Computers & Geosciences 86 (2016) 109–119

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Case study

Interoperable cross-domain semantic and geospatial framework for automatic change detection

Chiao-Ling Kuo [a], Jung-Hong Hong [a,b,*]

[a] Department of Geomatics, National Cheng Kung University, No. 1, University Rd., Tainan, Taiwan
[b] Information Division, National Science and Technology Center for Disaster Reduction, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 25 December 2014
Received in revised form
5 June 2015
Accepted 18 October 2015
Available online 21 October 2015

Keywords:
Semantic interoperability
Bridge ontology
Change detection
Topographic map data
Land use data

ABSTRACT

With the increasingly diverse types of geospatial data established over the last few decades, semantic interoperability in integrated applications has attracted much interest in the field of Geographic Information System (GIS). This paper proposes a new strategy and framework to process cross-domain geodata at the semantic level. This framework leverages the semantic equivalence of concepts between domains through bridge ontology and facilitates the integrated use of different domain data, which has been long considered as an essential superiority of GIS, but is impeded by the lack of understanding about the semantics implicitly hidden in the data. We choose the task of change detection to demonstrate how the introduction of ontology concept can effectively make the integration possible. We analyze the common properties of geodata and change detection factors, then construct rules and summarize possible change scenario for making final decisions. The use of topographic map data to detect changes in land use shows promising success, as far as the improvement of efficiency and level of automation is concerned. We believe the ontology-oriented approach will enable a new way for data integration across different domains from the perspective of semantic interoperability, and even open a new dimensionality for the future GIS.

1. Introduction

Geodata are an essential component of Geographic Information System (GIS). In the past three decades, a huge volume and diverse types of geodata was established with the advances of mapping technology. Various GIS-based applications utilize the available data to meet specific domain needs. Despite the advantages of "sharing" has been well recognized, the use of geodata acquired from various domains frequently causes heterogeneity or interoperability obstacles during data integration (Bishr, 1998; George, 2005). In the past 10 years, many types of heterogeneity issues have been resolved by common standards or interfaces, but the issue of semantic heterogeneity still remains as a challenge to conquer (Bishr, 1998; Kavouras et al., 2005).

Many previous researches focused on clarifying the semantics and facilitating the semantic interoperability of GIS to ensure the correct use of application. The various methods proposed include semantic interpretation (Goddard and Wierzbicka, 2002; Kavouras and Kokla, 2007; Goddard, 2012), semantic transformation (Kuhn and Raubal, 2003; Baglatzi and Kuhn, 2013), and semantic similarity measurement (Janowicz et al., 2008; Schwering, 2008). The goals for these studies are to build the basis for semantic comparison, establish the conversion rules and determine the semantic relationships of concepts for semantic integration.

Another topic that has attracted considerable research interest is ontology-based semantic integration. Ontology is defined as *an explicit specification of a conceptualization* (Gruber, 1995). Ontology is a formal way to present data concepts, relationships, restrictions, meanings, and knowledge, therefore, it have been widely used for presenting the semantics of data (Studer et al., 1998; Stevens et al., 2000; Keet, 2004). Noy (2004) proposed three dimensions for semantic integration based on ontology, namely, *mapping discovery*, *declarative formal representations of mappings*, and *reasoning with mappings*. The first dimension corresponds to the challenge of ontology integration. Various methods for addressing the relationships of concepts between ontologies have been proposed, for example, ontology mapping, ontology bridging, ontology alignment and ontology merging (Pinto and Martins, 2001; Wache et al., 2001; Xu et al., 2004; de Bruijn et al., 2006; Euzenat and Shvaiko, 2007; Leung et al., 2009; Amrouch and Mostefai, 2012; Shvaiko and Euzenat, 2013). Except the ontology merging approach combines several ontologies into an ontology, the other three approaches are all based on the analysis and presentation of the relationships between concepts of different

* Corresponding author.
E-mail addresses: chauoling@gmail.com (C.-L. Kuo), junghong@mail.ncku.edu.tw (J.-H. Hong).

http://dx.doi.org/10.1016/j.cageo.2015.10.011
0098-3004/

ontologies. The choice of ontology integration strategies is task dependent and often made by human experts.

For the second dimension, the major focus is to develop semantic relationships for presenting the relationships or similarity between concepts (Giunchiglia and Shvaiko, 2003; Cruz et al., 2004; Kokla and Kavouras, 2005; Heer et al., 2009). Despite different terms are used, the primitive relationships proposed by various studies in semantic matching (Giunchiglia and Shvaiko, 2003), ontology merge (Heer et al., 2009), and ontology alignment (Cruz et al., 2004) are in fact semantically equivalent. The only exception is the difference between "overlap"(Giunchiglia and Shvaiko, 2003; Kokla and Kavouras, 2005; Heer et al., 2009) and "approximate" (Cruz et al., 2004), the former is used to indicate two concepts are related, while the major interest of the latter case is to determine if two concepts are close enough to claim they are approximate.

Besides the exploitation of ontology integration and semantic representation, how to apply the result of ontology integration to geodata processing has also drawn lots of research interests. This is the third dimension raised by (Noy, 2004). For example, Fonseca et al. (2002) proposed an ODGIS architecture based on the three-level ontologies to identify the relationship of concepts and instance of remote sensing systems and geographic information systems, so the image could be processed together with GIS data. A domain ontology viewed as consensus between different fields is developed for data sharing for monitoring environment changes (Pundt and Bishr, 2002). Uitermark et al. (2005) developed a conceptual framework to explicitly present the relationships between instance and further proposed a reference model to check the consistence of instances from two different topographic data sets. The GeoMergeP system (layer-based) that involves the semantic enrichment and the merging of integrating geographic sources is proposed in (Buccella et al., 2011). The common features for the frameworks proposed in the above studies are to generate a global ontology via domain ontologies integration, and to retrieve data or to find out the corresponding instances based on the global ontology. Through the literature reviews, the ontology integration with semantic relationships is essential for achieving geographic sources integration. While most of previous literatures focused on only the semantic interoperability for the similar domains, the semantic interoperability of cross-domain applications is also very important because GIS typically has to simultaneously deal with a number of datasets acquired from different resources. The lack understanding of the semantics of the geodata being used certainly becomes a major obstacle for making correct decisions.

The phenomena in reality may change continuously over time. Therefore, geodata must be updated regularly to maintain a consistency relationship with the real world. Change detection is a crucial step in the data update procedure. The area changed must first be identified and then evaluated to determine if the data can be updated to the correct status. Most of the change detection studies are based on either remote sensing techniques or on integrated GIS and remote sensing methods. The remote sensing approach (Bruzzone and Prieto, 2000; Zhang, 2004; Im et al., 2008; Bouziani et al., 2010; Wright and Wimberly, 2013) demonstrated its superior capability for establishing rough categories of data, such as vegetation, buildings, water body, etc. Its capability is, however, rather restricted in terms of classifying detailed categories of data, such as hospitals or police offices. As to the approaches of combining GIS and remote sensing (Walter, 2004; Zhang and Couloigner, 2004; Shalaby and Tateishi, 2007; Matikainen et al., 2010; Tian et al., 2014), they can raise the processing efficiency as the shape and range of processed object derived from GIS vector data is relatively definite. The geometry serves a definite boundary for change detection. In (Walter, 2004; Zhang and Couloigner, 2004; Matikainen et al., 2010), they used the classified area from images to evaluate if an object from GIS vector data is changed or not and then update the object in the database. However, the result of change detection is determined by the classification result and the threshold of matched area and object. Besides, these approaches are suitable for the rough category change detection, such as water, forest, building, road, etc.

In addition to the change area assessment, the semantics of classes of referenced classification system used for change detection has been considered. Butkiewicz et al. (2008) used the semantic filters to locate a certain category in change detection procedure. Ahlqvist (2008) exploited the semantic change of classification category of land cover based on the definitions of classes via fuzzy sets based approach through semantic similarity metrics. The change results can be indicated by the semantic change image. These studies reveal that the semantics clarification of classes of classification system for change detection is vital and once the semantics of the overlap area is changed, it indicates that the area is changed.

According to the discussion above, to facilitate a new way of development for cross-domain applications from the semantic perspective, an ontology-based framework is proposed in this paper. The goals of the present work are as follows:

1. Developing a *framework for the semantic interoperability* of geodata.
2. Using this framework to develop an *automatic change detection mechanism*.

This framework is designed for cross-domain applications to enable the presentation of data semantics and to facilitate data processing for a certain application based on domain knowledge. We adopt the bridge ontology approach presented in the study conducted by Hong and Kuo (2015) to establish the semantic relationships of concepts among different domains for semantic interpretation. The semantic interoperability of geodata can then be achieved automatically by integrating this semantic relationship with domain knowledge. To develop a mechanism that can detect changes and to utilize the abundant existing geodata, we develop an automatic change detection mechanism and demonstrate the feasibility of the proposed framework. Through vector-based geodata analysis, we propose three types of change detection factors: spatial, temporal, and semantic. These factors serve as the basis for defining change detection knowledge (formulated as rules). This paper also summarizes six primitive types of change scenarios for demonstrating the change detection results, which can then be used for assessing further actions..

To demonstrate the usage of the framework, topographic map data are chosen to analyze its capability to update land use data. Despite the fact that they were established independently, our experimental results and statistics highlight the implications of each change type and the effectiveness of change detection between these two domains. To the best of our knowledge, the current study presents a new insight to address the issues of change detection and automatic updating based on a semantic consideration using existing data from other domains. The same principle is applicable to other cross-domain applications. The remainder of this paper is organized as follows: the next section introduces the conceptual framework of semantic interoperability and the workflow of automatic change detection. Section 3 discusses the use of topographic map data to detect changes in land use data with the proposed approach. Finally, section 4 concludes our major findings.

2. Methodology

To achieve data interoperability at the semantic level, we first

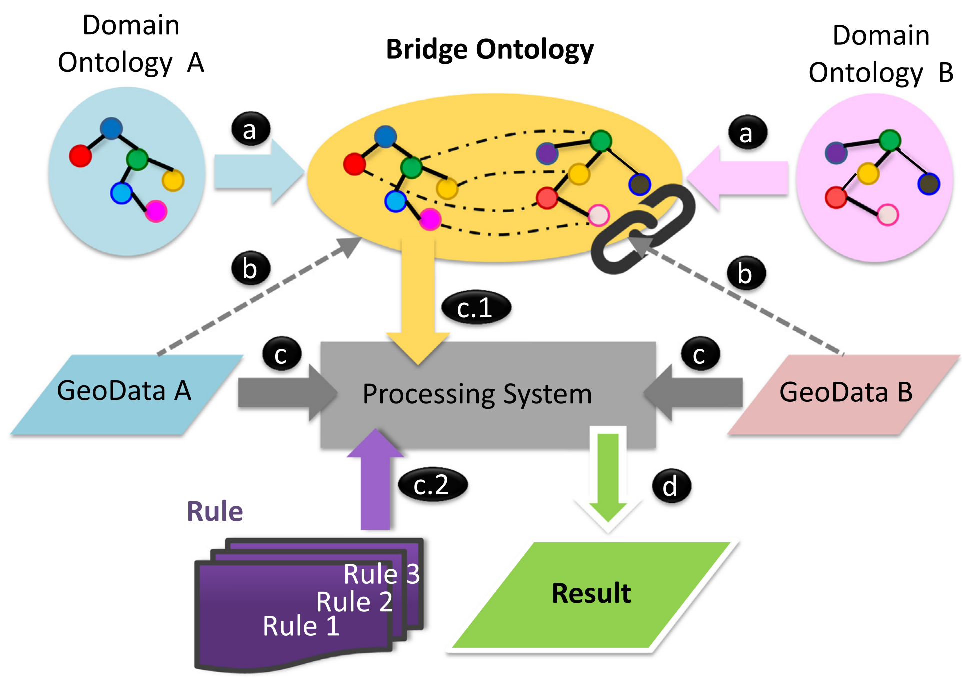

Fig. 1. Conceptual framework for semantic and interoperable application.

propose a conceptual framework based on bridge ontology to present the data semantics and to facilitate the integration of data from different domains. Second, the framework will serve as the basis for further developing an automatic change detection mechanism.

2.1. Conceptual framework

Fig. 1 illustrates the conceptual framework for developing semantic-based cross-domain interoperable applications. This framework is designed to record and infer the semantic relationships between the concepts of the two chosen domains comprehensively and to process geodata following rules that consider both the semantics and other properties of the geodata according to the application needs. This framework consists of six components, namely, domain ontology, bridge ontology, geodata, rule, processing system, and result. Four steps [steps a, b, c (including c.1 and c.2), and d] are designed to facilitate the interactions between these components.

1. *Domain Ontology* presents the semantics of the chosen domains. It serves as the basis for establishing bridge ontology between the two domains.
2. *Bridge Ontology* is an ontology being developed to formally present the semantic relationships between the concepts of the two domains. Bridge ontology is oriented; thus, the source and target ontology that constitute the bridge ontology must be decided in advance. Following the bridge ontology approach proposed by Hong and Kuo (2015), five semantic relationships between two concepts are considered: *sem_exact*, *sem_subset*, *sem_superset*, *sem_overlap,* and *sem_null* (the prefix *sem_* denotes semantic usage only).
   - *Sem_exact*: The semantics of the two compared concepts are identical, even if they are presented by different vocabularies in different domain ontologies.
   - *Sem_subset*: The semantics of the source concept is part of the target concept. The target concept contains additional concepts that the source concept does not possess.
   - *Sem_superset*: This relationship and the relationship of *sem_subset* are converse relationships. The semantics of the source concept include the target concept, and the source concept has additional concepts that the target concept does not possess.
   - *Sem_overlap*: The semantics of the two concepts share common concepts, but each concept has its own distinguished concepts the other does not have.
   - *Sem_null*: The semantics of the two concepts are completely different. This usually implies that these two concepts are irrelevant.

If the concepts of a particular domain are structured as a tree, the leaf nodes present the most detailed concepts of the domain. The parent–child relationships between nodes at different levels therefore can be seen as the sem_superset and sem_subset relationships between concepts. Moreover, the relationships between a node and its siblings are Sem_null. The bridge ontology records the semantic relationships between any pair of leaf nodes from the two domains using one of the five semantic relationships listed above. To enable further inference, the construction priority is ordered as follows: *sem_exact* > *sem_subset* > *sem_superset*=*sem_overlap* > *sem_null*. *Sem_exact* has the highest priority, as once a concept in source domain has *sem_exact* relationship with respect to a concept in the target domain, we can easily infer that the relationships between this concept and the siblings of the compared concept in the target domain are *sem_null*. The loading of comparison can then be reduced. *Sem_superset* and *sem_overlap* have the same priority, as a concept can relate to two different sibling concepts in the target domain simultaneously. Once the semantic relationship between a pair of nodes is determined, more semantic relationships can be inferred according to the Eq. (1).

For example (Fig. 2), if the relationship between concept F and concept M is *sem_exact*, then the relationships between concept F and concept P and concept N can be respectively inferred as *sem_subset* and *sem_null*. If the relationship between concept C and concept S is *sem_superset* and that between concept C and concept T is *sem_overlap*, then the relationship between concept A, which are the siblings of concept C, and concept S is inferred as *sem_null*. The same rule can be also used for inferring the

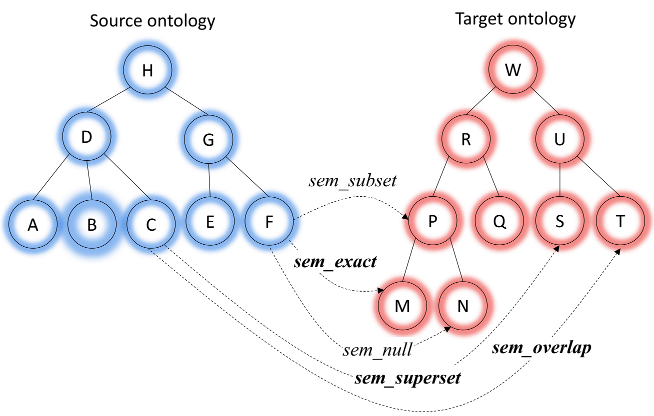

Fig. 2. Establishment of semantic relationships among concepts in two different domains.

relationship between concept B and concept S.

$$f(x,y)=\begin{cases} sem\_subset(x,\ y\text{'s parent}) \quad \text{if } sem\_exact(x,\ y) \\ \qquad = \text{True or } sem\_subset(x,\ y) = \text{True}, \\ sem\_superset(x\text{'s parent},\ y) \quad \text{if } sem\_exact(x,\ y) \\ \qquad = \text{True or } sem\_superset(x,\ y) = \text{True}, \\ sem\_null(x,\ y\text{'s siblings}) \quad \text{if } sem\_exact(x,\ y) \\ \qquad = \text{True or } sem\_subset(x,\ y) = \text{True}, \\ sem\_null(x\text{'s siblings},\ y) \quad \text{if } sem\_superset(x,\ y) \\ \qquad = \text{True}, \end{cases} \tag{1}$$

where (x, y) denotes the concepts from two different ontologies.

3. *Geodata* are the data for the two chosen domains. The semantics of every dataset must refer to a particular concept in the domain ontology.
4. *Rule* specifies the operation criteria for the desired applications. We define rules following the base form of *antecedent* ⇒ *consequent* in Semantic Web Rule Language (SWRL)[1]. Both *antecedent* and *consequent* are conjunctions of atoms. Given that space and time are two important components of geodata (Chrisman, 2001) and that semantics must be clarified during processing, a general form of rules is proposed below (the attribute of geodata is not considered because its content varies with data). The *spatial relationship*, *required condition*, and *semantic relationship* atoms are *antecedent*, and the application *outcome* is *consequent*. A rule may consist of one or more atoms depending on the applications.

$$\boxed{\begin{array}{l} \textit{Form of Rule}: \\ \quad \textit{atoms of spatial relationship} \quad \ldots \wedge \\ \quad \textit{atoms of required condition} \quad \ldots \wedge \\ \quad \textit{atoms of semantic relationship} \quad \ldots \\ \Rightarrow \\ \quad \textit{atoms of outcome} \end{array}}$$

a. In the current study, the *spatial relationship* is restricted to the primitive topological relations widely discussed in GIS literature (Egenhofer and Franzosa, 1991; Clementini et al., 1993) based on the same coordinate reference system (CRS). Each domain establishes geodata in its own way. If two features from different domains intersect geometrically, then they are regarded as "related features" from the spatial perspective.
b. A *required condition* is the atom specified as the constraints on the spatial and temporal properties of the geodata for specific applications. Typical examples include the positional accuracy, mapping time, and reference data time. The *required condition* ensures that the analyzed geodata meets the specific requirements for the intended applications.
c. The *semantic relationship* presents the semantic relationships between the concepts of the two respective domains. The relationship is application independent. Thus, application requirements must be considered to generate criteria. Different semantic relationships lead to different outcomes, and each semantic relationship in the application must be comprehensively considered and designed.
d. The *outcome* is determined by assessing all types of atoms. The different constraints of the three types of atoms, especially those based on the semantic relationships, generate a corresponding *outcome* for a rule.

5. *Processing System* processes the geodata according to the rules. The input data include both sets of geodata, bridge ontology, and rules. Rules control how the geodata and bridge ontology work together to generate the final outcomes.
6. *Result* shows the output of the processing system.

The four steps of the framework are designed to support semantic interoperability. In the first step (Fig. 1, step a), the bridge ontology of the two domain ontologies is established. Step b follows by establishing the corresponding relationships between the geodata and the bridge ontology. Steps c, c.1, and c.2 continue to import all the processed data, the semantic relationships, and rules into the processing system. This system then automatically applies rules onto the geodata in accordance with the spatial relationships, properties, and semantic relationship information. Finally, step d outputs the result.

2.2. Automatic change detection mechanism

In this section, we propose an automatic and efficient change detection mechanism based on the concept of semantic interoperability and the reuse of existing geodata. The basic idea of change detection is based on the comparison of two datasets, the source data (refer to a later date) and the target data (refer to an earlier date), to determine if changes occurred in the target domain. A distinguishable characteristic of our approach is that the source and target data are obtained from existing datasets of two different domains. Fig. 3 illustrates the change detection scenario. Fig. 3(a) displays a feature in target data *Td*. In Fig. 3(b), a source data is found to geometrically intersect with the feature in the target data; thus, the intersection area *Ia* (yellow area) is simultaneously interpreted by the two respective datasets. Since their statuses are referred to different time, the analysis on the semantics and properties can serve as the basis for change detection [Fig. 3(c)]. The following discussion further explains the detailed steps.

2.2.1. Analysis of change detection factors

Change detection is an application with specific geodata requirements. Four factors, namely, *geometry*, *positional accuracy*, *time*, and *semantics*, must be considered during a change detection task.

[1] http://www.w3.org/Submission/SWRL/

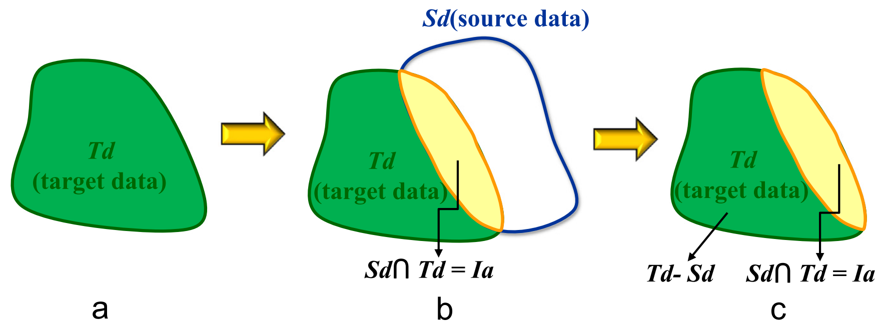

Fig. 3. Change detection scenario. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

1. *Geometry*: Each geodata has a geometric component to describe its location and spatial dimensionality. The location must refer to a specific CRS. It frequently happens that the geodata used in an application may correspond to different CRSs. The change detection tasks require all of the selected geodata to refer to the same CRS. If not, the coordinates must be transformed to the same CRS without deteriorating the positional accuracy of the source data. In the change detection task, only the geometric intersection area of related features is considered (refer to Fig. 3(c)).
2. *Positional accuracy*: Geometric intersection is critical to our approach. As the source and the target data are established independently, the difference of positional accuracy may influence the outcomes of the analysis. The positional accuracy of the source data is expected to be of the same level as or even higher than that of the target data. Otherwise, the data update decision is questionable. This principle remains valid regardless of the type of coordinate transformation performed. By considering positional accuracy, intersection areas that are smaller than the specified threshold are skipped to avoid processing gaps or of slivers between analyzed polygons.
3. *Time*: The content of the geodata refers to the real world status at a specific time. The time of the source data must be later than that of the target data to conduct meaningful comparisons in change detection. The constraints on temporal difference are usually specified by the decision makers.
4. *Semantics*: Geodata are designed and generated from a specific perspective. Therefore, the semantic issues between different domains must be clarified for cross-domain integrated applications. The change detection result is determined based on the semantic relationship of the source and target data.

The intersection area in the *geometry* factor belongs to the *spatial relationship* atom, whereas the *semantics* factor belongs to the *semantic relationship* atom. Positional accuracy, and time constitute the *required conditions* for change detection in Eq. (2)

$$\begin{cases} Sd.\ \ Positional\ \ Accuracy\ \ \geq Td.\ \ Positional\ \ Accuracy \\ Sd.\ \ Time\ \ \geq Td.\ \ Time \end{cases}, \tag{2}$$

where *Sd* corresponds to the source data and *Td* denotes the target data.

2.2.2. Change types and rules for change detection

The simplest outcome for change detection is straightforward, that is, the target data is either changed or not (Radke et al., 2005). In this paper, we propose six primitive change types according to eight rules. The simplest case for detecting changes is the analysis shows that the semantic relationship between the two statuses of the intersected area is *sem_null*. For example, if the status of the intersection area are “farm” and “road” for the source and target data respectively, it most likely indicates a new road is constructed on a land which was previously used for farm. On the other hand, if the semantic relationship is *sem_exact* or *sem_subset*, then the status of the target data remains unchanged because the source data either presents the same semantics or is a special case to the concept of the target data. Six primitive change types between the two compared datasets are proposed as follows:

1. *Non_process (NP)*: The source data does not meet the essential requirements; thus, no suggestion is made and no action is taken.
2. *No_change (NC)*: The analysis confirms that the status of the target data remains unchanged. Although the action may be similar to that of *NP*, the interpretation is completely different because the updated results are confirmed in this case. Typical examples are the instances in which the semantic relationship between the source and target data is *sem_exact* or *sem_subset*.
3. *Uncertain_I (UI)*: The two analyzed features are related, but the evidence for determining whether or not the status of the intersection area changes is insufficient. A typical example is when the source data meets the essential requirements and the semantic relationship between the related features is *sem_overlap*.
4. *Uncertain_II (UII)*: The semantic relationship changes from *sem_overlap* to *sem_superset*. Similar to *UI*, the intersection areas of both domains are semantically related; however, this change type lacks the evidence that determines whether or not the status changes.
5. *Change_To Concept (CTC)*: The analysis confirms changes occurs and suggests the updated status of the target data. A typical example is when the semantic relationship between the source and the target data is *sem_null*. The concept of source data can be transformed into a corresponding concept in the target domain with the semantic relationship of either *sem_exact* or *sem_subset*.
6. *Change_No Concept (CNC)*: Although the analysis confirms the presence of changes, it nonetheless fails to suggest the updated status to the target data. This situation is mainly caused by the lack of *sem_exact* or *sem_subset* between the concepts of the source and target data.

Table 1 summarizes the relationships of the proposed change types and the rules based on the atoms of *spatial relationship*, *required condition*, and *semantic relationship*. The essential requirement for the atoms of *spatial relationship* is that the area of the geometric intersection must exceed the specified threshold. Otherwise, this pair of geodata is not processed further (Rule 1). If the source and the target data for the atoms of the *required condition* do not meet the required condition of Eq. (2), then this pair of geodata is not processed, either. If neither of the two conditions is met, then the change type is determined as *NP*, regardless of their semantic relationships (Rule 2).

Both requirements of *spatial relationship* and *required condition* must be met before the atoms of *semantic relationship* are

Table 1
Change types and rules.

Spatial relationship	Required condition	Semantic relationship		Change type	Rule No.
False	Any	Any		*NP*	Rule 1
Any	False	Any		*NP*	Rule 2
True	True	*sem_exact*		*NC*	Rule 3
		sem_subset		*NC*	Rule 4
		sem_overlap		*UI*	Rule 5
		sem_superset		*UII*	Rule 6
		sem_null	Source data have a concept that corresponds to *sem_exact* or *sem_subset* relationships in the target domain	*CTC*	Rule 7
			Source data do not possess a concept that corresponds to *sem_exact* or *sem_subset* in the target domain	*CNC*	Rule 8

considered. Rules 3–8 are developed to determine the change types according to the semantic relationships between the source and target data. The proposed change types have important implications for the change detection actions for users. The change types *NC*, *CTC,* and *CNC* belong to the category of *confirmed_result*. The first two change types can even update the status of the target data. Despite that the change type of *CNC* cannot provide the updated suggestions accurately, it nevertheless indicates that changes occurred and that the authorized agencies can use other reference data, such as remote sensing images, for further inspection. The successful and automatic making of these three types of decisions saves much time in finding regions of changes.

For the other three change types, the change types of *UI* and *UII* belong to the *uncertain_change* category, and the change type of *NP* belongs to the *unprocessed category*. Neither of these categories can confirm if change occurred. The *uncertain_change category* suggests that the semantic relationships alone are not sufficient for drawing a confirmed conclusion. If the majority of concept analysis in the bridge ontology belongs to this category, then the concepts of the source data are not good candidates for updating the target data. Meanwhile, the *unprocessed category* indicates that the chosen dataset is inappropriate for updating purposes because it fails to meet the essential requirements.

The mechanism operates on the basis of rules. In the following example, the italicized texts with "?" are variables (following the rules of SWRL), ?*Sd* and ?*Td* denote the source and target data from the respective domains, and ?*Ia* corresponds to the intersection area of ?*Sd* and ?*Td*. The object property *Spatial_intersect* calculates whether ?*Sd* and ?*Td* intersect geometrically. If so, the object property of *hasIntersectionArea* indicates that such relationships exist among ?*Ia*, ?*Sd*, and ?*Td*. The object properties of *hasPositionalAccuracy,* and *hasMappingTime* are used to record the basic properties of ?*Sd* and ?*Td*. The comparison operations of *greaterThanOrEqual*, *lessThanOrEqual,* and *equal* are used to compare the recording values of the properties. The object property *Sem_exact* is used to record the semantic relationships between ?*Sd* and ?*Td*. Finally, the detection result is identified as *NC* if all of the conditions above and the test results are valid or true.

$$
\begin{aligned}
&\textit{Example Rule:}\\
&\quad Spatial\_intersect(?Sd,?Td)\ \wedge\ hasIntersectionArea(?Sd,?Ia)\ \wedge\\
&\quad hasIntersectionArea(?Td,?Ia)\ \wedge\ greaterThanOrEqual(?Ia,threshold)\ \wedge\\
&\quad hasPositionalAccuracy(?Sd,?SdPA)\ \wedge\ hasPositionalAccuracy(?Td,?TdPA)\ \wedge\\
&\quad lessThanOrEqual(?SdPA,?TdPA)\ \wedge\ hasMappingTime(?Sd,?SdMT)\ \wedge\\
&\quad hasMappingTime(?Td,?TdMT)\ \wedge\ greaterThanOrEqual(?SdMT,?TdMT)\ \wedge\\
&\quad Sem\_exact(?Sd,?Td)\\
&\Rightarrow\\
&\quad No\_change(?Ia)
\end{aligned}
$$

3. Experimental results and discussion

This section discusses the use of topographic map data as source data to detect changes and update the status of land use data following the mechanism proposed in Section 2. Both types of geodata are vital references to the reality, but are costly to maintain. The content of the topographic map data follows a detailed taxonomy framework and model reality on the basis of individual features; therefore, they serve as a useful reference for the update of land use data.

The change detection mechanism is feasible on each level of nodes, as long as the semantic relationship between the pair of concepts can be determined. In this paper, the change detection result is based on the analysis at the level of the leaf node. Table 2 summarizes the bridging results of the leaf nodes of the topographic map concepts to land use concepts. The topographic map and the land use taxonomy framework contain 327 and 103 leaf nodes, respectively. The analysis shows that the relationships between 24 pairs of leaf nodes are *sem_exact*, and there are 122 and 51 topographic map concepts respectively correspond to 53 concepts of the land use data with *sem_subset* and *sem_overlap* relationships. Some of these concepts are many-to-one relationships. Meanwhile, nine concepts of topographic map data correspond to 13 land use concepts with *sem_superset* relationship. The ratio of successful bridging for topographic map and land use data are 63.00% and 87.38%, respectively, which implies that most of the land use concepts are related to topographic map concepts.

Table 2
Bridging of the numbers and rates of topographic map and land use concepts.

	sem_exact	*sem_subset*	*sem_overlap*	*sem_superset*	Bridging rate
Topographic map	24	122	51	9	206/327 = 63.00%
Land use	24	53[a]		13	90/103 = 87.38%

[a] Many-to-one semantic relationships: *sem_subset* and *sem_overlap*.

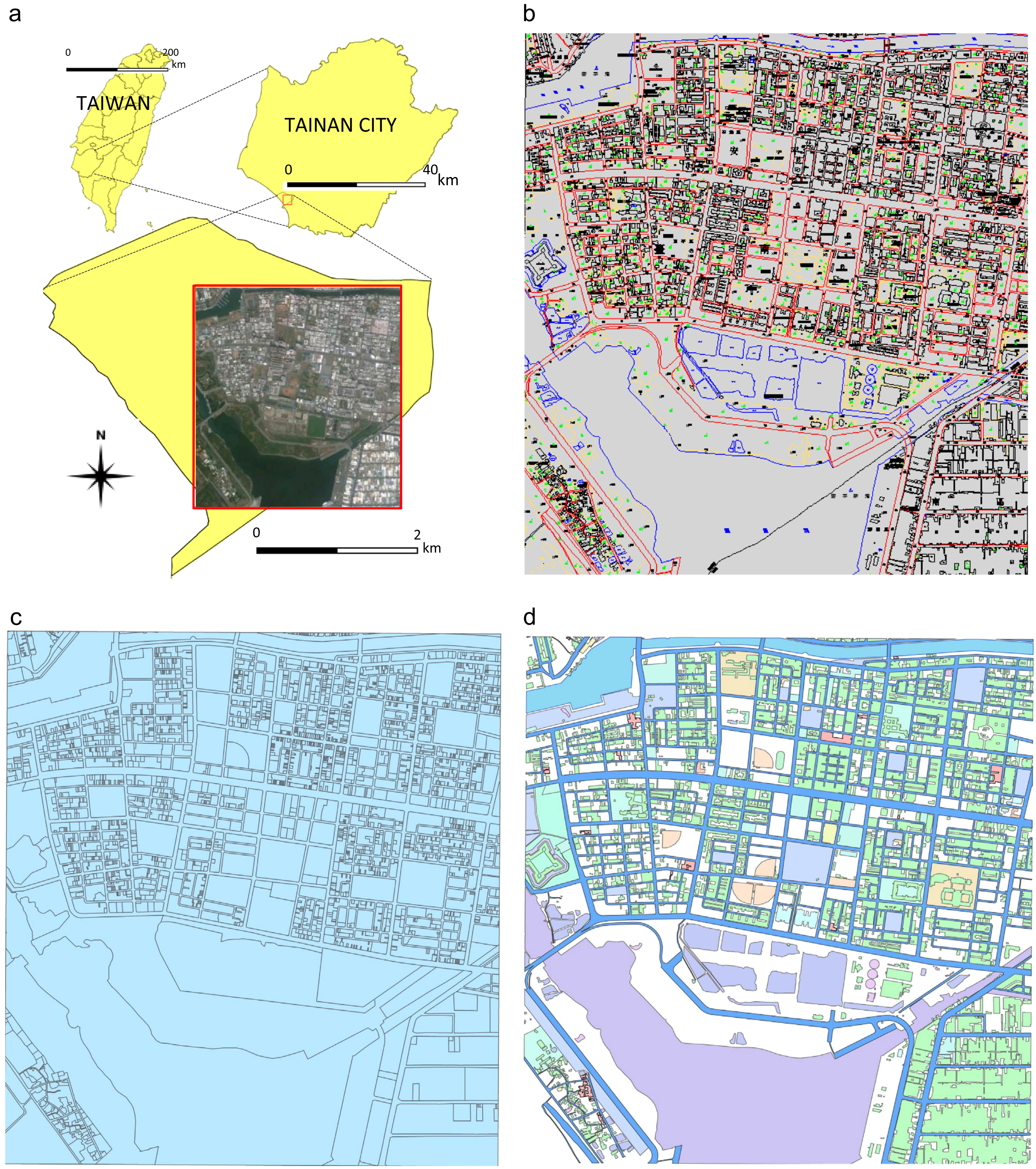

Fig. 4. Maps of the study area: (a) the location of the study area (satellite image is obtained from Google maps); (b) topographic map data (source data); (c) land use data (target data); (d) features retrieved from the topographic map for change detection.

We choose one sheet of a 1/5000 topographic map (7.1 km^2) of Tainan City as our test area, as depicted in Fig. 4(a). This region is a residential area near the seashore, having houses, historic sites, schools, government institutions, public construction, lawns, parking lots, and gas stations, etc. Fig. 4(b) and (c) illustrate the topographic map and land use data. Both datasets are produced rigorously according to their domain specifications. Topographic map data is composed of a point, line, and polygon features, while the land use data is composed of polygon features only. Collectively, 32 categories and 1325 polygon features are acquired from the topographic maps [Fig. 4(d)]. Based on their respective coding rules, each topographic map concept is represented by a five-digit

Table 3
Specifics of the change detection factors for the topographic map and land use data.

Data	Topographic map	Land use
Coordinate system	TWD97TM2	TWD97TM2
Absolute positional accuracy	1.5 m	25 m
Mapping time	2012/03	2008/09, 2008/12
Referenced image time	2010/05	2006/06
Ontology	Topographic map ontology	Land use ontology
	Bridge ontology of the topographic map and land use	

number, while each land use concept is represented by a six-digit number. Table 3 summarizes the detection factors of the two selected datasets. Referring to the positional accuracy requirements in the specification of topographic maps, the threshold value of the intersection area is set at 2.25 m^2.

Fig. 5 illustrates the change detection results. The yellow area in Fig. 5(a) shows an *NC* region. The dark blue and purple areas represent *UI* and *UII* regions [Fig. 5(b)], respectively. The *CTC* area is depicted as an orange polygon, and the *CNC* region is depicted as a red polygon in Fig. 5(c). The *NP* area is marked in white in Fig. 5.

Three cases of *CTC*-type and *CNC*-type change areas are discussed further below.

1. *Newly developed region*: This type of region often includes new constructions. The area labeled by A in Fig. 5(c) was a fishing port (*030402*) in the land use data. The topographic maps indicate that many new objects in this region differ semantically from the concept of "fishing port" and suggest it to be updated with polygons of urban roads (*94213)*, other reservoirs (*040303)*, and aquaculture (*010200)*. In the area labeled as B, one part of the unused land (*090801*) area is changed to *070203* (sports area) because a new skating rink was built there. In the area labeled as C, another unused land is changed to a traffic-relevant facility (*030304*) because a new parking lot was constructed there. Some unused land and areas under construction (*090802*) are also changed because the topographic maps suggest that buildings (*93110*) have been constructed. Thus, the new versions of topographic maps are useful references for detecting the change in the land use data of newly developed regions.
2. *Classification change*: The different domain perspectives may influence the outcomes. In Fig. 5(c), changing the land use status of the area labeled by D from *040700* (sea) to *040103* (canal) is recommended because the topographic map indicates it is a canal. The status of the area labeled E is recommended to change from park and square (*070201*) to certifiable cultural heritage (*070101*) due to similar reasons. In both cases, the phenomena in reality do not change, but we found they were interpreted by two domains differently (the same area is interpreted as sea and canal in the two domains). The comparison is thus helpful for decision makers to find possible mistakes in data (e.g., canal wrongly interpreted as sea) or ambiguous and conflict definitions between different domains.
3. *Revelation of additional details*: Many objects in the real world are not recorded in land use data, but are included in topographic map data because of its finer granularity of concepts. For example, if a small cistern is constructed within the plot of a building area, in a topographic map although the map depicts only a type of land use (*060600* or environmental protection facility) in the land use data [labeled as F in Fig. 5(c)]. The analysis indicates that the area changes in type from *060600* to *040303*. This result implies that changes caused by granularity can be detected.

Cases A, B, and E are explained in detail and step-by-step in Fig. 5(c). Taking area E as an example, the region with a fluorescent blue border in the first image represents an area interpreted as park and squire (*070201*) in the target dataset (original land use data). In the second image, the source data (topographic map, depicted in light blue color) are superimposed. The overlay result shows that one part of this area is historical site (*99431*). The third image shows that the light-blue intersection area in the second image changes from park and square (*070201*) to certifiable cultural heritage (*070101*) following rule 7 in Table 1 because *99431* is *sem_null* to *070201* and *99431* is *sem_subset* to *070101*.

The experimental results demonstrate that automatic change detection is feasible. For this particular test case, the percentages for the *CTC*, *CNC*, *NC*, *UI*, *UII*, and *NP* change types are 18%, 9%, 10%, 24%, 7%, and 32%, respectively. The percentages of the three major categories for change detection, that is, *confirmed_result*, *uncertain_change,* and *non_process,* are 37%, 31%, and 32%, respectively. Although these numbers are for reference only because they may vary dramatically given different test areas or datasets, it is clear that if increased instances of *sem_null*, *sem_exact* or *sem_subset* semantic relationships are detected for the intersection area, we can optimistically expect an increased percentage for the *confirmed_result* category.

Another strategy to improve the detection result involves reducing the areas for the *non_process* and *uncertain_change* categories. For *non_process* areas, other types of data besides polygon data, such as point or line features, may be used for analysis. Although these features cannot suggest the exact change boundary, they can still be used to suggest potential changes. To reduce the area of the *uncertain_change* category, a possible method involves analyzing the *sem_overlap* and *sem_supersert* semantic relationships and converting them into *sem_exact* or *sem_subset* semantic relationships whenever possible through a careful cross-domain examination. For example, the semantic relationship between urban road (*94213)* in the topographic map data and general road (*030303*) in the land use data is determined as *sem_overlap*. Although urban roads can be viewed as a type of general road by definition, this relationship is interpreted as such because the definition of general road has an additional restriction in terms of road width. If the same restriction can be incorporated into the definition of urban road as well, then this semantic relationship will become *sem_subset* instead. Such cross-domain tasks improve outcomes; for instance, the land use code for a new road [(Fig. 6(c)] is determined as *030303*, and the area of the urban road marked in blue [Fig. 6(d)] is identified as the *NC* change type.

The following section uses an updated version of land use data generated in 2012 as ground truth data with which to validate the outcomes of the *confirmed_result* category. Table 4 presents the accuracy rates of the *NC*, *CTC,* and *CNC* types. The number of areas is generated by the intersection of *confirmed_result* category areas with ground truth data. In general, the change detection results using the proposed method are consistent with those obtained from the new land use data. Three major reasons for inconsistency are discussed below:

1. Cognition: According to the definition of cognition[2], the semantics of a concept are influenced by the cognition of each domain. For example, area A in Fig. 6 is recorded as a lake in the topographic map data, whereas it is a part of the sea in the land use data. The detection results suggest that this area should be changed from sea (*040700*) to the lake (*040302*). However, the land use data for the year of 2008 and 2012 both represent this

[2] Cognition is defined as "*conscious mental activities*", as per http://www.merriam-webster.com/.

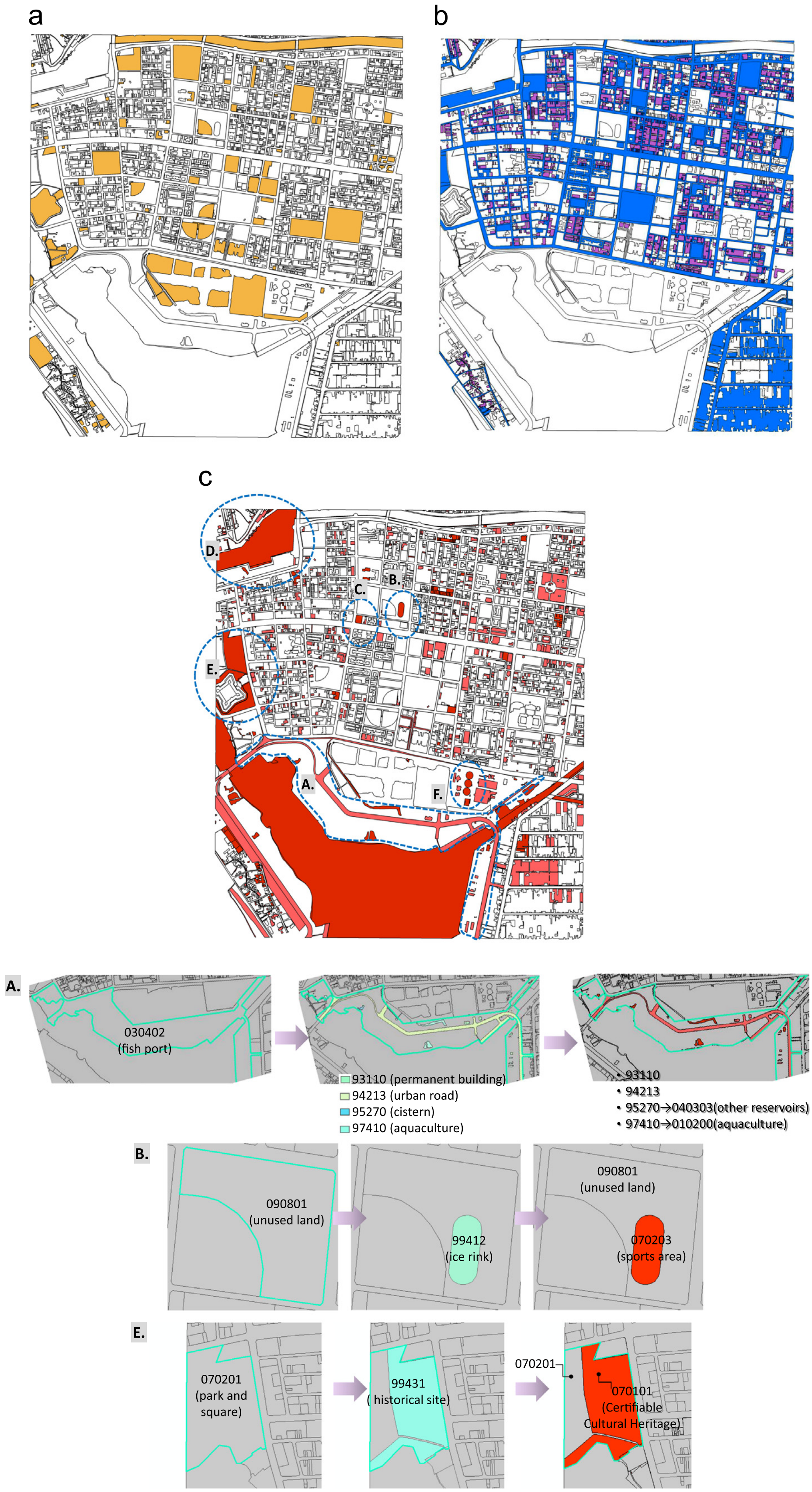

Fig. 5. Results of change detection: (a) *NC* type (yellow area); (b) *Uncertain_change* category: *UI* (dark blue area) and *UII* types (purple area); (c) *CTC* (orange area) and *CNC* types (red area). The three-step image exhibits the change detection procedure (the area of the blue border is the target): original land use→overlaying topographic map data→the change detection result of target. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

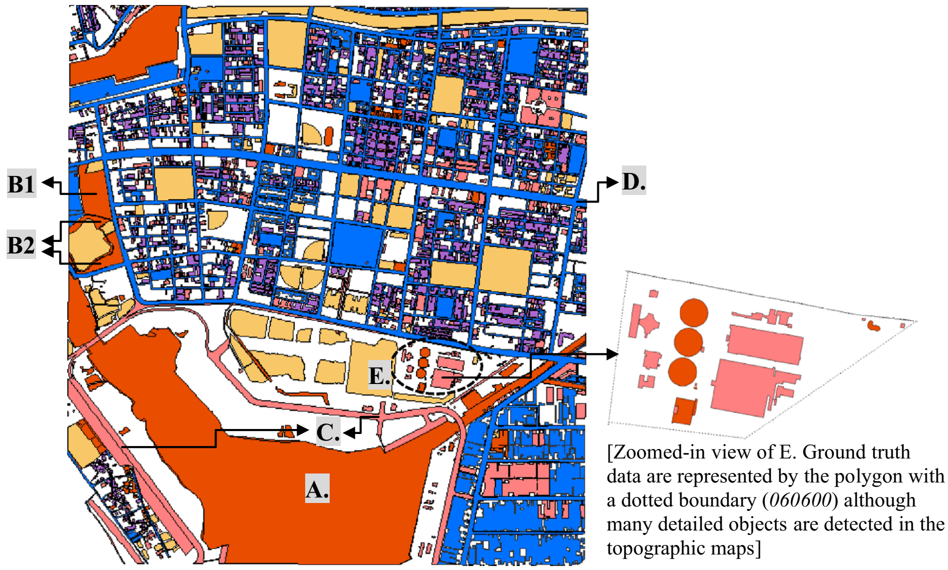

Fig. 6. Results of change detection analysis. A: change from *040700* to *040302*; B1 and B2: change from *070201* to *070101*; C: construction of new roads; D: urban road with an *sem_overlap* semantic relationship; and E: change from *060600* to *040303*.

Table 4
Accuracy rates of the *confirmed_result category* (compared with land use ground truth data).

Test result	Ground truth data		
Test type	Change (number of area)	No change (number of area)	Accuracy Rate (%)
NC	59	87	87/146 = 59.59
CTC	276	113	276/195 = 70.95
CNC	2202	361	2202/2561 = 86.91

region as sea. Since product specifications are domain-dependent, inconsistencies or even contradictions may be inevitable. A cross-domain examination of related semantic relationships is thus necessary to clarify their differences and to improve the operation results.

2. Boundary/scope: This issue is also related to the cognition issue concerning how domains determine the scope of features. For example, areas B1 and B2 in Fig. 6 (orange areas) are detected as a *CTC* type and recommended to change from *070201* to *070101* because the topographic maps show that these regions are heritage sites. However, the ground truth data indicate that areas B1 and B2 are *NC* type because they are lawns. Moreover, the scope of the heritage site covers only the main building. Therefore, the inclusion of the area surrounding the main building generates a different scope of features and influences change analysis. Understanding how each domain defines the scope of features is thus important.
3. Granularity: Various modeling perspectives and levels of granularity differentiate the data from one another. The granularity of the topographic map data is generally finer than that of the land use data. Hence, the topographic map data possesses features that the land use data does not have. For example, three cisterns are presented within a wastewater recycling center [Fig. 6(e)] in the topographic map data. The code for cistern is *95270*. Therefore, the intersection area of the target data is determined as a *CTC* type and recommended to change from environmental protection facility (*060600*) to other reservoirs (*040303*). The semantic relationship between cistern (*95270*) and (*040303*) is *sem_subset*. However, the ground truth data considers this relationship to be *NC* because the cistern is not considered in the land use data. Other details in this area are inconsistent with the ground truth data for the same reason.

4. Conclusion and future work

GIS interoperability is challenged in terms of achieving semantic interoperability with efficient processing strategies. In this study, a conceptual framework is designed for the semantic and interoperable applications of geodata. The framework includes six components and follows four steps to realize semantic processing based on bridge ontology and application-dependent rules. The bridge ontology records the semantic relationship of concepts formally and benefits from the comparison of concepts among domains. Rules based on bridge ontology also facilitate automatic and intelligent processing. We successfully applied our framework in the development of an automatic change detection mechanism. Six change types were proposed based on spatial, temporal, and semantic geodata, and our findings conclude two helpful categories of change detection results. The *confirmed_result* category includes both changed and unchanged instances. The *uncertain_change* category reflects areas that may require further examination. In our experiments, the topographic map data are used to detect land use data and to generate change detection results. These results demonstrate that our framework realizes the semantic interoperability of geodata obtained from different domains and can be used to automatically detect changes or even to provide update suggestions.

We expect this mechanism to reveal an entirely new prospect for change detection analysis based on existing geodata from other domains. We also believe that the proposed framework will contribute to numerous novel applications of semantic integration. Our future work will apply point and line features to change detection and establish a robust mechanism to reduce the number of *NP* areas.

References

Ahlqvist, O., 2008. Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: a study of 1992 and 2001 US National Land Cover Database changes. Remote. Sens. Environ. 112, 1226–1241.

Amrouch, S., Mostefai, S., 2012. Survey on the literature of ontology mapping, alignment and merging. In International Conference on Information Technology and e-Services (ICITeS), 1–5. IEEE.

Baglatzi, A., Kuhn, A.W., 2013. On the formulation of conceptual spaces for land cover classification systems. In: Vandenbroucke, D., Bucher, B., Crompvoets, J. (Eds.), Geographic Information Science at the Heart of Europe. Springer International Publishing, Cham, pp. 173–188.

Bishr, Y., 1998. Overcoming the semantic and other barriers to GIS interoperability. Int. J. Geogr. Inf. Sci. 12, 299–314. http://dx.doi.org/10.1080/136588198241806.

Bouziani, M., Goïta, K., He, D.-C., 2010. Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge. ISPRS J. Photogramm. Remote Sens. 65, 143–153.

Bruzzone, L., Prieto, D.F., 2000. Automatic analysis of the difference image for unsupervised change detection. Geosci. Remote Sens. IEEE Trans. 38, 1171–1182. http://dx.doi.org/10.1109/36.843009.

Buccella, A., Cechich, A., Gendarmi, D., Lanubile, F., Semeraro, G., Colagrossi, A., 2011. Building a global normalized ontology for integrating geographic data sources. Comput. Geosci. 37, 893–916. http://dx.doi.org/10.1016/j.cageo.2011.02.022.

Butkiewicz, T., Chang, R., Wartell, Z., Ribarsky, W., 2008. Visual analysis and semantic exploration of urban LIDAR change detection. In *Computer Graphics Forum,* pp. 903–910. Wiley Online Library.

Chrisman, N., 2001. Exploring Geographical Information Systems. Wiley Publishing.

Clementini, E., Di Felice, P., and van Oosterom, P., 1993. A small set of formal topological relationships suitable for end-user interaction. In *Advances in Spatial Databases*, 277–295. Springer.

Cruz, I.F., Sunna, W., and Chaudhry, A., 2004. Semi-automatic ontology alignment for geospatial data integration. In *Geographic Information Science*, 51–66. Springer.

de Bruijn, J., Ehrig, M., Feier, C., Martin-Recuerda, F., Scharffe, F., Weiten, M., 2006. Ontology mediation, merging and aligning. Semant. Web Technol., 95–113.

Egenhofer, M.J., Franzosa, R.D., 1991. Point-set topological spatial relations. Int. J. Geogr. Inf. Sci. 5, 161–174.

Euzenat, J., Shvaiko, P., 2007. Ontology Matching. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-49612-0.

Fonseca, F.T., Egenhofer, M.J., Agouris, P., Camara, G., 2002. Using ontologies for integrated geographic information systems. Trans. GIS 6, 231–257. http://dx.doi.org/10.1111/1467-9671.00109.

George, D., 2005. Understanding structural and semantic heterogeneity in the context of database schema integration. J. Dep. Comput. UCLAN 4, 29–44.

Giunchiglia, F., Shvaiko, P., 2003. Semantic Matching. The Knowledge Engineering Review, 18, 265–280.

Goddard, C., 2012. Semantic primes, semantic molecules, semantic templates: Key Concepts in the NSM Approach to Lexical Typology.

Goddard, C., Wierzbicka, A., 2002. Meaning and Universal Grammar: Theory and Empirical Findings, vol. 2. John Benjamins Publishing, Amsterdam.

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing? Int. J. Hum. Comput. Stud. 43, 907–928. http://dx.doi.org/10.1006/ijhc.1995.1081.

Heer, T., Retkowitz, D., and Kraft, B., 2009. Tool Support for the Integration of Light-Weight Ontologies. In *Enterprise Information Systems*, pp. 175–187. Springer.

Hong, J.-H., Kuo, C.-L., 2015. A semi-automatic lightweight ontology bridging for the semantic integration of cross-domain geospatial information. Int. J. Geogr. Inf. Sci. 29, 2223–2247. http://dx.doi.org/10.1080/13658816.2015.1072200.

Im, J., Jensen, J., Tullis, J., 2008. Object-based change detection using correlation image analysis and image segmentation. Int. J. Remote Sens. 29, 399–423.

Janowicz, K., Raubal, M., Schwering, A., Kuhn, W., 2008. Semantic similarity measurement and geospatial applications. Trans. GIS 12, 651–659.

Kavouras, M., Kokla, M., 2007. Theories of Geographic Concepts: Ontological Approaches to Semantic Integration. CRC Press, Boca Raton, FL.

Kavouras, M., Kokla, M., Tomai, E., 2005. Comparing categories among geographic ontologies. Comput. Geosci. 31, 145–154. http://dx.doi.org/10.1016/j.cageo.2004.07.010.

Keet, C.M., 2004. Aspects of ontology integration, literature research & background information for the PhD proposal entitled: bottom-up development of ontologies and ontology integration in the subject domain of ecology. *School of Computing*. School of Computing, Napier University, Edinburgh.

Kokla, M., Kavouras, M., 2005. Semantic information in geo-ontologies: extraction, comparison, and reconciliation. In *Journal on Data Semantics III,* 125–142. Springer.

Kuhn, W., Raubal, M., 2003. Implementing semantic reference systems. In: Gould, M., Laurini, R., Coulondre, S. (Eds.), *The 6th AGILE International Conference on Geographic Information Science*. Presses Polytechniques et Universitaires Romandes, Lyon, pp. 63-72.

Leung, N.K., Kang, S.H., Lau, S.K., Fan, J., 2009. Ontology Matching Techniques: a 3-Tier Classification Framework.

Matikainen, L., Hyyppä, J., Ahokas, E., Markelin, L., Kaartinen, H., 2010. Automatic detection of buildings and changes in buildings for updating of maps. Remote Sens. 2, 1217–1248.

Noy, N.F., 2004. Semantic integration: a survey of ontology-based approaches. ACM Sigmod Rec. 33, 65–70. http://dx.doi.org/10.1145/1041410.

Pinto, H.S., Martins, J.P., 2001. A methodology for ontology integration. In: *Proceedings of the 1st International Conference on Knowledge Capture*, ACM, pp. 131-138.

Pundt, H., Bishr, Y., 2002. Domain ontologies for data sharing–an example from environmental monitoring using field GIS. Comput. Geosci. 28, 95–102.

Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B., 2005. Image change detection algorithms: a systematic survey. Image Process. IEEE Trans. 14, 294–307.

Schwering, A., 2008. Approaches to semantic similarity measurement for geo-spatial data: a survey. Trans. GIS 12, 5–29.

Shalaby, A., Tateishi, R., 2007. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. Appl. Geogr. 27, 28–41.

Shvaiko, P., Euzenat, J., 2013. Ontology Matching: State of the Art and Future Challenges IEEE Transactions on Knowledge and Data Engineering (TKDE). 25, pp. 158-176. 〈http://dx.doi.org/10.1109/TKDE.2011.253〉.

Stevens, R., Goble, C.A., Bechhofer, S., 2000. Ontology-based knowledge representation for bioinformatics. Brief. Bioinform. 1, 398–414. http://dx.doi.org/10.1093/bib/1.4.398.

Studer, R., Benjamins, V.R., Fensel, D., 1998. Knowledge engineering: principles and methods. Data Knowl. Eng. 25, 161–197. http://dx.doi.org/10.1016/S0169-023X(97)00056-6.

Tian, J., Cui, S., Reinartz, P., 2014. Building change detection based on satellite stereo imagery and digital surface models. geosci. Remote Sens. IEEE Trans. 52, 406–417.

Uitermark, H.T., van Oosterom, P.J.M., Mars, N.J.I., Molenaar, M., 2005. Ontology-based integration of topographic data sets. Int. J. Appl. Earth Obs. Geoinform. 7, 97–106. http://dx.doi.org/10.1016/j. jag.2005.03.002.

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. In *IJCAI-01 Workshop: Ontologies and Information Sharing*, 108–117. Citeseer.

Walter, V., 2004. Object-based classification of remote sensing data for change detection. ISPRS J. Photogramm. Remote Sens. 58, 225–238.

Wright, C.K., Wimberly, M.C., 2013. Recent land use change in the Western Corn Belt threatens grasslands and wetlands. Proc. Natl. Acad. Sci. 110, 4134–4139.

Xu, B., Wang, P., Lu, J., Li, Y., Kang, D., 2004. Bridge ontology and its role in semantic annotation. *In: International Conference on Cyberworlds*. IEEE, Tokyo, pp. 329-334.

Zhang, C., 2004. Towards an operational system for automated updating of road databases by integration of imagery and geodata. ISPRS J. Photogramm. Remote Sens. 58, 166–186.

Zhang, Q., Couloigner, I., 2004. A framework for road change detection and map updating. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 35, 720–734.