

# The Geography Database - a data management environment for point-based geocoding and data aggregation in Statistics Sweden

The “Geography database” (GDB) is the main environment for geocoding and industrialised production of geospatial statistics in Statistics Sweden.

In essence, the GDB is an index database comprising a set of tables linking location data objects and geographies together and provides a structured foundation for geocoding and data aggregation. The GDB does not host any geospatial data types, nor does it host any statistical micro data. The GDB is linked with other databases, storing geospatial data and statistical microdata, through database views in order to avoid duplication of data.

The purpose of the setup is to provide a stable and efficient data management environment for geocoding of statistical micro data and aggregation of data using a predefined set of common geographies. The GDB contains three main tables: an index table for cadastral parcels, address locations and for buildings. These index tables are imported from the Real Property Register, which is the master data repository for location data objects. The Real Property Register is a copy of the Cadaster maintained by the NMCA, which is synchronised automatically by means of weekly notifications.

Each object in the index tables in the GDB is spatially related (point-in-polygon) to a pre-defined set of geographies, such as administrative and statistical geographies and grids (in different sizes), stored as polygon geometries in a data repository for geographies. Each location object, such as an address location or a building, can thus be identified by its unique PID and its membership of a specific regional unit or a grid cell.

The index tables are updated every year and an annual version of the GDB is prepared and stored. The unique identifiers of the location data object (cadastral parcels, address locations and buildings) correspond with PIDs used in statistical micro data (such as the Population Register or the Business register). Geocoding statistical micro data can be conducted as a simple join, or creation of database views, between the index tables in the GDB and unit record data. Aggregation of data for any output geographies present in the index tables is then a straightforward process and can be conducted without any involvement of any additional geospatial tools.

In case aggregations are needed for output geographies (custom geographies) not contained in the GDB, the index tables need to be spatially matched against these geographies before aggregation. This can be done directly within SQL server, using spatial syntaxes, or externally using desktop GIS software.

Figure 1: The Geography Database (GDB) and its relations to other data sources

