Introduction to the Data Catalogue

This catalogue brings together the key datasets needed to build and estimate a **Quantitative Spatial Model (QSM)** of cities. QSMs require information on where people live and work, how they move through the city, the availability and cost of land and housing, and the structure of the transport network.

Each dataset included here helps fill one of these critical data needs — from **population and employment counts**, to **road networks and travel times**, to **land availability and building footprints**. We highlight both **global, openly available sources** (useful when local data are scarce) and **specialized datasets** that provide more detailed, city-level insights.

The goal is to give policymakers, researchers, and urban planners a **practical reference**: what data exist, what each source contains, the countries or regions it covers, and how it can be used to calibrate and run QSMs. With this catalogue, users can more easily assemble the inputs needed to analyze and simulate urban policies.

World Bank Employment Dataset

The World Bank produces a range of labor market datasets that provide counts of workers by sector, gender, and sometimes by detailed geographic unit such as districts or municipalities. These data are crucial for understanding **where jobs are located** and how they are distributed across a metropolitan area. For a QSM, employment data are often used to calibrate the **productivity of locations**: if a district has unusually high employment in tradable sectors and offers high wages, the model can infer that it is relatively productive.

Coverage is strongest in **low- and middle-income countries**, where national labor force surveys and Living Standards Measurement Study (LSMS) datasets are integrated. Most Sub-Saharan African countries, South Asian economies, and many in Latin America are included, though the spatial resolution may vary. In richer countries, national statistical agencies typically provide higher-resolution employment counts that can complement or replace the World Bank data.

Kontur Population Dataset

The Kontur dataset provides **global, high-resolution population counts** on a uniform grid (roughly 400 m by 400 m). It blends census data, household surveys, and machine learning techniques that use satellite imagery to predict population density in areas where official data are sparse or outdated.

For QSM estimation, Kontur's main role is to define the **residential side of the spatial equilibrium**: how many people live in each grid cell or administrative unit. Because it is gridded and global, it allows researchers to work consistently across cities in very different countries, even where official small-area population data are unavailable or old.

OpenStreetMap (OSM)

OpenStreetMap is a **crowdsourced, open geographic database** with global coverage. It includes road networks, footpaths, bike lanes, rail and bus lines, as well as points of interest such as schools, hospitals, and shops. In many urban areas it also provides **building footprints**, sometimes with height or use tags (residential, commercial, industrial).

In QSMs, OSM plays a double role. First, it provides the base network needed to construct a **travel time matrix** — the cost or time of commuting between any two points in the city. Second, the building footprints and land use tags can be used to estimate the **amount of existing floorspace** when rental price data are missing, giving a proxy for the supply side of the housing market. Coverage is strongest in Europe, North America, and major cities worldwide, though data quality can be uneven in rural and less mapped regions.

Global Road Inventory Project (GRIP)

The Global Road Inventory Project provides a harmonized, global database of **road networks classified by type and quality**, including highways, primary, secondary, and tertiary roads. Unlike OSM, GRIP includes standardized **speed class estimates**, making it especially valuable for estimating realistic **commuting costs** when local traffic data are not available.

Its coverage is **worldwide** and particularly valuable in **low- and middle-income countries** where official traffic or speed datasets are scarce or inconsistent. In a QSM, GRIP is often used to **augment or validate** the road data from OSM and to calibrate the **time it takes to travel between locations**.

Satellite Data on Land Development

Satellite-based products such as the **Global Human Settlement Layer (GHSL)** or the **European Space Agency's WorldCover** provide information on **built-up areas, land cover, and sometimes building height or volume**. This is crucial for understanding the **supply side** of urban land markets.

By identifying where land is already developed and where it is vacant or agricultural, these datasets help define the **amount of land available for construction** in each spatial unit of the model. Building height or intensity estimates (available for some regions) are particularly useful when floorspace price data are missing, because they allow researchers to infer how much space exists in multi-story structures. GHSL provides global time series back to 1975, which is valuable for studying **urban expansion over time**.

Protected Areas Databases

The **World Database on Protected Areas (WDPA)** and similar global inventories list national parks, reserves, cultural heritage zones, and other legally protected or restricted areas. Including this information in a QSM is important to avoid **overstating the land supply**: not all land in a spatial cell can be developed.

By subtracting protected or legally restricted zones, researchers create a more realistic estimate of the **effective buildable area** available to the construction sector in each location. This adjustment is crucial in cities with large amounts of green belts, wetlands, or heritage zones.

Mobile Phone and App-Based Mobility Data

Anonymized mobility data from cell towers (Call Detail Records) or app-based providers (e.g., Meta Data for Good, SafeGraph, Cuebiq) offer near real-time **origin–destination flows** of people within a city. These data reveal **commuting patterns**, trip lengths, and even the timing of flows during the day.

For a QSM, mobility data are invaluable for **validating or calibrating the travel time matrix**. They provide observed trip distributions and can be used to estimate **mode shares** (driving, transit, walking, biking) indirectly when combined with transport networks. In data-poor settings, they can substitute for expensive household travel surveys and offer far more frequent updates.

How These Datasets Fit Together

To estimate a Quantitative Spatial Model, we need to **map the city into spatial units** and describe both the demand and supply sides of its economy:

- **Residents:** counted using **Kontur population grids** or census microdata.
- **Jobs and productivity:** mapped using **World Bank employment data** or national labor statistics.
- **Commuting costs:** built from **OpenStreetMap** and **GRIP** networks, refined or validated with **mobile phone mobility flows**.
- **Land supply and use:** derived from **satellite imagery** (built-up areas, heights) while subtracting **protected areas**.

- **Floorspace markets:** informed by OSM building footprints or rental data where available; inferred where prices are missing.

These inputs feed the **forward problem** of the model (predicting wages, rents, and densities) and allow **model inversion**— recovering hidden fundamentals such as productivity and amenities from observed patterns.