



MORE GRANULAR, MORE TIMELY: THE NEW STATISTICAL LANDSCAPE

International Conference of Statistics Flanders, 10 May 2022

Paul Schreyer, OECD Chief Statistician



@OECD_Stat



www.oecd.org/sdd



www.stats.oecd.org



Contents

1. Setting the scene
2. Spatial and granular: examples from the OECD data ecosystem
3. Three key ingredients: harmonisation, guidance, co-investment
4. To conclude

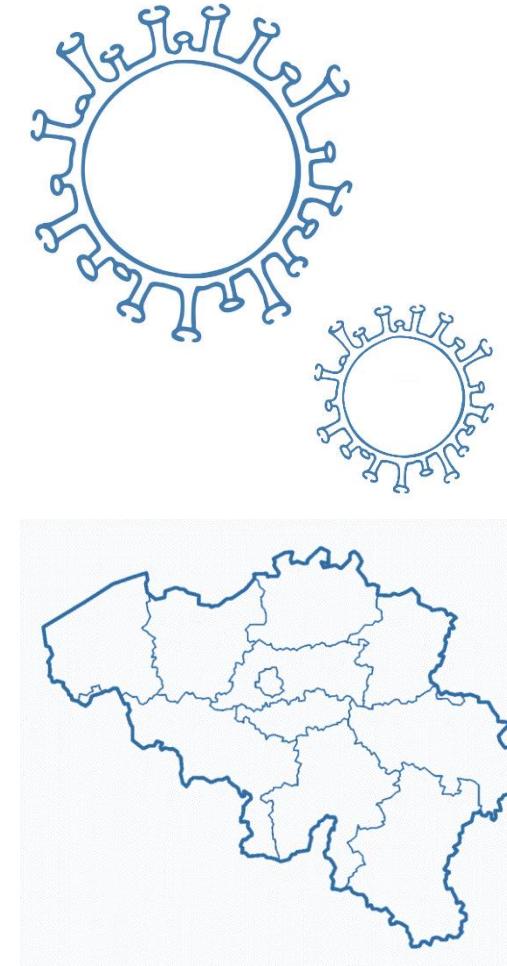
SETTING THE SCENE





Increasing demand for **more timely** and **more granular** data

- Recent crises have underlined the need for **more timely**, **more granular**, and **trustworthy statistics**, in order for policymakers to make swift and effective decisions
- There is increasing recognition that socio-economic shocks do not impact all of society, or all regions, in the same way
- Citizens also increasingly expect to see themselves, or their street, in their statistics
- With these new challenges comes the need for **new guidance**, **new definitions**, **new data** and **new tools**

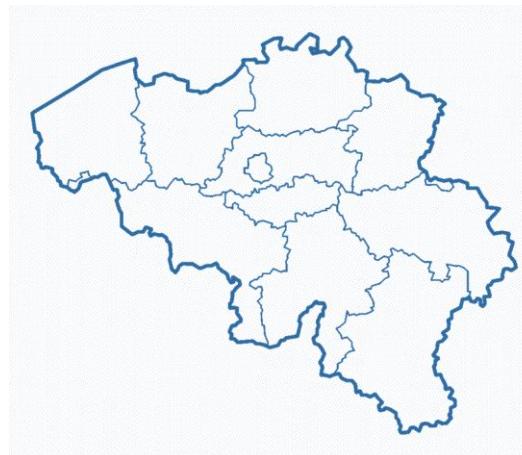
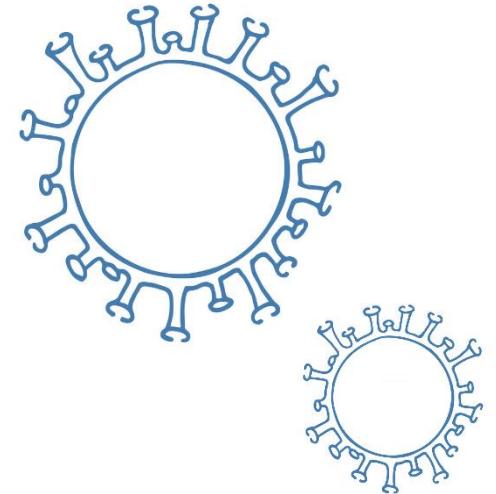




A quickly evolving data ecosystem



- New sources and new tools from digital transformation
- NSO monopoly as data provider is gone
- NSOs now producers *and* users of data
- Data often fragmented, non-communicative between sources, inaccessible for users



Emerging data stewardship roles (1)



- Data Stewardship = bringing governance to data ecosystem
- Sharing and collaboratively improving data interoperability and production processes, know-how and technology, literacy and communication, ease-of-access to data for evidence



Statistics Canada:

« Whole of government approach to creating, protecting, using, managing and sharing data as a strategic asset enabling informed decisions that lead to better outcomes and services for Canadians »

Emerging data stewardship roles (2)

- No single model but typically involves:
- *Other Producers of Official Statistics*: standards , guidelines, inter-operability
- *Administrative data/statistics producers*; data steward role in administrative data collection or processing
- *Third Party private data*: promotion of standards, terminology, classifications and technology, labelling
- ‘Once Only’ principle in data collection
- *New tasks*



Example United Kingdom: ONS COVID-19 Infection Survey

- # of people across England, Wales, Northern Ireland and Scotland that test positive for COVID at a given point in time, regardless of whether they report experiencing symptoms
- Representative sample of private households and characteristics that are tested and surveyed
- Timely, granular, trusted



The logo for the Office for National Statistics (ONS) Data Stewardship. It features the word "data" in a blue sans-serif font, "stewardship" in a larger blue serif font, and "ONS" in a smaller blue sans-serif font. To the left of "data" is a small blue square icon containing a white stylized 'd'. Below the main text is a cloud of smaller, semi-transparent words in various colors (blue, green, yellow, red) including "confidence", "network", "teamwork", "influence", "mission", "power", "open data", "vision", "success", "data privacy", "mentor", "capital", "team", "collaborate", "capacity", "public private partnership", and "movement".

Office for
National Statistics

English (EN) | [Cymraeg \(CY\)](#)

[Release calendar](#) [Methodology](#) [Media](#) [About](#) [Blog](#)

[Home](#) [Business, industry and trade](#) [Economy](#) [Employment and labour market](#) [People, population and community](#) [Taking part in a survey?](#)

Search for a keyword(s) or time series ID

[Home](#) > [People, population and community](#) > [Health and social care](#) > [Coronavirus \(COVID-19\)](#) > [Coronavirus \(COVID-19\) Infection Survey, UK](#)

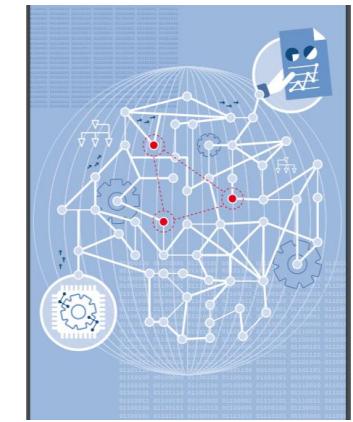
Coronavirus (COVID-19) Infection Survey, UK: 10 December 2021

Example Switzerland: DSCC

The screenshot shows the homepage of the Federal Statistical Office (FSO) of Switzerland. At the top left is the Swiss coat of arms and the text "Schweizerische Eidgenossenschaft", "Confédération suisse", "Confederazione Svizzera", and "Confederaziun svizra". The top center has the text "Federal Statistical Office". On the right is a search bar with the placeholder "Search...". Below the header is a navigation menu with links: News, Look for statistics, Services, Basics and Surveys, Registers, National data management NaDB, Data Science Competence Center, and The FSO. The "Data Science Competence Center" link is highlighted. The main content area shows the breadcrumb navigation "Federal Statistical Office > Data Science Competence Center >". Below this is a backlink to "Federal Statistical Office" and a "Return to overview" link. The main title "Data Science Competence Center (DSCC)" is displayed.



- **Services:**
 - Developing quality standards, guidelines on data protection
 - Consulting on innovative data science methods
 - Methodological support and coaching in implementation
 - Execution of data science requests
 - Application-oriented training



<http://www.experimental.bfs.admin.ch/>

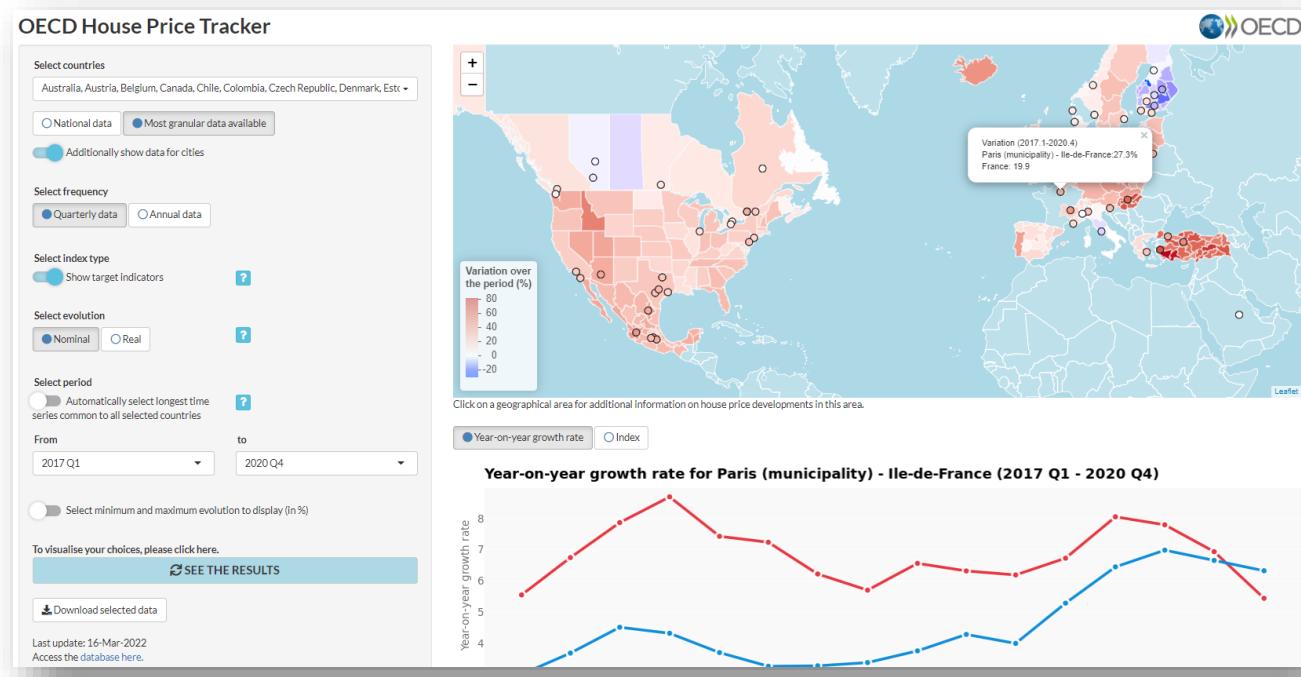
SPATIAL AND GRANULAR: Examples from the OECD data ecosystem





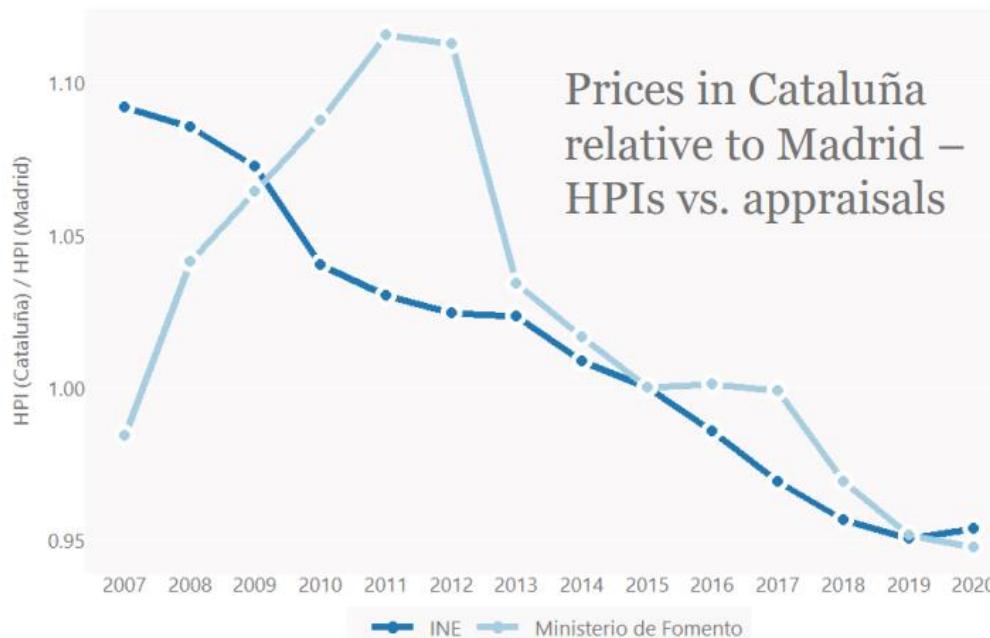
Ex. 1: OECD Regional House Price Tracker

- **Official data from statistical agencies**, allowing users to monitor how house prices have evolved over time in different countries, regions and cities



Ex. 1 (ctd): From indices to levels

- OECD database on national and regional house price *indices*
- Just as interesting: house price *levels*
- Rarely available from official sources – risk of inconsistency



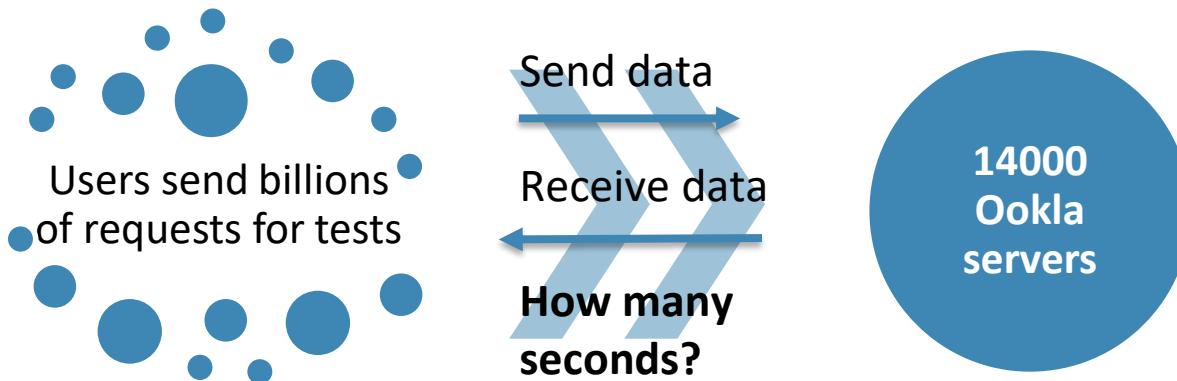
- OECD model uses both sources to construct consistent measure of sub-national levels and indices

Source: Pionnier and Schuffels, [OECD Statistics Working Papers 2021/03](#)



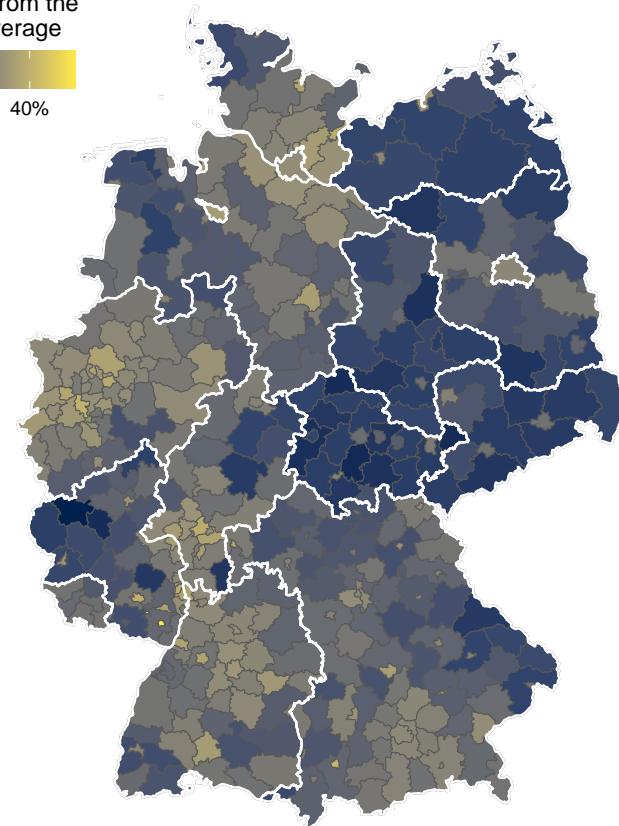
Ex. 2: Inequalities: The rural-urban connectivity divide

- OECD partnership with Ookla
- Traditional collection and self-reporting yield **poor data on internet access and reliability**
- Users conduct speed tests automatically or deliberately on **speedtest.com by Ookla**



Gaps in fixed download speeds experienced by users, (2020)

Deviation from the national average
-40% 0% 40%

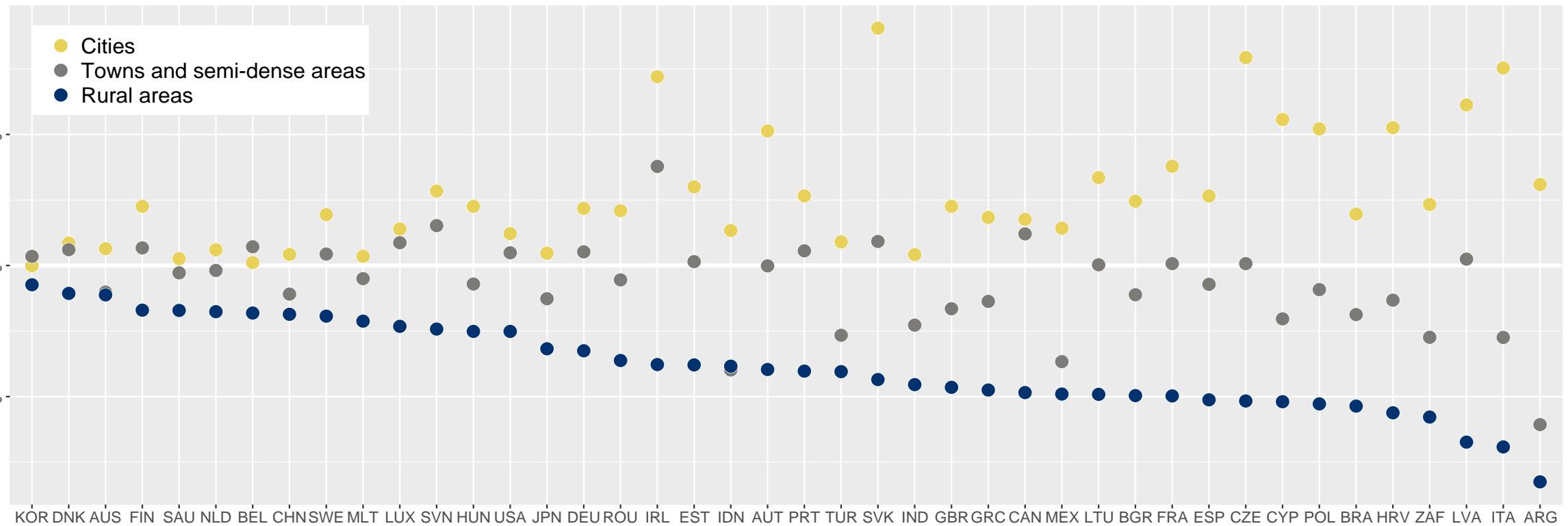




Ex 2 (ctd): New timely and granular data show disparities in internet speeds

Gaps in fixed download speeds experienced by users, by degree of urbanization (2020)

Deviation from the national average (in percentage points)

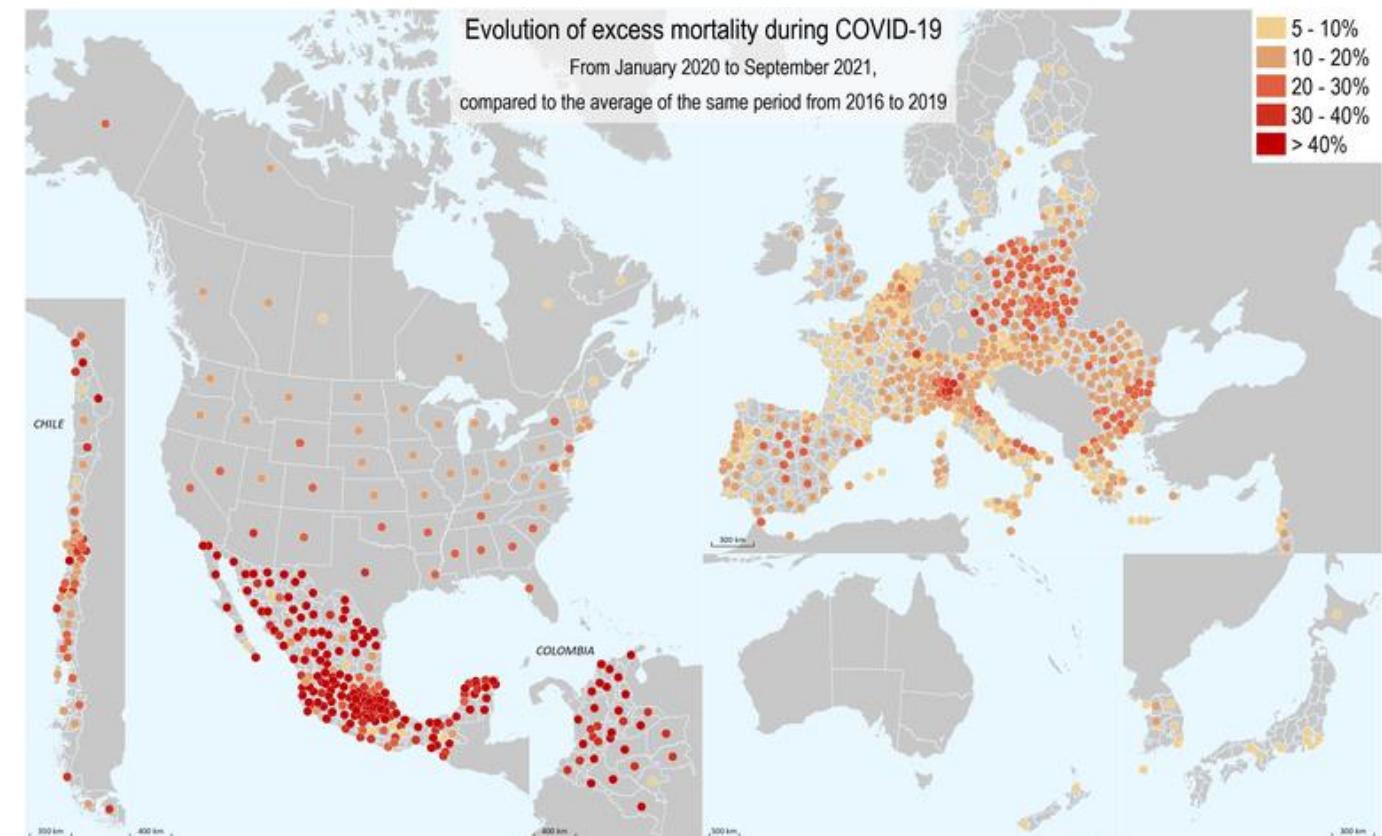


Source: OECD [Bridging digital divides in G20 countries](#)



Ex. 3: COVID-19: Tracking excess mortality

- **Excess mortality during COVID-19** had a significant spatial dimension
- OECD analysis combines **official deaths registers** from NSOs or health ministries with information from the **OECD Regional Statistics Database** and other sources



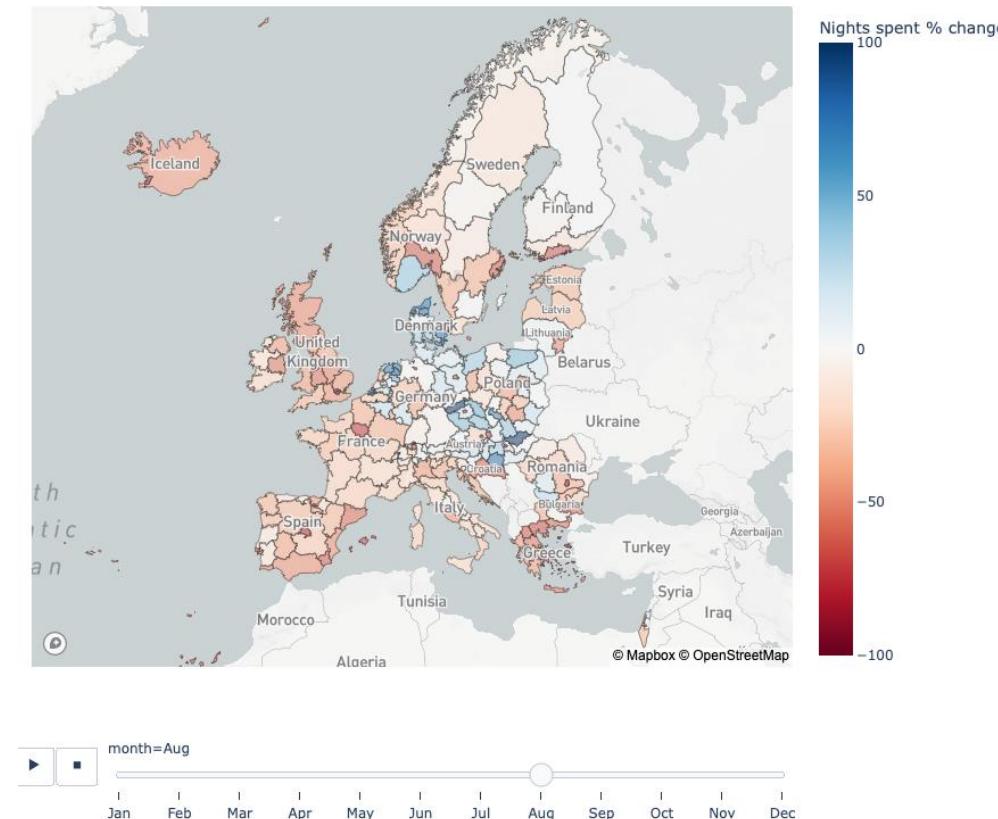


Ex. 4: COVID-19: The impact on tourism trends

- Provisional analysis combines official data with web scraped data on TripAdvisor to assess the impact of COVID-19 on tourism in regions and cities at a high granularity
- Clearance to publish not obvious

Indicators	nights spent number of arrivals occupancy rates
Coverage	38 OECD countries Plus 5 key partners
Granularity	TL2, TL3 and FUA level Monthly and quarterly

Monthly nights spent year-on-year change
2019-2020, TL2 regions



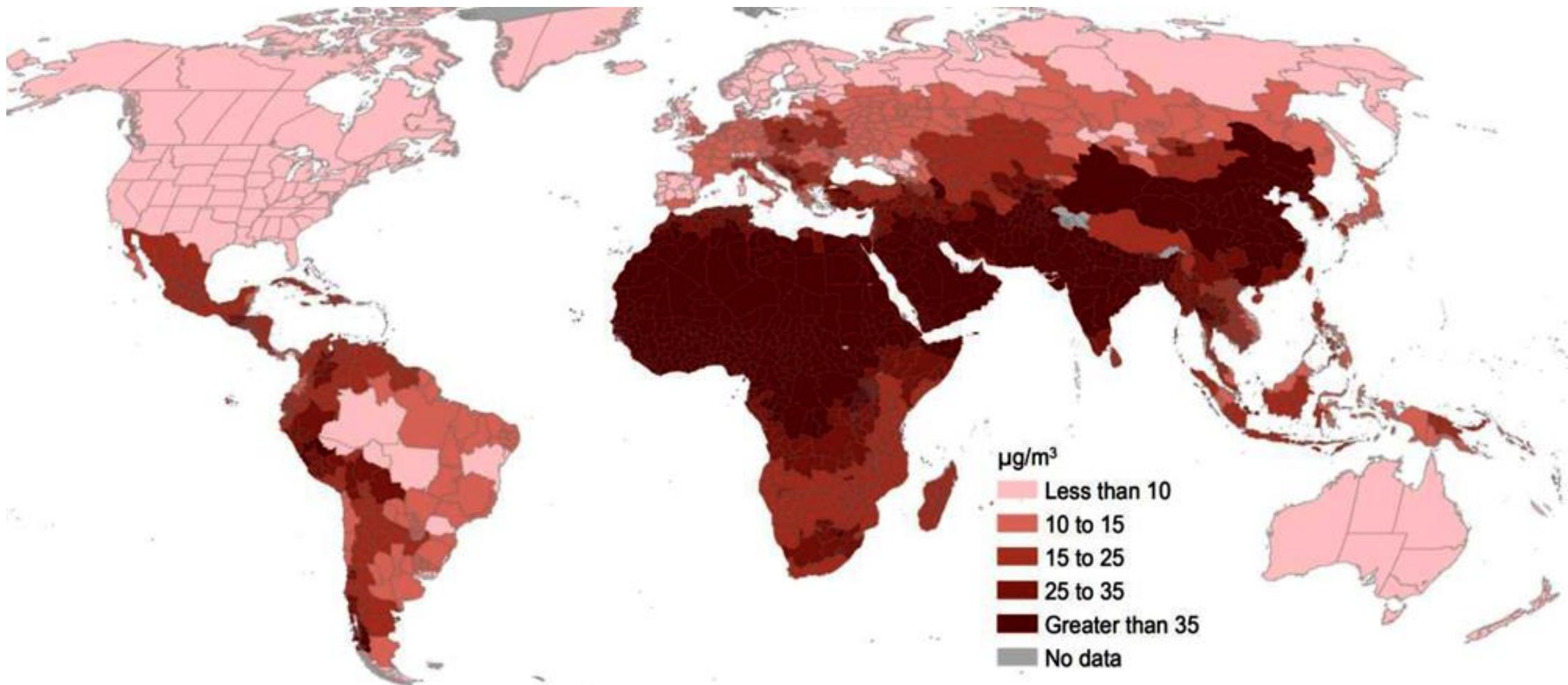


Ex. 5: Climate change: Population exposure to pollution

- Combining **new satellite data** on air pollution (PM2.5) concentration with **traditional census, survey or other spatial data** on population distribution
- Some countries with relatively high-country levels of pollution exposure have regions where exposure is relatively low (e.g. China, India, and Brazil), and vice versa (e.g. Mexico and the United States)
- There have been **improvements in most OECD countries** and a **deterioration in many fast-growing Asian economies**



Ex 5 (ctd):Population exposure to outdoor PM2.5, 2019



THREE INGREDIENTS



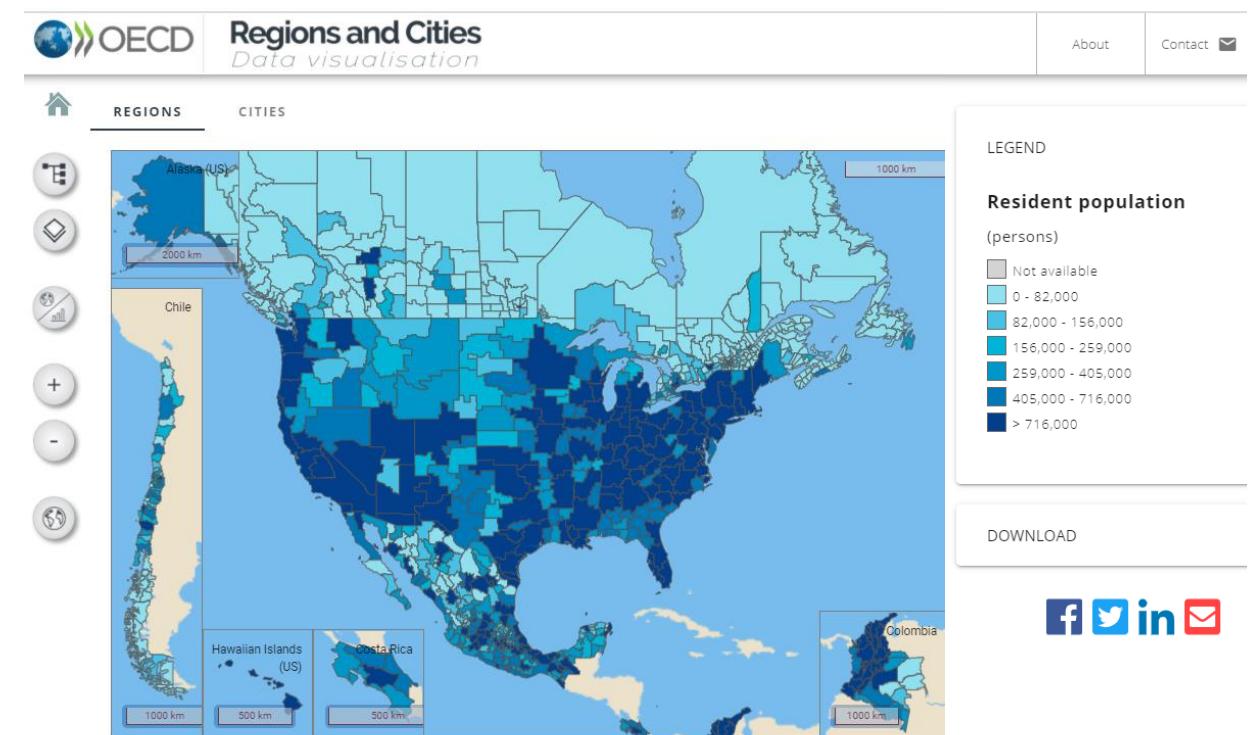
1. Harmonisation





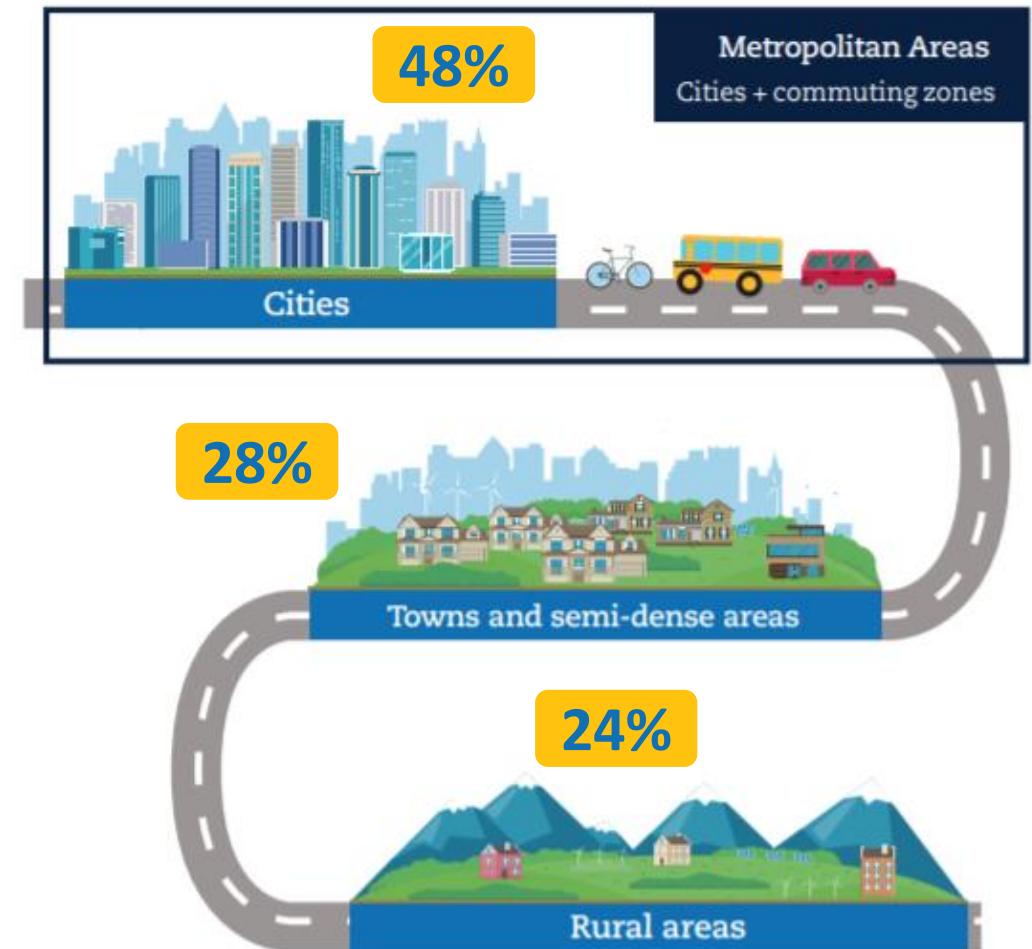
From subnational regions – administrative definition...

- The **OECD Territorial Grid** lists small (over 2300 TL3 regions) and large regions (over 400 TL2 regions)
- Used in **OECD's Regional Database**, covering **80 statistical indicators**
- Demography, economic accounts, labour market, education, social and innovation themes



...to harmonised definitions of cities, towns, villages and rural areas

- **City populations have doubled** over the last 40 years, and are projected to increase further to 55% of the world population by 2050
- OECD/European Commission developed **consistent, global definitions**
- **Degree of Urbanisation** and **Functional Urban Area = cities + commuting zones**



2. Guidance



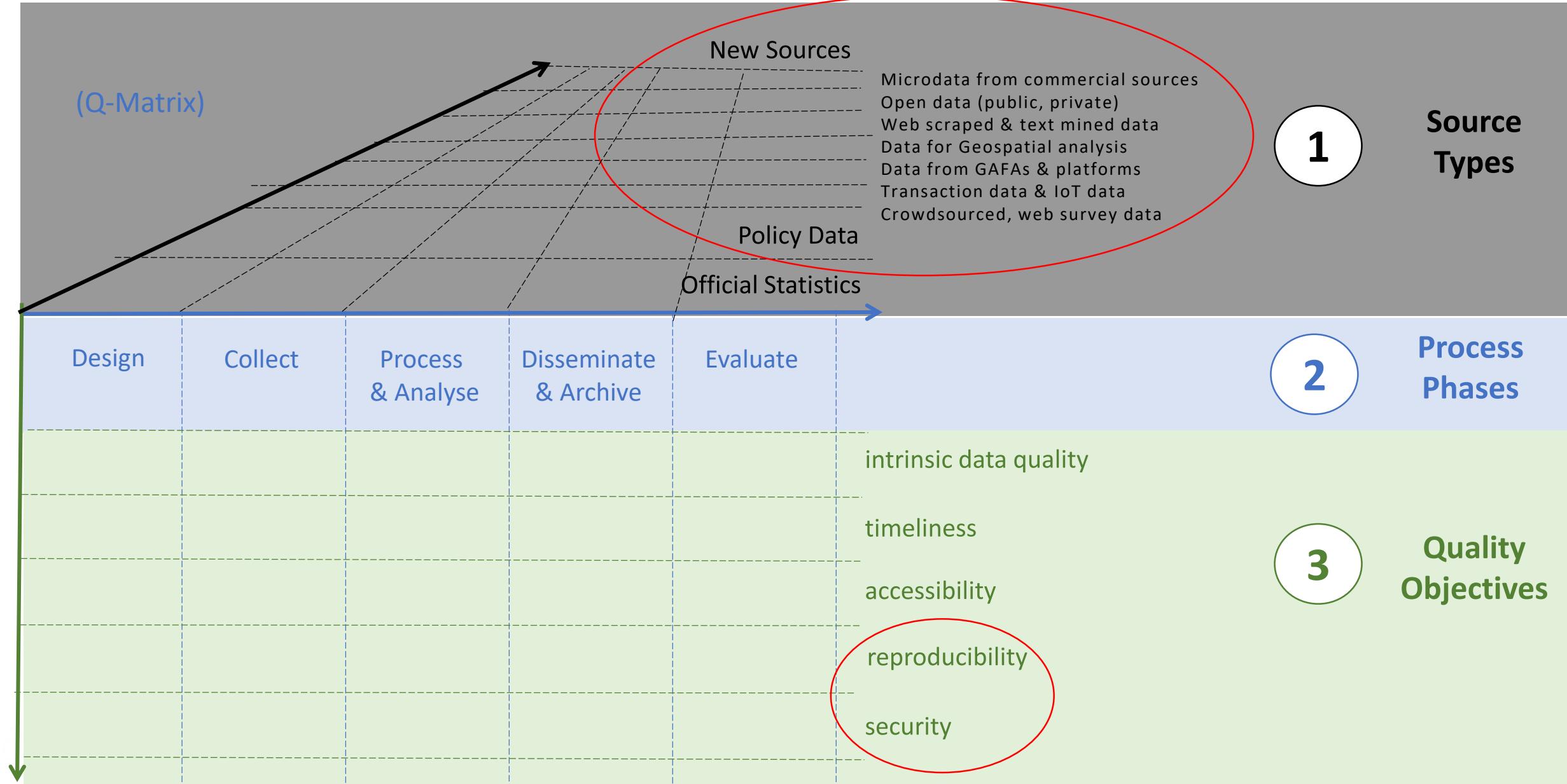


Maintaining a threshold for quality and privacy

- With more granularity comes heightened challenges in maintaining the necessary threshold for **quality and privacy**
- The revised **OECD data quality framework** outlines the dimensions of data quality (**intrinsic data quality, timeliness, accessibility, reproducibility, security**) for all stages of the data lifecycle
- In addition, the OECD has worked to outline a number of **good practices in the use and presentation of sub-national data**



RISK CONTAINMENT (2) - DATA QUALITY FRAMEWORK



3. Co-investment





Internal: OECD Laboratory for Geospatial Analysis



- **Encourages widespread adoption of geospatial analyses** in shaping policies that tackle today's place-based problems
- **Cuts across institutions and scientific programs** to create long-lasting policy and global research networks that integrate geospatial information, statistical data and spatial modelling



Fostering transparency and reproducibility of open data pipelines



Promoting ethical and responsible use of geospatial data



The nexus of sustainability and economic development: utilising geospatial data to integrate environmental and economic planning



Reconciling mismatches between official statistics and earth observation data



Data fusion and integration between public and private sources for nationwide monitoring and diagnosing

External: HLG-MOS

High-Level Group for the Modernisation of Official Statistics

Created by Steven Vale, last modified by Taeke Gijtema on 20 Oct, 2021



Group of committed Chief Statisticians to steer the modernisation of statistical organisations.
CAN, AUS, IRL, ITA, MEX, NLD, NZL, POL, KOR, UK, Eurostat, OECD

<https://statswiki.unece.org/display/hlgbas>

- *Supporting standards*, e.g., Generic Statistical Business Process Model (GSPBM)
- *Capabilities and communications*, e.g., change management, communicating modernisation
- *Blue Sky Thinking Network*, e.g., synthetic data sets guidance and experience
- *Group for Machine Learning*, led by UK ONS: research, skills and sharing resources on ML applications for official statistics



SIS-CC
Statistical Information System
Collaboration Community

eurostat

PARiS 21
Partnership in statistics
for development
in the 21st century

United Nations
Statistics Division

3 Partners

NationalBank
OF BELGIUM
Eurosysteem



Pacific
Community
Communauté
du Pacifique

Stats NZ
Tatauranga Aotearoa



International
Labour
Organization



**Australian
Bureau of
Statistics**

INE
Instituto Nacional de
Estadísticas • Chile

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Eidgenössisches Departement des Innern EDI
Département fédéral de l'intérieur DFI
Dipartimento federale dell'interno DFI
Federal Department of Home Affairs FDHA
Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Federal Statistical Office FSO

المركز الاتحادي
للتنافسية والإحصاء
FEDERAL COMPETITIVENESS
AND STATISTICS CENTRE



احصائيات تونس
STATISTIQUES TUNISIE

UK Data Service

Istat
Istituto Nazionale
di Statistica

STATEC
Luxembourg

OECD

15 Members

External: Development Data Partnership



- **Objective:** promote and facilitate the use of third-party data in research and international development
- **Means:**
 - Legal foundations: template data license agreements
 - Responsible and ethical data use: data governance principles
 - Multi-disciplinary teams (incl. data scientists, sector domains experts, legal counsels, communication specialists...)
 - Centralised and secured IT architecture and processes
 - Web-based Data Partnership Management portal
 - Accessible repositories for derived data products and algorithms





In conclusion



- A whole new world of data for evidence
- Ubiquity of data and pressing problems create enormous demand for trusted and granular evidence
- Spatial disaggregation is probably the single most important aspect of granularity
- NSOs, official statistics and IOs have an important role to play



Thank you!



ANNEX



How do we feed OECD's regional databases?

- Indicators are compiled primarily from **official statistics**
- In cases where the information is not available, the indicators are modelled using a variety of **modelling and GIS techniques**



Official Statistics

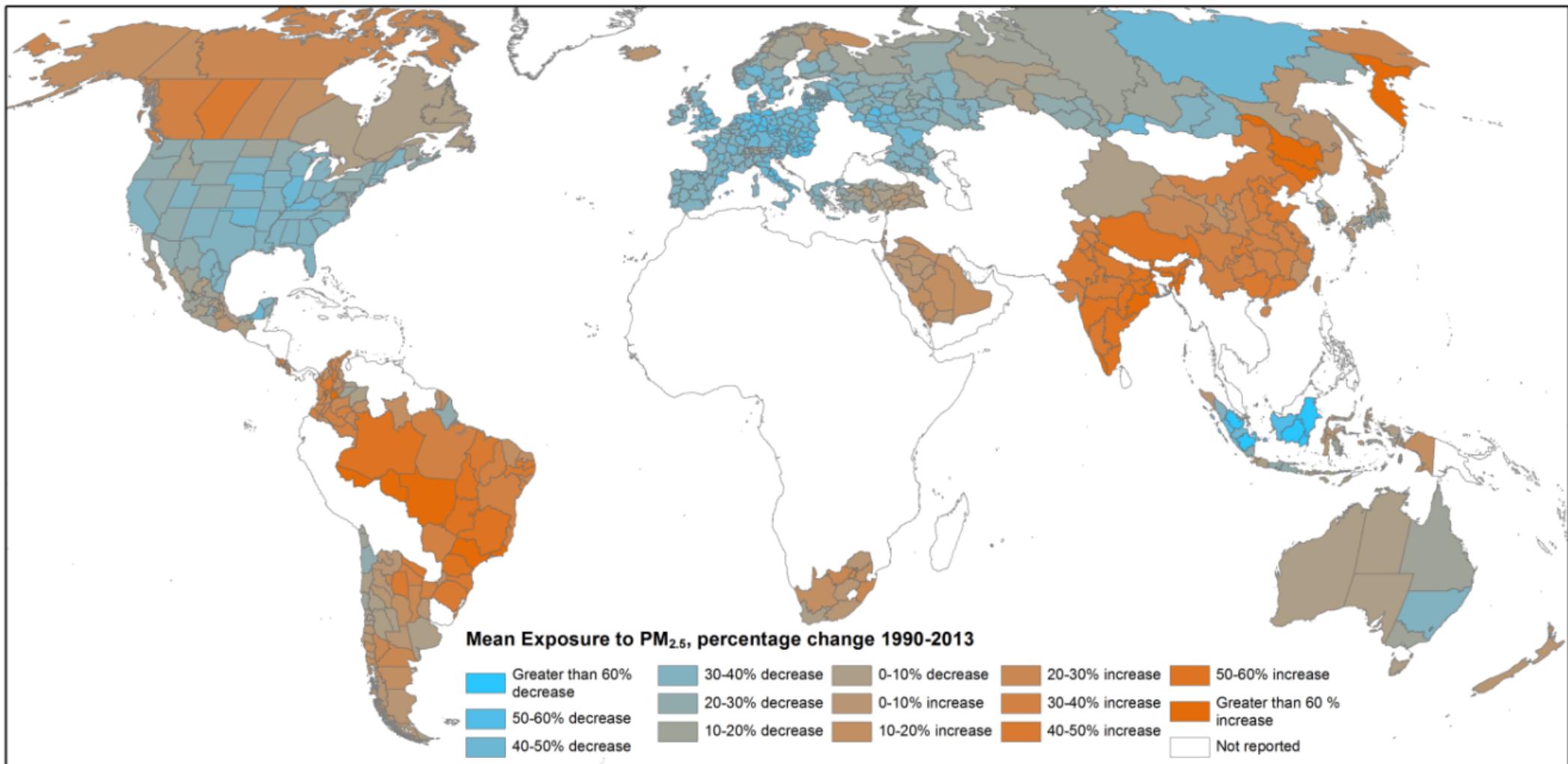
- Obtained from questionnaires filled by delegations
- Obtained from National Statistical Agencies

Modelling techniques

- Derived from official statistics
- Estimates based on unconventional sources (e.g. GIS techniques, big data, web-scraping, satellite imagery, administrative data, etc.)



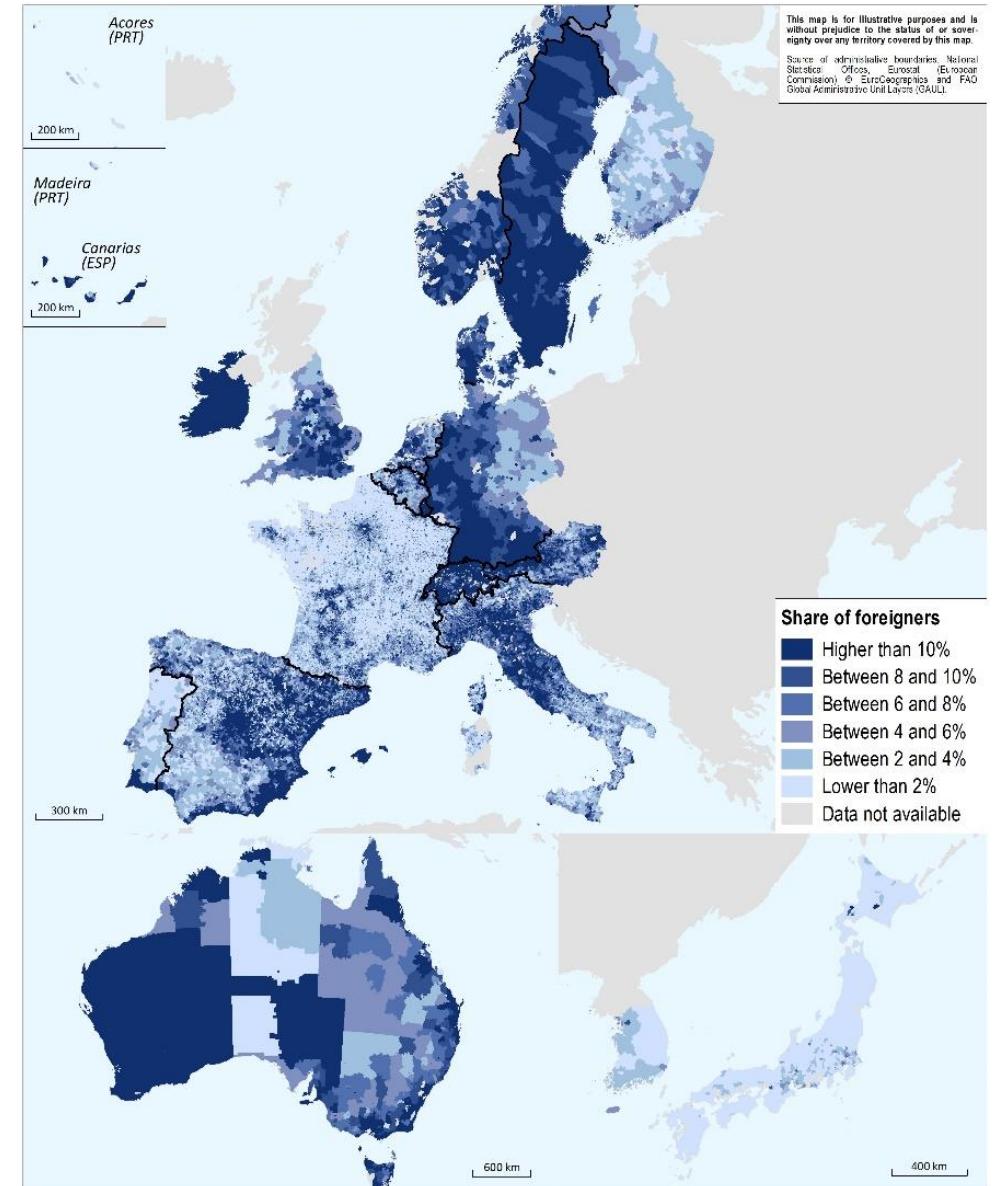
Change in exposure to PM_{2.5} in macro-regions, 1990-2013





Inequalities

- OECD hosted PARIS21 and Colombia's national statistics office are using **satellite imagery to get a more detailed picture of poverty**
- The **OECD Municipal Migration Database** uses data from continuous population registers and surveys to offer detailed geographic information on migration

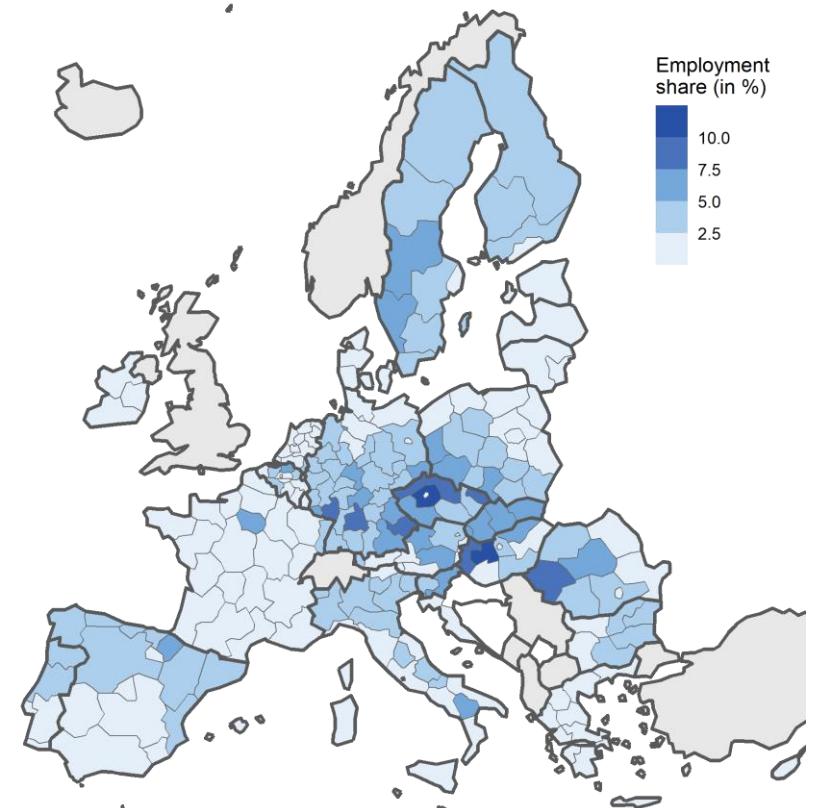




Climate change: The green transition

- The green transition is a global challenge with local implications and requires **transformation across every industry**
- Creation of **new job types, loss of old jobs**, and a **shift in skills** required in many jobs
- This trend is **extremely localised** and we **need good regional data** to best understand risks and opportunities

Employment in high-emission manufacturing sectors



Source: Regional Industrial Transitions to Climate Neutrality: Identifying vulnerable regions.



Land-use in OECD cities

- Land and built-up area are major **environmental and economic factors**
- Monitoring in **near real-time land-use** in OECD functional urban areas
- By using public **Sentinel satellite imagery** data and **Deep Learning models** trained on the Copernicus urban atlas
- Applications: **Urban expansion, Land artificialisation**

Model predictions on Amsterdam

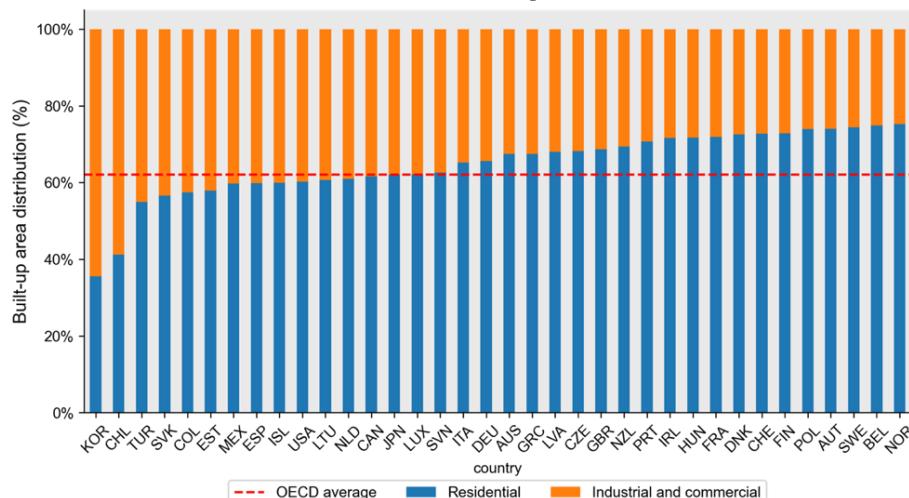
Sentinel-2 RGB



Prediction



Distribution in built-up land use, 2021





OECD good practices for using sub-national data

Interpretability: Provide clear information on sources and methods including geographic scope, territorial units, and geographic structure used for collection and presentation of statistics.

Coherence: Compile subnational data according to internationally agreed standards (e.g. OECD Territorial Grid, Functional Urban Areas) in particular with respect to classifications of territorial units.

Comparability: Sub-national estimates do not always align with national estimates. Clear information on sources and disaggregation methods should be provided to users, otherwise not presented together.

Accuracy: Describe coverage of results and assess the impact of any potential coverage issues and misclassifications. An estimate of the likely impact of non-response on final estimates is recommended.

Reporting by countries: Encourage countries to provide gridded data (data disaggregated by grid cells of consistent size), allowing aggregation and production of indicators for specific policy issues.

Modernisation and statistical governance: Encourage countries to develop or modernise a system of information and statistics at sub-national level integrated to the national statistical system.



OECD visualisation tool for SDGs in regions and cities

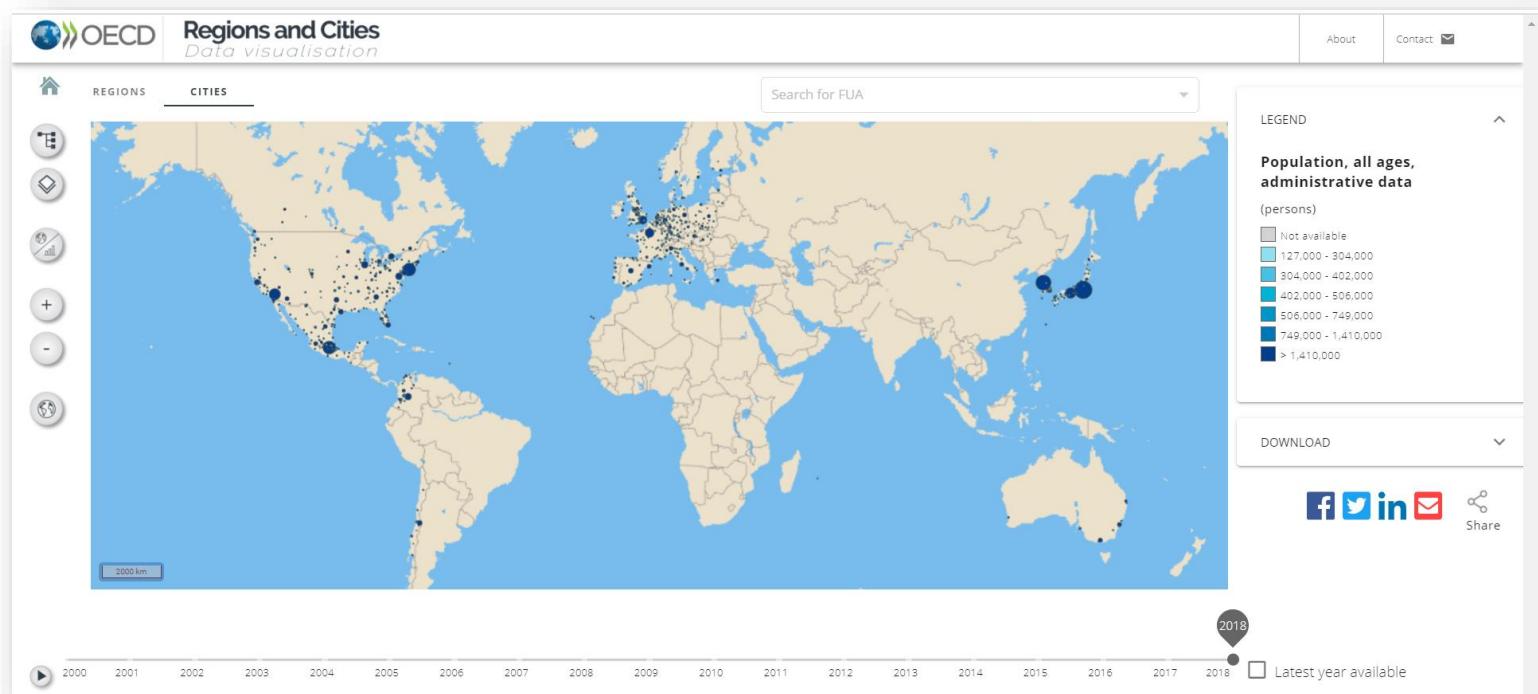
- With its **more than 100 indicators**, this OECD tool measures the **distance towards the SDGs** in **more than 600 regions and 600 cities** of OECD and partner countries

The screenshot shows the homepage of the OECD SDG visualization tool. At the top, there are navigation links for 'Home', 'About', and 'Data'. The OECD logo is in the top right corner. The main title 'Measuring the distance to the SDGs in regions and cities' is displayed, followed by a subtitle explaining the tool's scope: 'With its more than 100 indicators, this OECD tool measures the distance towards the SDGs in more than 600 regions and 600 cities of OECD and partner countries.' Below this is a world map with various regions highlighted by size, indicating their proximity to the SDGs. A zoom control (+/-) is visible on the left side of the map. To the right of the map is a sidebar titled 'Start to explore' with tabs for 'Regions' (which is selected) and 'Cities'. The 'Regions' tab includes a definition of what a region is and a list of regions: 'Australian Capital Territory', 'Flemish Region', 'Wallonia', 'Canada', 'Alberta', 'British Columbia', and 'Manitoba'. The 'Cities' tab is currently inactive.



OECD Regions and Cities Statistical Atlas

- The OECD's Regions and Cities Atlas allows users to browse indicators across **demography, economy, labour, innovation and social and environment** at different levels of granularity





OECD Regional Well-Being web tool

- This tool allows users to measure well-being in their region and compare it with **402 other OECD regions** based on eleven topics central to the quality of our lives

