GeoE3_A4_2_2 Implementation guidelines for including tabular data in the Geospatially Enabled Ecosystem

Co-financed by the Connecting Europe Facility of the European Union

Publishing Information/Date

Revision History

Version	Status	Author	Modifications
0.9	Updated	Pekka Latvala, Panu Muhli	
0.1	First version	Panu Muhli	Initial version

Distribution

Version	Status	Recipient
1.0	Final	Public

Table of Contents

Figures

1. Introduction

The tabular data (e.g., statistical data) is typically distributed via web-based APIs. This document describes the guidelines for integrating tabular data from the PxWeb API with geospatial datasets by using the proof-of-concept data joining service [1], developed in the GeoE3 project. The document contains technical and functional descriptions of the developed service and guidelines for its use.

The data joining process is managed from the web-based user interface. The user interface contains functionalities for retrieving the statistical table metadata and data from the PxWeb API together with functionalities for performing the data joining operation. The data integration process is executed via a prototype application that is based on the OGC API – Joins draft standard [2]. The results of the data joining operation are visualized on the user interface that includes a map-based preview together with links where the joined dataset can be downloaded.

2. Guidelines

This section contains guidelines for the use of the proof-of-concept data joining service, information on the used geospatial datasets and statistical datasets, and technical implementation of the service. This section also contains instructions on the use of the PxWeb API of Statistics Finland.

2.1. Geospatial Data

The following geospatial datasets are used in the GeoE3 data joining service:

- Finnish municipalities (2017-2022)
- Finnish regions (2017-2022)

Each dataset has a yearly version available from the years 2017-2022. The geospatial datasets were loaded into PostgreSQL / PostGIS database, and their metadata were configured to be served via the OGC API – Joins service as collections. The datasets were named in the OGC API – Joins service so that the names contain the area division name in Finnish and the year, to which the dataset applies to, separated by hyphen (example: 'kunnat-2021', for the 2021 municipalities dataset).

The following dataset fields were configured to be key fields in the OGC API – Joins service. The key fields contain unique values, and they can be used for joining the tabular data with the geospatial data.

- Area codes
- Area names in English
- area names in Finnish
- area names in Swedish

2.2. Statistical Data

The PxWeb is an application for publishing statistical tables online. The PxWeb API is a web interface that can be used for querying the PxWeb statistical tables. The statistical table metadata and data from the Statistics Finland's PxWeb API [3] were included to the GeoE3 data joining service.

The database options from the PxWeb API were limited to 'Kokeelliset_tilastot' (experimental statistics) and 'Kuntien_avainluvut' (municipality key figures). The experimental statistics database contains statistics relating to traffic network coverage, game animal collisions, population by type of activity and deaths by week. The municipal key figures database contains various key figures relating to Finnish municipalities, including population, share of Swedish speakers of the population, number of families and employment rate.

2.3. PxWeb API

A list of available databases in the PxWeb API of Statistics Finland can be retrieved by making a HTTP GET request to address:

https://pxweb2.stat.fi/PxWeb/api/v1/{language}

where {language} is the one of the following: 'en' for English, 'fi' for Finnish and 'sv' for Swedish.

A list of topics in a specific statistical database can be retrieved by making a HTTP GET request to address

https://pxweb2.stat.fi/PxWeb/api/v1/{language}/{database-id}

where {language} is the requested language and the {database-id} is the identifier of the statistical database from the response of the statistical databases query.

A list of statistical tables under a topic level can be retrieved by making a HTTP GET request to address:

https://pxweb2.stat.fi/PxWeb/api/v1/{language}/{database-id}/{levels}

where {language} is the requested language, {database-id} is the identifier of the statistical database and {levels} is the identifier of a topic.

The metadata of a specific statistical table can be retrieved by making a HTTP GET request to address

https://pxweb2.stat.fi/PxWeb/api/v1/{language}/{database-id}/{levels}/{table-id}

where {language} is the requested language, {database-id} is the identifier of the statistical database, {levels} is the identifier of the topic and {table-id} is the identifier of the statistical table.

Some of the statistical topics in the PxWeb API may contain additional statistics level that is related to a topic. A list of statistics levels can be retrieved by making a HTTP GET request to address (not used in the poc service):

https://pxweb2.stat.fi/PxWeb/api/v1/{language}/{database-id}/{levels}

where {language} is the requested language, {database-id} is the identifier of the statistical database and {levels} is the identifier of a topic.

A list of available statistical tables in a statistics level can be retrieved by making a HTTP GET request to address (not used in the poc service):

https://pxweb2.stat.fi/PxWeb/api/v1/{language}/{database-id}/{levels}/{levels}

where {language} is the requested language, {database-id} is the identifier of the statistical database, the first {levels} is the identifier of the topic and the second {levels} is the identifier of the statistics level.

The metadata of a specific table under a statistics level can be retrieved by making a HTTP GET request to address (not used in the poc service):

https://pxweb2.stat.fi/PxWeb/api/v1/{language}/{database-id}/{levels}/{levels}/{table-id}

where {language} is the requested language, {database-id} is the identifier of the statistical database, the first {levels} is the identifier of the topic, the second {levels} is the identifier of the statistics level and {table-id} is the identifier of a statistical table.

A more detailed instructions for the use of the Statistics Finland's PxWeb API can be found from the PxWeb API Help document [4].

2.4. User Interface

2.4.1. Selecting a Statistical Table

The use of the data joining service begins with the process of the user selecting a statistical table from the user interface. All metadata relating to the statistical tables are retrieved from the PxWeb API of the Statistics Finland in the JSON format.

In the user interface, the process of retrieving a statistical table is executed by selecting values from various select lists and retrieving metadata from the PxWeb API based on the user's selections (Figure 1). The use of the service starts with the user selecting a language. The language selection influences the language of the whole user interface and the language of the retrieved statistical data. The use of the service continues by the user selecting a statistical database, a statistical topic, and a statistical table.

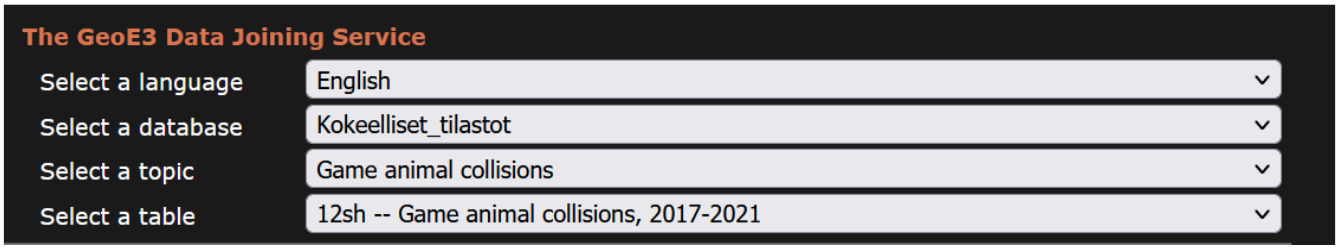

Figure 1: Selecting a Statistical Table in the User Interface.

2.4.2. Selecting Statistical Table Variables

After the user has selected a statistical table and its metadata have been retrieved from the PxWeb API, the contents of the metadata are listed in the user interface (Figure 2). The metadata contains a title and a list of statistical variables with their selectable values.

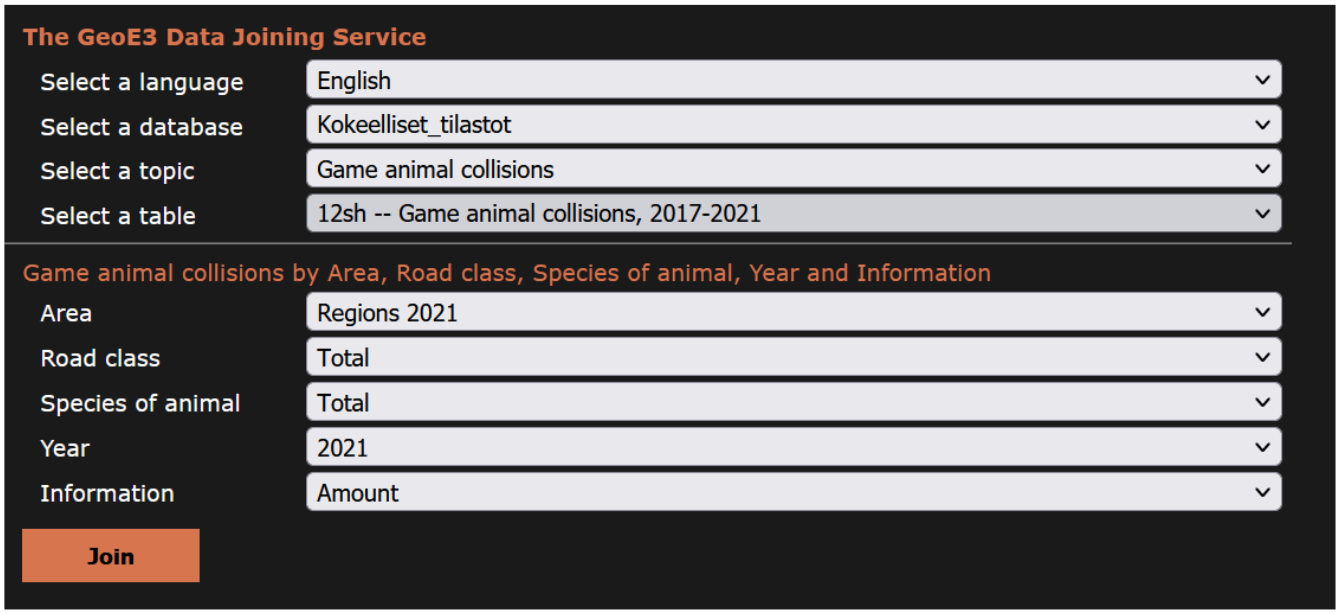

Figure 2: Selecting Values for the Statistical Variables of the Selected Statistical Table in the User Interface.

The values of the statistical variables determine the statistical data that can be retrieved from the PxWeb API. A 'map' element was added to the area variable response of the statistical table metadata for indicating the yearly version of the area division, to which the statistical data corresponds to. The options for the area selection were limited to municipalities and regions that correspond to the geospatial datasets that are available in the service. The individual area codes that belong to those area divisions are grouped under these options, so that when a specific area division is selected, all area codes that belong to that group are selected.

The OGC API – Joins service requires that the values that are related to an individual geographical area unit, are presented in a single row in the CSV file. If multiple values are selected for other statistical variables that area code variable, the PxWeb API of Statistics Finland returns the values in multiple rows. Therefore, the value selection for other statistical variables was limited to a single value.

2.4.3. Retrieving Statistical Data

After the user has selected values for the statistical variables and clicked the Join-button, the statistical data are retrieved via a HTTP POST request from the PxWeb API. The request contains information on the selected values for each statistical variable and information on the request response format. The statistical data are requested in the CSV format.

In some cases, the key values of the returned statistical areas don't' match with the key values used in geospatial datasets and need to be processed to create matching key values before the data joining query can be executed.

2.4.4. Data Joining

After the potential editing of the key values, a data joining request is made to the OGC API - Joins service that joins the statistical data with the geospatial dataset. The geospatial dataset used in the operation can be determined automatically, based on the selected area division and the value of the 'map' variable that indicates the yearly version of the dataset.

The response of the OGC API – Joins service is parsed in the user interface and the results page is created from the response's contents (Figure 3). The results page contains general information about the join, information on the

successfulness of the join operation, map preview, download links for the joined dataset in various formats and a map legend. The map preview was created with the Oskari application's RPC functionality [5]. The map styling was created by classifying the data values into four classes.

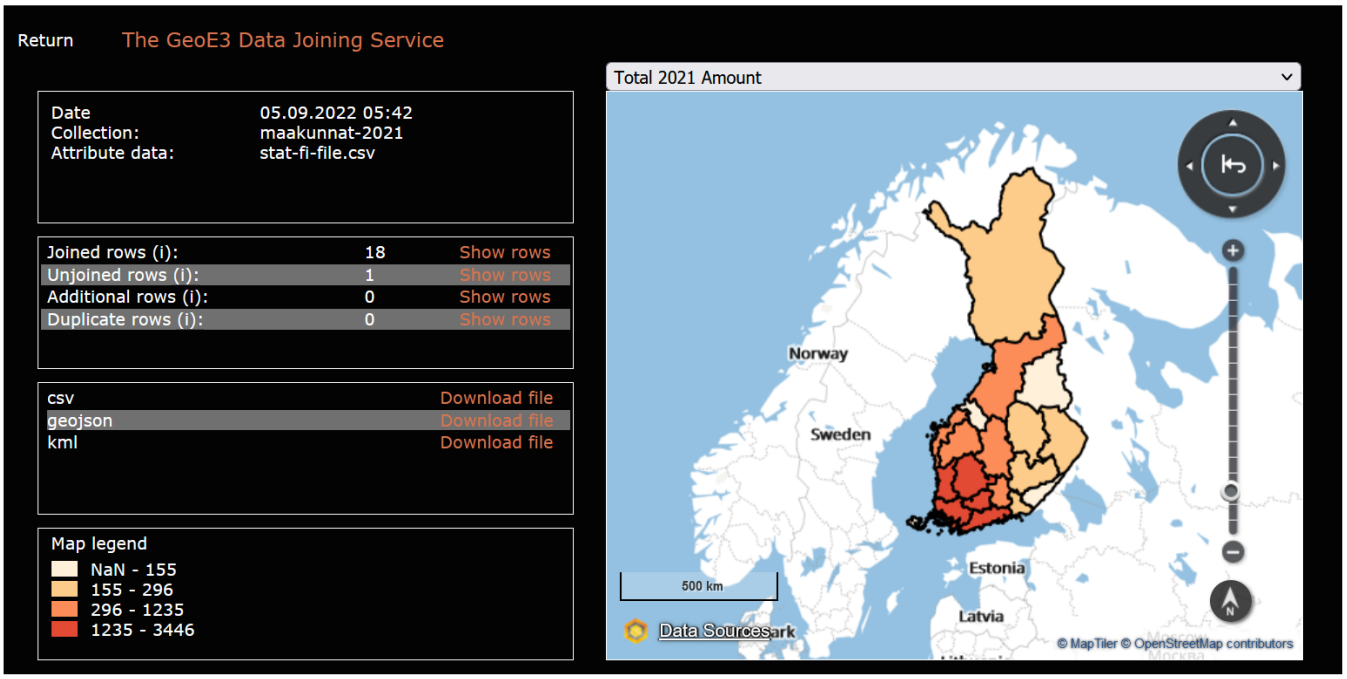

Figure 3: The Results Page of the User Interface.3

2.5. Service Architecture

The architecture of the data joining service is presented in (Figure 4). The user interface is used for managing the entire data joining process. The Oskari application is used for producing the map preview functionality and it is maintained by the National Land Survey of Finland. The PxWeb API is used for requesting the statistical metadata and data and it is maintained by Statistics Finland.

The geospatial datasets are hosted in the PostgreSQL / PostGIS database. The metadata of the geospatial datasets are published via the OGC API – Joins Service. The joined data outputs are published from the database to the GeoServer application where they can be retrieved in several output formats.

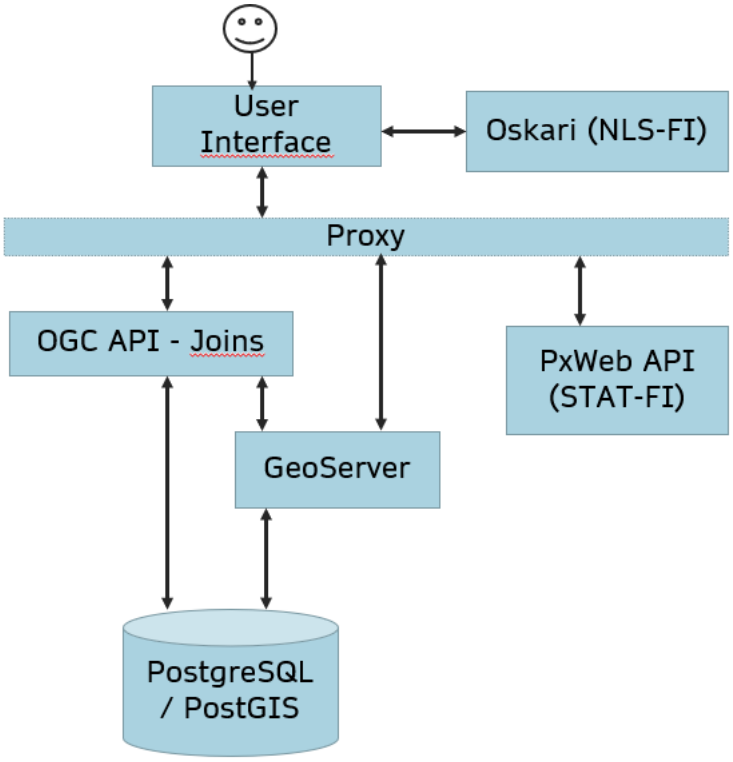

Figure 4: Architecture of the Data Joining Service.

2.6. Hyperlinks

[1]: The URL of the developed data joining service: https://vm3897.kaj.pouta.csc.fi/djs/

[2]: The draft for the OGC API - Joins standard in GitHub: https://github.com/opengeospatial/ogcapi-joins

[3]: The URL of the PxWeb API of Statistics Finland, used in the data joining service: https://pxweb2.stat.fi/PxWeb/api/v1/en/

[4]: Statistics Finland's PxWeb API Help document: https://www.stat.fi/static/media/uploads/org_en/avoindata/px-web_api-help.pdf

[5]: Oskari.org web page: https://oskari.org