

Data requirements for SEI Tallinn study regions

To run Realize v2 at its full capacity, Spare requires specific spatially-distributed data from SEI Tallinn's study countries. These should ideally all be at the same resolution (preferably at **500 m resolution or higher**), and in **latitude/longitude point format** (rather than shapefiles). They should cover the **entire country** in a single dataset if possible.

Many countries report socio-economic data in the format of census polygons. These are usually given at a variety of resolutions (e.g. regional, city, municipal, neighborhood), and the higher the resolution, the better. Many countries also provide 'centers of population' or 'census polygon centroids' which attribute a single representative latitude/longitude as the centrepoint of a census polygon. [Dataset repositories such as HDX](#) provide population estimates in CSV or TIFF format, which can be converted to lat/long point datasets.

Each latitude/longitude centroid should ideally be associated with the following information:

- Population count (#)
- Job count (#)
- Proportion of population below poverty level (%)
- Proportion of population educated to high school level or less (%)
- Proportion of zero-car households (%)
- Proportion of population that is poor *and* disabled (%)
- Proportion senior population (%)
- Proportion child/youth population (%)

Additionally, to properly assess the intersection of on-demand transit with wider public transportation, Spare would need, for each given area of interest:

- GTFS data of transit timetabling

Below we provide a more detailed description of each data source, as well as its importance. 'Critical' importance signals Realize v2 cannot be used without the data; 'Important' signals that a lot of Realize's value is unlocked by having the data in question; 'Optional' signals that the data can be omitted, but would provide value to users if it was included.

Data source	Unit	Description	Status
Population count	# (absolute count)	Total number of residents in the census polygon.	Critical
Job count	# (absolute count)	Total number of jobs in the census polygon.	Critical
Proportion of population below poverty level	%	Proportion of residents in the census polygon whose household income is less than the nationally-determined poverty level.	Important
Proportion of population educated to high school level or less	%	Proportion of residents in the census polygon whose <i>highest level of education</i> is high school, or less.	Important
Proportion of zero-car households	%	Proportion of households in the census polygon with no access to their own car.	Important
Proportion of population that is poor <i>and</i> disabled	%	Proportion of residents in the census polygon that both have at least one disability <i>and</i> live below the poverty level.	Important
Proportion senior population	%	Proportion of residents in the census polygon who are aged 65 or more.	Important
Proportion child/youth population	%	Proportion of residents in the census polygon who are aged 18 years or less.	Optional
GTFS data	Standard GTFS zip folder	Full GTFS data (a folder of multiple interrelated .txt files) of most recent transit offering for an area of interest	Important