Vector Tile Table Join Service (VTJS) - cartographic support for the communication of AI results

Markus Jobst [a,b] *, Sharon Chawanji [a], Georg Gartner [a]

[a] Departement of Geodesy and Geoinformation at the Vienna University of Technology, markus@jobstmedia.at,
Sharon Chuwanji, schawanji@gmail.com, Georg Gartner, georg.gartner@tuwien.ac.at

[b] Federal Office of Metrology and Surveying, markus.jobst@bev.gv.at

* Corresponding author

Keywords: Vector Tile Cache, Table Join Service, Basemap, Communication

Abstract:

The growth of the Web has been characterised by publication and establishment of linked open data on the Web powered by technologies such as RDF and SPARQL. Much of this data are embedded with geographic content that links them to a position on Earth's surface. This data can be extracted from sources such as social media, remote sensing and government portals. The abundance of the data on the Web create a huge pool of data. By employing geospatial artificial intelligence (geoAI), knowledge can be extracted from the spatial big data repositories that can answer questions to real world phenomena. At least the communication of spatial related phenomena will ask for a graphical representation, a map, which can be integrated with the results of big data analysis.

Spatial Data logistics
Standardized interfaces in order to access spatial features (aka geometry objects or geographies) have been established in the recent years. One of the most popular is OGC WFS (web feature service), which is in a renewal process to establish the OGC API features (OGC 2021). The main aim of theses interfaces is to provide unambiguous access to the geographical feature, which is needed for the data integration process of thematic data and therefore create the fundament of map representation. Latest data integration implementations on the basis of WFS (Yu et al 2020) demonstrate that this service-related access and extraction of a database also consists of some drawbacks in terms of spatial data logistics. Whenever a digital near real-time map is created on the basis of spatial data services, spatial data logistics will help that the requested content is complete according to the request, of appropriate quality and delivered in an acceptable period of time. A WFS may not fulfil these aspects as part of its design: (1) requested content may not be complete because too many objects were requested or the network connection has been injured. (2) The WFS delivers any quality that is stored in the database. It does not adopt the geometric quality to the scale of the request and therefore overflowingly uses bandwidth and even could negatively influence processing times. (3) Too detailed geometries, embedded tags and the amount of "unindexed" geometries result in long response times, which are not accepted at the user side, when consuming a map. In order to overcome these disadvantages for the presentation tier (map), vector tile caches (de Beukelaar 2018) have been widely established. Thus a pre-rendered collection of vector tiles from a spatial data repository can be accessed, without permanent stressing the database. The pre-rendered collection follows visualization rules of a standard display resolutions concerning geometric density. From a spatial data logistics point of view, a vector tile cache seems to be an optimal source of geographies for a real-time data-integration of "big data extractions".

Real-time data integration support
Maps with its topographic information and the power of spatial structuring help to communicate facts of space. Whenever these spatial framework data are in use for AI results, it may deal with high access rates (requests per minute) in order to provide appropriate fundamental information for the geospatial knowledge creation. Therefore caching mechanisms, like vector tile caches, may help to enhance performance and to reduce permanent DB access. In order to communicate the results of geospatial AI and machine learning, persistent relations between a feature ID, a geographic object identifier and the cache is needed for any further processing upon the Vector Tile Cache (VTC) and its styling. In order to establish real-time data integration on the basis of VTC the set of applicable and persistent ID´s (PID) needs to be known. Any failure of the PID will lead to inconsistent and wrong data integrations and therefore disrupt the resulting map. This is also true for "non-matching" timestamps, which result in loss of data for the resulting spatial information.

The importance of a standardized feature ID enrichment

The requirement for persistent and time-consistent ID´s (PID) is well distributed in the statistical community (Golodoniuc 2017), which extensively makes use of data integration for statistical maps. The topic of PID is anchored for any spatial data infrastructure, thematic data community (data space, Jarke 2017) and even at the level of United Nations (Haldorson and Moström 2018). Too often the definitions of PID are created in their data silos (data spaces) and cannot reach out to other communities. One possible solution are open registries or register networks for PID and their naming schemas, so that changes over time are transparent and any enrichment of fundamental geographies with PID from other data spaces could be standardized. The topic of PID was inevitable for the vector tile table join service (VTJS), but could not be solved in the run of this contribution.

A general architecture of VTJS

The main role of the vector tile cache (VTC) is discharging the database connection and needed to enhance performance of the spatial data service (WMTS of VTC, the web map tile service of the vector tiles). This means that a VTC is at the outermost border of the IT infrastructure that is accessed by the map application or user client. With this condition, a VTJS has to be even nearer to the application or user as it uses the VTC for data integration. The VTJS is a spatial data service that integrates results from big data analysis or SPARQL queries, in their simplest form as CSV (comma separated values), with fundamental geographies provided as vector tiles. In fact the VTJS relates features in the VTC with PID´s in the thematic tables. Both access points have to be specified in the request (Schawanji 2020).

What is the role of the VTC against general spatial framework data?

Spatial framework data are repositories of fundamental geographies (UN-GGIM 2019), which are used for data integration. In the VTC these spatial framework data are transformed into "display-friendly" tiles according to zoom levels and the tile specification. The vector tile cache is seeded and updated with actual spatial information from the backend datastore. A preseeding or on-demand updating is part of the VTC configuration (e.g. in geoserver). This means that the VTC does in general not contain different spatial information then the spatial data in the data store, although the VTC transformation into display tiles (Balog 2017) is a irreversible processing for the visual representation.

The vector tile table join service (VTJS) is part of the visualization tier. It delivers thematically enriched fundamental geographies for direct use in (thematic) maps. The data integration of the VTJS georeferences tables and fundamental geographies in the VTC with their PID. As result a performant thematic map layer with flexible styling is brought into map applications and transmit facts that were derived from big data analysis, SPARQL endpoints or AI.

When it comes to the communication of spatial information, the role of cartography is clear. The simplification methods within the VTC help to configure a clear and unambiguous perception of geographies. The open and demand-based styling of the VTC further enhances the success of these VTC maps (Linwood 2020). Enabling a VTJS adds to the cartographic advantages of VTC for a map representation on screen. It adds a near real-time integration of thematic data or a georeferencing of thematic data with fundamental geographies as long as a persistent ID can be identified. Any embedding of additional attributes expands cartographic expressiveness on one hand, but could also help to expand fundamental geographies with additional PID´s on the other hand. The VTJS is applied at the last step of spatial information consumption. Therefore the VTJS is a non permanent intervention on the data, but results in enriched and advanced thematic maps.