



Geospatial data ontology: the semantic foundation of geospatial data integration and sharing

Kai Sun, Yunqiang Zhu, Peng Pan, Zhiwei Hou, Dongxu Wang, Weirong Li & Jia Song

To cite this article: Kai Sun, Yunqiang Zhu, Peng Pan, Zhiwei Hou, Dongxu Wang, Weirong Li & Jia Song (2019) Geospatial data ontology: the semantic foundation of geospatial data integration and sharing, Big Earth Data, 3:3, 269-296, DOI: [10.1080/20964471.2019.1661662](https://doi.org/10.1080/20964471.2019.1661662)

To link to this article: <https://doi.org/10.1080/20964471.2019.1661662>



© 2019 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the CASEarth Strategic Priority Research Programme.



Published online: 16 Sep 2019.



Submit your article to this journal [↗](#)



Article views: 6704



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 16 View citing articles [↗](#)

RESEARCH ARTICLE



Geospatial data ontology: the semantic foundation of geospatial data integration and sharing

Kai Sun^{a,b}, Yunqiang Zhu^{a,c,d}, Peng Pan^{e,f}, Zhiwei Hou^{a,b}, Dongxu Wang^g, Weirong Li^{a,b} and Jia Song^{a,c,d}

^aState Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, People's Republic of China; ^bUniversity of Chinese Academy of Sciences, Beijing, People's Republic of China; ^cJiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, People's Republic of China; ^dCenter for Collaborative Innovation in Baiyangdian River Basin Ecological Protection and Beijing-Tianjin-Hebei Sustainable Development, Baoding, People's Republic of China; ^eAppraisal Center for Environment and Engineering, Ministry of Ecology and Environment, Beijing, People's Republic of China; ^fState Environmental Protection Key Laboratory of Numerical Modeling for Environment Impact Assessment, Beijing, People's Republic of China; ^gBeijing GEOWAY Software Co., Ltd, Beijing, People's Republic of China

ABSTRACT

Effective integration and wide sharing of geospatial data is an important and basic premise to facilitate the research and applications of geographic information science. However, the semantic heterogeneity of geospatial data is a major problem that significantly hinders geospatial data integration and sharing. Ontologies are regarded as a promising way to solve semantic problems by providing a formalized representation of geographic entities and relationships between them in a manner understandable to machines. Thus, many efforts have been made to explore ontology-based geospatial data integration and sharing. However, there is a lack of a specialized ontology that would provide a unified description for geospatial data. In this paper, with a focus on the characteristics of geospatial data, we propose a unified framework for geospatial data ontology, denoted GeoDataOnt, to establish a semantic foundation for geospatial data integration and sharing. First, we provide a characteristics hierarchy of geospatial data. Next, we analyze the semantic problems for each characteristic of geospatial data. Subsequently, we propose the general framework of GeoDataOnt, targeting these problems according to the characteristics of geospatial data. GeoDataOnt is then divided into multiple modules, and we show a detailed design and implementation for each module. Key limitations and challenges of GeoDataOnt are identified, and broad applications of GeoDataOnt are discussed.

ARTICLE HISTORY

Received 30 May 2019
Accepted 25 August 2019

KEYWORDS

Geospatial data; ontology; geospatial semantics; semantic heterogeneity; data integration; data sharing

1. Introduction

Rapid access to desired geospatial data is the key to data-driven geography in the context of big Earth data (Miller & Goodchild, 2015), which requires efficient geospatial data integration and sharing. Unfortunately, the processes of integration and sharing face many

challenges. One of the challenges is semantic heterogeneity caused by the characteristics of multiple sources, types, and forms of geospatial data. Many efforts for geospatial data integration and sharing, based on the idea of metadata standards, have been made in the past. Examples include the Federal Geographic Data Committee (FGDC)¹, the National Spatial Data Infrastructure (NSDI)², the International Organization for Standardization (ISO) 19115³, and the National Map from the United States Geological Survey (USGS) (Budak Arpinar et al., 2006). However, these efforts only partially addressed the semantic issues (Baglioni, Giovannetti, Masserotti, Renso, & Spinsanti, 2008; Lutz & Klien, 2006).

A more promising approach to solve heterogeneity problems is to develop and use ontologies (Baglioni, Giovannetti, Masserotti, Renso, & Spinsanti, 2008; Hu, 2017). Ontologies are formal and explicit specifications of shared concepts in a machine-readable way (Gruber, 1993; Studer, Benjamins, & Fensel, 1998). Ontologies can be used to provide a semantic description for geospatial data and help computers to understand the semantic meaning implied in the content of geospatial data. Ontologies can also be used to describe the relationships between semantic entities, and the reasoning mechanism of ontologies can help to discover more implicit relationships (Peuquet, 2002). Thanks to these advantages, ontologies are among the best solutions to implement geospatial data integration and sharing on the semantic level.

In the 1990s, ontologies were introduced in the domain of geographic information science (Deliiska, 2007; Hahmann & Stephen, 2018; Hu et al., 2013; Li et al., 2017a; Liu, Li, Shen, Yang, & Luo, 2018; Reitsma, Laxton, Ballard, Kuhn, & Abdelmoty, 2009; Stock et al., 2012; Sun, Wang, Ranjan, & Wu, 2015; Winter, 2001). These past decades have witnessed a rapid development of geo-ontologies, namely extensions of ontologies in geographic information science (GIScience). The insights from the previous studies can be organized into two aspects (Agarwal, 2005). The first aspect is trying to provide a unified framework of shared geographic concepts by constructing a top-level geo-ontology, such as the seven-level hierarchy of geo-ontology (Couclelis, 2010), 5-tier geo-ontology (Frank, 2001), and a semantic reference system (Kuhn, 2003). The second aspect is to develop a domain-level geo-ontology to facilitate specific tasks. Many task-oriented geo-ontologies focused on different applications, including information integration (Bittner, Donnelly, & Smith, 2009; Buccella, Cechich, & Fillottrani, 2009; Buccella et al., 2011; Chen, Sabri, Rajabifard, & Agunbiade, 2018; Comber, Fisher, & Wadsworth, 2011; Fonseca, Egenhofer, Agouris, & Camara, 2002; Hong & Kuo, 2015; Uitermark, van Oosterom, Mars, & Molenaar, 2005; Wang, Ma, & Chen, 2018; Zhao, Zhang, Wei, & Peng, 2008), information retrieval (Baglioni et al., 2008; Gui et al., 2013; Li, Goodchild, & Raskin, 2014; Lutz & Klien, 2006; Peachavanish & Karimi, 2007; Sun et al., 2015; Wiegand & García, 2007), data interoperability (Brodeur, Bedard, Edwards, & Moulin, 2010; Kuo & Hong, 2016), semantic representation of geospatial data (Dean, 2007; Fonseca & Llano, 2011; Schuurman & Leszczynski, 2006; Sun et al., 2015), geospatial clustering (Wang, Gu, Ziebelin, & Hamilton, 2010), spatial decision support (Li, Raskin, Goodchild, & Janowicz, 2012), disaster management and response (Xu, Nyerges, & Nie, 2014; Zhang, Zhao, & Li, 2010), web service discovery (Chen, Chen, Hu, & Di, 2011; Klien, Lutz, & Kuhn, 2006), composition of geoprocessing services (Li et al., 2015; Lutz, 2007), urban environment analysis (Fonseca, Egenhofer, Davis, & Borges, 2000), and environment modeling (Fallahi, Frank, Mesgari, & Rajabifard, 2008). This paper mainly focuses on the studies relevant to geospatial data integration and sharing.

An integration process includes two steps: semantic enrichment and mapping discovery (Buccella et al., 2009). Semantic enrichment is to annotate the data with essential semantic information. Mapping discovery is to find the mappings of semantic annotations of different data. According to the roles of ontologies in these steps, ontology-based methods for data integration can be classified into three categories: single-ontology methods, multiple-ontologies methods, and hybrid methods (Wache et al., 2001). A single top-level ontology was developed to describe all the relations between different kinds of basic entities that were frequently used to annotate the relations between data in the integration process (Bittner et al., 2009). Hong and Kuo (2015) established multiple bridge ontologies to determine the relations of concepts for cross-domain geospatial data integration. Chen et al. (2018) also transformed domain knowledge into multiple domain ontologies that were used to bond the data and ontology via semantic enrichment. Buccella et al. (2011) annotated geospatial data with multiple domain ontologies at the semantic enrichment step, and then a global ontology was introduced to complete the semantic mapping by combining the domain ontologies.

To enable geospatial data sharing, ontologies were used to implement geospatial data discovery at the semantic level in the previous studies. Generally, an ontology was developed to provide a formal and hierarchical semantic annotation for geospatial data, and users can carry out data discovery on the deeper semantic details of geospatial data based on the developed ontology (Stock et al., 2013). Lutz and Klien (2006) first proposed a relatively complete framework, including components for creating an ontology, registration mappings to describe the relationships between the ontology and feature types; additionally, a user interface was implemented so that the users could conduct data discovery. Andrade, Baptista, and Davis (2014) also used an ontology to improve data discovery in terms of space, time, and theme. Wiegand and García (2007) formalized the relationships between different tasks and information on data sources for each task using an ontology; thus, data sources required by a specific task could be found by reasoning on the ontology.

Another focus of geospatial data sharing is to enhance semantic interoperability of geospatial data (Fallahi et al., 2008). An ontology-based conceptual framework to describe the different configurations involved in geospatial data interoperability was proposed by Brodeur et al. (2010), which included five ontological phases (including reality, a cognitive model of reality, and a set of conceptual representations, etc.). Kuo and Hong (2016) leveraged a bridge ontology to generate semantic mappings of cross-domain concepts to facilitate geospatial data interoperability.

Geospatial metadata was also an important aspect for data sharing. Some studies focused on improving the quality of geospatial metadata using ontologies. Sun et al. (2015) initially quantitatively measured two types of metadata uncertainty (incompleteness and inaccuracy) via possibilistic logic and probabilistic statistics; subsequently, an ontology that includes these uncertainties was developed to improve the quality of metadata. Schuurman and Leszczynski (2006) proposed ontology-based geospatial metadata, which added some non-spatial fields to the existing metadata schemas to describe geospatial data, such as sampling methodologies or a measurement specification.

The aforementioned studies greatly facilitated the progress in this field. However, most of these studies developed general ontologies for GIScience domain. The ontologies developed for geospatial data integration and sharing did not consider all aspects of semantic heterogeneity. Only spatial, temporal, and thematic information of

geospatial data was used, ignoring the information of provenance and morphology. Hence, integration and sharing of heterogeneous geospatial data require a specialized geospatial data ontology to help deal with the existing semantic issues.

To address these requirements, we initially propose a multilevel characteristics hierarchy of geospatial data and analyze the semantic problems in geospatial data integration and sharing. Next, we construct an integral ontological framework, named “an ontology for integrating and sharing geospatial data” (hereafter called GeoDataOnt), to gather and integrate semantic knowledge of geospatial data. GeoDataOnt is a specialized geospatial data ontology that is specifically designed and constructed with a focus on characteristics of geospatial data. GeoDataOnt is more complete and can provide a unified and standardized representation for semantic information of geospatial data. Thus, the proposed ontology will lay the semantic foundation for achieving a common understanding of semantic information of geospatial data and enable a seamless integration and full sharing of geospatial data. The main contributions of this paper are as follows:

- a systematic analysis of the semantic problems in geospatial data integration and sharing;
- an integral general framework to provide the overall structure and composition of GeoDataOnt; and
- a multilayer modular GeoDataOnt base that includes the semantic knowledge about geospatial data.

The remainder of this paper is structured as follows. [Section 2](#) describes the characteristics hierarchy of geospatial data. [Section 3](#) presents a systematic analysis of the semantic problems in geospatial data integration and sharing. The general framework of GeoDataOnt is presented in [Section 4](#). [Section 5](#) shows the detailed design and implementation of GeoDataOnt. [Section 6](#) discusses the key limitations, challenges, and broad applications of GeoDataOnt. [Section 7](#) summarizes this paper.

2. Characteristics hierarchy of geospatial data

The design and construction of GeoDataOnt is based on the characteristics hierarchy of geospatial data. Thus, in this section, the characteristics hierarchy is initially proposed according to the lifecycle of geospatial data. The typical lifecycle for geospatial data can be divided into six stages:

- (1) Data acquisition. Geospatial data are acquired by observing geographic phenomena or measuring geographic features with specific tools or instruments by data producers.
- (2) Data processing. Some processing steps will be adopted to make the data more standardized or easier to use.
- (3) Data storage. The data will be represented in a suitable data structure and format.
- (4) Data management. The metadata of the processed data will be recorded according to the data content. The data will be managed using a professional database or data catalog.

- (5) Data sharing. The data will be shared via a web platform for data sharing or a hard copy.
- (6) Applying data. The users will retrieve and download the desired data from the sharing platform. Next, the users will apply GIS operations on geospatial data to complete an application task.

Multiple characteristics of geospatial data are formed gradually in the first three steps, as shown in Figure 1. The characteristics hierarchy of geospatial data includes three levels: overall level, compound level, and elementary level. The overall characteristics can be separated into three types at the compound level: essential, morphologic, and provenance characteristics. Before starting to acquire geospatial data, the three basic questions “what, where, and when” about the data – namely, theme, spatial and temporal coverage – should be clarified (Zhu et al., 2017). These three elementary characteristics constitute the essential characteristics. These characteristics are called essential because they are the identity of geospatial data and can be used to distinguish geospatial data from one another.

Morphologic characteristics describe the internal structure and external shape of geospatial data. Elementary morphologic characteristics include spatial accuracy, spatial granularity, temporal granularity, coordinate reference system (CRS), temporal reference system, time format, units of measurement, language, map symbol, data type, and data format, etc. However, not all types of morphologic information will appear in all geospatial data (some of them may be missing).

Provenance characteristics describe how geospatial data have been derived and include elementary characteristics of tools or instruments used to collect data, algorithms or software employed to process data, and people who conduct these steps, etc.

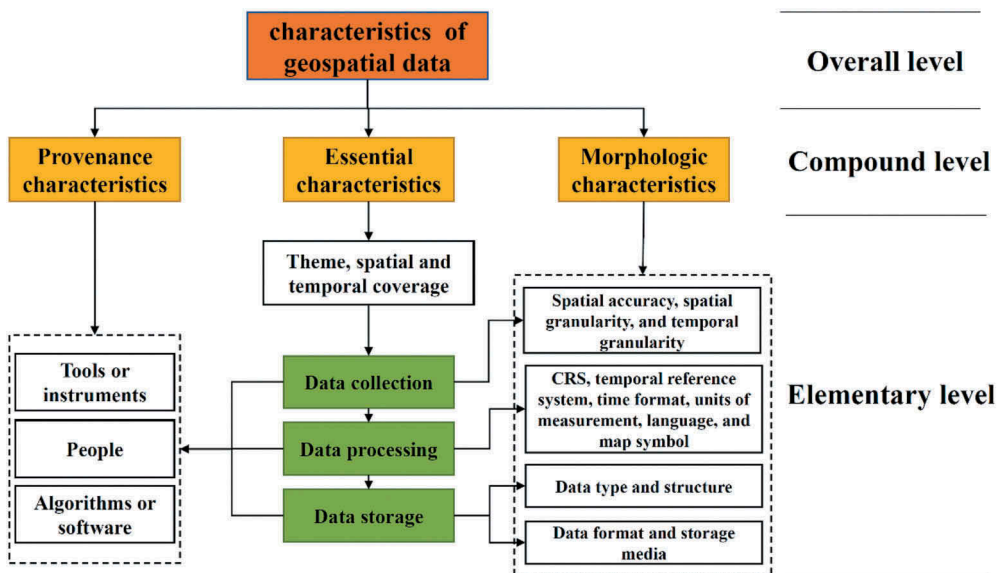


Figure 1. Characteristics hierarchy of geospatial data.

Geospatial data may be a result of reprocessing of source data. In this case, provenance information of the newly generated data will refer to the source data.

An example to show the characteristics hierarchy of geospatial data has been provided in Table 1. This example data is from the Chinese National Earth System Science Data Sharing Infrastructure⁴, and is titled “China 1KM Land Use/Cover Dataset (2000).”

We proposed a typical lifecycle and characteristics hierarchy of geospatial data because they can make the following design and implementation of GeoDataOnt more complete. However, this does not mean that GeoDataOnt is applicable only to geospatial data that has a complete lifecycle and all the described characteristics. GeoDataOnt can also apply to geospatial data with particular stages of the lifecycle and only the partial characteristics.

3. Analysis of semantic problems in geospatial data integration and sharing: a perspective of characteristics of geospatial data

Semantic heterogeneity refers to different interpretations caused by different representations of geospatial data. Hakimpour (2003) classified semantic heterogeneity of geospatial data into four categories. Taking the geospatial data about road features as an example, the possible four types of heterogeneity can be explained as follows:

- 1. Heterogeneity in the conceptual model: A road is represented as an object class or a relation, respectively, in different geospatial datasets.
- 2. Heterogeneity in the spatial model: The roads are represented by polygons or lines, respectively, in different geospatial datasets.
- 3. Heterogeneity in the structure: The roads are represented by lines in two geospatial datasets. However, these two datasets use different formats.
- 4. Heterogeneity in semantics: Due to different definitions of roads, one feature that seems to be a road may be either regarded as a road or assigned to other feature types.

Table 1. An example for the characteristics hierarchy of geospatial data.

Compound characteristics	Elementary characteristics	Instances
Essential characteristics	Spatial characteristics	China
	Temporal characteristics	2000
	Thematic characteristics	Land user/cover
Morphological characteristics	Spatial accuracy	1 km
	Data format	ArcView GIS ASCII
	Projection	Albers
	Units	meter
	Spheroid	Krasovsky
Provenance characteristics	Data source	Remote-sensing image of SPOT4, MODIS, AVHRR, etc.
	Algorithms or tools	GIS operations of clip, mosaic or transforming vector to raster

Goh (1997) also identified three types of semantic heterogeneity:

1. Confounding heterogeneity: two terms seem to have the same meaning, but refer to different meanings in reality, because of certain reasons such as under different contexts.
2. Scaling heterogeneity: the geographic features are measured under different geographic reference systems.
3. Naming heterogeneity: the features are named using different naming schemes.

These two insights for semantic heterogeneity in geospatial data are all proposed in a relatively macroscopical perspective and lack a systematic analysis of detailed semantic issues in a microcosmic perspective. For the specific task of geospatial data integration and sharing, they are too general, not sufficiently pertinent, and incomplete. For example, they did not identify the semantic heterogeneity in provenance and morphologic characteristics of geospatial data.

Therefore, from a perspective of characteristics of geospatial data, we propose a detailed analysis of semantic problems in geospatial data integration and sharing, as follows.

(1) Heterogeneity in essential characteristics

Heterogeneity in essential characteristics can be divided into theme heterogeneity, spatial heterogeneity, and temporal heterogeneity. Thematic characteristics of geospatial data are usually described in two parts: theme terms and category. The same theme terms may have different meanings due to different definitions adopted by different systems or organizations. Geospatial data using different classification systems of categories, such as the classification systems of Global Change Master Directory⁵ and ISO19115, will lead to heterogeneity in theme. Different classification systems may lead to difference in the categories in terms of granularity. Some classification systems provide high-resolution categories, and geospatial data are classified in a fine-grained manner. In contrast, other classification systems classify geospatial data in a coarse-grained way. Even using the same classification system, geospatial data are possibly assigned to different categories. There are usually some category hierarchies with different levels of granularity in a classification system. In different situations, geospatial data may be assigned to categories from different hierarchies in one classification system, which may also lead to heterogeneity in theme of geospatial data.

Spatial coverage of geospatial data is usually described by feature names. Two aspects of heterogeneity in spatial coverage can be considered: the feature names themselves and the spatial footprints corresponding to the feature names. Heterogeneity in the feature names is caused by different forms (such as codes or aliases) of feature names or by feature names with different levels of completeness, such as abbreviations. For a certain feature, its spatial footprint may be different for some reasons. First, the spatial footprint of a feature may change over time. Second, different levels of details for feature footprints may be employed under different scales. Third, some places are with a vague boundary because different people may understand them differently.

Heterogeneity in temporal coverage can also be inspected in two aspects. First, there are three types of time to represent temporal coverage of geospatial data. Event time

refers to the time when a phenomenon or event exists or occurs. Database time (or transaction time) refers to the time when a database record for an event is created, updated or deleted. Data time refers to the time when a phenomenon or event is observed or collected. Second, temporal coverage may be recorded with different levels of detail.

(2) Heterogeneity in morphologic characteristics

Heterogeneity in morphologic characteristics will be identified according to the order in which sub-characteristics are formed in the workflow of geospatial data. The difference in spatial accuracy is closely related to the data type of geospatial data. The spatial accuracy of vector data is represented by a map scale, such as 1:100,000, while that of raster data is represented with spatial resolution, such as 1 m. The spatial accuracies of different geospatial data with the same data type can also be different. Heterogeneity in spatial accuracy may lead to a difference in the level of details and map symbols of the features. One feature will be represented in different geometries under different map scales. For example, on a map of a small scale, one road is generally denoted in a single-line format, and one county is generally denoted with point geometry. However, on a map of a large scale, one road may be denoted in a double-line format, and polygon geometry is used to represent one county. Spatial and temporal granularities refer to statistical caliber of geospatial data, and different levels of granularity may also cause difference in geospatial data.

In the stage of expressing and visualizing the collected data, different CRSs and temporal reference systems can be employed. The difference in CRS may occur in datum, spheroid, or projection, etc. Different calendars or temporal origins may contribute to a difference in temporal coordinate systems. Even for the time with the same temporal coordinate system, difference also occurs due to different formats employed to represent the same time, such as yyyy-mm-dd of ISO 8601⁶ and mm/dd/yyyy in America (yyyy, mm, and dd represent year, month, and day, respectively). Month may be represented as a name, as a number (1–12), or as a Roman numeral (I–XII). The units of measurement and the languages may also be different at this stage. Features are measured with different units of measurement. For example, the water level may be measured in meters or centimeters. Geospatial data may be described in different languages. For maps, using different map symbol systems may cause heterogeneity.

The collected data will be organized and stored in a machine-readable manner using some software in the following stage. Heterogeneity occurred at this stage is reflected at different levels. The first level of heterogeneity is the difference in data type. There are different data types which can be used to organize geospatial data. For one specific data type, different data structures can be employed to organize data. Thus, difference in data structure is the second level. The third and fourth levels of heterogeneity refer to the data format and storage media, respectively. Specifically, different data formats with the same data structure may be used to formalize the data, and formalized data may be stored in different storage media.

(3) Heterogeneity in provenance characteristics

Heterogeneity can also be exhibited in the provenance characteristics of geospatial data. Different tools or instruments can be used to collect or produce data, which may cause difference in data. When data is produced by reprocessing the original data using some algorithms (e.g., spatial interpolation), different original datasets or algorithms may also cause difference in data.

4. General design of GeoDataOnt

The ontological framework is presented in two levels: general and detailed design. This section describes the general design from two parts: the logic model and general framework of GeoDataOnt.

4.1. Logic model of GeoDataOnt

GeoDataOnt can be defined as a set of geographic entities concerning geospatial data and the possible relations between entities, as defined in Equation (1).

$$\text{GeoDataOnt} = \left\{ \left(E, R_{(E_i, E_j)} \right) \mid E_i, E_j \in E, 0 \leq i \leq n, 0 \leq j \leq n \right\} \quad (1)$$

Here, E is the set of all relevant entities, n is the size of this set, and R represents the relations between entities from this set. The relations between entities might have three types of property: symmetry, transitivity, and reversibility, which can help to enhance the reasoning mechanism of an ontology.

The entities set includes three types of entities and can be defined as in Equation (2).

$$E = \{C \cup P \cup I \mid C, P, I \subset D\} \quad (2)$$

The following notation is used:

- C is the finite set of classes. Classes represent the types of objects or kinds of things.
- P is the finite set of properties. There are two types of property: datatype property and object property. Datatype property refers to characteristics or parameters that are associated to a class. Object property refers to relations between classes.
- I is the set of instances. Instances represent the objects of classes.

4.2. General framework of GeoDataOnt

We adopted a method proposed by Noy and McGuinness (2001) to guide the design of the general framework. In this method, the general framework was designed via enumerating top-level relevant terms. To respond to the semantic issues in geospatial data integration and sharing listed in Section 2, we enumerated top-level terms from the perspective of geospatial data characteristics, namely essential, morphologic, and provenance characteristics. These terms will guide the following detailed design of GeoDataOnt. Therefore, GeoDataOnt can be separated into three compound modules: essential ontology, morphology ontology, and provenance ontology. This is the general framework in the dimension of the ontology content. GeoDataOnt can also be divided into three levels: general level, domain level, and application level, in the dimension of the ontological category. The general ontology is to represent the general terms that are common across all domain-level and application-level ontologies. Domain-level ontologies are to represent the domain-specific terms that belong to a certain aspect of geospatial data. The application ontologies are task-oriented. In the dimension of the ontology logic, GeoDataOnt is a 4-tuple structure: classes, relations, properties, and instances, which is consistent with the logic model of GeoDataOnt. Thus, the general framework of GeoDataOnt is designed from three different dimensions (Figure 2).

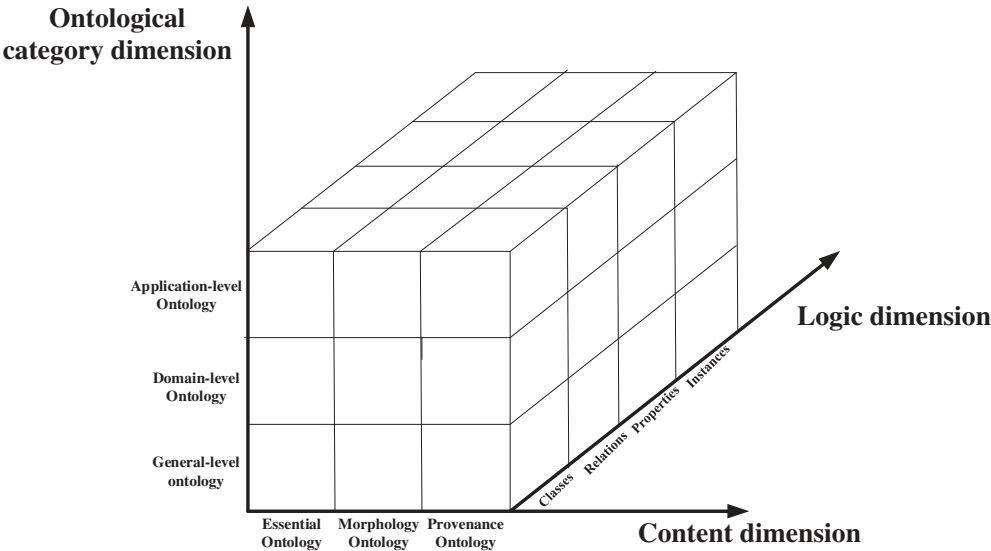


Figure 2. General framework of GeoDataOnt.

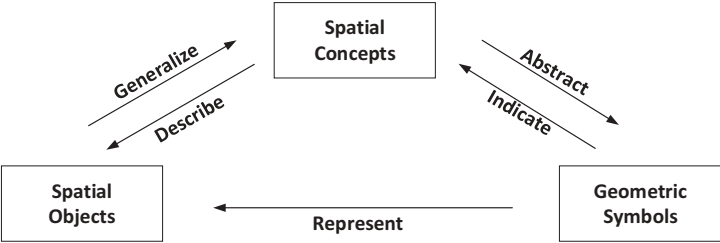


Figure 3. Conceptual model of geospatial data spatial ontology.

5. Detailed design and implementation of GeoDataOnt

A basic principle of top-down design and bottom-up development of ontology is used in the stage of design and development of GeoDataOnt (Kalbasi, Janowicz, Reitsma, Boerboom, & Alesheikh, 2014; Varanka & Usery, 2018). In the previous section, GeoDataOnt has been divided into three compound modules: essential ontology, morphology ontology, and provenance ontology from the perspective of content, which is consistent with the characteristics hierarchy of geospatial data. In this section, the detailed design for each module will be provided in a unified organizational form and content structure. Each module will be presented in four parts: conceptual model, classes hierarchy, relations, and core properties. The conceptual model is to provide an abstract definition for each module, and the last three components are to define the knowledge framework for each module. In the process of detailed design, we reuse some existing ontologies to make GeoDataOnt more complete and convincing.

5.1. Geospatial data essential ontology

Geospatial data essential ontology can be further divided into three ontology modules: spatial ontology, temporal ontology, and thematic ontology.

5.1.1. Geospatial data spatial ontology

(1) Conceptual model

The conceptual model of spatial ontology is shown in [Figure 3](#). Spatial objects are understood and identified based on their properties. According to the similarities and differences of properties, spatial objects are generalized and divided into different categories, therefore deriving spatial concepts. In turn, spatial concepts can be used to describe spatial objects. In a certain spatial scale, spatial objects can be represented with a geometric symbol according to their geometry features.

(2) Classes hierarchy

Many hierarchies of feature types for spatial entities have been proposed and applied. Some of them are derived from the public gazetteers such as GeoNames⁷ and Alexandria Digital Library gazetteer (ADL). Some previous studies also proposed their feature types hierarchy (Wang, Zhu, Pan, Luo, & Hou, 2016). The feature types hierarchy from GeoNames is a highly recognized and commonly used classification; thus, we use it as a classes hierarchy in the spatial ontology. The spatial entities are classified into 9 first-level categories and more than 600 second-level categories, as shown in [Table 2](#).

(3) Spatial relations

There are three types of spatial relations: topology relation, direction relation, and measure relation. The sub-relations for each type are shown in [Table 3](#).

(4) Core properties

The properties of spatial entities can be summarized from five aspects: materiality, reason for formation, space, time, and functionality. Materiality describes the physical

Table 2. Classes hierarchy of spatial ontology.

Num	First-level categories	Examples of second-level categories
1	Administrative Boundary (A)	A.ADM1, A.ADM1H, A.ADM2, A.ADM2H, A.ADM3, A.ADM3H, A.ADM4 ...
2	Hydrographic (H)	H.AIRS, H.ANCH, H.BAY, H.BAYS, H.BGHT, H.BNK, H.BNKR, H.BNKX ...
3	Area (A)	L.AGRC, L.AMUS, L.AREA, L.BSND, L.BSNP, L.BTL, L.CLG, L.CMN, L.CNS ...
4	Populated place (P)	P.PPL, P.PPLA, P.PPLA2, P.PPLA3, P.PPLA4, P.PPLC, P.PPLCH, P.PPLF ...
5	Road or Railroad (R)	R.PTGE, R.RD, R.RDA, R.RDB, R.RDCUT, R.RDJCT, R.RJCT, R.RR, R.RRQ ...
6	Spot (S)	S.ADMF, S.AGRF, S.AIRB, S.AIRF, S.AIRH, S.AIRP, S.AIRQ, S.AMTH, S.ANS ...
7	Hypsographic (T)	T.ASPH, T.ATOL, T.BAR, T.BCH, T.BCHS, T.BDLD, T.BLDR, T.BLHL ...
8	Undersea (U)	U.APNU, U.ARCU, U.ARRU, U.BDLU, U.BKSU, U.BNKU, U.BSNU, U.CDAU ...
9	Vegetation (V)	V.BUSH, V.CULT, V.FRST, V.FRSTF, V.GRSLD, V.GRVC, V.GRVO, V.GRVP ...

Table 3. Spatial relations.

Relation type	Instances of relations
Topology relation	Equal, intersect, contain, within, touch, separate, overlap, and overlapedby
Direction relation	Orientation East, South, West, North, NorthEast, SouthEast, SouthWest, and NorthWest
	Vertical direction Vertically upward and vertically downward
	Horizontal direction Left and right
	Angle direction Clockwise and counter-clockwise
Measure relation	Distance Far and near
	Length Long and short
	Height High and short

characteristics of spatial entities. The reason for formation reveals the formation mechanism of spatial entities. Spatial property refers to the size, form, and location. Temporal property presents the periodicity of spatial entities in some aspects. Functionality indicates how spatial entities can be used in human activity. Different classes of spatial entities have different properties in these five aspects. Table 4 shows the properties of spatial entities with the class of rivers as an example.

5.1.2. Geospatial data temporal ontology

(1) Conceptual model

We referred to the time ontology proposed by Zhang, Cao, Sui, & Wu. (2011) and proposed the conceptual model of geospatial data temporal ontology, which has a 5-tuple structure (Figure 4) (Hou et al., 2015). Temporal concepts (TC) are the foundation of this model, which include the shared temporal classes and instances. Temporal metric (TM) is the set of concepts used to measure time. Temporal relations (TR) refer to the relations between temporal entities. Both TM and TR are to support temporal formalization (TF), which is to formalize temporal

Table 4. Properties of rivers.

Classes	Property type	Properties	Instances of property
Rivers	Materiality	Forms of features Runoff	Point, line, and polygon Value of runoff
	Reason of formation	Reason for river formation	Natural river, man-made river, etc.
	Space	Types of main stream and tributaries Flow direction	Main stream and tributaries Inland rivers and outflow rivers
	Time	Types of ephemeral river and intermittent river	Ephemeral river, perennial river, etc.
	Functionality	Functionality of river	Transportation, water supply, etc.

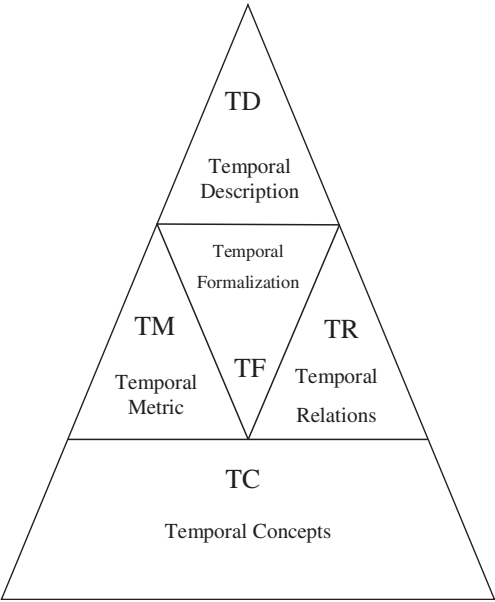


Figure 4. Conceptual model of geospatial data temporal ontology.

entities. Temporal description (TD) refers to the logical structure of temporal classes and instances.

(2) Classes hierarchy

Temporal entities can be classified into two categories: solar calendar and some basic classes (Figure 5). Solar calendar is most commonly used to describe the temporal characteristics of geospatial data and can be represented in two forms: date and clock. Basic classes include the concepts of time zone, season, etc.

(3) Relations

The relations between temporal entities are called temporal topology relations (Hou et al., 2015). According to algebraic theory of temporal entities, there are 14 temporal topology relations. Possible topology relations among temporal entities are closely related to the temporal entities type. There are two basic types of temporal entities: instant and interval. Compound temporal entity is a combination of instant and interval. Table 5 shows all possible topology relations for different combinations of temporal entities.

(4) Core properties

For temporal entities described by the solar calendar, their properties include unit and value of time. The properties of the time zone class contain offset hours, offset direction, and central meridian.

5.1.3. Geospatial data thematic ontology

We develop geospatial data thematic ontology by directly representing an existing classification system of themes of geospatial data in an ontological format. Thus, we did not propose another new conceptual model of classification system. We select

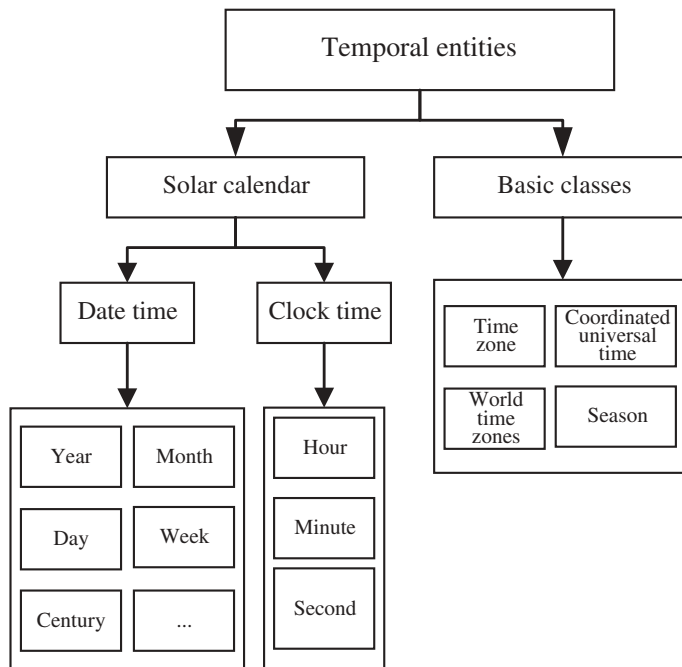


Figure 5. Classes hierarchy of temporal ontology.

Table 5. Temporal topological relations between different temporal entities.

Temporal entities	Temporal topological relation						
	Before	After	Contain	During	Finish	FinishedBy	Meet
instant – instant	✓	✓					
instant – interval	✓	✓	✓	✓	✓		
interval – interval	✓	✓	✓	✓	✓	✓	✓
complex entity	✓	✓	✓	✓	✓	✓	✓
Temporal entities	Metby	Overlap	OverlapedBy	Start	StartedBy	NonOverlap	Equal
instant – instant				✓			✓
instant – interval				✓			✓
interval – interval	✓	✓	✓	✓	✓	✓	✓
complex entity	✓	✓	✓	✓	✓	✓	✓

GCMD 8.7⁸ as the standard from many existing classification systems. The relationships of thematic ontology only include ancestor-descendant, parent-children, and sibling relationships between two themes in the classification hierarchy. A theme entity has two properties: theme definition that explains the meaning and scope of this theme, and “id” that is its unique identifier.

5.2. Geospatial data morphology ontology

(1) Conceptual model

Geospatial data morphology can be divided into two categories: external morphology that describes the external form of geospatial data, and internal morphology that presents the internal structure of geospatial data (Figure 6) (Sun et al., 2016). The relationships between external morphology and internal morphology can be further elaborated as follows. (a) Internal morphology describes intrinsic logic and organization of geospatial data. It determines the extrinsic expression form of data to some extent. This means that internal morphology determines external morphology. (b) In contrast, external morphology of geospatial data reflects its internal morphology in different forms. This means that external morphology characterizes internal morphology.

(2) Classes hierarchy

The definition and scope of morphologic characteristics of geospatial data were provided in Section 2. As shown in Figure 7, external morphology includes data type, data format, storage media, time format, units of measurement, language, and map symbol. Internal morphology includes data structure, CRS, temporal reference system, spatial accuracy, spatial granularity, and temporal granularity.

(3) Relations

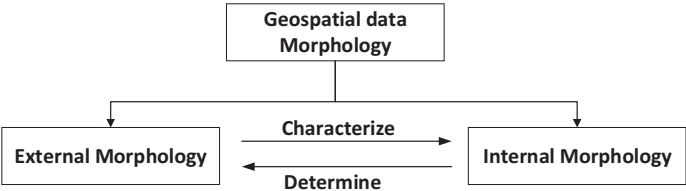


Figure 6. The conceptual model of geospatial data morphology ontology.

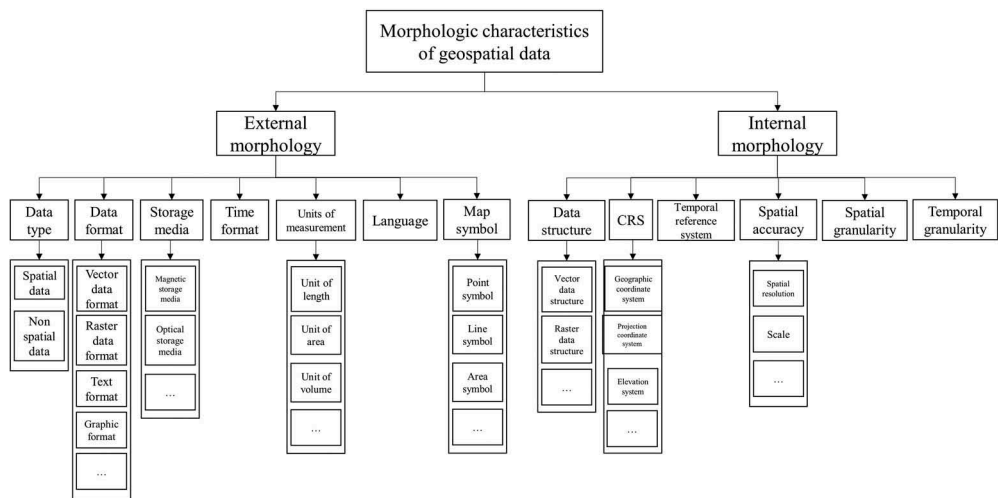


Figure 7. Classes hierarchy of morphology ontology.

The relations in each class of elementary morphologic characteristics are different. Some important relations of the classes of data format, data type, data structure, resolution, scale, storage media, and map symbol are listed in Table 6. For example, the relations of data format class include newerVersion, olderVersion, extension, restriction, containment, theSameFamily, and differentFamily. Specially, theSameFamily and differentFamily relations are distinguished according to whether data formats are produced with the same software (Zhu et al., 2017). For example, the formats of ArcInfo coverage and Shapefile are published by ArcGIS. Therefore, ArcInfo coverage has a theSameFamily relation with Shapefile.

(4) Core properties

The properties of some classes have been shown in Table 7. For example, the data format class has seven properties, which are derived from the sustainability factors of data formats proposed by the Library of Congress⁹.

5.3. Geospatial data provenance ontology

(1) Conceptual model

Table 6. Relations of morphology ontology.

Classes	Relations
Data format	NewerVersion, OlderVersion Extension, Restriction Containment TheSameFamily, different Family
Data type	Contain
Data structure	Contain
Resolution	Higher Resolution, Lower Resolution
Scale	Larger Scale, Smaller Scale
Storage media	Faster Read speed, Slower Read speed Bigger Capacity, Smaller Capacity
Map symbol	Topological Relations

Table 7. Properties of morphology ontology.

Classes	Properties	Classes	Properties
Data format	Extension	Scale	State
	Version	Storage media	Forms of expression
	Publisher		Style
	Disclosure		Read speed
	Adoption	Measuring unit	Capacity
	Self-documentation		Name
Geographic coordinate system	External dependencies		Type
	Reference ellipsoid	Language	Symbol
	Prime meridian		Name
	Unit of angle		Code
Projection coordinate system	GCS	Map symbol	Country/Region
	Name		Name
	Scale factor		Color
	Offset		Geometric element

A conceptual model for geospatial data provenance ontology is proposed based on the previous studies (Li, Zhu, Song, Sun, & Yang, 2017b; Ram & Liu, 2009). There are three principal components in this model: data sources, activity, and agent (Figure 8). Data sources refer to the data used for generating new data. Activity includes the operations to be employed on data sources to generate new data or to produce new data directly. When and where an activity happens, and which tools (or algorithms) an activity uses are the three elements of activity. Agent is the executer of activity.

(2) Classes hierarchy

The classes hierarchy of provenance ontology (Figure 9) is based on the provenance model. Organizations and individuals are two types of agents. Activity is classified into four categories: collection, processing, management, and distribution of data. Data sources include data of text, image, map, etc.

(3) Relations

There are many possible relations between different combining pairs of data, data source, activity, and agent. The key relations between data and data source, data and activity, data and agent, activity and activity, activity and agent, as well as agent and agent are listed in Table 8.

(4) Core properties

There are many properties for each class in provenance ontology. We only consider the class of agent as an example to show the properties (Table 9).

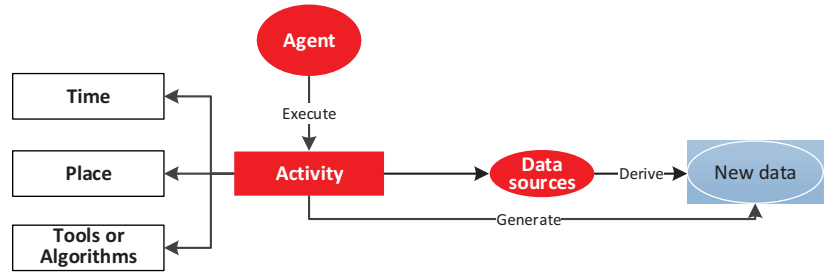


Figure 8. Conceptual model of geospatial data provenance ontology.

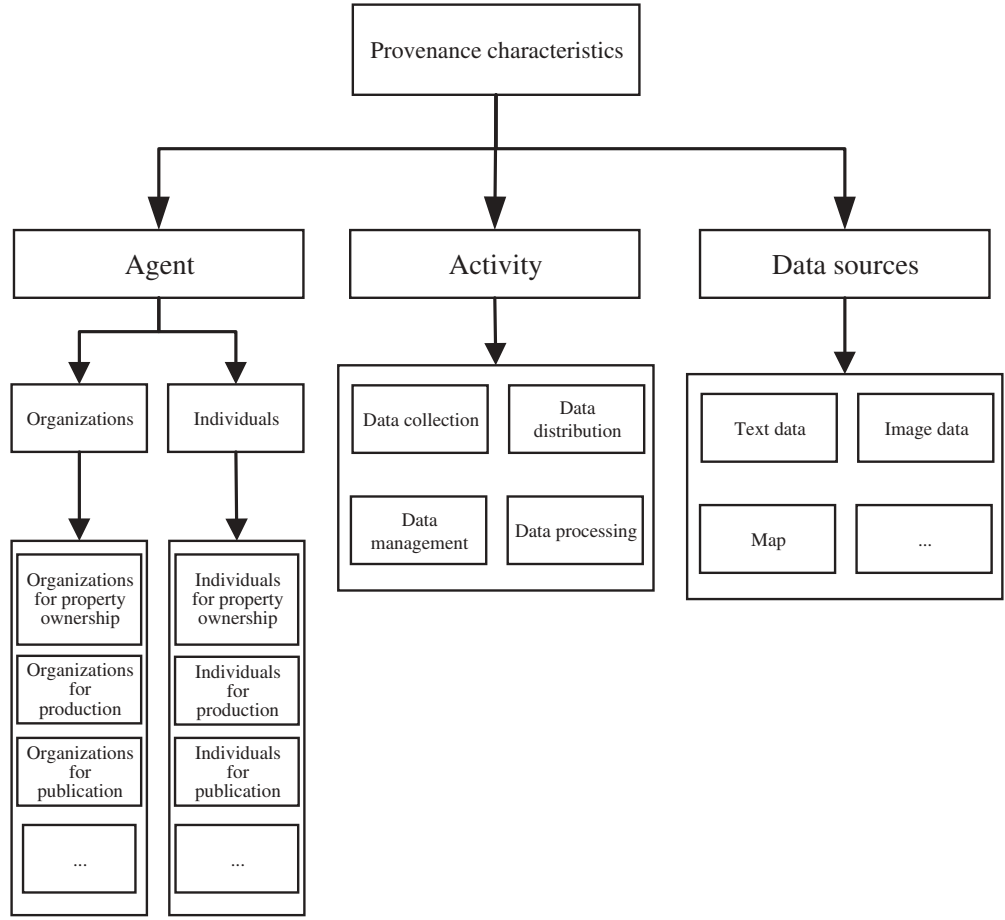


Figure 9. Classes hierarchy of provenance ontology.

Table 8. Relations of provenance ontology.

Classes	Relations	Descriptions of relation
data–data source	Cite	Multiple data sources are merged into a new dataset, and the new dataset includes the contents of old dataset.
	Update	A new dataset is generated by adding new information to the existing data.
	Fuse	Multiple data sources are fused into a new dataset, and the new dataset does not contain the contents of old dataset.
	Revise	A new dataset is generated by correcting errors in the existing data.
	Derive	A new dataset is derived from one data source via some activities.
data–activity	Use	Activity is employed on the existing data.
	Generate	A completely new dataset is generated by an activity. Before the activity, this dataset has no existence.
data–agent	Contribute	The agent takes part in the activity and makes contributions to generation of new dataset.
activity–activity	Belong to	The agent has the ownership of a dataset.
	Symbiosis	There are multiple activities involved in the process of data production, and they are indispensable to each other.
activity–agent	Be responsible	The agent is responsible for a dataset.
agent–agent	Authorize	An agent entrusts another agent to execute activity.

Table 9. Properties of agent.

Classes	Sub Classes	Properties
Agent	Individuals	Position Entry time Contact information Contact address Employee number Department
	Organizations	Legal person Establishment time Type Scale Scope of business Service time

5.4. Implementation of GeoDataOnt

The implementation of GeoDataOnt is to represent the aforementioned detailed knowledge framework in an ontological form with ontology representation languages and ontology development editors (Budak Arpinar et al., 2006). We selected Web Ontology Language (OWL) as the ontology representation language of GeoDataOnt and chose Protégé¹⁰, developed by Stanford University, as the ontology editor.

To make GeoDataOnt extensible and easily updated, a modular method is used to develop GeoDataOnt. According to the proposed knowledge framework, GeoDataOnt is divided into many small ontological modules. Initially, each module is developed independently as a sub-ontology, and 208 sub-ontologies are built in total. Then, all the sub-ontologies are integrated, eventually forming the constructed GeoDataOnt base. Figures 10–14, respectively, show the screenshots of the constructed spatial ontology, temporal ontology, thematic ontology, morphology ontology, and provenance ontology.

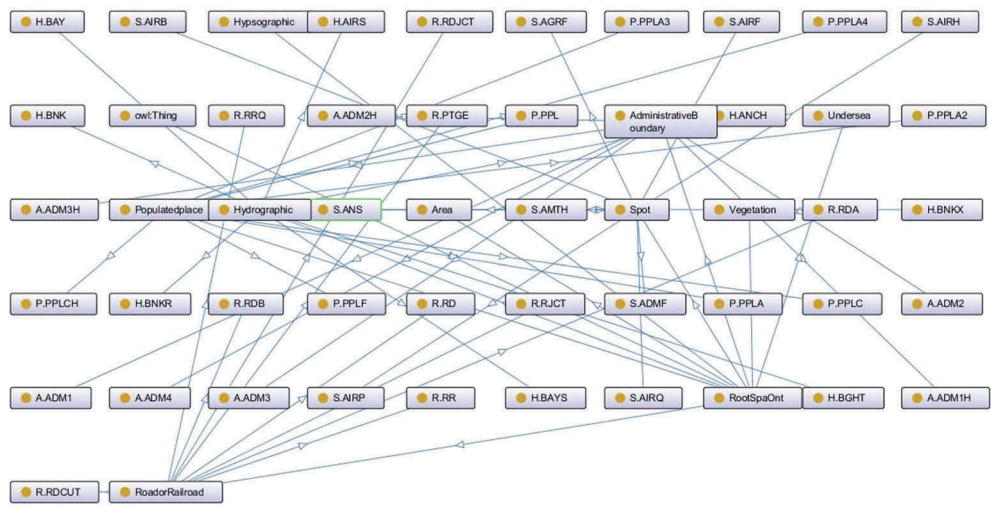


Figure 10. Screenshot of spatial ontology.

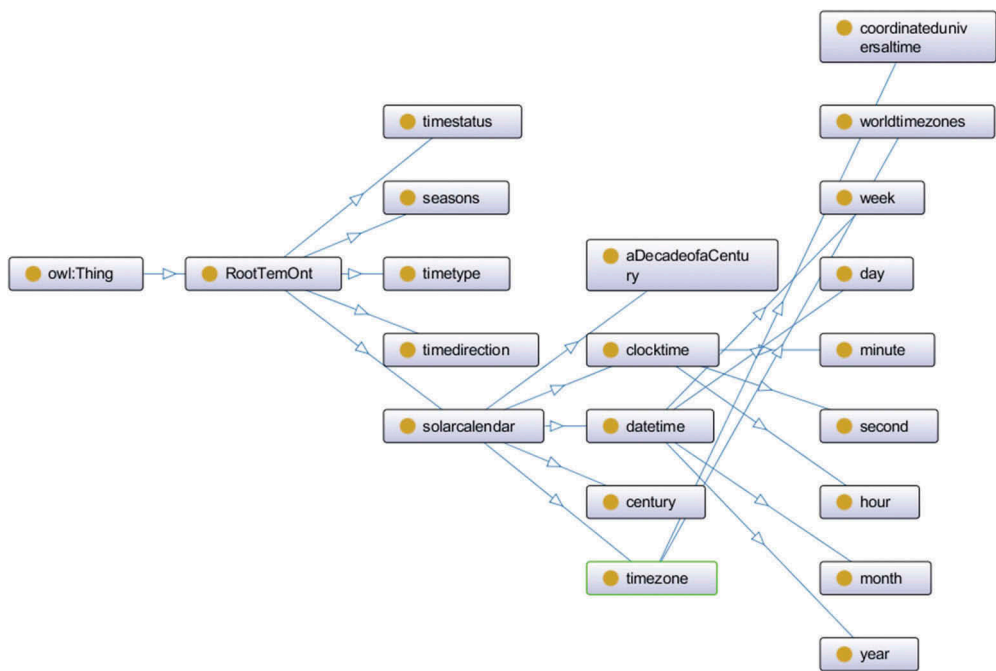


Figure 11. Screenshot of temporal ontology.

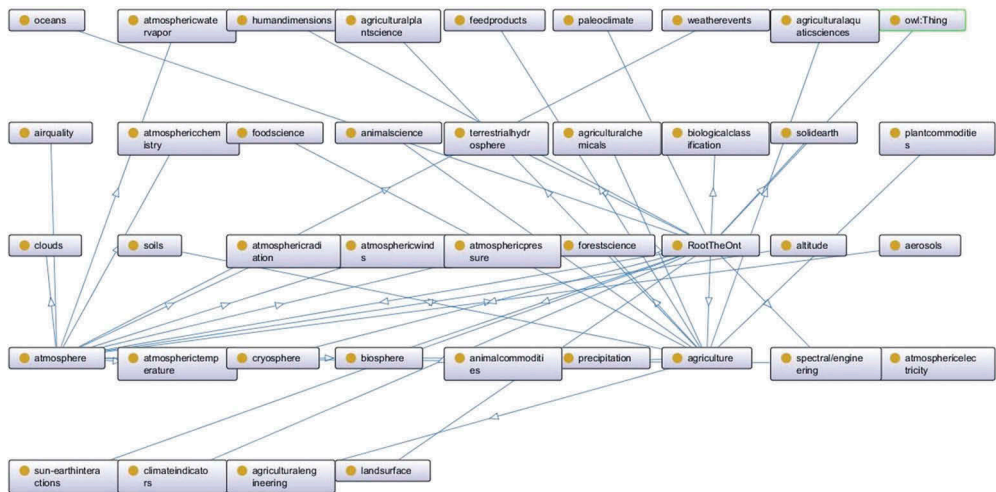


Figure 12. Screenshot of thematic ontology.

6. Discussion

6.1. Key limitations and challenges of GeoDataOnt

(1) Quality assessment of GeoDataOnt

The quality of GeoDataOnt has a major impact on whether it can effectively help to solve the semantic issues of geospatial data. Current GeoDataOnt is built manually, and

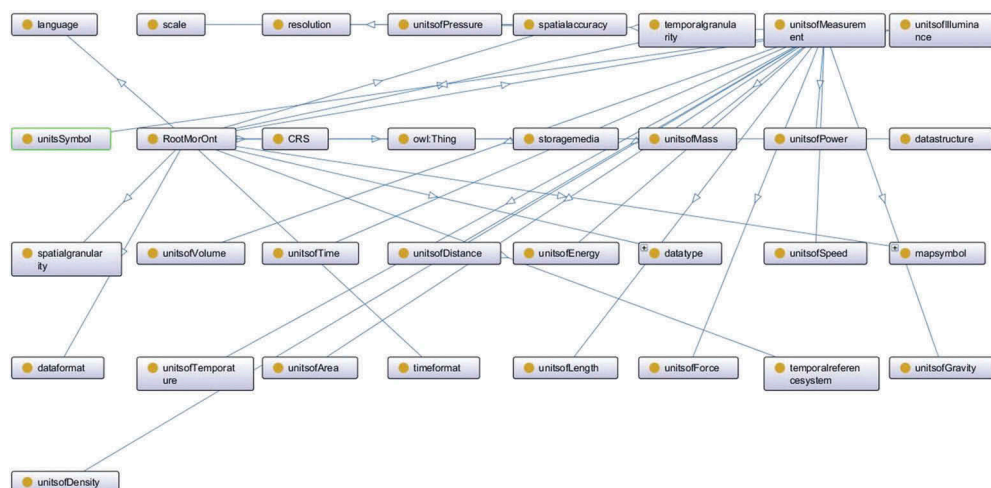


Figure 13. Screenshot of morphology ontology.

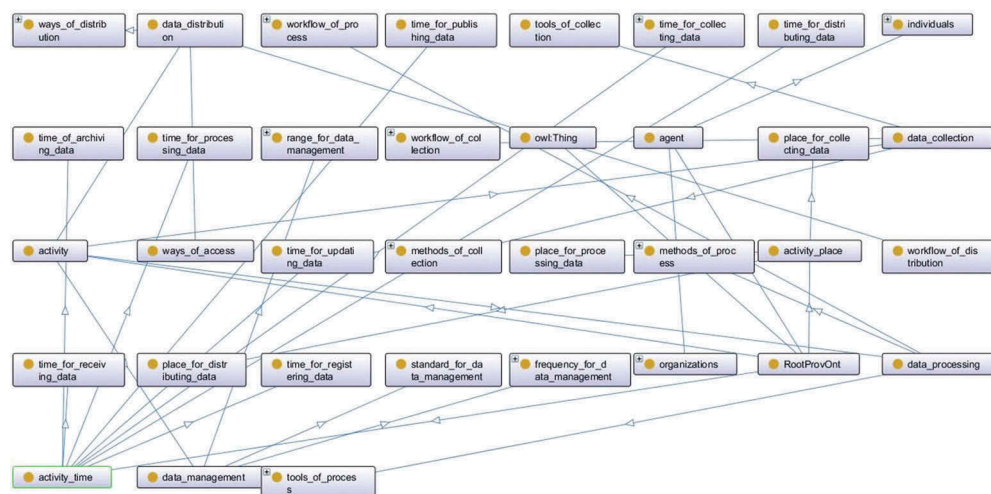


Figure 14. Screenshot of provenance ontology.

some errors may be created. Thus, the quality of knowledge contained in GeoDataOnt needs to be assessed quantitatively. Some studies focused on quality assessment of geographic knowledge have been conducted (Senaratne, Mobasher, Ali, Capineri, & Haklay, 2017), and various quality indicators have been proposed, including position accuracy, temporal accuracy, thematic accuracy, consistency, and completeness. However, these approaches can only be used for quality assessment of spatial modules in GeoDataOnt. There is a lack of effective measures and techniques to evaluate the quality of other modules, such as temporal modules, thematic modules, morphologic modules, and provenance modules. Thus, more measures or indicators to handle quality assessment of these modules should be investigated.

(2) Making GeoDataOnt more complete, accurate, and consistent

Complete, accurate, and consistent GeoDataOnt is essential for geospatial data integration and sharing. The quality assessment techniques can help identify the problems of GeoDataOnt in terms of its completeness, accuracy, and consistency. To solve these problems, two types of methods can be leveraged to improve the quality of GeoDataOnt at both the atomic and holistic scales. They use different types of available geographic information, respectively. To make full use of structured information, such as existing knowledge bases, methods of ontology alignment and integration can be employed to improve the quality of GeoDataOnt. Moreover, the methods of knowledge extraction generate new knowledge from the unstructured information.

(3) Aligning and integrating GeoDataOnt with existing knowledge bases

There are many open-source geographic knowledge bases, such as GeoNames and OpenStreetMap¹¹, and general knowledge bases, such as YAGO¹², DBpedia¹³, and WordNet¹⁴. The approach to improve GeoDataOnt using existing knowledge bases generally includes two steps: finding the relations and establishing mappings between entities from GeoDataOnt and other knowledge bases, and then integrating these entities based on the found relations and established mappings. Entity alignment is an effective method to find correspondences of entities from different ontologies (Sun, Zhu, & Song, 2019). Many methods of geographic entity alignment have been proposed in the previous studies. However, most of them are in the stage of experiment or prototype, and therefore cannot be applied in practical usage.

Leveraging the found mappings of entities at the step of alignment, GeoDataOnt can be supplemented via integrating with existing knowledge bases. Some knowledge gap of GeoDataOnt will be filled by knowledge integration, and GeoDataOnt will be more complete. For example, some missing properties of entities in spatial ontology might be filled.

(4) Knowledge extraction from unstructured geographic information

There are two types of unstructured geographic information containing huge volume of geographic knowledge, namely geo-text and geo-image. The corresponding knowledge can help to improve the richness and freshness of GeoDataOnt. Geo-text data refer to the combination of geographic location and natural language texts, including web texts (e.g., geotagged tweets published in social media platforms), news reports, and academic literatures (Hu, 2018). Extracting knowledge from geo-text data is an interdisciplinary study, which combines GIScience and natural language processing (NLP) innovatively. Many previous studies contributed to extracting geographic knowledge from geo-text data (Goldberg, Wilson, & Knoblock, 2009; Hu, Mao, & McKenzie, 2018; Zhang et al., 2019). Geo-image data include remote sensing images, aerial photos, scanned historical maps, and internet photos (e.g., geotagged photos taken with smartphones). There are also some studies conducted to extract knowledge from geo-images (Hu et al., 2015). However, the quality of extracted knowledge from geo-text and geo-image varies greatly, and therefore it cannot be applied directly. The techniques of evaluating and improving the quality of extracted knowledge need to be investigated.

(5) Dynamic update, automatic construction, and reuse of GeoDataOnt

The knowledge relevant to GeoDataOnt may change over time. Hence, GeoDataOnt needs to be updated continuously to ensure that this ontology is sufficiently up-to-date to support GIScience applications. However, it might be difficult and labor-intensive to

update GeoDataOnt manually. Thus, developing techniques on updating GeoDataOnt and filling new knowledge to GeoDataOnt is necessary. Meanwhile, the methods of constructing GeoDataOnt automatically also need to be studied. There are three approaches to automatically construct GeoDataOnt. The first one is to extract knowledge from available data and represent extracted knowledge in an ontological form (Goldberg et al., 2009). The second one is to crowdsource knowledge from volunteers (Fast & Rinner, 2018). The last one is to reuse the existing ontologies via alignment techniques. Additionally, it is necessary to improve the reusability of GeoDataOnt for easy reuse by other researchers (Fernández-López, Poveda-Villalón, Suárez-Figueroa, & Gómez-Pérez, 2018).

(6) Extracting and annotating semantic information of geospatial data based on GeoDataOnt automatically

GeoDataOnt is an ontology specifically designed and constructed for geospatial data. The initial step to employ it in geospatial data integration and sharing is to extract and annotate semantic information of geospatial data based on GeoDataOnt. However, it is difficult and labor-intensive to complete this step manually, and the result may be relatively inaccurate. First, there are many types of geospatial data, and they vary greatly in detail. Thus, it is difficult for non-professionals to understand all of them; therefore, GIScience experts may be required to complete it. Second, it is difficult to find accurate terms from GeoDataOnt to describe geospatial data due to massive knowledge contained in GeoDataOnt. Thus, methods of automatic extraction and annotation of semantic information of geospatial data based on GeoDataOnt need to be developed.

6.2. Applications of GeoDataOnt

(1) Interpreting geospatial data based on GeoDataOnt

An understanding and analysis of geospatial data is required to employ geospatial data in practical applications. Geospatial data is currently described using metadata. Under different contexts, however, metadata of geospatial data may be provided with different meanings or forms, which may cause semantic issues when applying data (Macário, de Sousa, & Medeiros, 2009). Instead of unstructured metadata, semantic-enabled annotation of geospatial data can be provided in a unified and formalized manner based on GeoDataOnt, which can help geospatial data to be understood equally under different contexts. This will lay the foundation for integration, discovery, association, and recommendation of geospatial data in a semantic level.

(2) Integrating geospatial data based on GeoDataOnt

The mappings between geospatial data on a characteristic level can be discovered under the support of their semantic annotations. GeoDataOnt and the inference rules can help to reveal more relations between geospatial data. With the found mappings and relations, geospatial data can be integrated effectively in different levels. For example, geospatial data with the same theme but different spatial and temporal characteristics can be gathered together. Geospatial data with the same theme but different units of measurement can be integrated via unit conversion.

(3) Semantic-enabled linking, discovering, and recommending geospatial data based on GeoDataOnt

Entities from different modules of GeoDataOnt are linked by object properties, and these links show the logical relationships between classes. These logical relationships will be computed further by running a reasoner on the triples in GeoDataOnt, thereby generating a reasoned GeoDataOnt model with new reasoned triples. The reasoned model can help identify links between geospatial data based on precise and unambiguous semantic annotation of geospatial data. The main way to identify links is semantic expansion, which refers to finding those terms that are semantically relevant to a certain term in GeoDataOnt. Representing those found links in a machine-readable way can help to build a semantically rich and interconnected data network, namely linked geospatial data (Zhu et al., 2017). The constructed data network can facilitate discovery and recommendation of geospatial data.

(4) Discovering and composing geoprocessing web services based on GeoDataOnt

GeoDataOnt can also be used to provide semantic description for geoprocessing web services, thereby achieving semantic-based discovery of services. To handle more complex tasks, a composition of services is necessary. The input and output of services will initially be described formally based on GeoDataOnt. The services chain will then be constructed by matching the output of a preceding service to the input of a succeeding service.

7. Conclusions

Ontologies provide semantic descriptions for heterogeneous information sources in a machine-interpretable way, thereby solving semantic issues. Thus, ontologies have attracted significant attention in the field of semantic-enabled geospatial data integration and sharing. In this paper, we proposed an ontological framework, named GeoDataOnt, to provide the semantic foundation for geospatial data integration and sharing. To make this ontological framework more targeted, we initially defined the characteristics hierarchy of geospatial data and analyzed the semantic problems in geospatial data integration and sharing from a new perspective of characteristics. Next, a simple logic model for GeoDataOnt was defined to explain what knowledge should be contained in this ontology. The general framework of GeoDataOnt was also proposed, which can be understood in three dimensions: content dimension, ontological category dimension, and logic dimension. According to the characteristics of the compound level, GeoDataOnt was divided into three compound modules: essential ontology, morphology ontology, and provenance ontology. Each compound module was further divided into the elementary modules based on the characteristics of elementary level. Each module was designed in detail including a conceptual model, core classes, relations, and properties. All modules were then implemented in Protégé software and integrated to form an ontology base. Some important insights about the key limitations and challenges, as well as the applications of GeoDataOnt were discussed.

The contributions of this study can be seen from two perspectives. From the perspective of methodology, this paper continued to complement and perfect the ontology-based methods of geospatial data integration and sharing. From the perspective of geo-ontology, a relatively complete geo-ontology that is targeted at geospatial data was developed.

GeoDataOnt does not contain all detailed knowledge about geospatial data. It only provides a general semantic foundation for geospatial data integration and sharing.

Thus, it can only resolve the semantic issues at a coarse-grained level. A more detailed GeoDataOnt is necessary to deal with the semantic issues at a fine-grained level. Moreover, the descriptions of certain characteristics of some geospatial data may be not accurate enough, leading to some limitations on solving the semantic issues in corresponding characteristics using an ontology. Therefore, there is still a long way to solve these problems completely. Future work will focus on overcoming these limitations and challenges and on extending the applications of GeoDataOnt.

Notes

1. <https://www.fgdc.gov/>.
2. <https://www.fgdc.gov/nsdi>.
3. <https://www.iso.org/standard/26020.html>.
4. <http://www.geodata.cn/>.
5. <https://gcmd.nasa.gov/>.
6. <https://www.iso.org/iso-8601-date-and-time-format.html>.
7. <https://www.geonames.org/>.
8. <https://gcmdservices.gsfc.nasa.gov/static/kms/sciencekeywords/sciencekeywords.xml>.
9. <https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml>.
10. <https://protege.stanford.edu/>.
11. <https://www.openstreetmap.org/about>.
12. <https://github.com/yago-naga/yago3>.
13. <https://wiki.dbpedia.org/>.
14. <https://wordnet.princeton.edu/>.

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Funding

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences [grant number XDA23100100]; National Natural Science Foundation of China [grant number 41771430], [grant number 41631177]; China Scholarship Council [grant number 201804910732].

ORCID

Kai Sun  <http://orcid.org/0000-0002-0343-6032>

Jia Song  <http://orcid.org/0000-0002-9051-1925>

References

- Agarwal, P. (2005). Ontological considerations in GIScience. *International Journal of Geographical Information Science*, 19(5), 501–536.
- Andrade, F. G. D., Baptista, C. D. S., & Davis, C. A. (2014). Improving geographic information retrieval in spatial data infrastructures. *Geoinformatica*, 18(4), 793–818.
- Baglioni, M., Giovannetti, E., Masserotti, M. V., Renso, C., & Spinsanti, L. (2008). Ontology supported querying of geographical databases. *Transactions in GIS*, 12, 31–44.
- Bittner, T., Donnelly, M., & Smith, B. (2009). A spatio-temporal ontology for geographic information integration. *International Journal of Geographical Information Science*, 23(6), 765–798.
- Brodeur, J., Bedard, Y., Edwards, G., & Moulin, B. (2010). Revisiting the concept of geospatial data interoperability within the scope of human communication processes. *Transactions in GIS*, 7(2), 243–265.
- Buccella, A., Cechich, A., & Fillottrani, P. (2009). Ontology-driven geographic information integration: A survey of current approaches. *Computers and Geosciences*, 35(4), 710–723.
- Buccella, A., Cechich, A., Gendarmi, D., Lanubile, F., Semeraro, G., & Colagrossi, A. (2011). Building a global normalized ontology for integrating geographic data sources. *Computers & Geosciences*, 37(7), 893–916.
- Budak Arpinar, I., Sheth, A., Ramakrishnan, C., Lynn Usery, E., Azami, M., & Kwan, M.-P. (2006). Geospatial ontology development and semantic analytics. *Transactions in GIS*, 10(4), 551–575.
- Chen, N., Chen, Z., Hu, C., & Di, L. (2011). A capability matching and ontology reasoning method for high precision OGC web service discovery. *International Journal of Digital Earth*, 4(6), 449–470.
- Chen, Y., Sabri, S., Rajabifard, A., & Agunbiade, M. E. (2018). An ontology-based spatial data harmonisation for urban analytics. *Computers, Environment and Urban Systems*, 72, 177–190.
- Comber, A., Fisher, P., & Wadsworth, R. (2011). Integrating land-cover data with different ontologies: Identifying change from inconsistency. *International Journal of Geographical Information Science*, 18(7), 691–708.
- Couclelis, H. (2010). Ontologies of geographic information. *International Journal of Geographical Information Science*, 24(12), 1785–1809.
- Dean, D. J. (2007). Characterizing spatial databases via their derivation: A complement to content ontologies. *Transactions in GIS*, 11(3), 399–412.
- Deliiska, B. (2007). Thesaurus and domain ontology of geoinformatics. *Transactions in GIS*, 11(4), 637–651.
- Fallahi, G. R., Frank, A. U., Mesgari, M. S., & Rajabifard, A. (2008). An ontological structure for semantic interoperability of GIS and environmental modeling. *International Journal of Applied Earth Observations and Geoinformation*, 10(3), 342–357.
- Fast, V., & Rinner, C. (2018). Toward a participatory VGI methodology: Crowdsourcing information on regional food assets. *International Journal of Geographical Information Science*, 32(11), 2209–2224.
- Fernández-López, M., Poveda-Villalón, M., Suárez-Figueroa, M. C., & Gómez-Pérez, A. (2018). Why are ontologies not reused across the same domain? *Journal of Web Semantics*, 57.
- Fonseca, F. T., Egenhofer, M. J., Agouris, P., & Camara, G. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3), 231–257.
- Fonseca, F. T., Egenhofer, M. J., Davis, C. A., & Borges, K. A. V. (2000). Ontologies and knowledge sharing in urban GIS. *Computers Environment & Urban Systems*, 24(3), 251–272.
- Fonseca, R. L., & Llano, E. G. (2011). Automatic representation of geographical data from a semantic point of view through a new ontology and classification techniques. *Transactions in GIS*, 15(1), 61–85.
- Frank, A. U. (2001). Tiers of ontology and consistency constraints in geographical information systems. *International Journal of Geographical Information Systems*, 15(7), 667–678.
- Goh, C. H. (1997). *Representing and reasoning about semantic conflicts in heterogeneous information systems*. Cambridge: Massachusetts Institute of Technology.

- Goldberg, D. W., Wilson, J. P., & Knoblock, C. A. (2009). Extracting geographic features from the internet to automatically build detailed regional gazetteers. *International Journal of Geographical Information Science*, 23(1), 93–128.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220.
- Gui, Z., Yang, C., Xia, J., Liu, K., Xu, C., Li, J., & Lostritto, P. (2013). A performance, semantic and service quality-enhanced distributed search engine for improving geospatial resource discovery. *International Journal of Geographical Information Science*, 27(6), 1109–1132.
- Hahmann, T., & Stephen, S. (2018). Using a hydro-reference ontology to provide improved computer-interpretable semantics for the groundwater markup language (GWML2). *International Journal of Geographical Information Science*, 32(6), 1138–1171.
- Hakimpour, F. (2003). *Using ontologies to resolve semantic heterogeneity for integrating spatial database schemata*. Zurich, Switzerland: University of Zurich.
- Hong, J.-H., & Kuo, C.-L. (2015). A semi-automatic lightweight ontology bridging for the semantic integration of cross-domain geospatial information. *International Journal of Geographical Information Science*, 29(12), 2223–2247.
- Hou, Z., Zhu, Y., Gao, X., Pan, P., Luo, K., & Wang, D. (2015). Time-ontology and its application in geodata retrieval. *Journal of Geo-information Science*, 17(4), 379–390. in Chinese.
- Hu, Y., Janowicz, K., Carral, D., Scheider, S., Kuhn, W., Berg-Cross, G., ... Kolas, D. (2013). A geo-ontology design pattern for semantic trajectories. *International Conference on Spatial Information Theory*, 8116, 438–456.
- Hu, Y. (2017). Geospatial semantics. In B. Huang, Ed., *Comprehensive geographic information systems* (pp. 80–94). Oxford: Elsevier.
- Hu, Y. (2018). Geo-text data and data-driven geospatial semantics. *Geography Compass*, 12(11), e12404.
- Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, 54, 240–254.
- Hu, Y., Mao, H., & McKenzie, G. (2018). A natural language processing and geospatial clustering framework for harvesting local place names from geotagged housing advertisements. *International Journal of Geographical Information Science*, 33(4), 714–738.
- Kalbasi, R., Janowicz, K., Reitsma, F., Boerboom, L., & Alesheikh, A. (2014). Collaborative ontology development for the geosciences. *Transactions in GIS*, 18(6), 834–851.
- Klien, E., Lutz, M., & Kuhn, W. (2006). Ontology-based discovery of geographic information services —An application in disaster management. *Computers Environment & Urban Systems*, 30(1), 102–123.
- Kuhn, W. (2003). Semantic reference systems. *International Journal of Geographical Information Science*, 17(5), 405–409.
- Kuo, C. L., & Hong, J. H. (2016). Interoperable cross-domain semantic and geospatial framework for automatic change detection. *Computers & Geosciences*, 86, 109–119.
- Li, L., Liu, Y., Zhu, H., Ying, S., Luo, Q., Luo, H., ... Shen, H. (2017a). A bibliometric and visual analysis of global geo-ontology research. *Computers & Geosciences*, 99, 1–8.
- Li, N., Raskin, R., Goodchild, M., & Janowicz, K. (2012). An ontology driven framework and web portal for spatial decision support. *Transactions in GIS*, 16(3), 313–329.
- Li, W., Zhu, Y., Song, J., Sun, K., & Yang, J. (2017b). Geospatial data provenance-ontology and its application in data linking. *Journal of Geo-information Science*, 19(10), 1261–1269. in Chinese.
- Li, W., Goodchild, M. F., & Raskin, R. (2014). Towards geospatial semantic search: Exploiting latent semantic relations in geospatial data. *International Journal of Digital Earth*, 7(1), 1–21.
- Li, W., Yang, C., Nebert, D., Raskin, R., Houser, P., Wu, H., & Li, Z. (2015). Semantic-based web service discovery and chaining for building an Arctic spatial data infrastructure. *Computers & Geosciences*, 37(11), 1752–1762.
- Liu, Y., Li, L., Shen, H., Yang, H., & Luo, F. (2018). A co-citation and cluster analysis of scientometrics of geographic information ontology. *ISPRS International Journal of Geo-Information*, 7(3), 120–133.

- Lutz, M. (2007). Ontology-based descriptions for semantic discovery and composition of geoprocessing services. *Geoinformatica*, 11(1), 1–36.
- Lutz, M., & Klien, E. (2006). Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science*, 20(3), 233–260.
- Macário, C. G. N., de Sousa, S. R., & Medeiros, C. B. (2009). Annotating geospatial data based on its semantics. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 81–90), Seattle, WA.
- Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, 80(4), 449–461.
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Knowledge Systems Laboratory, Stanford University.
- Peachavanish, R., & Karimi, H. A. (2007). Ontological engineering for interpreting geospatial queries. *Transactions in GIS*, 11(1), 115–130.
- Peuquet, D. J. (2002). *Representations of space and time*. New York, NY: Guilford Press.
- Ram, S., & Liu, J. (2009). A new perspective on semantics of data provenance. *Proceedings of the First International Conference on Semantic Web in Provenance Management-Volume 526* (pp. 35–40), Washington, DC.
- Reitsma, F., Laxton, J., Ballard, S., Kuhn, W., & Abdelmoty, A. (2009). Semantics, ontologies and eScience for the geosciences. *Computers & Geosciences*, 35(4), 706–709.
- Schuurman, N., & Leszczynski, A. (2006). Ontology-based metadata. *Transactions in GIS*, 10(5), 709–726.
- Senaratne, H., Mobasher, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographic Information Science*, 31(1), 139–167.
- Stock, K., Karasova, V., Robertson, A., Roger, G., Small, M., Bishr, M., ... Gardner, Z. (2013). Finding science with science: Evaluating a domain and scientific ontology user interface for the discovery of scientific resources. *Transactions in GIS*, 17(4), 612–639.
- Stock, K., Stojanovic, T., Reitsma, F., Ou, Y., Bishr, M., Ortmann, J., & Robertson, A. (2012). To ontologise or not to ontologise: An information model for a geospatial knowledge infrastructure. *Computers & Geosciences*, 45(4), 98–108.
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & Knowledge Engineering*, 25(1), 161–197.
- Sun, K., Zhu, Y., Pan, P., Luo, K., Wang, D., & Hou, Z. (2016). Research on morphology-ontology and its application in geospatial data discovery. *Journal of Geo-information Science*, 18(8), 1011–1021. in Chinese.
- Sun, K., Zhu, Y., & Song, J. (2019). Progress and challenges on entity alignment of geographic knowledge bases. *ISPRS International Journal of Geo-Information*, 8(2), 77–101.
- Sun, S., Wang, L., Ranjan, R., & Wu, A. (2015). Semantic analysis and retrieval of spatial data based on the uncertain ontology model in digital earth. *International Journal of Digital Earth*, 8(1), 3–16.
- Uitermark, H. T., van Oosterom, P. J. M., Mars, N. J. I., & Molenaar, M. (2005). Ontology-based integration of topographic data sets. *International Journal of Applied Earth Observations & Geoinformation*, 7(2), 97–106.
- Varanka, D. E., & Usery, E. L. (2018). The map as knowledge base. *International Journal of Cartography*, 4(2), 201–223.
- Wache, H., Voegele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., & Hübner, S. (2001). *Ontology-based integration of information – A survey of existing approaches*. Proceedings of IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, WA.
- Wang, C., Ma, X., & Chen, J. (2018). Ontology-driven data integration and visualization for exploring regional geologic time and paleontological information. *Computers & Geosciences*, 115, 12–19.
- Wang, D., Zhu, Y., Pan, P., Luo, K., & Hou, Z. (2016). Construction of geodata spatial ontology and its application in data retrieval. *Journal of Geo-information Science*, 18(4), 443–452. in Chinese.
- Wang, X., Gu, W., Ziebelin, D., & Hamilton, H. (2010). An ontology-based framework for geospatial clustering. *International Journal of Geographical Information Science*, 24(11), 1601–1630.

- Wiegand, N., & García, C. (2007). A task-based ontology approach to automate geospatial data retrieval. *Transactions in GIS*, 11(3), 355–376.
- Winter, S. (2001). Ontology: Buzzword or paradigm shift in GI science? *International Journal of Geographical Information Science*, 15(7), 587–590.
- Xu, J., Nyerges, T. L., & Nie, G. (2014). Modeling and representation for earthquake emergency response knowledge: perspective for working with geo-ontology. *International Journal of Geographical Information Science*, 28(1), 185–205.
- Zhang, C., Cao, C., Sui, Y., & Wu, X. (2011). A Chinese time ontology for the semantic web. *Knowledge-Based Systems*, 24(7), 1057–1074.
- Zhang, C., Zhao, T., & Li, W. (2010). Automatic search of geospatial features for disaster and emergency management. *International Journal of Applied Earth Observation and Geoinformation*, 12(6), 409–418.
- Zhang, Y., Ma, Q., Chiang, Y.-Y., Knoblock, C., Zhang, X., Yang, P., ... Hu, X. (2019). Extracting geographic features from the Internet: A geographic information mining framework. *Knowledge-Based Systems*, 174, 57–72.
- Zhao, T., Zhang, C., Wei, M., & Peng, Z.-R. (2008). Ontology-based geospatial data query and integration. *International Conference on Geographic Information Science* (pp. 370–392), Park City, UT.
- Zhu, Y., Zhu, A.-X., Song, J., Yang, J., Feng, M., Sun, K., ... Zhao, H. (2017). Multidimensional and quantitative interlinking approach for Linked Geospatial Data. *International Journal of Digital Earth*, 10(9), 923–943.