

Article

# Discovering Links between Geospatial Data Sources in the Web of Data: The Open Geospatial Engine Approach

Lianlian He <sup>1,\*</sup> and Ruixiang Liu <sup>2</sup> 

<sup>1</sup> School of Mathematics and Statistics, Hubei University of Education, #129 Second Gaoxin Road, East Lake Hi-Tech Zone, Wuhan 430205, China

<sup>2</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; lrx\_lucky@whu.edu.cn

\* Correspondence: helianlian@hue.edu.cn

**Abstract:** The Web of Data has been fueled significantly by geospatial data over the last few years. In the current link discovery frameworks, there is still a lack of robust support for finding geospatial-aware links between geospatial data sources in the Web of Data. They are also limited in efficient association capabilities for large-scale datasets. This paper extends the data integration capability based on the spatial metrics in the open geospatial engine OGE. These metrics include topological relationships and spatial matching between geospatial entities within multiple geospatial data sources. Thus, the tool can be employed by data publishers to set geospatial-aware links to facilitate geospatial data and knowledge discovery in the Web of Data. Several geospatial data sources are used to demonstrate the usability and effectiveness of the approach and tool implementation.

**Keywords:** Web of Data; geospatial Linked Open Data; Linked Data; spatial matching; link discovery

## 1. Introduction

The advent of Linked Data shows great promise for effectively sharing and interlinking Web resources on a global scale [1]. Linked Data plays a pivotal role in advancing data interoperability and facilitating the creation of a global knowledge graph that enhances data integration, accessibility, and discoverability, thereby fostering interdisciplinary research and innovation across various domains [2]. It follows a set of recommended best practices for exposing, sharing, and connecting data organized in the Resource Description Framework (RDF) [3]. Linked Data can create one global database for all data and offer great opportunities for the wide sharing and integration of isolated and heterogeneous data, such as the integration of spatial proteomics [4]. The recent progress in natural language processing (NLP), specifically with large language models (LLMs), has demonstrated significant potential for automating a wide range of tasks. Therefore, there are current research efforts combining Linked Data and large language models (LLM), such as using the GPT-3 language model to answer natural language questions over Linked Data [5]. In the geospatial domain, Linked Data results in a paradigm shift, from distributed complex databases accessed through Web services to knowledge bases represented as RDF graphs [6].

Two basic ideas are involved in building the Web of Data: publishing structured data on the Web using the RDF data model and establishing RDF links between different data sources [7]. To use the Web as a single global data space, setting interlinks between diverse data sources, including those geospatial data sources, is a crucial issue. It will bring a new dimension to the connectivity of the Web of Data when taking into account geospatial attributes to create RDF links. On the one hand, they can be employed to establish links between spatial relationships, such as topological, directional, and distance relationships; on the other hand, they can be weighted with other properties in similarity metrics to generate identity links. For example, linguistic difference often hinders matches of different URIs identifying the same geospatial entity. In such cases, the geospatial properties of these



**Citation:** He, L.; Liu, R. Discovering Links between Geospatial Data Sources in the Web of Data: The Open Geospatial Engine Approach. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 143. <https://doi.org/10.3390/ijgi13050143>

Academic Editors: Wolfgang Kainz and Dev Raj Paudyal

Received: 25 January 2024

Revised: 25 April 2024

Accepted: 26 April 2024

Published: 28 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

entities will contribute to the similarity calculation using spatial relationship metrics. In addition, using geospatial information when creating links will also improve the accuracy of similarity matching and avoid semantic mis-matches, as different geospatial entities may have the same lexical information and classification in terms of an ontology yet have totally different locations.

There are already several link discovery frameworks available to achieve connections of entities in one dataset to entities in another, such as Silk [8], LIMES [9], and LinQL [10]. These existing link discovery tools, however, lack support for spatial matching functions, while Silk and LIMES, despite supporting such features, have lower matching efficiency. As the scale of geospatial data continues to expand, researchers are increasingly focusing on association efficiency. For instance, recent notable association frameworks such as Geo-L [11] and JedAI-spatial [12] have demonstrated advanced performance. Table 1 presents a comparison of different link discovery frameworks (including ours). From these points of view, this paper is motivated to support parallel geospatial link discovery for the Web of Data by integrating spatial relation computation and matching methods in a link discovery framework. This paper takes an open geospatial engine (OGE) as an example, enriching it with geospatial metrics. The OGE system features three key aspects following the principles of open science [13] and open GIS [14]: open-source architecture, adherence to OGC open standards and APIs, and system openness with scalability. The tool, named the OGE knowledge graph component (OGE-KG), thus can be employed by data publishers to set geospatial-aware links to facilitate geospatial data and knowledge discovery in the Web of Data. Several geospatial data sources in the Linked Open Data (LOD) cloud [15] are used to demonstrate the usability and effectiveness of the approach.

**Table 1.** Comparison of different link discovery frameworks.

Framework	Spatial Supported	Parallel Computing Supported	Efficiency
LinQL	No	No	Low
Silk	Partial	No	Low
LIMES	Yes	No	Low
Geo-L	Yes	No	High
JedAI-spatial	Yes	Yes	High
OGE-KG	Yes	Yes	High

The contribution of this article is summarized as follows: (1) The paper addresses the gap in existing link discovery frameworks by integrating spatial relation computation and matching methods, including relationship links and identity links. (2) This paper enables parallel geospatial link discovery for the Web of Data, improving the efficiency of matching functions and thus enhancing the connectivity of diverse data sources. (3) The paper introduces the OGE knowledge graph component (OGE-KG), an extension of the open geospatial engine (OGE). OGE-KG is enriched with geospatial metrics, allowing data publishers to establish geospatial-aware links and facilitate geospatial data and knowledge discovery in the Web of Data.

This paper is structured as follows. Section 2 describes the background and related work. Section 3 introduces the approach to integrating geospatial metrics in the OGE-KG data interlinking platform. The implementation of the geospatial extension based on the OGE-KG is given in Section 4. Several use cases are provided in Section 5 to demonstrate the usability and effectiveness of the implementation. Section 6 provides the discussion. Conclusions and future work are given in Section 7.

## 2. Background and Related Work

### 2.1. Geospatial Data Sources in LOD

Linked Data offers great opportunities in the geospatial domain, since conventionally isolated and heterogeneous geospatial data could be exposed as Linked Data on the Web, thus promoting the wide sharing and integration of geospatial information [16]. Statistics

show that about a quarter (25.05%) of datasets in the LOD cloud use the WGS84 vocabulary [17], which demonstrates the significance of geo-referencing data on the Web. There is a large body of work dedicated to publishing geographically-related data on the Web. Table 2 summarizes some prominent geospatial data sources in the LOD cloud. GeoNames is the first case to provide geographical entities as Linked Data and linked by a large number of datasets on the Web [18]. By adding a spatial dimension to the Web of Data, LinkedGeoData transforms OpenStreetMap data into the RDF data model and maps them to other spatial datasets [19]. GADM, the Database of Global Administrative Areas, is a spatial database of locations of the world's administrative areas. It provides administrative boundaries and hierarchical relationships among administrative divisions. GADM is published as Linked Data, named GADM-RDF [20]. NUTS, the Nomenclature of Units for Territorial Statistics, provides geospatial regions in the European Union as Linked Data for statistical and policy purposes [21].

**Table 2.** Geospatial data sources and related statistics.

Data Source	Size
GeoNames <sup>1</sup>	11,985,741 features about 182 million triples
LinkedGeoData <sup>2</sup>	20,000,000,000 triples
GADM <sup>3</sup>	400,276 administrative areas
NUTS <sup>4</sup>	316,238 triples
Ordnance Survey NHD <sup>5</sup>	36,773,687 triples 20,000,000 triples
GeoLinkedData.es	21,564,199 triples

<sup>1</sup> <http://www.geonames.org/> (accessed on 21 January 2024). <sup>2</sup> <http://linkedgeodata.org/> (accessed on 21 January 2024). <sup>3</sup> <http://gadm.geovocab.org/> (accessed on 21 January 2024). <sup>4</sup> <http://nuts.geovocab.org/> (accessed on 21 January 2024). <sup>5</sup> <https://www.usgs.gov/national-hydrography/national-hydrography-dataset> (accessed on 21 January 2024).

In the geospatial Linked Open Data cloud, there are also increasing governmental efforts. Ordnance Survey, the national mapping agency of Great Britain, provides an up-to-date geospatial RDF dataset of Great Britain [22]. The Geological Survey of the United States (USGS) makes various geospatial and environmental datasets accessible as RDF data, such as the National Hydrography Dataset (NHD) [23]. Another linked dataset, GeoLinked Data, publishes Spanish geospatial data on the Web [24].

## 2.2. Geospatial Ontology Modeling

In order to publish geospatial resources as Linked Data on the Web, various ontologies or vocabularies are developed in the LOD communities. These ontologies and vocabularies define geospatial concepts and their relationships and help geospatial data integration in the Web of Data. The W3C Geo Vocabulary, now updated by the Geospatial Incubator Group as Geo OWL [25], is one of the most used vocabularies across multiple domains. The Open Geospatial Consortium (OGC) develops the GeoSPARQL standard (version 1.1) to represent and query geospatial resources on the Semantic Web [26]. The NeoGeo vocabulary (geometry and spatial vocabularies) is another effort to representing geospatial data and their relationships [27]. In addition, some geospatial data providers also contribute to the ontologies. For example, Ordnance Survey develops both geometry and spatial relations ontologies that are widely used by the LOD communities. A detailed comparison of these vocabularies can be found in [28].

## 2.3. Geospatial Links in LOD

There are three important types of RDF links: relationship links, identity links, and vocabulary links [7]. Relationship links connect entities in one dataset to entities in another, which can provide more information to the source dataset. Identity links aim at constructing interlinks between different URIs identifying the same entity. Both relationship and

identify links allow data consumers to discover more data by following links, thus possibly tapping much more potential for data utilization. Vocabulary links map relationships between terms from different vocabularies. Such links, for example, “owl:subClassOf” and “owl:subPropertyOf”, can benefit data integration using different datasets, since they help machines to understand terms from different vocabularies.

Some efforts have studied the interconnection of linked geospatial data by extending linking criteria from the general information domain, such as a hybrid use of distance and semantic criteria [29–31]. A geographic entity can be described using either a geographical name in the text form or a geographical feature with location represented in coordinates. Both spatial scores and name similarity scores can then be calculated using various distance algorithms, including Jaco-Winkler [32], to determine the linkage of entities, such as the “owl:sameAs” relationship. In some cases, distance computation may be employed to establish relationships such as “nearBy” between two geographical features. In the LinkedGeoData, datasets (e.g., roads and lakes) are interlinked at the feature level using different distance measures. In addition, topological relations among geospatial features can be determined and presented as links. For example, some efforts employ the topological predicate, “spatial:EQ”, to integrate the NUTS and GADM datasets [33,34]. Ma [35] describes conversion from vector objects and raster data to RDF. Vector data are encoded in Geography Markup Language (GML), from which topological relations are pre-computed and then converted into RDF triples.

### 3. Incorporating Geospatial Metrics: The OGE Approach

In the previous section, we listed different vocabularies for representing geospatial features, geometries, and their relationships. While there is a high variety in expressing geo-referencing data and their spatial relations, we adopt the GeoSPARQL vocabularies in the development process for the future possibility of spatial reasoning and follow the Dimensionally Extended Nine Intersection Model (DE-9IM) specified by OGC. To integrate geospatial metrics into the link discovery framework, the spatial properties of geospatial datasets must be fully taken into account. First, the spatial dimension of LOD cloud datasets can be computed by topological operators to detect spatial relationship links between them. Second, they can be compared by the geometry-based metric to establish identity links between them.

#### 3.1. Topological Predicates

Topological relations for geospatial features are used to make links between different geographic datasets. This kind of relationship link can be established by extracting geographic coordinates encoded in GML or Well Known Text (WKT), computing spatial relations using the encoding, and then leveraging appropriate vocabularies to explicitly describe topological relations.

The GeoSPARQL standard [26] provides vocabularies for representing geospatial information and defines different families of topological relations between spatial objects, including simple features, Region Connection Calculus (RCC), and Egenhofer relations. The Simple Features Specification (SFS) adopts the DE-9IM model and defines eight topological predicates including Equals, Disjoint, Intersects, Touches, Crosses, Within, Contains, and Overlaps. The topological predicates are Boolean functions that return TRUE (T) if a comparison meets the function criteria and FALSE (F) otherwise. These binary predicates make topological comparisons rather than pointwise comparisons and can be described by related DE-9IM patterns. If “I” represents the interior of a geometry, “B” represents the boundary of a geometry, and “E” represents the exterior of a geometry, then the DE-9IM model of two geometries is represented by a nine-character string composed of F/T/\*, where, from left to right, it represents the following:  $I(a) \cap I(b)$ ,  $I(a) \cap B(b)$ ,  $I(a) \cap E(b)$ ,  $B(a) \cap I(b)$ ,  $B(a) \cap B(b)$ ,  $B(a) \cap E(b)$ ,  $E(a) \cap I(b)$ ,  $E(a) \cap B(b)$ , and  $E(a) \cap E(b)$ . For example, the pattern matrix of the “Within” predicate accounts for the fact that the predicate returns true (T) when the interiors of two geometries intersect and false (F) when the

interior and boundary of a geometry intersect the exterior of the other geometry. All other conditions do not matter (\*) whether an intersection exists or not. Also, the pattern matrix of the “Intersects” signifies that either the interiors of two geometries intersect (T\*\*\*\*\*), the interior of one geometry intersects with the boundary of another geometry (\*T\*\*\*\*\* or \*\*\*T\*\*\*\*\*), or the boundaries of two geometries intersect (\*\*\*\*T\*\*\*\*), constituting an intersection. The pattern matrix of the “Equals” represents that  $I(a) \cap I(b) = T$ ,  $I(a) \cap B(b) = F$ ,  $I(a) \cap E(b) = F$ ,  $B(a) \cap I(b) = F$ ,  $B(a) \cap B(b) = T$ ,  $B(a) \cap E(b) = F$ ,  $E(a) \cap I(b) = F$ ,  $E(a) \cap B(b) = F$ , and  $E(a) \cap E(b) = T$ . Table 3 lists the applicable geometry types and DE-9IM intersection patterns of SFS topological relations. “Applicable Dimensions” refers to the types of geometries for which simple feature topological relations can be applied. The symbol P is used to refer to 0-dimensional geometries (e.g., points), L to 1-dimensional geometries (e.g., lines), and A to 2-dimensional geometries (e.g., polygons).

**Table 3.** Simple feature topological relations.

Relation Name	Applicable Dimensions	DE-9IM Pattern
Equals	All	TFFFFTFFFT
Disjoint	All	FF*FF**** T***** *T*****
Intersects	All	***T***** ****T**** FT*****
Touches	All except P/P	F**T***** F**T****
Within	All	T*F**F***
Contains	All	T****F**
Overlaps	P/P, A/A, L/L	T*T***T** (for P/P, A/A) 1*T***T** (for L/L)
Crosses	P/L, P/A, L/A, L/L	T*T***T** (for P/L, P/A, L/A) 0***** (for L/L)

### 3.2. Geometry-Based Metric

Currently, most identity links are generated using string similarity metrics over the non-spatial properties. This may result in semantic mis-matches, especially in the geospatial domain, since it is often the case that different entities may have the same non-spatial properties yet totally different spatial properties. Therefore, it is necessary to involve spatial attributes when building identity links between geospatial datasets.

Before building spatial equivalences between geospatial entities, it is noted that the geometric shape of the same spatial feature may be measured at varying resolutions. For example, there are different geometric descriptions of the administrative geography of Berlin from official data of the German government and vague data published by international survey agents. Hence, existing methods for determining similarities between two geometries are needed. For example, the Hausdorff distance is a frequently used distance measure for comparing the similarity of two geometric shapes. The measured value can be normalized to lie in the range [0, 1], where the higher value indicates a greater degree of similarity. The input geometries are considered to be a matching shape if the measure is within a given tolerance with respect to the Hausdorff distance.

### 3.3. High-Performance Geospatial Data Linking

In recent years, the advancement of parallel computing technology has provided solutions to high-performance geographic computing issues and has become a research hotspot in the fields of big data analysis and data mining. Efficient spatial algorithms for real-time processing of massive amounts of geospatial data have enabled the simulation and analysis of geospatial phenomena on a global scale and over extended time periods, which were previously challenging to compute. Spatial metrics are typical attributes of

geographically associated data. As an essential part of geospatial reasoning, enhancing their computational efficiency is crucial for constructing vast, wide-ranging, multi-scale geographical knowledge graphs.

Geospatial parallel computation can be divided into two types: data-intensive and compute-intensive. Data-intensive computing processes different geographic data in a Single Instruction Multiple Data (SIMD) manner, with the core characteristic being that the geometric objects are mutually independent during computation. For instance, establishing topological predicates between large-scale heterogeneous geographic data sources and determining topological relationships fall under data-intensive computations. Compute-intensive calculations are conducted when complex intersection relationships exist between polygons. They primarily involve operations like intersection, difference, union, negation of intersection, amalgamation, updating, identification, and spatial connections, exhibiting typical features of high algorithmic complexity and intensive computation. For example, when determining spatial equivalence relationships like the Hausdorff distance, it is necessary to compute the distance between different pairs of points inside two polygons, which is also a compute-intensive operation.

### 3.4. Adding Geospatial Metrics into OGE-KG

The OGE represents a comprehensive platform dedicated to the analysis of large-scale spatial-temporal data. The OGE-KG Data Interlinking Workbench, embedded within the overarching framework of the OGE knowledge graph, is a Web application that enables users to create links between two datasets in an interactive way. It provides three components: workspace browser, linkage rule editor, and evaluation. The workspace component provides a tree view of all projects and allows users to customize data sources and link tasks for each project. The linkage rule editor is a graphical interface that enables users to generate linkage rules by dragging and dropping its built-in operators (transformations, comparators, and aggregators). The evaluation component allows users to evaluate the links generated by the current linkage rule.

Compared with common data-linking tools Silk and LIMES, the spatial extension to the OGE-KG includes enriching comparators with topological operators and geometry-based metrics. The extension framework is illustrated in Figure 1. First, the source dataset and target dataset are inputted into the workspace browser, and an association task should be created with the two datasets. Then, the RDF path selector is utilized in the linkage rule editor to further filter the data that need to be associated. Subsequently, functions are employed in the transformations module to preprocess the data, such as renaming and filtering. Afterward, various comparator operators, such as topological association, string association, and geometric operators, can be used to concatenate data in OGE-KG. Additionally, aggregator operators can be utilized for aggregation operations. Each operator is regarded as a plugin that can be embedded into various operators: transformations, comparators, and aggregators. Finally, the system will execute computations for the linkage workflow using efficient parallel computing. Upon completion of the computation, results can be exported in the evaluation module, where attributes such as similarity of association scores can be viewed and results can be exported.

Using the extension framework, the OGE-KG Data Interlinking Workbench is able to find topological relationships between entities within different geospatial data sources and supports the generation of identity links based on geometry similarity.

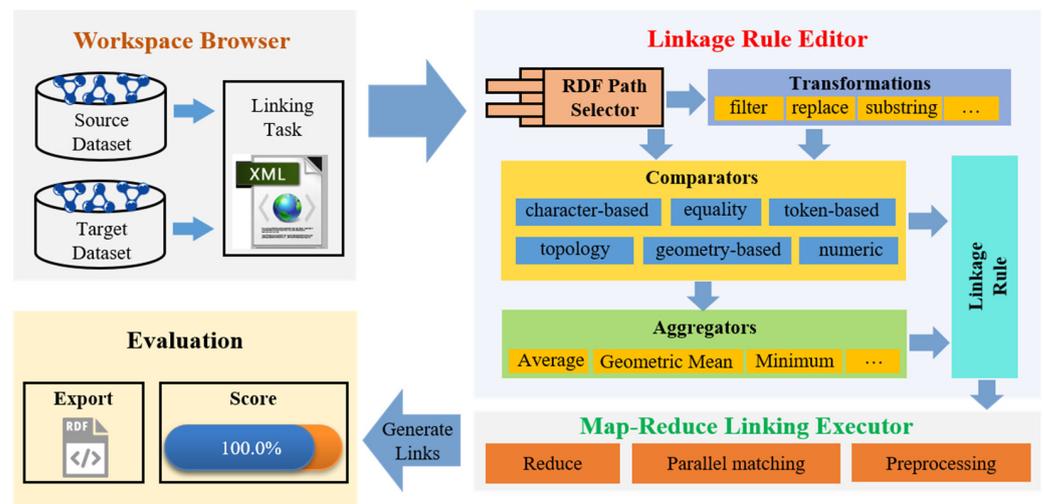


Figure 1. The framework for the geospatial extension to OGE-KG.

#### 4. Implementation of Geospatial Extension in OGE-KG

This section describes the implementation of geospatial extension in OGE-KG. In the development process, the JTS Topological Suite is used to provide spatial data operations required in the OGE-KG framework. The JTS is an open-source Java API that provides the implementation of spatial predicates and functions described in the OpenGIS Simple Features Specification [36]. To speed up the process of data linking in the OGE-KG framework, we parallelize it using the MapReduce approach in a single machine. MapReduce is a programming model and an associated implementation introduced by Google to process and generate large datasets [37]. Apache Spark is an open-source, distributed computing system that is designed to be fast and general purpose, making it suitable for a wide range of tasks from batch processing to real-time data processing and advanced analytics [38].

##### 4.1. Topological Operators

Given two geometries  $g_1$  and  $g_2$ , which are created by the JTS WKTRReader from two string parameters  $s_1$  and  $s_2$ , the topological operators are a set of binary predicates that compute whether a certain topological relationship exists between the two geometries. For example, if the statement “ $g_1.within(g_2)$ ” returns true, it means every point of  $g_1$  is a point of  $g_2$ , and the interiors of  $g_1$  and  $g_2$  have at least one point in common. Hence, we can describe the topological relationship between  $g_1$  and  $g_2$  using the GeoSPARQL vocabulary “ $geo:sfWithin$ ”. Table 4 describes all of the topological operators added in the OGE-KG framework and their associated GeoSPARQL vocabularies.

Table 4. Topological operators extended in the OGE-KG framework.

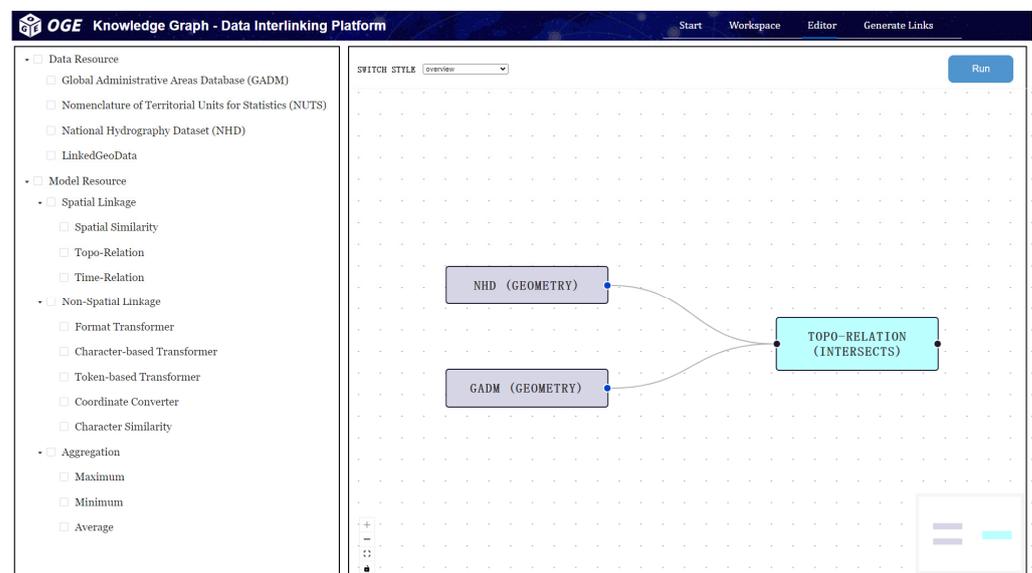
Plugin Label	Relation URI	Description
Equals	geo:sfEquals	Returns true if two geometries are topologically equal.
Disjoint	geo:sfDisjoint	Returns true if the intersection of the two geometries is an empty set.
Intersects	geo:sfIntersects	Returns true if two geometries have at least one point in common.
Touches	geo:sfTouches	Returns true if two geometries have at least one boundary point in common but no interior points.
Within	geo:sfWithin	Returns true if the first geometry lies in the interior of the second geometry.
Contains	geo:sfContains	Returns true if the second geometry lies in the interior of the first geometry.
Overlaps	geo:sfOverlaps	Returns true if two geometries of the same dimension share some but not all points in common and the intersection of the interiors of the two geometries has the same dimension as the geometries themselves.
Crosses	geo:sfCrosses	Returns true if two geometries share some but not all interior points and the dimension of their intersection is less than the maximum dimension of the two source geometries.

#### 4.2. Geometry-Based Similarity Operator

We have implemented the Hausdorff similarity measurement. There are various methods of computing the Hausdorff distance between two geometric shapes. The JTS computes the Hausdorff distance (HD) based on a discretization of the input geometries, and the discrete Hausdorff distance (DHD) is less than or equal to the standard HD for all geometries. In order to increase the accuracy of the result, the input geometries are densified by a factor of 0.25. When the densify factor tends to zero, the DHD value will approach the true HD. Next, the DHD value is normalized by dividing it by the diagonal distance across the envelope of the combined geometries.

#### 4.3. Extended OGE-KG Workbench

As illustrated in Figure 2, the topological predicates and Hausdorff metric can be dragged and dropped as built-in comparison operators in the OGE-KG Data Interlinking Workbench. The topological operators can be used to find topological relationships between two entities within different geospatial datasets. The Hausdorff metric can be used to recognize the spatial equivalence of two geometries and then aid the establishment of identity links.



**Figure 2.** An interface of the extended OGE-KG Workbench.

### 5. Experiments in OGE-KG

To assess the usability and effectiveness of geospatial enrichment in the OGE-KG, we have conducted two kinds of experiments: finding spatial relationship links (Section 5.1) and building identity links (Section 5.2). Three use cases are reported involving four geospatial databases including LinkedGeoData, GADM, NUTS, and NHD. The XML namespace prefixes used in the experiments are summarized in Table 5. It is worth noting that to reflect the association efficiency, we calculated the serial and parallel durations for the following three experiments. All linking experiments were implemented on a 64-bit desktop system with a 2.10 GHz CPU, 16 GB memory, and 12 Cores environment. In OGE-KG, Apache Spark uses 12 cores, with data partitioned into 24 partitions. For the single-threaded experiment, 1 core is utilized, with data partitioned into 24 partitions as well. We also conducted comparative experiments with two frameworks, Silk and LIMES (both equipped with spatial association capabilities), within the same experimental environment to highlight our efficiency advantage. Additionally, we conducted a scalability analysis using the topological relation. To demonstrate scalability, we not only implemented comparative experiments with two recent open-source frameworks, Geo-L and JedAI-

spatial, but also performed a detailed evaluation of the parallel efficiency of the OGE-KG framework in a multi-core server environment.

**Table 5.** XML namespaces used in the experiments.

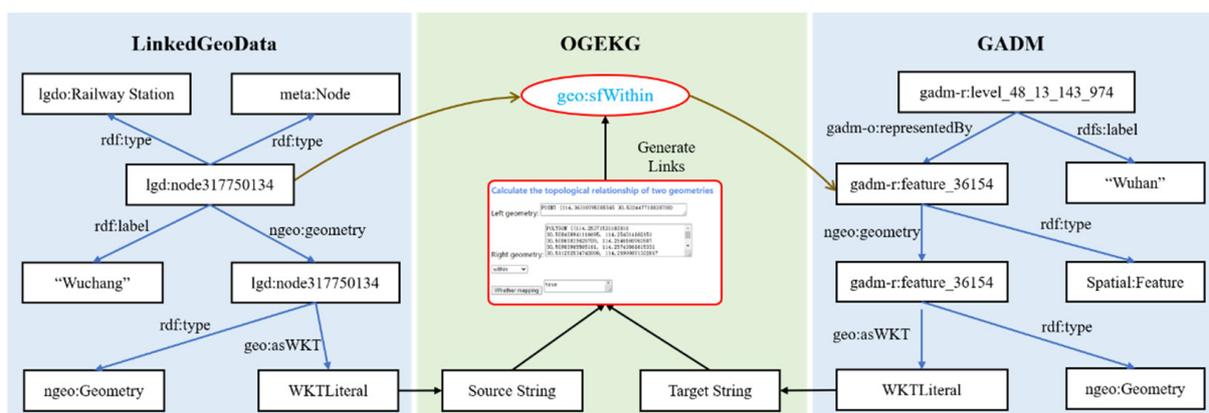
Prefix	Namespace
rdf	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a> (accessed on 21 January 2024)
rdfs	<a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a> (accessed on 21 January 2024)
ngeo	<a href="http://geovocab.org/geometry#">http://geovocab.org/geometry#</a> (accessed on 21 January 2024)
spatial	<a href="http://geovocab.org/spatial#">http://geovocab.org/spatial#</a> (accessed on 21 January 2024)
geo	<a href="http://www.opengis.net/ont/geosparql#">http://www.opengis.net/ont/geosparql#</a> (accessed on 21 January 2024)
lgdo	<a href="http://linkedgeodata.org/ontology/">http://linkedgeodata.org/ontology/</a> (accessed on 21 January 2024)
lgd	<a href="http://linkedgeodata.org/triplify/">http://linkedgeodata.org/triplify/</a> (accessed on 21 January 2024)
lgdg	<a href="http://linkedgeodata.org/geometry/">http://linkedgeodata.org/geometry/</a> (accessed on 21 January 2024)
meta	<a href="http://linkedgeodata.org/meta/">http://linkedgeodata.org/meta/</a> (accessed on 21 January 2024)
gadm-o	<a href="http://linkedgeodata.org/ld/gadm2/ontology/">http://linkedgeodata.org/ld/gadm2/ontology/</a> (accessed on 21 January 2024)
gadm-r	<a href="http://linkedgeodata.org/ld/gadm2/resource/">http://linkedgeodata.org/ld/gadm2/resource/</a> (accessed on 21 January 2024)
nhd-o	<a href="http://cegis.usgs.gov/rdf/nhd#">http://cegis.usgs.gov/rdf/nhd#</a> (accessed on 21 January 2024)
ramon	<a href="http://rdfdata.eionet.europa.eu/ramon/ontology/">http://rdfdata.eionet.europa.eu/ramon/ontology/</a> (accessed on 21 January 2024)

### 5.1. Spatial Relationship Links

We employed topological operators extended in the OGE-KG Data Interlinking Workbench to find spatial relationships between different geospatial datasets. The following two topological operators are selected as examples to demonstrate the workflow of discovering spatial relationship links in the LOD cloud.

#### 5.1.1. Within Operator

In this case, we want to discover railway stations in the LinkedGeoData that are “Within” the Hubei Province, China, in the GADM data source. We configure the <Link-Type> to be geo:sfWithin. Figure 3 presents the process of generating geo:sfWithin relationship links between lgd:node317750134 in LinkedGeoData and gadm-r:feature\_36153 in GADM. Figure 4 shows the screenshot of linking results in the OGE-KG Data Interlinking Workbench. Dataset statistics and linking results are given in Table 6. The discovery process takes only 5 s approximately with a parallel program. The same datasets were also tested using the Contains (the inverse of Within) operator. The establishment of Contains relationship links costs almost equal time as that of Within links. Additionally, we tested the time of Silk and LIMES using the same dataset and experimental environment. Silk and LIMES may have employed optimizations for efficiency; hence, their single-threaded times may be faster. However, the efficiency of using MapReduce remains significantly higher, with substantial improvements in the parallel computation of spatial relationship links.



**Figure 3.** An example application of the “Within” operator.

Source	Target	Score	Correct
<a href="http://linkedgeo.org/triplify/node1703084243">http://linkedgeo.org/triplify/node1703084243</a>	<a href="http://geos.wlu.edu.cn/resource/gadm/3/50">http://geos.wlu.edu.cn/resource/gadm/3/50</a>	100%	✓
<a href="http://linkedgeo.org/triplify/node2381028362">http://linkedgeo.org/triplify/node2381028362</a>	<a href="http://geos.wlu.edu.cn/resource/gadm/3/24">http://geos.wlu.edu.cn/resource/gadm/3/24</a>	100%	✓
<a href="http://linkedgeo.org/triplify/node2398956501">http://linkedgeo.org/triplify/node2398956501</a>	<a href="http://geos.wlu.edu.cn/resource/gadm/3/50">http://geos.wlu.edu.cn/resource/gadm/3/50</a>	100%	✓
<a href="http://linkedgeo.org/triplify/node2662698249">http://linkedgeo.org/triplify/node2662698249</a>	<a href="http://geos.wlu.edu.cn/resource/gadm/3/62">http://geos.wlu.edu.cn/resource/gadm/3/62</a>	100%	✓
<a href="http://linkedgeo.org/triplify/node3097123132">http://linkedgeo.org/triplify/node3097123132</a>	<a href="http://geos.wlu.edu.cn/resource/gadm/3/20">http://geos.wlu.edu.cn/resource/gadm/3/20</a>	100%	✓
<a href="http://linkedgeo.org/triplify/node3182603523">http://linkedgeo.org/triplify/node3182603523</a>	<a href="http://geos.wlu.edu.cn/resource/gadm/3/22">http://geos.wlu.edu.cn/resource/gadm/3/22</a>	100%	✓

```

<interlink>
  <Namespace prefix="rdf">http://www.w3.org/1999/02/22-rdf-syntax-ns#</Namespace>
  <Namespace prefix="geo">http://www.opengis.net/ont/geosparql#</Namespace>
  <LinkType>topoRelation</LinkType>
  <SourceDataset var="a">
    <RestrictTo> ?a rdf:type geo:Feature </RestrictTo>
  </SourceDataset>
  <TargetDataset var="b">
    <RestrictTo> ?b rdf:type geo:Feature </RestrictTo>
  </TargetDataset>
  <LinkCondition type="singleCompare" metric="">
    <Compare metric="within" weight="">
      <SourceParam name="c">
        <RestrictTo> ?a geo:hasGeometry ?point </RestrictTo>
        <RestrictTo> ?point geo:asWKT ?c </RestrictTo>
      </SourceParam>
      <TargetParam name="d">
        <RestrictTo> ?b geo:hasGeometry ?multipolygon </RestrictTo>
        <RestrictTo> ?multipolygon geo:asWKT ?d </RestrictTo>
      </TargetParam>
    </Compare>
  </LinkCondition>

```

**Figure 4.** Result of interlinking LinkedGeoData and GADM datasets.

**Table 6.** Linking railway stations in LinkedGeoData and Hubei administrative areas in GADM.

Dataset Statistics	Linking Result
Triples of railway stations in LinkedGeoData	15,297,705
Triples of GADM (Hubei province)	720
Links discovered	366
Time used (single-threaded program)	19.6 s
Time used (Apache Spark MapReduce program)	5 s
Time used—Silk	13.1 s
Time used—LIMES	9.5 s

### 5.1.2. Intersects Operator

In this case, we use NHD as the source dataset and GADM as the target dataset. NHD represents the water drainage network of the United States with features such as rivers, streams, canals, lakes, ponds, coastlines, dams, and stream gages. GADM is a high-precision global administrative boundary database. It encompasses administrative boundaries data of multiple levels including national, provincial, municipal, and district boundaries for all countries and regions worldwide. In the use case, we want to find rivers in the NHD that are “Intersects” with the administrative regions of Missouri State in GADM. We configure the <LinkType> to be geo:sfIntersects. Figure 5 gives the steps of setting geo:sfIntersects relationships between NHD and GADM. As mentioned in Figure 1 before, OGE-KG allows users to select a path in the RDF graph around a particular resource. For example, the path “?geomtry/geo:asWKT” would select the value of WKTLiteral associated with a geometry. Therefore, if we want to set spatial relationship links between features, the source and target paths in this example should be “?a/nhd-o:geometryProperty” and “?b/ngeo:geomtry/geo:asWKT” respectively. Then, we utilize the “geo:sfIntersects” in the GeoSPARQL vocabulary to link the two datasets. Table 7 shows the number of datasets triples and amount of discovered links. We also conducted comparative experiments. The computation times in Silk and LIMES are 30.8 s and 17.6 s, respectively, faster than the single-threaded time. It takes about 9.4 s to finish the linking process with a parallel program, which is the fastest. Similar to the conclusions drawn with “Within”, there is a significant improvement in computational efficiency when parallelized with MapReduce.

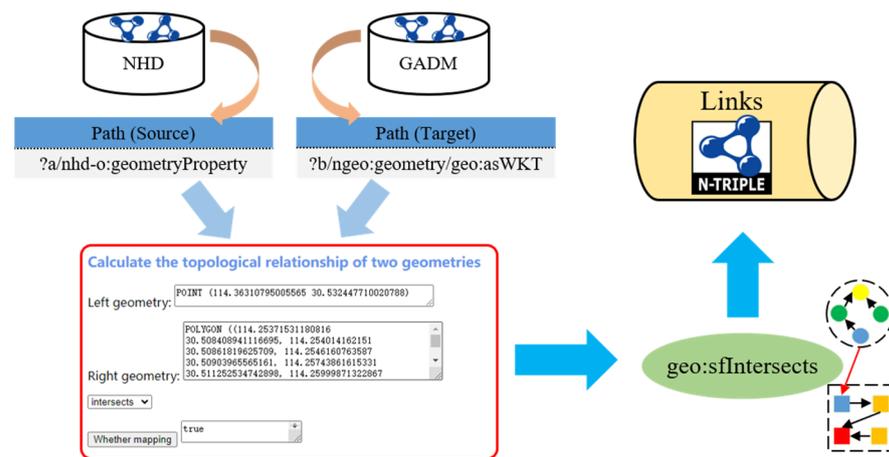


Figure 5. An example of generating “Intersects” links.

Table 7. Linking NHD and GADM-Missouri with the “Intersects” relationship.

Dataset Statistics	Linking Result
Triples of NHD	1,579,951
Triples of GADM (Missouri)	917
Links discovered	73,541
Time used (single-threaded program)	41.3 s
Time used (Apache Spark MapReduce program)	9.4 s
Time used—Silk	30.8 s
Time used—LIMES	17.6 s

### 5.2. Identity Links

Building identity links between geospatial datasets using both spatial and non-spatial properties will improve the accuracy of linking results. Some efforts have been made to integrate NUTS and GADM datasets based on spatial properties using Linked Data technologies [39,40]. Now, this kind of task can be carried out in the OGE knowledge graph framework. Take the Berlin administrative region data for example. Figure 6 shows the incongruity of geometric shapes about Berlin from NUTS (low resolution) and GADM (high resolution). To find the spatial equivalence of Berlin within the NUTS and GADM datasets, a linkage rule using the “min” aggregation function is specified. It aggregates the scores of string similarity and geometric similarity, where their minimum values are set to 0.9 and 0.7, respectively. The linkage rule is implemented and executed in the OGE-KG Data Interlinking Workbench (Figure 7). Following the geographical equivalence rules set above, we conducted additional tests based on NUTS (level 0) and GADM (level 0) in the same environment. Dataset statistics and linking results are given in Table 8. The NUTS-0 dataset with 35 objects and the GADM-0 dataset with 36 objects are used as experimental data. Since Silk lacks the Hausdorff metric operator, it cannot be compared with OGE-KG regarding performance; we only compared OGE-KG with LIMES. The results indicate that even when the number of points for each geometry object is particularly large, MapReduce achieves a certain level of performance improvement compared to single-threaded processing. Additionally, despite LIMES’ performance optimization for the Hausdorff metric in its source code, its performance still falls short of MapReduce.

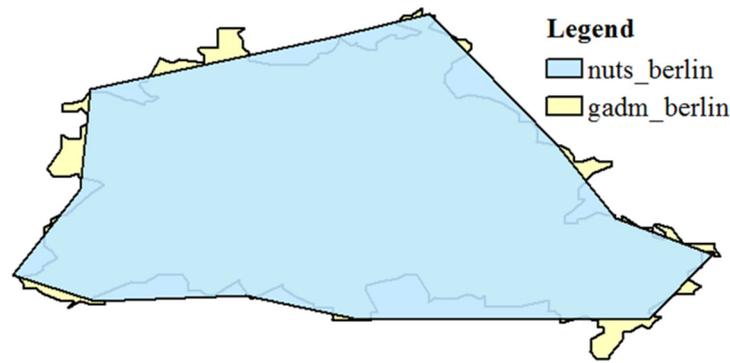


Figure 6. Incongruity of geometric shapes for Berlin between NUTS and GADM.

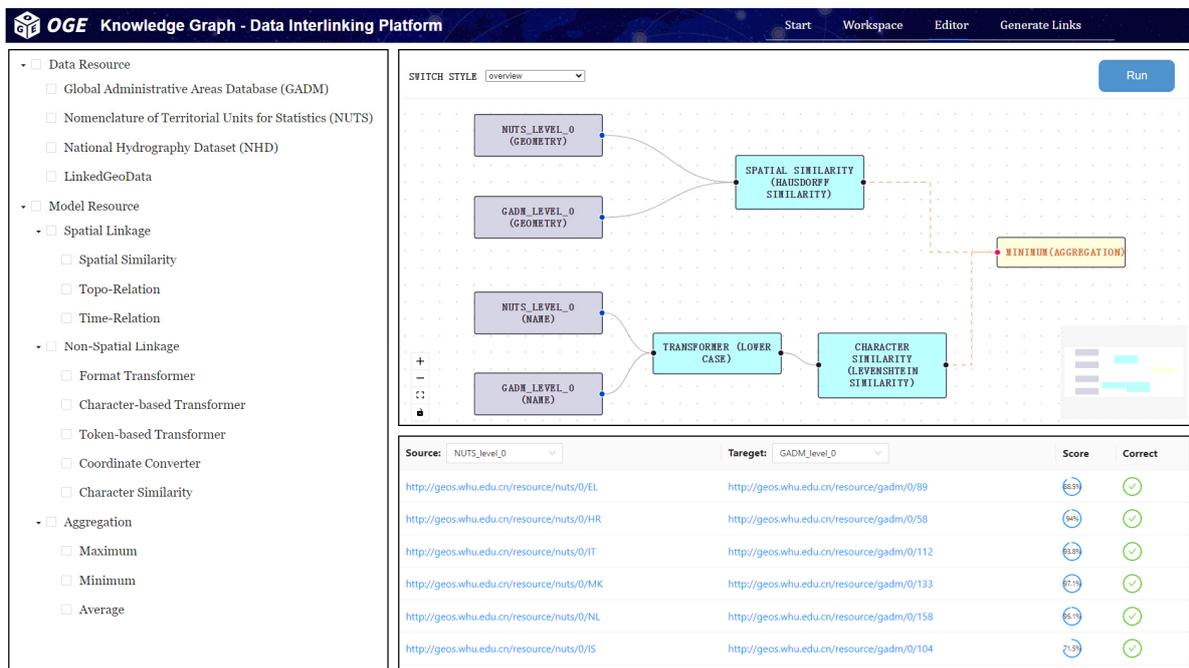


Figure 7. A screenshot of interlinking NUTS and GADM.

Table 8. Linking NUTS (level 0) and GADM (level 0) with the “Equals” relationship.

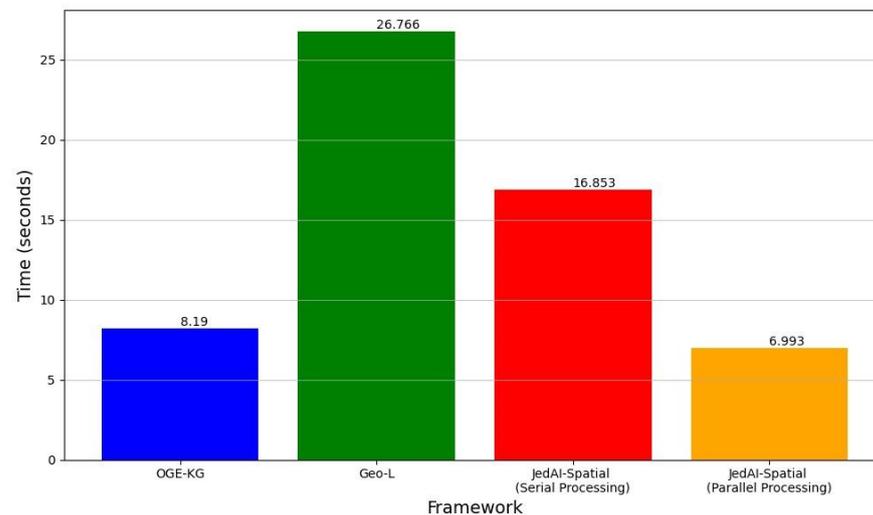
Dataset Statistics	Linking Result
Triples of NUTS (level 0)	77
Triples of GADM (level 0)	79
Links discovered	33
Time used (single-threaded program)	1307 s
Time used (Apache Spark MapReduce program)	823 s
Time used—LIMES	1165 s

### 5.3. Scalability Analysis

Traditional link discovery tools as mentioned above primarily utilize existing spatial operators for geographic linking to facilitate integration into their own system frameworks; hence, they often overlook scalability and efficiency. Considering this factor, recent research has mainly focused on optimizing these issues. Among these, Geo-L utilizes PostgreSQL and PostGIS for efficient indexing and spatial linking of geometric data, enhancing the efficiency of topological link discovery. JedAI-spatial is an open-source system that calculates topological relationships between datasets with geometric entities based on the DE9IM

model. Similar to OGE-KG, JedAI-spatial offers not only a serial version but also a parallel processing capability based on Apache Spark.

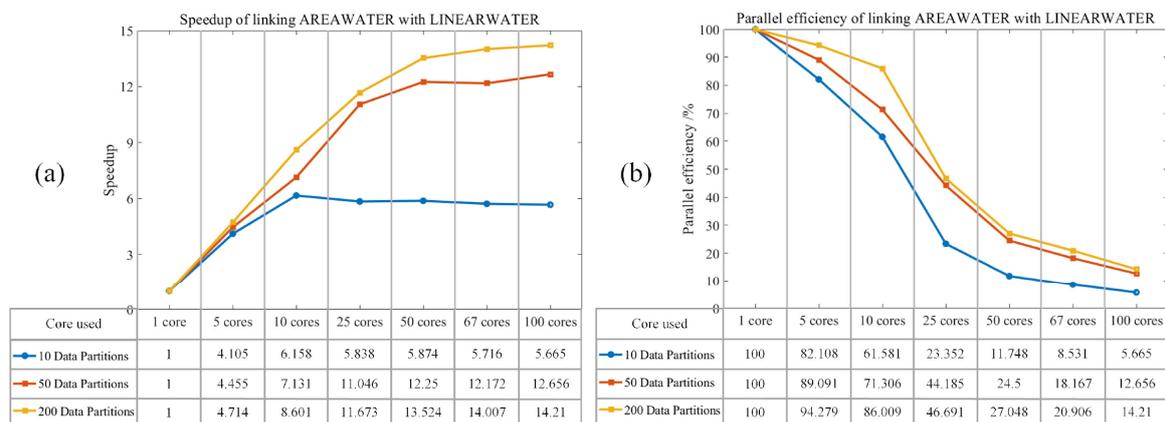
In this case, we first conducted tests of “Within” relation associations with Geo-L and JedAI-spatial (serial version) in the same environment using datasets of 165,000 Smart Points of Interest (SPOI) entities and 1782 NUTS entities as used in paper [11]. The comparative results are shown in Figure 8. In the OGE-KG program, the discovery process takes approximately 8 s, while Geo-L requires about 26 s, noting that this process only tests the mapping stage. Due to the use of the R-Tree-over-GiST index for managing geometric data in PostGIS during the data preprocessing stage, Geo-L can significantly improve association efficiency during the mapping stage. Compared to Geo-L, JedAI-spatial achieves higher association efficiency by incorporating not only tree-based algorithms but also grid-based and partition-based filters. However, according to the time-consuming results, MapReduce still maintains its processing advantage for large datasets. Additionally, OGE-KG’s processing time is slightly inferior to that of JedAI-spatial with the parallel scheme. This is due to its indexing filters effectively reducing computational overhead during the joining stage, although it requires configuring extra indices and grids for repartitioning during the preprocessing stage. Nevertheless, our framework still achieves efficient performance improvements.



**Figure 8.** Performance comparison of the “Within” spatial linking among different frameworks.

Next, we conducted speedup ratio and parallel efficiency calculations using parallel “intersects” tests to validate the applicability of the OGE-KG framework for large datasets. To provide a more intuitive comparison, we migrated the experiment environment to a multi-core server and tested the average processing time of each batch under different data partitioning (Case 1: 10 partitions; Case 2: 50 partitions; and Case 3: 200 partitions) and cluster computing resources. The server had specific parameters: Intel(R) Xeon(R) Gold 5220R CPU @ 2.20 GHz, 2 NUMA nodes, 100 cores, and 100 GB RAM. The datasets used included 2,292,766 entities of areal hydrologic data (AREAWATER) and 5,838,339 entities of linear hydrologic data (LINEARWATER) as utilized in the experiments of paper [12]. Figure 9 indicates the following: (1) The performance of Case 3 surpasses that of Case 1 and Case 2 because Case 3 has more partitions, allowing it to fully leverage the advantages of concurrency, especially in scenarios with a higher number of cores. (2) While using more cores can further reduce the time required for spatial correlation, excessive core allocation can lead to decreased parallel efficiency, resulting in decreased utilization efficiency of each core. Therefore, optimal core allocation is crucial in practical operations. (3) Increasing the number of partitions appropriately as the number of cores increases can enhance efficiency further. It is worth noting that in Case 3, the processing time under single-threaded conditions is approximately 4 h, whereas under optimal parallelization,

the processing time is reduced to 18 min. This significant reduction in processing time demonstrates the effectiveness of parallelization in maximizing computational efficiency. By efficiently utilizing the computational capacity of each core, the OGE-KG framework achieves remarkable improvements in processing time. Moreover, as computational resources increase and more data partitions are utilized, the framework demonstrates a higher degree of processing parallelization, further enhancing its scalability. Therefore, the results demonstrate the scalability of the OGE-KG framework in handling large-scale geospatial datasets and interconnecting operations.



**Figure 9.** Performance evaluation with different computing resources: (a) speedup and (b) parallel efficiency.

## 6. Discussion

We implement our design by conforming to the OpenGIS SFS. However, the ways geometries are represented vary largely. The SFS model is only used in limited geospatial datasets in the Web of Data. There are also some geospatial datasets employing the GeoRSS feature model [41], which is adopted by the W3C Geospatial Incubator Group for representing geospatial concepts in GeoOWL ontology [25]. Thus, there are still several issues that need to be tackled to make our extension more generic. Additionally, we discuss the characteristics of the OGE-KG framework and compare its advantages and disadvantages with other frameworks.

### 6.1. Coordinate Reference Systems

The coordinate reference system (CRS) (also called the spatial reference system), composed of a coordinate system, an Earth ellipsoid, a geodetic datum, and a map projection, is the essential metadata of a geometry. Each CRS can have a unique spatial reference system identifier. For example, the World Geodetic System 1984 (WGS84), the most widely used CRS in the LOD cloud, is identified using EPSG:4326. However, other coordinate reference systems are often used by local geographical organizations. For instance, the Ordnance Survey uses EPSG:27700 to record its geographic data. Thus, comparing a geometry in the Ordnance Survey dataset with another one in the GeoNames dataset (using WGS84) will return an incorrect result in the current OGE knowledge graph framework. To make it work right, a conversion from EPSG:27700 to EPSG:4326 is required before the comparison.

### 6.2. Literals for Geometries

Setting geospatial links within the geospatial LOD is also hindered by the variety of encoding methods. When generating geometry literals, triple store vendors may choose either WKT serialization or GML serialization. These two serializations have different geometry types. Some datasets employ WKT geometry types (e.g., GADM) to implement their geospatial triple stores and some use GML geometry types (e.g., Ordnance Survey datasets), while others use both of them (e.g., NHD). These varieties prevent geometry literals from being compared easily by spatial operators in the OGE-KG framework. Therefore,

the transformation of geometry literals is needed in the OGE-KG framework to make it applicable to more geospatial datasets.

### 6.3. Literals or Ontologies for Geometric Representation

GeoSPARQL works in a compact way such that the entire description of geometry is contained in a single literal (WKT and GML literals). The geometry ontology developed by Ordnance Survey works the same way but only focuses on the GML literal. Some geospatial datasets describe the geometry shape using a collection of points, each of which is represented within a single RDF node identified by a latitude/longitude pair such as the NUTS. These cases illustrate a fundamental issue: Literals or ontologies for geometric representation may vary among different data providers, thereby complicating spatial linkage between data sources or rendering them non-interoperable. To set spatial links between geospatial data sources with different geometric structures, on the one hand, standardization activities should be taken for LOD data providers, including the specification of preferred syntaxes for modeling geospatial data [40]; on the other hand, for link generation tools, flexible plugins to bridge these gaps are encouraged.

### 6.4. Characteristics and Comparative Analysis of OGE-KG

OGE-KG stands out in comparison to other link discovery frameworks due to its unique focus on integrating spatial relation computation and matching methods, including both relationship links and identity links. Unlike LinQL, Silk, and LIMES, OGE-KG supports spatial matching functions, which significantly enhances its capability for geospatial-aware link discovery. Moreover, OGE-KG surpasses its counterparts by enabling parallel computing, ensuring high efficiency in matching functions. Compared to existing open-source geospatial link frameworks, OGE-KG exhibits a parallel advantage in scalability, allowing it to handle larger datasets and more complex queries more effectively. Even when compared to the most advanced frameworks with spatial topological linking capabilities, the OGE-KG framework maintains its computational efficiency advantage while ensuring flexibility. Specifically, it requires simple data and process configuration to achieve parallel association without the need for complex constraint settings. Therefore, within the OGE-KG framework, it is possible to accurately describe features and their spatial relationships more rapidly using models such as SFS and RCC8. This provides a knowledge graph foundation for subsequent spatial association, reasoning, and even discovery of patterns hidden within the data. It aids in better understanding the intrinsic structure and characteristics of geographic spatial datasets.

One limitation of OGE-KG is its reliance on traditional methods and lack of integration with machine learning techniques. The current trend in link discovery moves beyond isolated spatial and semantic computations, transitioning toward direct association using neural networks [42]. This approach leverages the power of neural networks to capture complex relationships and patterns within data, potentially offering more holistic and efficient linkage solutions. Furthermore, the framework employs an automated identification and skipping mechanism to ensure computational accuracy in handling inconsistent, incomplete, or erroneous data. There is potential for further enhancement in this regard. Overall, the impact of data quality on linkage discovery results will bolster the applicability of this research to real-world datasets. Currently, manual judgment is utilized for assessing linkage quality, which represents a limitation of our approach.

Additionally, OGE-KG's adherence to open science principles and open GIS standards ensures its scalability and usability. We embrace the principles of FOSS4G (Free and Open-Source Software for Geospatial) with the goal of advancing the development and utilization of open-source geospatial software [43].

## 7. Conclusions and Future Work

In this paper, we demonstrate how to enrich the OGE-KG framework with geospatial awareness. The extended framework supports discovering spatial relationship links and

building spatial identity links within geospatial Linked Open Data. It is used to detect topological relations between different geospatial data sources on the LOD cloud and help build identity links between geospatial features within different datasets based on their geometries. The results indicate that, compared to other frameworks, OGE-KG significantly improves the efficiency of linking different geospatial data sources using MapReduce.

The future work will focus on the issues proposed in the discussion by implementing geospatial transformation operators in the OGE-KG framework. Handling various geometry representations will be the first step followed by the conversion function between different spatial reference systems. In addition, we will continue to pay attention to the efficiency issue as well to achieving better runtime performance.

**Author Contributions:** Conceptualization, Lianlian He and Ruixiang Liu; methodology, Lianlian He; software, Ruixiang Liu; validation, Lianlian He and Ruixiang Liu; formal analysis, Lianlian He; investigation, Lianlian He; resources, Lianlian He; data curation, Lianlian He; writing—original draft preparation, Lianlian He and Ruixiang Liu; writing—review and editing, Lianlian He and Ruixiang Liu; visualization, Lianlian He; supervision, Lianlian He; project administration, Lianlian He. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was supported by the Educational Commission of Hubei Province of China.

**Data Availability Statement:** The data for the experiments in Section 5 are available online. Table 2 presents links to all datasets used in the experiments, including LinkedGeoData, GADM, NUTS, and NHD.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Yue, P.; Guo, X.; Zhang, M.; Jiang, L.; Zhai, X. Linked data and SDI: The case on Web geoprocessing workflows. *ISPRS-J. Photogramm. Remote Sens.* **2016**, *114*, 245–257. [CrossRef]
2. Bizer, C. The emerging web of linked data. *IEEE Intell. Syst.* **2009**, *24*, 87–92. [CrossRef]
3. Bizer, C.; Heath, T.; Berners-Lee, T. Linked data—The story so far. *Int. J. Semant. Web Inf. Syst.* **2009**, *79*, 637–638. [CrossRef]
4. Lundberg, E.; Borner, G.H.H. Spatial proteomics: A powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 285–302. [CrossRef] [PubMed]
5. Faria, B.; Perdigão, D.; Oliveira, H.G. Question answering over linked data with GPT-3. In *12th Symposium on Languages, Applications and Technologies (SLATE 2023)*; Vila do Conde, Portugal, 26–28 June 2023; Simões, A., Berón, M.M., Portela, F., Eds.; Schloss Dagstuhl: Wadern, Germany, 2023; Volume 113, pp. 1–15.
6. Zhu, R.; Janowicz, K.; Cai, L.; Mai, G. Reasoning over higher-order qualitative spatial relations via spatially explicit neural networks. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 2194–2225. [CrossRef]
7. Heath, T.; Bizer, C. *Linked data: Evolving the Web into a Global Data Space*; Springer Nature: Berlin, Germany, 2022.
8. Volz, J.; Bizer, C.; Gaedke, M.; Kobilarov, G. Silk—A link discovery framework for the web of data. In Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, 20–24 April 2009.
9. Ngomo, A.C.N.; Auer, S. Limes—A time-efficient approach for large-scale link discovery on the web of data. In Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona Catalonia, Spain, 16–22 July 2011.
10. Hassanzadeh, O.; Lim, L.; Kementsietsidis, A.; Wang, M. A declarative framework for semantic link discovery over relational data. In Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, 20–24 April 2009.
11. Zinke-Wehlmann, C.; Kirschenbaum, A. Geo-L: Topological Link Discovery for Geospatial Linked Data Made Easy. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 712. [CrossRef]
12. Papamichalopoulos, M.; Papadakis, G.; Mandilaras, G.; Siampou, M.; Mamoulis, N.; Koubarakis, M. Three-dimensional geospatial interlinking with jedai-spatial. *J. Web Semant.* **2024**, *81*, 100817. [CrossRef]
13. Vicente-Saez, R.; Martinez-Fuentes, C. Open Science now: A systematic literature review for an integrated definition. *J. Bus. Res.* **2018**, *88*, 428–436. [CrossRef]
14. Sui, D. Opportunities and impediments for open GIS. *Trans. GIS.* **2014**, *18*, 1–24. [CrossRef]
15. The Linked Open Data Cloud. Available online: <https://lod-cloud.net/> (accessed on 21 January 2024).
16. Zhang, X.; Chen, N.; Chen, Z.; Wu, L.; Li, X.; Zhang, L.; Di, L.; Gong, J.; Li, D. Geospatial sensor web: A cyber-physical infrastructure for geoscience research and application. *Earth-Sci. Rev.* **2018**, *185*, 684–703. [CrossRef]
17. Schmachtenberg, M.; Bizer, C.; Paulheim, H. Adoption of the linked data best practices in different topical domains. In Proceedings of the 13th International Semantic Web Conference, Riva del Garda, Italy, 19–23 October 2014.
18. GeoNames. Available online: <http://www.geonames.org/> (accessed on 21 January 2024).

19. Stadler, C.; Lehmann, J.; Höffner, K.; Auer, S. Linkedgeodata: A core for a web of spatial open data. *Semant. Web.* **2012**, *3*, 333–354. [CrossRef]
20. GADM-RDF. Available online: <http://gadm.geovocab.org/> (accessed on 21 January 2024).
21. Wilke, A.; Ngonga Ngomo, A.C. LauNuts: A Knowledge Graph to identify and compare geographic regions in the European Union. In Proceedings of the 20th European Semantic Web Conference, Hersonissos, Crete, Greece, 28 May–1 June 2023.
22. Goodwin, J.; Dolbear, C.; Hart, G. Geographical linked data: The administrative geography of great britain on the semantic web. *Trans. GIS.* **2008**, *12*, 19–30. [CrossRef]
23. Access National Hydrography Products. Available online: <https://www.usgs.gov/national-hydrography/access-national-hydrography-products> (accessed on 21 January 2024).
24. Ronzhin, S.; Folmer, E.; Lemmens, R.; Mellum, R.; von Brasch, T.E.; Martin, E.; Romero, E.L.; Kytö, S.; Hietanen, E.; Latvala, P. Next generation of spatial data infrastructure: Lessons from linked data implementations across Europe. *Int. J. Spat. Data Infrastruct. Res.* **2019**, *14*, 83–107.
25. Lieberman, J.; Singh, R.; Goad, C. *W3C Geospatial Vocabulary—W3C Incubator Group Report*; W3C Geospatial Incubator Group: Wakefield, MA, USA, 23 October 2007.
26. Perry, M.; Herring, H. *OGC GeoSPARQL—A Geographic Query Language for RDF Data*; Version 1.0, OGC 11-052r4; Open Geospatial Consortium, Inc.: Wayland, MA, USA, 2012.
27. Salas, J.M.; Harth, A. *NeoGeo Vocabulary Specification—Madrid Edition. Public draft*; Open Geospatial Consortium, Inc.: Wayland, MA, USA, 07 February 2012.
28. Atemezing, G.A.; Troncy, R. Comparing vocabularies for representing geographical features and their geometry. In Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web, Boston, MA, USA, 12 November 2012.
29. Auer, S.; Lehmann, J.; Hellmann, S. Linkedgeodata: Adding a spatial dimension to the web of data. In Proceedings of the 8th International Semantic Web Conference, ISWC 2009, Chantilly, VA, USA, 25–29 October 2009.
30. Patni, H.; Henson, C.; Sheth, A. Linked sensor data. In Proceedings of the 2010 International Symposium on Collaborative Technologies and Systems, Chicago, IL, USA, 17–21 May 2010.
31. Yuan, J.; Yue, P.; Gong, J.; Zhang, M. A linked data approach for geospatial data provenance. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 5105–5112. [CrossRef]
32. Winkler, W.E. String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In Proceedings of the Survey Research Methods Section, Alexandria, VA, USA, 6–9 August 1990.
33. Neumaier, S.; Polleres, A. Enabling spatio-temporal search in open data. *J. Web Semant.* **2019**, *55*, 21–36. [CrossRef]
34. Zhang, F.; Lu, Q.; Du, Z.; Chen, X.; Cao, C. A comprehensive overview of RDF for spatial and spatiotemporal data management. *Knowl. Eng. Rev.* **2021**, *36*, e10. [CrossRef]
35. Ma, X. Linked Geoscience Data in practice: Where W3C standards meet domain knowledge, data visualization and OGC standards. *Earth Sci. Inform.* **2017**, *10*, 429–441. [CrossRef]
36. Aquino, J.; Davis, M. *JTS Topology Suite Technical Specifications*, version 1.4; Vivid Solution, Inc.: Victoria, BC, Canada, 2003.
37. Dean, J.; Ghemawat, S. MapReduce: Simplified data processing on large clusters. *Commun. ACM.* **2008**, *51*, 107–113. [CrossRef]
38. Zaharia, M.; Chowdhury, M.; Franklin, M.J.; Shenker, S.; Stoica, I. Spark: Cluster computing with working sets. In Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing, Berkeley, CA, USA, 22–25 June 2010.
39. Salas, J.; Harth, A. Finding spatial equivalences across multiple RDF datasets. In Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web, Bonn, Germany, 23 October 2011.
40. Harth, A.; Gil, Y. Geospatial data integration with linked data and provenance tracking. In Proceedings of the W3C/OGC Linking Geospatial Data Workshop, London, UK, 5–6 March 2014.
41. Reed, C. *An Introduction to GeoRSS: A Standards Based Approach for Geo-Enabling RSS Feeds*; Version 1.0.0; OGC 06-050r3; Open Geospatial Consortium, Inc.: Wayland, MA, USA, 2006.
42. He, L.; Li, H.; Zhang, R. A Semantic-Spatial Aware Data Conflation Approach for Place Knowledge Graphs. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 106. [CrossRef]
43. Brovelli, M.A.; Minghini, M.; Moreno-Sanchez, R.; Oliveira, R. Free and open source software for geospatial applications (FOSS4G) to support Future Earth. *Int. J. Digit. Earth.* **2017**, *10*, 386–404. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.