

# The GeoLOD Catalog and Recommender for Spatial Linked Data: A Demonstration

Vasilis Kopsachilis<sup>a</sup>, Michail Vaitis<sup>a</sup>

<sup>a</sup>University of the Aegean, Lofos Panepistimiou, Mytilene, 81100, Greece

## Abstract

This is a demo paper aiming to demonstrate GeoLOD main features. GeoLOD is a web catalog of Linked Open Cloud (LOD) spatial datasets and classes and a recommender for relevant datasets and classes for link discovery processes. GeoLOD catalog parses, maintains and generates metadata about datasets and classes provided by SPARQL endpoints that contain georeferenced point instances. It offers text and map-based dataset search functionality and dataset descriptions in GeoVoID, a spatial dataset metadata template that extends VoID. GeoLOD recommender pre-computes and maintains, for all classes in the catalog, ranked lists of relevant classes for link discovery. In addition, the on-the-fly recommender allows users to define an uncatalogued SPARQL endpoint, a GeoJSON file or a Shapefile and get class recommendations in real time. Generated recommendations can be automatically exported in Silk and LIMES configuration files in order to be used for a link discovery task. GeoLOD also offers a REST API for the automated execution of the above tasks.

## Keywords

Spatial Linked Data, Dataset Catalog, Class Recommender

## 1. Introduction

LOD cloud expansion offers more options to users, but at the same time, complicates the discovery of data sources that meet user needs. While several tools address linked data search and discovery problems [1, 2, 3], none is specialized in geographical datasets exploration. To this end, we have developed GeoLOD, a web catalog and a dataset recommender, that on the one hand focuses on LOD spatial datasets and on the other hand exploits their spatial characteristics to aid dataset exploration. GeoLOD addresses the following scenarios:

1. A user searches for datasets that cover a specific geographical area (e.g., a country);
2. A linked data publisher searches for datasets containing georeferenced information in order to georeference their data;
3. A linked data publisher searches for datasets that contain related instances to their own datasets in order to establish links between instances; and
4. A geographical information systems (GIS) professional wants to enrich their spatial data with linked data.

---


*GeoLD 2022: 5th International Workshop on Geospatial Linked Data co-located with ESWC, May 30 2022, Hersonissos, Greece*

✉ vkopsachilis@geo.aegean.gr (V. Kopsachilis); vaitis@aegean.gr (M. Vaitis)

🆔 0000-0003-3824-3932 (V. Kopsachilis); 0000-0002-1269-6071 (M. Vaitis)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

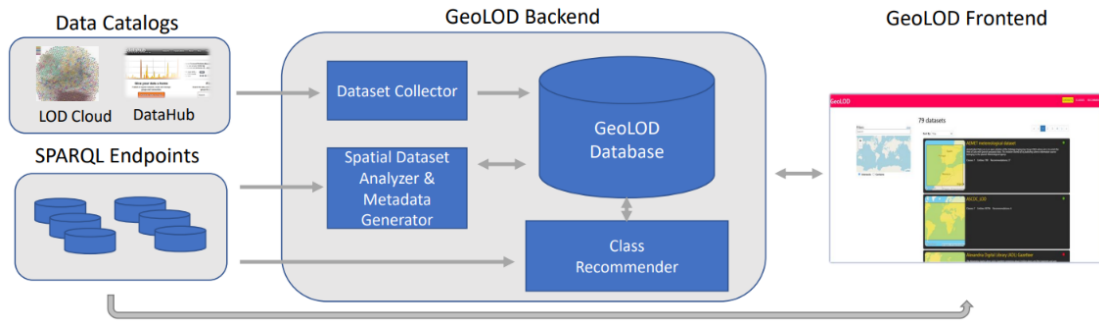
The GeoLOD catalog parses the LOD cloud [1] and the Datahub [2] in order to identify spatial datasets and their spatial classes, that is, datasets and classes that contain georeferenced point instances. Then, it extracts their metadata and generates additional metadata that capture spatial aspects, such as their bounding box, number of georeferenced instances and associated spatial vocabularies, and exposes them in GeoVoID, a vocabulary that extends VoID [4]. The catalog allows access to the lists of linked data spatial datasets and classes through a user interface and provides text and map-based search functionality, thus addressing scenarios 1 and 2.

The GeoLOD recommender generates ranked lists of spatial datasets and classes that may contain related instances with a given dataset or class, so as to be further examined in link discovery processes for the establishment of owl:sameAs links or related links. The recommendation algorithm, fully presented in [5], is based on the hypothesis that “pairs of classes whose instances present similar spatial distribution are more related than pairs of classes whose instances present dissimilar spatial distribution, in the sense that the former are more likely to contain semantically related instances”, and thus are better candidates to be recommended for a link discovery process. In a nutshell, the algorithm builds and stores summaries for all identified spatial classes that capture their spatial extent and index the location of their spatial instances on a QuadTree, and then apply metrics that compute the similarity of the class summaries. Pairs of classes whose summaries present high similarity are recommended for link discovery. GeoLOD maintains recommendations for each class in the catalog and allows the exploration for related classes and datasets through the user interface. It also generates pre-configured Silk [6] and LIMEs [7] configuration files for a selected pair of recommended classes that can be directly used for link discovery. Additionally, it allows on-the-fly recommendations for classes provided through a user-defined SPARQL endpoint, not listed in the catalog, and for GeoJSON and Shapefile datasets, which are typical GIS file formats, thus addressing scenarios 3 and 4.

## 2. Related Work

GeoLOD is related with two categories of systems: (a) dataset catalogs and (b) dataset recommenders for link discovery. Dataset catalogs provide single entry points for discovering available linked-data datasets and the most prominent examples are arguably the LOD cloud and the Datahub. Other data catalogs include the LODAtlas [3], which provide faceted dataset exploration and dataset statistics, LOD Laundromat [8], which provide search services for “cleaned” datasets, and SPARQLES [9], which is an application for SPARQL endpoints monitoring. Compared to the above systems, GeoLOD is the first catalog of exclusively spatial datasets and classes that computes and maintains metadata about their spatial characteristics and offers services for map-based dataset search. These metadata are exposed in GeoVoID, an RDF vocabulary for expressing spatial metadata and statistics of datasets that was designed during GeoLOD development and extends VoID.

Dataset recommenders for link discovery automatically recommend triplesets (e.g., datasets or classes) that may contain related instances to a given tripleset in order to be used as input in a link discovery process. Although several methodologies have been proposed to address the problem, only few are implemented in tools and web applications, such as TRTML [10] and LODSynthesis [11]. The main difference between our work and the above is that their recommendation



**Figure 1:** GeoLOD Architecture.

algorithms are based on information about existing links between datasets, while ours is based on the similarity of the spatial distribution of datasets and classes instances. GeoLOD novelties also include: (a) a pre-computed list of class recommendations for link discovery for each identified spatial class in the web of data, (b) on-the-fly class recommendation for new SPARQL endpoints and spatial datasets in GIS formats (e.g., ESRI shapefiles) and (c) the export of a recommended pair of classes to Silk and LINES configuration files for direct use in a link discovery process.

### 3. Implementation Methods

Figure 1 depicts the overall architecture of the GeoLOD system, which consists of a backend and a frontend module. At the backend, the *Dataset Collector* component parses the LOD cloud and the Datahub catalogs to discover datasets provided through SPARQL endpoints. Specifically, LOD cloud exposes its list of datasets and their metadata in an online JSON file and Datahub using the CKAN API. The component extracts basic dataset metadata, such as their title and endpoint URL, and stores them to the *GeoLOD database*. Then, the *Spatial Dataset Analyzer & Metadata Generator* component sends ASK SPARQL queries to identify spatial datasets, that is, datasets that contain terms from well-known spatial vocabularies, such as W3C Basic Geo and GeoSPARQL, for capturing the geographic position of instances. The component also sends SELECT SPARQL queries to retrieve the list of spatial classes for each identified spatial dataset and to generate additional metadata that capture the number of spatial instances, the bounding box and the spatial vocabularies found in the datasets and stores them in the GeoLOD database. The third backend component is the *Class Recommender*, which recommends relevant classes for link discovery to a given class by first building “spatial” summaries of classes and then computing their similarity [5, 12]. The component is executed both offline, for existing classes in the Catalog, and real-time, for uncatalogued classes and GIS datasets. The frontend module is the GeoLOD web application that provides the user functionality, described in Section 4.

The GeoLOD backend API was developed in Node.js and the frontend application in React. The GeoLOD database is a PostgreSQL with the PostGIS extension for spatial data management. GeoLOD is hosted in a Ubuntu 18 LTS 4GB Virtual Machine, provided by okeanos, a GRNET cloud Infrastructure as a Service (IaaS) for Greek academic institutes. GeoLOD content, that is,

the list of spatial datasets and classes with their metadata and the recommendation lists for all classes, is updated automatically every two months, as a background process.

## 4. Demonstration

GeoLOD web application is available at <http://geolod.net> and is described in detail in [12]. In the GeoLOD frontend, users can browse the catalog of spatial datasets and classes and filter them using textual and spatial criteria. Upon entering a keyword in the Filters dialog box, GeoLOD searches in datasets and classes titles and descriptions, and upon selecting an area in the interactive map, GeoLOD returns datasets and classes whose bounding box intersect or are contained in the selected area. Upon selecting an item (a dataset or a class), users can view its full description and perform some actions. On a dataset description page (Fig. 2), users can view its title, description, SPARQL endpoint URL, its bounding box on a thumbnail, the spatial vocabularies it uses, the number of spatial instances and classes it contains, the number of recommendations and navigate in the list of dataset spatial classes. An icon indicates whether the SPARQL endpoint is currently available (green) or unavailable (red). In addition, users can download its VoID file (if available), its GeoVoID description and the dataset recommendations list in JSON. The latter can be used for as input in batch link discovery processes and consists of all recommendations for dataset classes. On a class description page, users can view its label, description, URI, the dataset it belongs to, its bounding box on a thumbnail, the spatial vocabulary it uses, the number of its spatial instances and the list of recommended classes and export the list of recommended classes in JSON. Furthermore, they can download live copies of class instances (extracted on the fly from the SPARQL endpoint) in RDF, JSON, and GeoJSON or browse class spatial instances on an interactive map. GeoJSON downloads are transformed in order to be readily consumable by a geographic information system (GIS) software, such as QGIS.

A snapshot of the recommendation list for a given class (e.g., for the Point class of the AEMET<sup>1</sup> dataset that contains information about meteorological stations) is depicted in Fig. 3. Users can navigate through the list, view details for a recommended class, such as the number of estimated related instances and the ranking order, and export Silk and LIMES configuration files for the pair of classes for direct use in a link discovery process. The configuration files are automatically generated using as input the source (in this example Point) and the selected target class SPARQL endpoint URLs and URIs and configured to apply simple string and spatial rules for instance matching. GeoLOD also provides an on-the-fly recommender user interface for generating recommendations for datasets that are not listed in the GeoLOD catalog. Initially, users select the type of the input dataset that can be a SPARQL endpoint, a GeoJSON file, or a Shapefile. For the first case, they enter the URL of the endpoint and select a class from the automatically populated list; for the other cases, they upload the corresponding files. GeoLOD parses the input dataset, builds in real time the required metadata and summaries and searches in the catalog to return the list of recommended classes for link discovery.


GeoLOD also provides a REST API to serve its content in well-known templates and formats, enabling software-based consumption. Specifically, it provides services that expose catalog

---

<sup>1</sup>[http://www.aemet.es/en/datos\\_abiertos](http://www.aemet.es/en/datos_abiertos)

**GeoLOD** DATASETS CLASSES RECOMMENDER

## AEMET metereological dataset



*AemetLinked Data (es) is an open initiative of the Ontology Engineering Group (OEG) whose aim is to enrich the Web of Data with Spanish geospatial data. This initiative started off by publishing diverse information sources belonging to the Spanish Meteorological Agency.*

**SPARQL Endpoint:** <http://aemet.linkeddata.es/sparql>

**Spatial Ontologies:** W3C Basic Geo ([http://www.w3.org/2003/01/geo/wgs84\\_pos](http://www.w3.org/2003/01/geo/wgs84_pos))

**Entities:** 780

**Classes:** 1

**Recommendations:** 27

●

Download VOID
Export GeoVOID
Export Recommendations

**Classes**

10 ▾ Showing 1 to 1 of 1 classes 1

Label ▾	URI ▾	Ontology ▾	Entities ▾	Recommendations ▾
<input type="text" value="Enter Label..."/>				
Point	<a href="http://www.w3.org/2003/01/geo/wgs84_pos#Point">http://www.w3.org/2003/01/geo/wgs84_pos#Point</a>	W3C Basic Geo	260	27

**Figure 2:** The AEMET dataset description page.

**GeoLOD** DATASETS **CLASSES** RECOMMENDER

### Recommendations

10 ▾ Showing 1 to 10 of 27 classes 1 2 3 >

Label ▾	URI ▾	Dataset ▾	Common Entities ▾	Rank ▾	Link Discovery
<input type="text" value="Enter Label..."/>		<input type="text" value="Enter Dataset..."/>			
DragLift	<a href="http://linkedgedata.org/ontology/DragLift">http://linkedgedata.org/ontology/DragLift</a>	LinkLion - A Link Repository for the Web of Data	3	1	<span style="color: red;">S</span> <span style="color: red;">L</span>
Windsock	<a href="http://linkedgedata.org/ontology/Windsock">http://linkedgedata.org/ontology/Windsock</a>	LinkLion - A Link Repository for the Web of Data	4	2	<span style="color: red;">S</span> <span style="color: red;">L</span>
Windsock	<a href="http://linkedgedata.org/ontology/Windsock">http://linkedgedata.org/ontology/Windsock</a>	LinkedGeoData	5	3	<span style="color: red;">S</span> <span style="color: red;">L</span>
AirportGate	<a href="http://linkedgedata.org/ontology/AirportGate">http://linkedgedata.org/ontology/AirportGate</a>	LinkLion - A Link Repository for the Web of Data	4	4	<span style="color: red;">S</span> <span style="color: red;">L</span>
AirportGate	<a href="http://linkedgedata.org/ontology/AirportGate">http://linkedgedata.org/ontology/AirportGate</a>	LinkedGeoData	4	5	<span style="color: red;">S</span> <span style="color: red;">L</span>
Beacon	<a href="http://linkedgedata.org/ontology/Beacon">http://linkedgedata.org/ontology/Beacon</a>	LinkLion - A Link Repository for the Web of Data	5	6	<span style="color: red;">S</span> <span style="color: red;">L</span>
Beacon	<a href="http://linkedgedata.org/ontology/Beacon">http://linkedgedata.org/ontology/Beacon</a>	LinkedGeoData	6	7	<span style="color: red;">S</span> <span style="color: red;">L</span>
ParkingSpace	<a href="http://linkedgedata.org/ontology/ParkingSpace">http://linkedgedata.org/ontology/ParkingSpace</a>	LinkLion - A Link Repository for the Web of Data	3	8	<span style="color: red;">S</span> <span style="color: red;">L</span>
terms#Stop	<a href="http://vocab.gtfs.org/terms#Stop">http://vocab.gtfs.org/terms#Stop</a>	Aragon Interoperable Information Structure EIZA - aragon open data	12	9	<span style="color: red;">S</span> <span style="color: red;">L</span>
Lighthouse	<a href="http://linkedgedata.org/ontology/Lighthouse">http://linkedgedata.org/ontology/Lighthouse</a>	LinkLion - A Link Repository for the Web of Data	13	10	<span style="color: red;">S</span> <span style="color: red;">L</span>

**Figure 3:** The ranked class recommendations list for the Point class of the AEMET dataset.

content in DCAT [13] and JSON, dataset descriptions in GeoVOID and class link discovery recommendation lists for datasets and classes. The Table in Appendix lists the names, the request URI, and the descriptions of the main services.

## 5. Discussion and Conclusion

GeoLOD has identified and currently lists 82 spatial datasets (out of 629 unique SPARQL endpoints URLs found in the LOD cloud and the Datahub) and 5,218 spatial classes. Regarding their geographic coverage, most datasets are “global”, that is, they contain instances all over the world, and most non-global datasets are located in and around Europe. GeoLOD provides approximately 90,000 class recommendations and on average it recommends 25 relevant classes per class. Regarding the performance of the recommendation algorithm executed in a 4GB RAM Ubuntu VM, it requires on average 18 minutes to generate the class recommendation list for each class (although the execution time depends on the source class size and spatial extent, ranging from a few seconds to several minutes), while the LIMES execution time for examining all possible pairs of spatial classes, that is, approximately 5,000, in order to generate instance links is approximately 5 hours. Possible lines of GeoLOD future development are the inclusion of datasets provided as RDF dumps, the further improvement of the recommendation algorithm in terms of efficiency and effectiveness, and the integration with Silk or LIMES web services for the instant generation of instance link matching.

## References

- [1] C. Bizer, A. Jentzsch, R. Cyganiak, State of the LOD Cloud, 2011. URL: <http://www4.wiwiw.fu-berlin.de/lodcloud/state/>.
- [2] Datahub, <https://old.datahub.io/>, 2013. Accessed: 2022-04-22.
- [3] E. Pietriga, H. Gözükan, C. Appert, M. Destandau, S. Cebiric, F. Goasdoué, I. Manolescu, Browsing linked data catalogs with lodatlas., in: D. Vrandecic, K. Bontcheva, M. C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, L.-A. Kaffee, E. Simperl (Eds.), International Semantic Web Conference (2), volume 11137 of *Lecture Notes in Computer Science*, Springer, 2018, pp. 137–153. URL: <http://dblp.uni-trier.de/db/conf/semweb/iswc2018-2.html#PietrigaGADCGM18>.
- [4] K. Alexander, R. Cyganiak, M. Hausenblas, J. Zhao, Describing Linked Datasets - On the Design and Usage of void, the ‘Vocabulary of Interlinked Datasets’, in: WWW 2009 Workshop: Linked Data on the Web (LDOW2009), Madrid, Spain, 2009.
- [5] V. Kopsachilis, M. Vaitis, N. Mamoulis, D. Kotzinos, Recommending geo-semantically related classes for link discovery, *Journal on Data Semantics* 9 (2020) 151–177. URL: <https://doi.org/10.1007/s13740-020-00117-4>. doi:10.1007/s13740-020-00117-4.
- [6] J. Volz, C. Bizer, M. Gaedke, G. Kobilarov, Silk - a link discovery framework for the web of data., in: C. Bizer, T. Heath, T. Berners-Lee, K. Idehen (Eds.), LDOW, volume 538 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2009. URL: <http://dblp.uni-trier.de/db/conf/www/ldow2009.html#VolzBGK09>.
- [7] A.-C. Ngonga Ngomo, M. A. Sherif, K. Georgala, M. Hassan, K. Dreßler, K. Lyko, D. Obraczka, T. Soru, LIMES - A Framework for Link Discovery on the Semantic Web, *KI-Künstliche Intelligenz, German Journal of Artificial Intelligence - Organ des Fachbereichs “Künstliche Intelligenz” der Gesellschaft für Informatik e.V.* (2021). URL: [https://papers.dice-research.org/2021/KI\\_LIMES/public.pdf](https://papers.dice-research.org/2021/KI_LIMES/public.pdf).

- [8] W. Beek, L. Rietveld, H. R. Bazoobandi, J. Wielemaker, S. Schlobach, Lod laundromat: A uniform way of publishing other people’s dirty data., in: P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. A. Knoblock, D. Vrandecic, P. Groth, N. F. Noy, K. Janowicz, C. A. Goble (Eds.), *International Semantic Web Conference (1)*, volume 8796 of *Lecture Notes in Computer Science*, Springer, 2014, pp. 213–228. URL: <http://dblp.uni-trier.de/db/conf/semweb/iswc2014-1.html#BeekRBWS14>.
- [9] P.-Y. Vandebussche, J. Umbrich, L. Matteis, A. Hogan, C. B. Aranda, Sparqls: Monitoring public sparql endpoints., *Semantic Web* 8 (2017) 1049–1065. URL: <http://dblp.uni-trier.de/db/journals/semweb/semweb8.html#VandebusscheUM17>.
- [10] A. A. M. Caraballo, N. M. Arruda, B. P. Nunes, G. R. Lopes, M. A. Casanova, Trtml - a triplest recommendation tool based on supervised learning algorithms, in: V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, A. Tordai (Eds.), *The Semantic Web: ESWC 2014 Satellite Events*, Springer International Publishing, Cham, 2014, pp. 413–417.
- [11] M. Mountantonakis, Y. Tzitzikas, Lodsynthesis: Global scale knowledge services, *Heritage* 1 (2018) 335–348. URL: <https://www.mdpi.com/2571-9408/1/2/23>.
- [12] V. Kopsachilis, M. Vaitis, Geolod: A spatial linked data catalog and recommender, *Big Data Cogn. Comput.* 5 (2021) 17. doi:10.3390/bdcc5020017.
- [13] W3C, Data catalog vocabulary (dcat), <https://www.w3.org/TR/2020/SPSD-vocab-dcat-20200204/>, 2014. Last accessed 20 April 2022.

## A. GeoLOD REST API

**Table 1**

GeoLOD services (the left part of the Request URI is <http://snf-661343.vm.okeanos.grnet.gr>).

Service Name	Request URI	Description
GeoLOD Description	<code>/api/download/dcat</code>	Returns a DCAT-compliant turtle file that contains general information about GeoLOD and the list of the datasets in the Catalog
Dataset List	<code>/api/datasets</code>	Returns, in JSON, the list of datasets with their metadata (including internal dataset IDs) in the GeoLOD Catalog
Dataset Description	<code>/api/datasets/&lt;ID&gt;</code>	Returns, in JSON, the specified dataset metadata with the list of its classes. The dataset ID is a variable corresponding to the internal dataset ID. (e.g., <a href="http://snf-661343.vm.okeanos.grnet.gr/api/datasets/915">http://snf-661343.vm.okeanos.grnet.gr/api/datasets/915</a> returns the metadata for the <i>AEMET</i> dataset)
Class List	<code>/api/classes</code>	Returns, in JSON, the list of classes with their metadata (including internal classes IDs) in the GeoLOD Catalog.
Class Description	<code>/api/classes/&lt;ID&gt;</code>	Returns, in JSON, the specified class metadata with the list of its recommended classes. The class ID is a variable corresponding to the internal class ID. (e.g., <a href="http://snf-661343.vm.okeanos.grnet.gr/api/classes/139090">http://snf-661343.vm.okeanos.grnet.gr/api/classes/139090</a> returns the metadata for the <i>CaveEntrance</i> class of <i>Linklion</i> dataset).
Dataset GeoVoid	<code>/api/download/geovoid/&lt;ID&gt;</code>	Returns, in turtle format, the GeoVoid description of the specified dataset.
Dataset Recommendations	<code>/api/download/datasetrecommendations/&lt;ID&gt;</code>	Returns, in JSON, the list of recommendations for all specified dataset classes.
Class Recommendations	<code>api/download/classesrecommendations/&lt;ID&gt;</code>	Returns, in JSON, the list of recommendations for the specified class.