AGILE: GIScience Series, 3, 34, 2022. https://doi.org/10.5194/agile-giss-3-34-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

A Geospatial Dashboard Prototype for Evaluating Spatial Datasets by using Semantic Data Concepts and Open Source Libraries

Heiko Figgemeier[1], Arne Rümmler[1], and Christin Henzen[1]

[1] Chair of Geoinformatics, Technische Universität Dresden, Germany

Correspondence: Heiko Figgemeier (heiko.figgemeier@tu-dresden.de)

Abstract. In today's research data management, experts discuss datasets to be FAIR, as they should become findable, accessible, interoperable and reusable (Lacagnia et al. 2021). In recent years, quality information and provenance information as well as dataset's general metadata have become important aspects to evaluate a dataset's fitness for use. In order to capture and process this meta-information in a systematic way, users need frameworks and meaningful user interfaces that allow them to interact with the information and to visualize them. Therefore, we provide a user-friendly and interactive geodashboard implementation as first prototype that supports the evaluation of spatial datasets with linked widgets by applying semantic concepts and using open source libraries.

Keywords. geodashboard, spatial data quality, provenance, metadata

1 Interface and Methodology

This poster presents a geodashboard prototype [1] that combines the components (1) Provenance, (2) Spatial Data Quality as well as (3) General Metadata in one user interface, built on the conceptual work of the geodashboard concept by Figgemeier et al. (2021). Provenance information are essential in Earth System Sciences, since they facilitate reviewing the genesis of research datasets and the involved processes (Magagna 2020). Moreover, spatial data quality information can be used to evaluate dataset's fitness for use for a specific purpose from the data users' perspective and thus, serve as an essential criterion in spatial dataset selection (Devillers 2010, RfII 2020).

In our case, the dashboard acts as a user interface component supporting the evaluation of datasets, which are managed in a metadata catalog, by utilizing dataset provenance and quality information. The required information is retrieved via SPARQL request against a triplestore. The triplestore manages the copied and semantic enabled metadata of the catalog. In the current version, the metadata is implemented by applying a GeoDCAT[2] profile that focuses provenance and quality. Provenance information on datasets and related processes are modelled with the provenance ontology (PROV-O[3]); geodata quality with data quality vocabulary (DQV[4]) by using quality metrics from an openly available register[5].

The dashboard prototype includes four different components that act as flexible and linked widgets (see Figure 1). When selecting a dataset in one of the dashboard's widgets, the contents of the other widgets will be automatically adapted.

The provenance graph serves as entry point for the evaluation of spatial datasets. It contains all datasets and processes that were involved in the selected dataset's creation. The visualization of the provenance graph is derived from a general-purpose RDF visualization application that is built upon the P5JS[6] library and tailored to PROV-O conform provenance information, using the

1 https://github.com/GeoinformationSystems/Geodashboard, May 11, 2022
[2] https://doi.org/10.5281/zenodo.4916698, April 22, 2022
[3] https://www.w3.org/TR/prov-o/, April 22, 2022
[4] https://www.w3.org/TR/vocab-dqv/, April 22, 2022
[5] https://geokur-dmp.geo.tu-dresden.de/quality-register, April 22, 2022
[6] https://p5js.org/, April 22, 2022

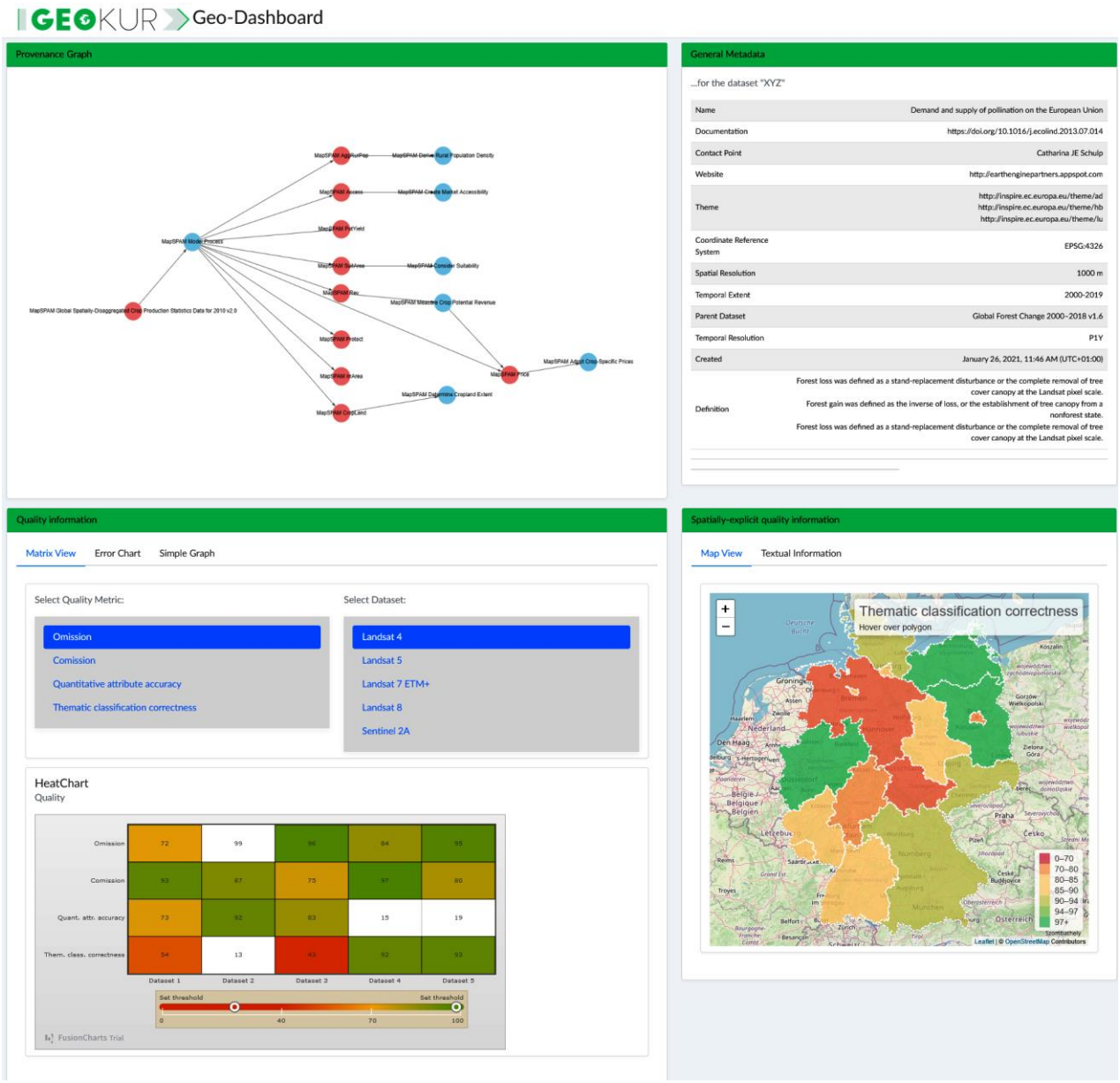

Figure 1. User interface of the geodashboard with the different widgets (1) Provenance Graph, (2) General Metadata, (3) Quality information and (4) MapView

common Starting Point Terms "Entities" and "Activities"[7].

The metadata widget shows the general metadata of the selected dataset that is provided via triplestore. This includes for example dataset provider, resolution, or themes.

A data user's decision-making is supported by meaningful metadata visualizations that are customizable and intuitive to use; e.g., widget 3 (see Figure 1, bottom-left) that facilitates the inspection of geodata quality. For this purpose, we provide three different quality visualizations in the dashboard: (1) A simple graph displays absolute quality metrics. The graph is useful for answering questions such as: How has a quality metric changed during the creation process? What is the difference of a quality metric for selected datasets? Since quality measures of spatial datasets often include uncertainties or value ranges, an (2) error chart allows the exploration of such uncertainties. Moreover, (3) a matrix view displays the quality information as a compact colored table and allows setting user-defined class boundaries. Here, all required metadata and information are retrieved from the

[7] https://www.w3.org/TR/prov-o/#description-starting-point-terms, May 11, 2022

triplestore and visualized with adapted diagrams from the open source library *Plotly*[8].

The fourth widget contains a map view. Here, a map serves as an information-bearing interactive user interface element. The map functionality is integrated by using the open source library Leaflet.js[9]. In the map, either the extents of the datasets (as preview) or spatially-explicit geodata quality information are displayed. If geodata quality is available at sub-dataset level, it can be resolved to the according polygons, e.g., official boundaries or spatial extents of biomes or climate zones. If necessary, a textual view of the spatially-variant quality information is also provided.

2 Conclusions and future work

The dashboard facilitates the inspection and evaluation of complex quality information and related information in a user-friendly interactive UI component. The dashboard requires an existing data basis for quality information and metadata, implemented as triplestore and adhere to a predefined metadata schema. The functionalities of the dashboard could be extended by further visualizations of more complex information as well as temporal and thematic variant quality information. Additionally, in the current version, the dashboard only queries a single curated triplestore. Further developments could open up the dashboard to the Linked Open Data Cloud by using federated queries. This requires handling of missing information (e.g., PROV-O conform dataset provenance) and/or schema conversions, but makes countless research datasets from different communities accessible.

References

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P and Shi, W.: Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities, Transactions in GIS, 14 (4), 387–400. https://doi.org/10.1111/j.1467-9671.2010.01212.x, 2010.

Magagna, B., Goldfarb, D., Martin, P., Atkinson, M., Koulouzis, S. and Zhao, Z.: Data Provenance. In: Zhiming Zhao und Margareta Hellström: Towards Interoperable Research Infrastructures for Environmental and Earth Sciences, Bd. 12003. Springer International Publishing, 208–225, 2020.

RfII – Rat für Informationsinfrastrukturen: The Data Quality Challenge - Recommendations for Sustainable Research in the Digital Turn. Göttingen, 2020.

Figgemeier, H., Henzen, C. and Rümmler, A.: A Geo-Dashboard Concept for the Interactively Linked Visualization of Provenance and Data Quality for Geospatial Datasets. In: AGILE GIScience Ser., 2, 1–8. https://doi.org/10.5194/agile-giss-2-25-2021, 2021.

Lacagnina, C., Ge, P., Ivanova, I., Downs, R., Ramapriyan, H., Moroni, D., Wei, Y., Wyborn, L., Jones, D. and Ganske, A.: International Community Guidelines for Sharing and Reusing Quality Information of Individual Earth Science Datasets, https://doi.org/10.1002/essoar.10508601.1, 2021.

[8] https://plotly.com/javascript/, April 22, 2022
[9] https://leafletjs.com/, April 22, 2022