Semantically Interoperable Geospatial Information Processing: A Review and A Canonical Model Approach

Sung-Gheel Jang and Tschangho John Kim
111 Temple Buell Hall, 611 Taft Dr. Champaign, IL 61820
University of Illinois at Urbana-Champaign, Illinois, USA
sjang3@uiuc.edu, tjohnkim@uiuc.edu

INTRODUCTION

Sharing geospatial data across multiple, distributed, and heterogeneous information sources is not trivial when there is little information on the structures and meanings of geospatial data that we wish to share across information sources. Even if geospatial data across multiple organizations has been successfully exchanged or transferred electronically using coordinated data file formats or transfer standards, challenges to integrate the exchanged geospatial data into new or existing information systems still remain. In recent years, the notion of *interoperability* has been widely recognized as an underpinning of pervasive information systems. While there also have been extensive works on the interoperability of geospatial information at various levels, however, there are few studies that provide a tangible and feasible conceptual framework for an interoperable geospatial information system that reflects recent distributed network technology.

This paper, first, surveys the interoperability researches of geospatial information systems from organizational, technical and cognitive perspectives. Further, this paper examines major approaches that can shed light on the development of a conceptual framework to solve issues on interoperability. Finally, this paper proposes a canonical model approach based upon the ISO 19100 series of international standards for geographic information as a feasible conceptual framework for an interoperable geospatial information system that accesses distributed and heterogeneous spatial information bases across jurisdictions in a decentralized fashion.

TWO PERSPECTIVES ON SHARING GEOSPATIAL INFORMATION

Table 1 shows the major issues that this paper wants to address from two perspectives on sharing geospatial information: 1) organizational and technological perspective, and 2) cognitive perspective. As shown in Table 1, this paper examines the two perspectives with the following framework: *who* wants to share the geographic information?; *what* geographic information is to be shared?; and *how* can geographic information be shared?

The Organizational and Technological Perspective

Who: the principal question is who wants to share the geographic information? Most previous studies have paid primary attention to the institutional entities that may want to share data between local, state, and federal organizations at either the inter- or intra-organizational levels. Therefore, they have conceived data sharing schemes based on the physical organization (e.g. city government, planning department) rather than an abstract domain (e.g. transportation planning, urban development). Efforts to ease the sharing of geospatial information have been focused on the elimination of technological and organizational barriers (Onsrud and Rushton 1995). Many studies conclude that the institutional, organizational, or behavioral aspects are more pertinent and significant than the technological aspects of sharing geographic information (Onsrud and Rushton 1995; Nedovic-Budic and Pinto 1999). Therefore, most research subjects are related to investigating the factors or impediments of the inter-organizational relationship or to coordination among organizations.

Table 1: Two perspectives on sharing geospatial information

	Organizational and Technological Perspective	**Cognitive Perspective**
Who	Organization or institutional body	Geospatial information community (GIC)
What	Spatial data (e.g., geodetic data, aerial photos, satellite imagery, elevation data, cadastral records, political boundaries, land use, etc.)	Spatial information focusing on the semantics (e.g., different classification systems for the same or similar themes, inconsistent view on the same themes among different GICs)
How	i) understanding the behaviors of organizations (organizational barriers: conceptual framework, mechanisms, factors or impediments, outcomes or benefits); ii) overcoming technical barriers (data transfer standards, metadata); iii) syntactic and schematic interoperability	i) pursuing schematic and semantic interoperability; ii) use of same conceptual model; iii) adopting ontologies

What: studies in the organizational context of sharing geographic information define the geographic information simply as geographic data or spatial data. A U.S. nation-wide survey (Warnecke et al. 1998) categorized spatial data types into: geodetic data, aerial photos, digital orthophotos, satellite imagery, elevation data, cadastral/land records, political boundaries, roads, hydrology, utilities, land-use/zoning, land cover, tree inventories, soils, wetlands, wildlife/habitat, and unique/natural areas. While researchers taking this approach have successfully exchanged geographic *data* by removing the organizational impediments and the technical barriers for the sharing, they did not succeed to share geographic *information*.

How: issues surrounding data sharing are usually classified as those pertaining to technology and those related to the organization. It is emphasized that these issues are mutually dependent. An extensive body of work has been done focusing on the following issues: 1) *conceptual framework*; 2) *mechanisms*; 3) *factors or impediments*; and 4) *outcomes or benefits* (Onsrud and Rushton 1995; Nedovic-Budic and Pinto 2000, 2001). Many studies also agree that standardization would play a pivotal role to overcome the technical impediments (Albrecht 1999; Kim 1999; Keller 1999).

The Cognitive Perspective

Who: abstract *information communities* are formed within domains, disciplines and universes of discourse without requiring the presence of an established institutionalized entity (e.g. transportation community, earth resources community). The geospatial information community (GIC) has been introduced as “a group of spatial data producers and users who share *ontology* of real-world phenomena.” (Bishr et al. 1999a) The work on sharing geospatial information from the cognitive perspective, therefore, focuses on how the GICs share the ontology or semantics of their domain within the same GICs or among different GICs.

What: the cognitive studies on sharing geospatial information focus on the semantics of geospatial information. The formal specification of the semantics of geospatial information is called ontology. Craig (1995) introduces the intrinsic problem of spatial data, called “idiosyncratic data.” It implies that an inconsistent classification or omission of certain data fields between two different GICs raises the need for a new research approach to reconcile the differences of semantics between the GICs. Users often realize, after sharing the data, that they can not use them due to the inconsistency of perspectives from the person who recorded them. In this case, we can simply state that data transfer was successful, but information was not shared (Bishr et al. 1999a).

How: people conceptualize the real world in different ways and as a result different specifications of the same phenomena may exist. We often face the problem of semantic heterogeneity in the sharing of geospatial information among different GICs. Bishr and Kuhn (2000) argue that any coherent information model needs to be based on an accepted ontological foundation to guarantee unambiguous interpretation, and that a model of reality and its ontological foundation should be accessible by computers for successful automatic interpretation of information. Ontology of geographic kinds is designed to yield a better understanding of the structure of the geographic world, and to support the development of geographic information systems that are conceptually sound (Pundt and Bishr 2002). Harvey et al. (1999) argue that semantic interoperability would be a substantive alternative to resolve the conflict of data sharing among different organizations and to improve access.

APPROACHES TO THE INTEROPERABILITY RESEARCH

A core question on the interoperability research would be how a computerized system can be developed while being syntactically, schematically, and semantically interoperable with heterogeneous spatial databases. This question on interoperable information systems has been pursued since the early 1980's even though the awareness on semantic interoperability is relatively new. Sheth (1999) provides excellent comparative analysis on the paradigm changes of the research and development of information systems while paying attention to the historical changes of resolving technologies to each interoperability issue such as syntactic, schematic, and semantic interoperability.

Based upon the Sheth (1999)'s observation, two salient approaches to integrating heterogeneous spatial databases including non-spatial databases are surveyed in this paper. Since the interoperability problems of the first generation have been mostly resolved by the advances of hardware and software technology, the primary focuses of the approaches lie in the schematic and semantic interoperability. The schema integration approach, first, has been intensively studied since mid 1980’s, and it still influences the most current technologies which must deal with distributed and heterogeneous databases in a variant way. More and more, the ontology approach is drawing big attentions from recent researchers since it is focusing on the semantic interoperability problem. After reviewing the two approaches, finally, we propose a canonical model approach based on a conceptual modeling process adopted by an international standardization organization for geographic information.

Schema Integration Approach

Since the mid 1980's in the database management and other related fields, there has been much consideration of integrating multiple heterogeneous and autonomous databases. The integrated database system is called a federated database system (FDBS), which is a collection of cooperating but autonomous component database systems (Sheth and Larson 1990). The main features of the FDBS are characterized as: distribution, heterogeneity, and autonomy. Researchers have investigated heterogeneity at the communication level (operating system, hardware, and communication system) as well as at the database system level. However, heterogeneity at the database system level continues to be a major research problem. Sheth and Larson (1990) categorize the heterogeneities at the database system level as: i) differences in DBMS (i.e., the data model which includes differences in structure, constraints, and query languages; and system level support), and ii) semantic heterogeneity. Interestingly, semantic heterogeneity in database systems -- the difficulty in identification, resolution, and use of DBMS schemas which do not provide enough semantics to interpret data consistently -- is known and acknowledged.

Schema integration is a very powerful and practical methodology for those who want to integrate different structures on spatial objects and spatial representation in order to allow the reuse of existing data sources in the sense of schematic interoperability. If a spatial information system has to be developed in a conventional manner, first, the modern system designer will identify what data is needed through application requirement analysis. After that, he/she will try to integrate the newly acquired data from various heterogeneous data sources into a single, uniform data store. While the issues and the resolutions of the schema integration methodology have been thoroughly investigated for conventional databases, there have been only a handful of works on systematic schema integration for spatial information systems.

In addition to the well-known issues of the schema integration for conventional databases, Laurini (1998) comprehensively identifies the unique challenges of the schema integration for spatial databases including matching various spatial representations, matching geometric discrepancies, and matching boundaries. As a modified version of Parent and Spaccapietra (1998)'s procedure, he suggests a general schema integration procedure for spatial multi-databases by including four more processes: geometric conflict resolution, boundary alignment, topological continuity, and spatial indexing. This implies that building a federated spatial database system is much more complex than the conventional federated database systems due to the special characteristics of spatial information. Devogele et al. (1998) demonstrate more tangible schema integration procedures by showing the issues and their resolutions in the process of building an integrated schema using two different spatial databases with different spatial representations due to differences in scale. Specifically, they use the Unified Modeling Language (UML) to delineate the data model that each spatial database employs while describing schema correlations between two local schemas using an inter-schema correspondence assertion.

Ontology Approach

Various kinds of information communities (i.e., domains) use a variety of different disciplines or universes of discourse such as transportation planning, urban development, etc. Each information community presumably has its own semantics for its universe of discourse, and thus the semantics between information communities are not the same. The similarity and/or dissimilarity of the semantics between information communities, however, need to be reconciled and integrated in order to achieve semantic interoperability in a heterogeneous information system.

Ontologies are considered to be an adequate methodology to support semantic interoperability in heterogeneous information systems including geospatial information systems since ontologies can provide a "common basis" for semantic mapping between information communities (Bishr et al. 1999b). Ontologies are explicit formal specifications of domains of discourse on the premise that ontologies facilitate knowledge sharing and reuse (Gruber 1993; Noy and Musen 2003). Guarino (1998) elaborates further the notion of an ontology as "a logical theory accounting for the intended meaning of a formal vocabulary, i.e., its ontological commitment to a particular conceptualization of the world." He significantly contributed to the domain of ontology research by proposing the four levels of ontologies, which are: top-level ontology, domain ontology, task ontology, and application ontology.

More importantly, Guarino (1998) addresses that it would be convenient to agree on a single top-level ontology rather than relying on agreements based on the intersection of different ontologies when needing to integrate ontology between two different information communities. This implies that a top-down approach would make less trouble to build domain or application ontologies than a bottom-up approach, which requires painstaking integration processes such as reconciling disparate ontologies and mapping ontologies. It is no wonder, therefore, that there are active movements to build standardized top level ontologies that domain experts can use to share and annotate information in their field: e.g., SNOMED, Unified Medical Language System in the medicine domain; the UNSPSC classification for products and services (Noy and Musen 2003). Major research issues in the study of ontology are: defining correlation between multiple ontologies; determining semantic similarity; integrating ontologies that have various approaches such as merging, transformation, alignment, and articulation; versioning ontologies and managing ontologies; generating ontologies; and developing an ontology editing tool (Wache et al. 2001; Noy and Musen 2003).

Since the end of the 1990's, there has active body of works have developed on employing an ontological approach to resolve the semantic heterogeneity issues in the geospatial information domain. The majority of previous works focus on showing the rigorousness of the ontology method to integrate different geospatial classification system such as land use and land cover classification using the Formal Concept Analysis (Kokla and Kavouras 2001; Kavouras and Kokla 2002), measuring semantic similarity (Fonseca et al. 2002; Rodriguez and Egenhofer 2003; Feng and Flewelling 2004), semantic translator or semantic mapper (Bishr 1998; Bishr et al. 1999b; Visser et al. 2002). There have been only a few efforts concerning the practical application of the ontology approach to a GIS. Those works are also rather limited to support more precise data collection in accordance with ontologies in a specific domain such as stream survey and geological field mapping (Pundt and Bishr 2002; Pundt 2002; Brodaric 2004).

Canonical Model Approach

The notion of a canonical (or common) model comes from the research on federate database systems (FDBS) (Sheth and Larson 1990; Pitoura et al. 1995). Jang and Kim (2006) introduce canonical model approach for modeling an interoperable GIS application based on authoritative conceptual models. Since the ISO 19100 series of international standards for geographic information provides a pool of standardized conceptual schemas and methods, semantically coherent and interoperable geographic information applications or services will be feasible if we accept the standardized models as canonical model. Jang and Kim (2006) elaborate how the conceptual modeling process of the ISO 19100 series ensures development of semantically interoperable GIS applications by showing its semantically coherent architecture of the standardized concepts as well as conceptual parallelism between the ISO 19100 series and ontology approach (See Figure 1).

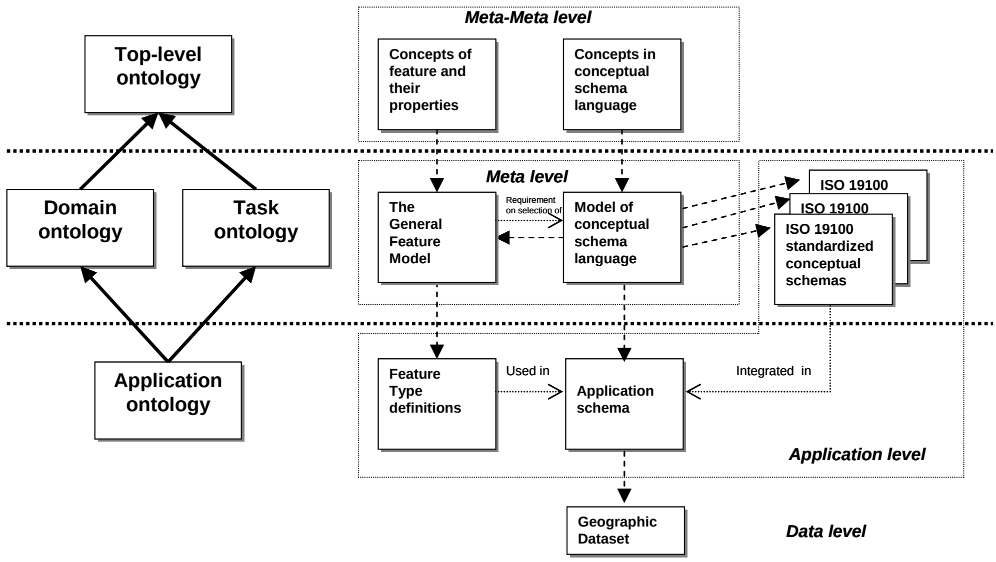

Figure 1: Correspondences between the four layers of the ISO 19100 series and the three types of Ontology (Jang and Kim 2006, p.117).

In the ISO 19100 series, geographic concepts are explicitly specified using a rigorous object-oriented formalism, i.e. conceptual schema language, which is a UML profile for the ISO 19100 series. Thus, we propose a framework to develop application schemas for an interoperable GIS data model using the conceptual schemas of the ISO 19100 series as *quasi*-ontologies. Using this method, any domain expert (e.g. transportation engineer) can formally describe target application schemas for a specific domain application both by conceptualizing the problems of the domain of interest and by adopting rules of application schema.

Based upon the ontology-based modeling process (Fonseca et al. 2003), we propose a conceptual modeling process using the standardized conceptual schemas of the ISO 19100 series that commit as the geospatial information domain ontologies, and the standardized conceptual schemas can be modeled to application schemas – formal descriptions about data required by a specific application – while taking into account particular domain knowledge (See Figure 2). As a result of the proposed process, semantically coherent application schemas for a particular application are derived in a top-down manner without causing semantic disparities among different levels of abstractions.

In short, effective conceptual modeling processes for a semantically interoperable GIS are as follows:

- Identify domain concepts (domain ontologies) for a particular application that uses transportation data,
- Identify standardized conceptual schemas in the ISO 19100 series that satisfy the domain concepts,
- Create application schemas (application ontologies) for the particular application domain that satisfy the domain concepts either by reusing or by extending the standardized conceptual schemas in a conceptual schema language (i.e. ISO 19103), and
- Create either platform neutral encodings (e.g. XML) or communication interfaces for varied computing platforms (e.g. SQL, OLE/COM, Java, etc.) from the application schemas.

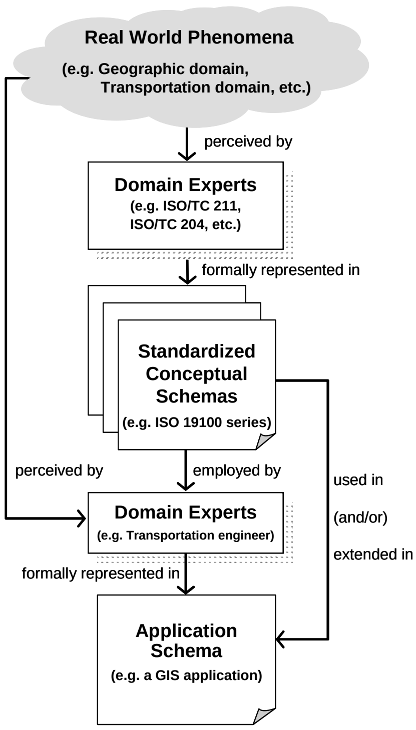

Figure 2: Proposed conceptual modeling process.

CONCLUSIONS

The integration of heterogeneous geospatial resources becomes a very challenging subject in the geographic information science field since it requires overcoming inherent impediments due to the use of heterogeneous geospatial information with different structures and meanings across multiple information sources. To come up with effective and feasible ways to overcome the integration issues of geographic information systems, it is important to have clear understandings on previous research efforts as well as research trends. In this regard, we reviewed the interoperability researches on geographic information systems from organizational and technological perspective and cognitive perspective. The *who*, *what*, and *how* framework has provided clear perspectives on sharing efforts for distributed and heterogeneous geographic information. We have proposed a canonical model approach based upon standardized concepts adopted in the ISO 19100 series of international standards in order to achieve unambiguous sharing of concepts in the geospatial information domain. In another research (Jang and Kim 2006), we have evaluated the proposed canonical model approach by: creating application schemas for a specific GIS application (e.g. multimodal location based services) where geographic information plays a crucial role; implementing a platform neutral encoding (i.e. ISO 19136 GML application schema) from the created application schemas; and developing a prototype to evaluate the achievement of a schematically and semantically interoperable GIS application utilizing web-based services.

In summary, Table 2 compares major characteristics of the approaches to the interoperability research including the proposed approach in terms of various implementation aspects of interoperable geospatial information processing: level of interoperability, key player, feasibility of implementation, and expected bottle-neck of implementation.

Table 2: Characteristics of the three approaches to interoperable geospatial information processing.

	Schema integration approach	**Ontology approach**	**Canonical model approach**
Level of interoperability	Syntactic, schematic	Syntactic, schematic, semantic	Syntactic, schematic, semantic
Key player	Modeler with assistance of GI domain experts	Ontologist with assistance of GI domain experts	GI domain experts
Feasibility of implementation	Well known approach, supports from many commercial products	Developing stage, esp., few experiences on implementing GI applications	Rapid development by creating profiles of the standards
Expected bottleneck of implementation	Creating integrated schema, mapping, esp., for geospatial data	Creating an authoritative ontology	Creating interfaces for existing systems

ACKNOWLEDGEMENTS

This work has been done partially supported by the Ministry of Construction and Transportation of Korea through the Korea Institute of Construction and Transportation Technology Evaluation and Planning (Project number: 05-Infrastructure-D02).

BIBLIOGRAPHY

Albrecht, J. (1999), Towards Interoperable Geo-Information Standards: A Comparison of Reference Models for Geo-Spatial Information. Annals of Regionl Science. 33, 151–169.

Bishr, Y. (1998), Overcoming the semantic and other barriers to GIS interoperability. International Journal of Geographical Information Science. Vo. 12, No. 4, 299–314.

Bishr, Y. A., et al. (1999a), Probing the Concept of Information Communities – a First Step toward Semantic Interoperability. In: Goodchild, M., et al. (eds): Interoperating geographic information systems. Kluwer Academic Publishers, MA, 55–69.

Bishr, Y. A., Pundt, H., and Ruther, C. (1999b), Proceeding on the Road of Semantic Interoperability - Design of a Semantic Mapper Based on a Case Study from Transportation. In: Vckovski, A. et al. (eds): Interoperating Geographic Information Systems. Proceedings of the 2nd

International Conference INTEROP'99. Zurich, Switzerland, March 10-12. Springer-Verlag, Berlin, 203–215.

Bishr, Y. and Kuhn, W. (2000), Ontology-Based Modelling of Geospatial Information. Proceedings of 3rd AGILE Conference on Geographic Information Science. Helsinki/Espoo, Finland, May 25–27, 24–27.

Brodaric, B. (2004), The design of GSC FieldLog: ontology-based software for computer aided geological field mapping. Computers & Geosciences, 30, 5–20.

Craig, W. J. (1995), Why We Can't Share Data: Institutional Inertia. In: Onsrud, H. J., Rushton G. (eds.): Sharing Geographic Information. the Center for Urban Policy Research, New Brunswick, NJ, 107–118.

Devogele, T., Parent, C., and Spaccapietra, S. (1998), On spatial database integration. International Journal of Geographical Information Science. Vol. 12, No. 4, 335–352.

Feng, C.-C. and Flewelling, D. M. (2004), Assessment of semantic similarity between land use/land cover classification systems. Computers, Environment and Urban Systems. 28, 229–246.

Fonseca, F. T., et al. (2002), Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS. Vol. 6, No. 3, 231–257.

Fonseca, F., Davis, C., and Camara, G. (2003), Bringing ontologies and conceptual schemas in geographic information integration. GeoInformatica. Vol. 7, No. 4, 355–378.

Gruber, T. R. (1993), Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In: Guarino, N., Poli, R. (eds): Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers, Deventer, The Netherlands.

Guarino, N. (1998), Formal Ontology and Information Systems. Proceedings of FOIS '98. Trento, Italy, June 1998, 3–15.

Harvey, F., et al. (1999), Semantic Interoperability: a Central Issue for Sharing Geographic Information. Annals of Regional Science. 33, 213–232.

Kavouras, M. and Kokla, M. (2002), A method for the formalization and integration of geographical categorizations. International Journal of Geographical Information Science. Vol. 16, No. 5, 439–453.

Keller, S. F. (1999), Modeling and Sharing Geographic Data with INTERLIS. Computers & Geosciences. 25, 49–59.

Kim, T. J. (1999), Metadata for geo-spatial data sharing: A comparative analysis. Annals of Regionl Science. 33 171–181.

Kokla, M. and Kavouras, M. (2001), Fusion of top-level and geographical domain ontologies based on context formation and complementarity. International Journal of Geographical Information Science. Vol. 15, No. 7, 679–687.

Laurini, R. (1998), Spatial multi-database topological continuity and indexing: a step towards seamless GIS data interoperability. International Journal of Geographical Information Science. Vol. 12, No. 4, 373–402.

Nedovic-Budic, Z. and Pinto, J. K. (1999), Understanding Interorganizational GIS Activities:a Conceptual Framework. Journal of Urban and Regional Information System Association. Vol. 11, No. 1, 53–64.

Nedovic-Budic, Z. and Pinto, J. K. (2000), Information Sharing in an Interorganizational GIS Environment. Environment and Planning B: Planning and Design, Vol. 27, No. 3, 455–474.

Nedovic-Budic, Z. and Pinto, J. K. (2001), Organizational (soft) GIS Interoperability: Lessons from the U.S.. International Journal of Applied Earth Observation and Geoinformation, Vol. 3, No. 3, 290–298.

Noy, N. F. and Musen, M. A. (2003), The PROMPT suite: interactive tools for ontology merging and mapping. International Journal of Human-Computer Studies. Vol. 59, No. 6, 983–1024.

Onsrud, H. J., and Rushton G., eds. (1995), Sharing Geographic Information. the Center for Urban Policy Research, New Brunswick, NJ.

Parent, C. and Spaccapietra, S. (1998), Issues and Approaches of Database Integration. Communications of the ACM. Vol. 41, No. 5es, 166–178.

Pitoura, E., Bukhres, O., and Elmagarmid, A. (1995), Object orrentation in multidatabase systems. ACM Computing Survey. Vol. 27, No. 2, 141–195.

Pundt, H. (2002), Field data collection with mobile GIS: dependencies between semantics and data quality. GeoInformatica. Vol. 6, No. 4, 363–380.

Pundt, H. and Bishr, Y. (2002), Domain Ontologies for Data Sharing – an Example from Environmental Monitoring using Field GIS. Computers & Geosciences. 28, 95–102.

Rodriguez, M. A. and Egenhofer, M. J. (2003), Determining semantic similarity among entity classes from different ontologies. IEEE Transactions on Knowledge and Data Engineering. Vol. 15, No. 2, 442–456.

Sheth, A. P. (1999), Changing Focus on Interoperability in Information Systems: from System, Syntax, Structure to Semantics. In: Goodchild, M., et al. (eds): Interoperating geographic information systems. Kluwer Academic Publishers, MA, 5–29.

Sheth, A. P. and Larson, J. A. (1990), Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. ACM Computing Survey. Vol. 22, No. 3, 183–236.

Jang, S.-G. and Kim, T. J. (2006), Modeling an Interoperable Multimodal Travel Guide System using the ISO 19100 Series of International Standards. Proceedings of the 14th ACM International Symposium on Advances in Geographic Information Systems, November 10-11, 2006, Arlington, Va., Silvia Nittel and Rolf A. de By (eds.), ACM Press, 115-122.

Visser, U., et al. (2002), Ontologies for geographic information processing. Computers & Geosciences, 28, 103–117.

Wache, H., et al. (2001), Ontology-Based Integration of Information – A Survey of Existing Approaches. In: Stuckenschmidt, H. (ed): IJCAI-01 Workshop: Ontologies and Information Sharing, 108–117.

Warnecke L. et al. (1998), Geographic Information Technology in Cities and Counties: A Nationwide Assessment. American Forests, Washington DC.