Merging statistics and geospatial information, 2020 projects - Finland

Statistics Explained

GSDIG – geospatial statistical data integration service; 2020 project; final report 28 October 2022

This article forms part of Eurostat 's statistical report on the *Integration of statistical and geospatial information* .

Problem

Data users have a need for a secured and reusable service enabling the integration of statistical unit-level data and geospatial data.

Objectives

The objectives of service implementation were to:

- achieve point-based statistical data integration across organisations as a reusable multidisciplinary geospatial statistical data integration service (GSDIG) application;
- provide this as a fully open-source application.

Method

The project started with an analysis of user needs/requirements and the specification of a user interface. The target was to produce an indexing-based process to integrate and aggregate statistical data from any domain in areal subdivisions, grids, standard or user-modified geographies and provide an integrated geocoding system and data supply. The project should also enable the reuse of location data for the original unit record data, usually buildings or cadastral parcels, and geographies, with further aggregation capabilities. The four basic functionalities that were identified are the following.

- Data integration in one areal classification.
- Data integration to another areal classification.
- Creating a new areal classification.
- Integration of register unit data and areal classification while respecting statistical disclosure control (privacy protection).
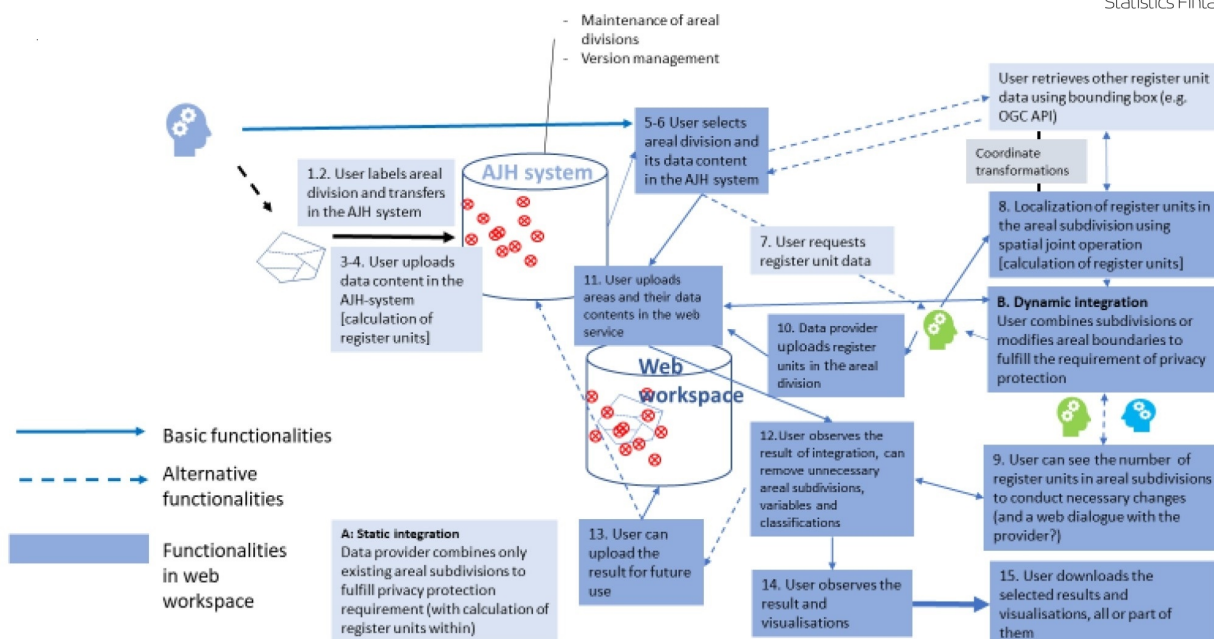
Figure 1: GSDIG user functionalities overview

A preliminary technical specification was developed with the aim that this should be reusable in other organisations and countries. This started by evaluating previous solutions, existing standards and statistical disclosure rules.

- The data linking in areal aggregates is based on areal indexing of record unit data using the location coordinates of record units by implementing a point-based system (of the global statistical geospatial framework / Earth System Grid Federation). The service applied enables a consistent and relatively simple process for users.

- It was concluded that it was too early to implement the SDMX 3.0 specification or the spatiotemporal asset catalog (STAC) for metadata. A custom metadata structure was designed for the project as well as for the search capability of the user interface.

- The table joining service (TJS) standard of the Open Geospatial Consortium (OGC) was implemented.

- The indexing service approach enables a consistent and relatively simple working package and process for users to implement a join while respecting statistical disclosure rules.

Privacy protection issues have a central role in the service developed. For this reason, a modular microservice was developed for record unit calculation (to check for data protection), data linking and data aggregations (on the record unit level). This resulted in the following logical function model of the service, whereby the user:

- selects or imports the data and variables desired and selects, modifies or imports the areal division/classification;

- runs the microservice to count the number of record units in each subarea of the areal subdivision/classification and modifies the area polygons or variables to reach the acceptable counts of record units in each subarea; the number of record units is visible to users only when the requirements of data protection have been reached;

- receives the aggregate data for each subarea from the statistical organisation;

- compiles the results for analysis or further integration.

Results

The technical solution was implemented. Two test cases (called KELA and LUKE) were used to verify the service specifications for implementation.

Methodology

- Final reports (EU Login required)

External links

- Statistics Finland

- National Land Survey of Finland