolls beyond the header, main navigation bar and menu, they find an example of engaging **Case Studies** that demonstrate a selection of applied examples (Figure 8).

Figure 9. Mapineq Link Web interface Case Studies.

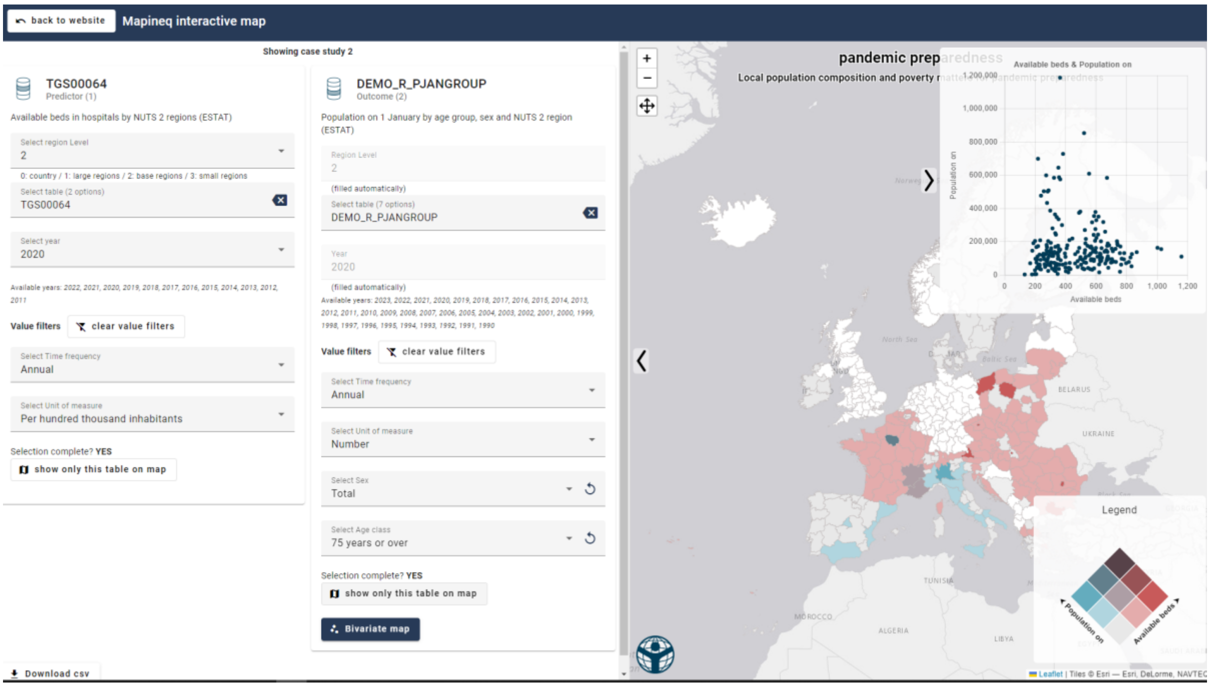
When users click on a particular case study, they find a **short blog** describing an example of how this data can be used from Mapineq projects and researchers and beyond (see Figure 9).

Figure 10. Mapineq Link Web interface Case Study Blog.



Within the description of the Case Study, users are invited to visit and explore the **interactive map, bivariate and association tools** that bring the story and data to life (see Figure 10). The figure shows the Alpha version of an interactive map that demonstrates the bivariate relationship between two geospatial indicators. The user can experiment and produce a geospatial map of one indicator, or as shown in Figure 10, explore the geospatial relationship between two indicators discussed in the case study. This unique bivariate option allows users to explore associations in a rapid and visual manner. In Figure 10, for instance we see the relationship between the number of available hospital beds by NUTS 2 regions by the population aged 75 years and older for both sexes. Users can select different options, such as changing the time period, unit of measurement, sex or age.

Figure 11. Mapineq Link Interactive Map.



As Figure 11a illustrates, when expanded (by clicking on the middle arrow), the right hand map showcases several elements. The colours on the map refer to the **bivariate association of the two selected indicators.**

Figure 12. Mapineq Link Detail of Interactive Map

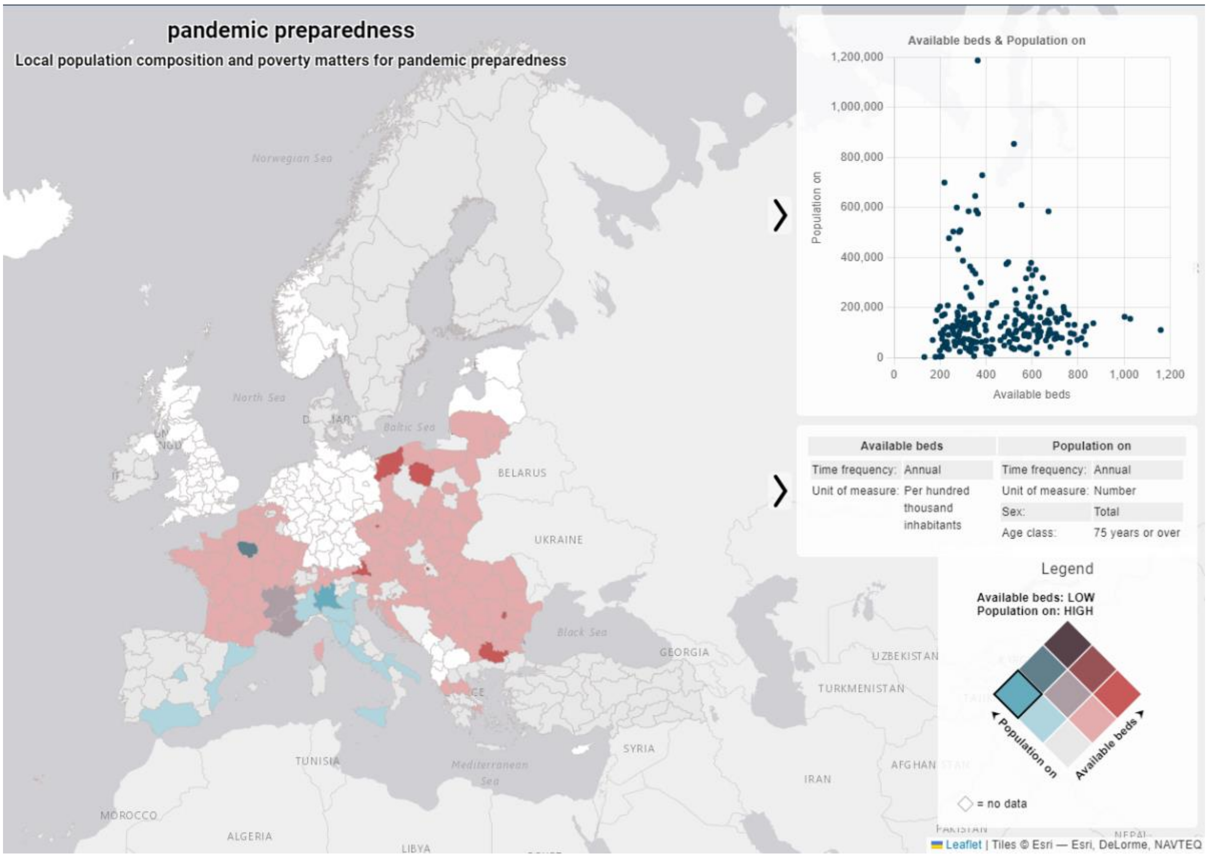
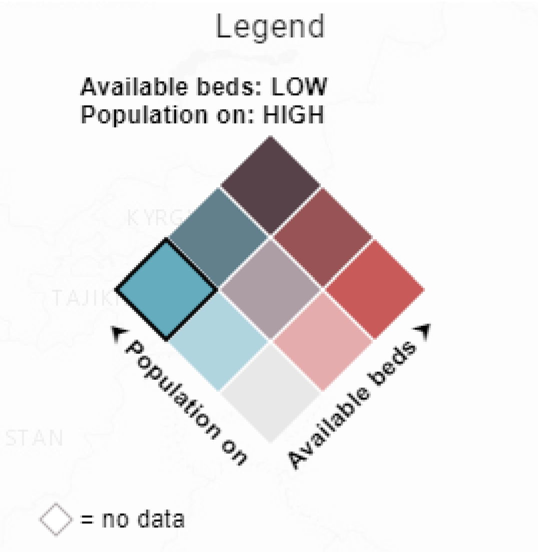


Figure 13. Mapineq Link Detail of Bivariate Legend.



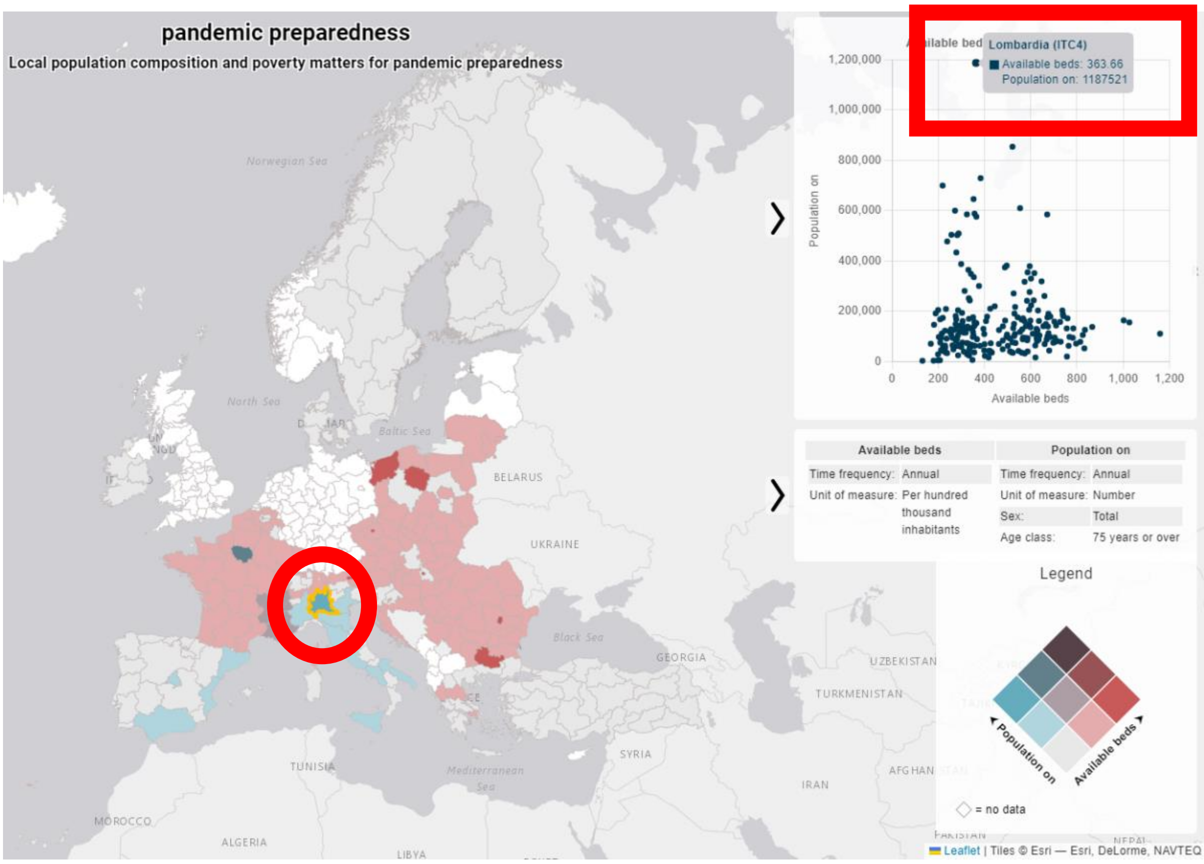
A **bivariate plot** graphs the relationship between two indicators that have been measured by a particular region or unit of analysis. This type of plot permits the user to see at a glance the degree and pattern of relation between the two indicators

If the user hovers over the blocks in the legend, they discover the meaning of the colour. Figure 11b illustrates that the brighter blue colour on the map is a region that has low available hospital beds and a high population who is 75 years and older.

In the upper right part of the Figure (see also Figure 12), we produced a **scatterplot**, where the dots represent the values for each of the two selected indicators. The position of each dot on the horizontal and vertical axis indicators a value for each region (in this example, each NUTS 2 region). Scatterplots are a quick way to view the relationship between two indicators. Here we see the association of the population aged 75 and older by the available hospital beds per 1,000 inhabitants. If the user hovers over a dot in the scatterplot, such as the example show in Figure 12, they see that the outlier at the top of the plot is a region with a very large number of individuals 75 years and older with a relatively lower number of hospital beds, which is the Italian region of Lombardia. By hovering over the dot (as emphasized by the red marker), the user sees the name of the region and measurements while also the area is simultaneously highlighted in yellow in the accompanying map (highlighted by the red circle). Users can also use the arrows to remove or include the scatterplot and accompanying information.

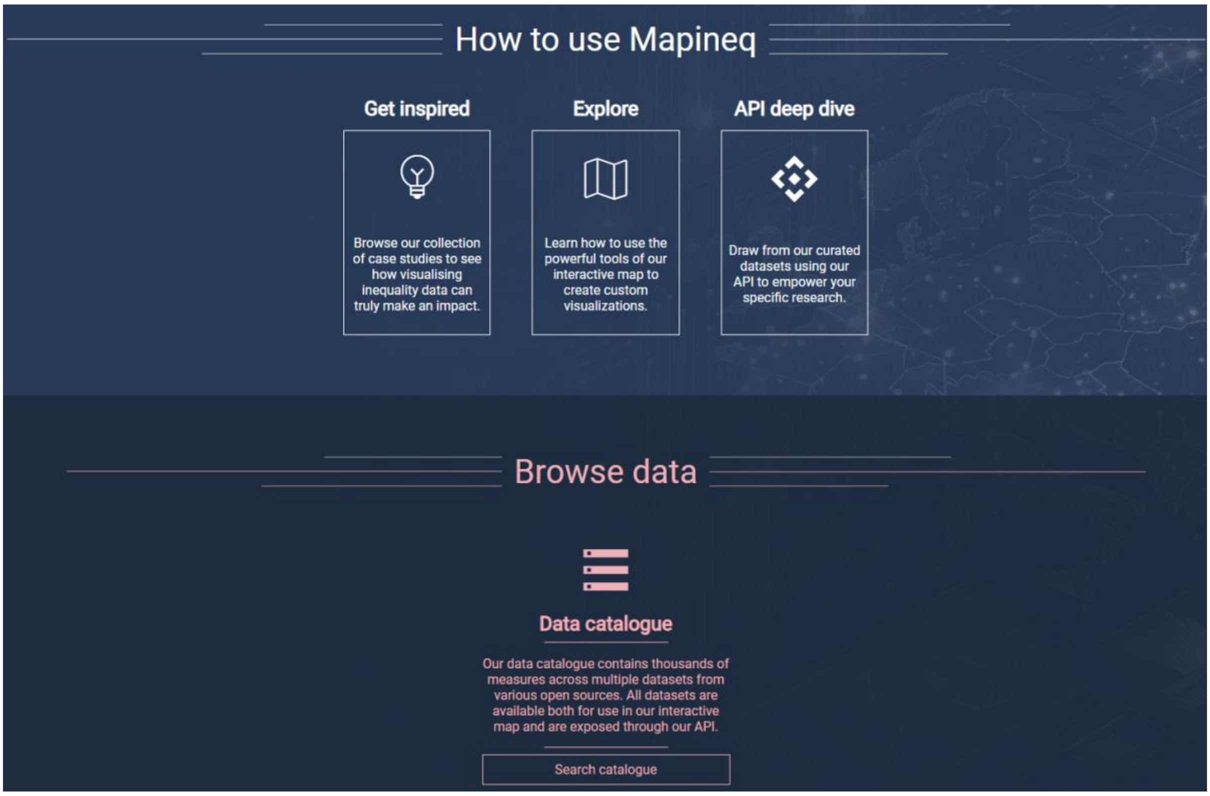
Funded by
the European Union

Figure 14.Mapineq Link Detail of Scatterplot functionality.



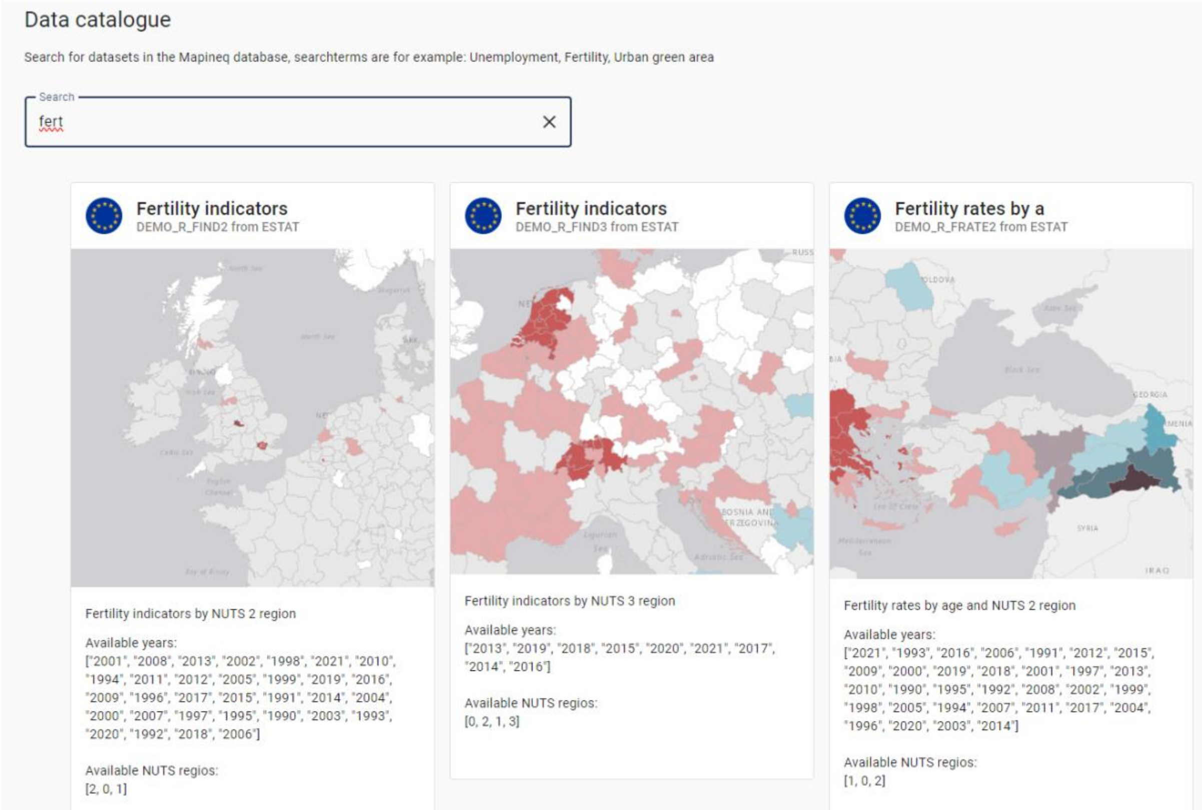
The user can exit the interactive map at any time to return to the web interface. If they scroll further down the landing page of the main site (see Figure 13), they can link to **additional educational material** to get inspired, explore the data or engage in an API deep dive. At the time of writing these resources, including written tutorials and videos are being developed together in co-design with user groups.

Figure 15. Mapineq Link Detail of How to Use the Web Interface and Data Catalogue



The bottom portion of Figure 13 shows the portal that allows users to enter the **Data Catalogue** to explore the different types of data that is available. If the user clicks on the 'Search catalogue' button they arrive at a new window shown in Figure 14. Here they can enter in keywords to explore which variables are currently available and their granularity in by geographical region and time period. The example shows that when you type 'fert' (to bring up fertility indicators), it shows are variety of indicators that are available (this figure only shows a small excerpt). They also find the variable names and source of the data.

Figure 16. Mapineq Link Detail of Data Catalogue search



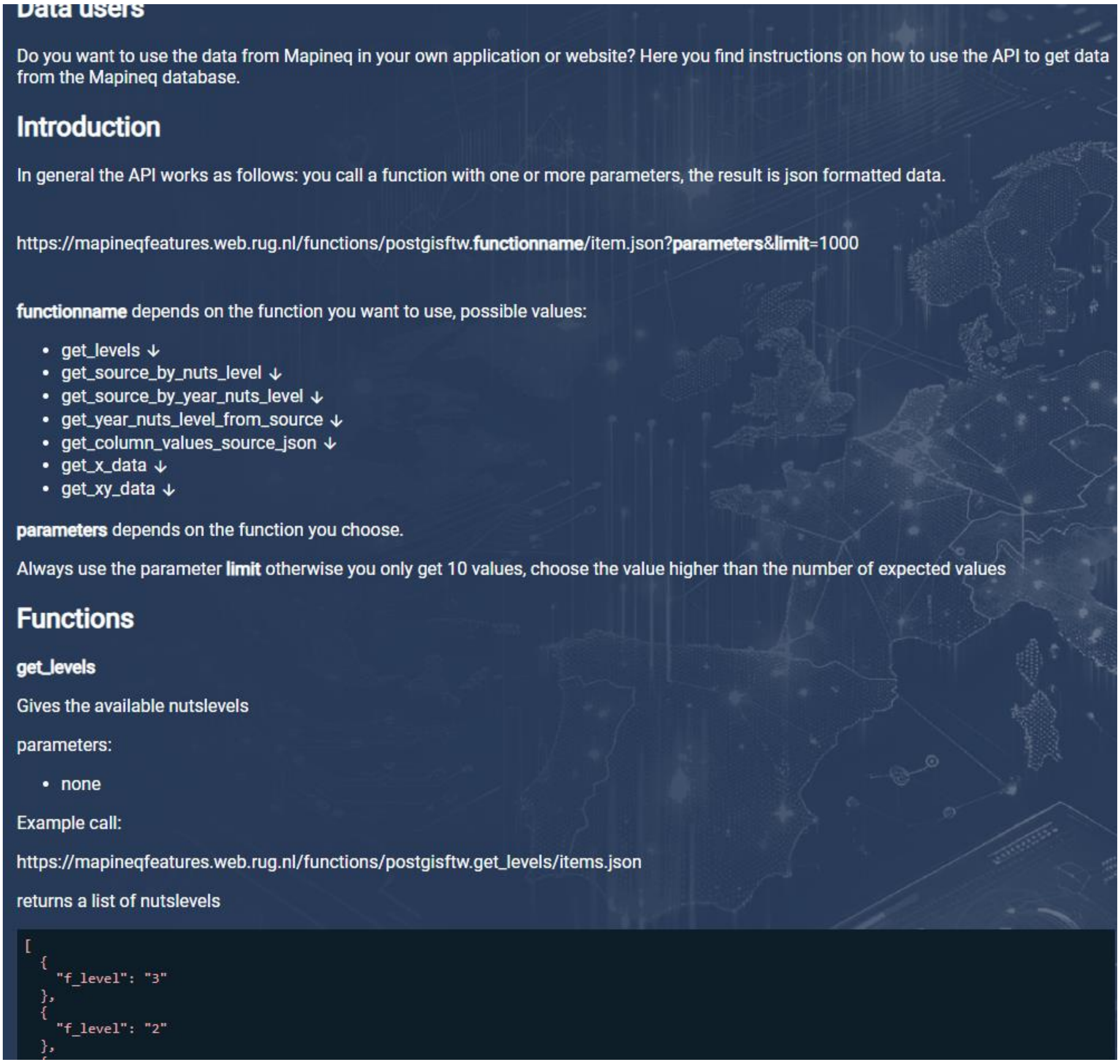
4.3. Mapineq Link API

Figure 15, provides an example of **API deep dive instructions** for programmers and data scientists to allow them to use the Mapineq data in their own work or to embed in their own website. This is a tangible example of our FAIR data principles in action. The Mapineq Link API serves three primary roles:

1. To serve as the intermediary for researchers to use the Mapineq web application(s) to query the database with user-defined parameters;
2. For statistical programmers and data scientists to retrieve data for project-specific locations or bespoke regions directly from within their data analysis environments in R, Python, or other programming languages;
3. For external web developers to easily integrate Mapineq data into their own interactive web applications.

The API pre-processes requests for data, queries the database, and formats the results into an easy-to-use format. It will perform quality checks on requests and results, communicating informative error messages back to the user where necessary and pointing them to relevant documentation to resolve the issue. To date we have developed the core infrastructure for the API using the Python Flask package and Nginx web server deployed in a Docker container for interoperability across operating systems and cloud computing platforms.

Figure 17. Mapineq Link Detail of API deep dive instructions (Alpha version).



Data users

Do you want to use the data from Mapineq in your own application or website? Here you find instructions on how to use the API to get data from the Mapineq database.

Introduction

In general the API works as follows: you call a function with one or more parameters, the result is json formatted data.

https://mapineqfeatures.web.rug.nl/functions/postgisftw.**functionname**/item.json?**parameters&limit**=1000

functionname depends on the function you want to use, possible values:

- get_levels ↓
- get_source_by_nuts_level ↓
- get_source_by_year_nuts_level ↓
- get_year_nuts_level_from_source ↓
- get_column_values_source_json ↓
- get_x_data ↓
- get_xy_data ↓

parameters depends on the function you choose.

Always use the parameter **limit** otherwise you only get 10 values, choose the value higher than the number of expected values

Functions

get_levels

Gives the available nutslevels

parameters:

- none

Example call:

https://mapineqfeatures.web.rug.nl/functions/postgisftw.get_levels/items.json

returns a list of nutslevels

```
[
  {
    "f_level": "3"
  },
  {
    "f_level": "2"
  },
```

We continue to focus on developing API "endpoints" for various types of database queries. Each endpoint is designed to support specific types of database queries and data formats. For example, an endpoint may be designed to accept geospatial coordinates defining user-defined research locations and return selected indicators or summary statistics for those locations. Other endpoints may be designed to simply return data for data visualisations in the Mapineq Link web application customised based on selections of data and locations by the user navigating the Mapineq Link web application.

4.4. FAIR Software Development

The Mapineq Link team is dedicated to FAIR principles to make our innovations Findable, Accessible, Interoperable, and Reusable (go-fair.org). The Mapineq Link database is powered by a PostgreSQL relational database server (postgresql.org) with a PostGIS extension to support geospatial functionality (postgis.net). We deployed this software using Docker containers (docker.com) to ensure compatibility across operating systems and cloud providers. The API is built using the Python Flask package and Nginx web server. We chose this stack of free and open source software because it not only provides a powerful geospatial toolbox and scalable data dashboard with broad interoperability, but it also maximises reusability of our innovations for other projects globally. We are hosting our code in a GitHub repository that will be released openly and published with a globally unique and persistent digital object identifier so that the source code is findable in perpetuity. The code will be openly available with a GNU GPL licence (https://www.gnu.org/licenses/gpl-3.0.en.html) which allows modifying and redistributing the source code. We deployed the Mapineq database onto a Paris-based server on Amazon Web Services which provides cloud-based computing infrastructure with scalable performance globally, but it is important to note that the source code is designed to be compatible with any cloud computing provider and/or local deployment on a desktop computer.

5. Mapineq Link Launch: October 7 2024

To mark the launch of the first Alpha version of Mapineq Link, we are holding the event "*Towards a New Equality of Place: Leveraging* Geospatial *Data for Policy Impact*". The event will take place on October 7 at the WZB Berlin Social Science Center and online. Participants will gain insights into the tool's capabilities for visualizing and addressing local disparities through geospatial data. In addition, a panel of experts will discuss the tool's potential for policy impact, followed by a reception for networking and further dialogue. As of 30 September 2024, we have received 271 registrations for the event.

The event will feature the following agenda and speakers:

17:00 - 17:10: Welcome & Overview of Mapineq Link and Its Purpose

Speaker: Melinda Mills, Professor of Data Science & Public Health Policy, University of Groningen and Director of the Leverhulme Centre for Demographic Science, Professor of Demography and Population Health at Oxford Population Health, University of Oxford.

17:10 - 17:20: Walk through of Mapineq Link Features and Capabilities

Speaker: Douglas Leasure, Senior Researcher and Data Scientist at the Leverhulme Centre for Demographic Science, University of Oxford.

17:20 - 18:00: Expert Panel: Insights and Q&A

Speakers:

Bastian Alm, Head of Division Regional Economic and Structural Policy, Federal Ministry for Economic Affairs and Climate Action.

Nicola Brandt, Head of OECD Berlin Center.

Paul Gellert, Co-Director of the Einstein Center Population Diversity and Associate Professor at the Institute of Medical Sociology and Rehabilitation Science, Charité - Universitätsmedizin Berlin.

Melinda Mills, as above

18:00: Reception for attendees to network and continue discussions.

6. Conclusions and future directions

This report provided an overview of the substantive content that is covered within the Mapineq Link database and dashboard, geospatial units and types of data as well as an overview of the Alpha version of our Minimum Viable Product (MVP) produced as of September 2024. This version includes primarily measures from Eurostat and OECD as well as key indicators to highlight each of the data types, such as internet speed (real-world data) and air pollution (environmental data). In future iterations and during user testing we will introduce additional measures. As new data becomes available, and if funding allows, it will remain a living database, where we will continue to include additional geographic unit frameworks as database development progresses. Given the European focus of Mapineq we also focus on Europe, with possibilities and the framework in place to expand in the future. Other extensions we are considering include a data donation component to allow users to suggest and prepare data to 'fill in the gaps' with data derived from their local area.

References

Eurostat. (2023a). Trainings & tutorials Eurostat API with SDMX, url: https://ec.europa.eu/eurostat/web/sdmx-infospace/trainings-tutorials

Eurostat. (2023b). Developer's Corner Eurostat, url: https://wikis.ec.europa.eu/display/EUROSTATHELP/Developer%27s+corner

Leasure, D., Mills, M.C., Wagner, S. & F. Verkroost (2023). Mapineq Link: Institutional policy and structural geospatial database. Turku: INVEST Research Flagship Centre/University of Turku. Available at: https://mapineq.eu/mapineqlink_v1_0/

Mills, M.C. (2024). Mapineq Link: Leveraging real-world data for research. Mapineq deliverables. Turku: INVEST Research Flagship Centre / University of Turku. Available at: https://mapineq.eu/mapineqlink_v3_0/

OECD. (2023a). Regional Statistics, url: https://www.oecd.org/regional/regional-statistics/

OECD. (2023b). OECD.Stat Regional Database, url: https://stats.oecd.org/Index.aspx?DataSetCode=REGION_DEMOGR and codebook: https://stats.oecd.org/Index.aspx?DataSetCode=REGION_DEMOGR

OECD. (2023c). OCED.Stat Metropolitan Database, url: https://stats.oecd.org/Index.aspx?DataSetCode=CITIES and codebook, url: https://stats.oecd.org/Index.aspx?DataSetCode=CITIES

OECD. (2023d). Territorial Grid, url: https://stats.oecd.org/wbos/fileview2.aspx?IDFile=cebce94d-9474-4ffc-b72a-d731fbdb75b9

OECD. (2023e). List of regions and typologies, url: https://stats.oecd.org/wbos/fileview2.aspx?IDFile=db68c5c3-5fd5-465c-b25b-b50aa14c2da1

Scotti-Bentivoglio, G., Leasure, D. & M.C. Mills. (2024). Mapineq Link: Demographic, socioeconomic, health and lifestyle indicators. Mapineq deliverables. Turku: INVEST Research Flagship Centre / University of Turku. Available at: https://mapineq.eu/mapineqlink_v4_0/

Verkroost, F., Scotti Bentivoglio, G., Zhoa, B., Leasure, D. & M.C. Mills. (2023). Mapineq Link: Environmental Geospatial Database. Mapineq deliverables. Turku: INVEST Research Flagship Centre / University of Turku. Available at: https://mapineq.eu/mapineqlink_v2_0/

Funded by
the European Union