Available online at www.sciencedirect.com

SCIENCE DIRECT®

Computers, Environment and Urban Systems
29 (2005) 669–698

Computers, Environment and Urban Systems

www.elsevier.com/locate/compenvurbsys

A geosemantic proximity-based prototype for the interoperability of geospatial data

Jean Brodeur [a,*,1], Yvan Bédard [b,1], Bernard Moulin [c,1]

[a] *Centre for Topographic Information, Natural Resources Canada, 2144 King Street West, Sherbrooke (QC), Canada JIJ 2E8*

[b] *Department of Geomatics Sciences, Pavillon Casault, Université Laval, Qué., Canada G1K 7P4*

[c] *Department of Computer Science, Pavillon Pouliot, Université Laval, Qué., Canada G1K 7P4*

Received 19 June 2003; accepted in revised form 21 April 2004

Abstract

The research agenda related to the interoperability of geospatial data is influenced by the increased accessibility of geospatial databases on the Internet, as well as their sharing and their integration. Although it is now possible to get and use geospatial data independently of their syntax and structure, it is still difficult for users to find the exact data they need as long as they do not know the precise vocabulary used by the organizations supporting geospatial databases. It is now a necessity to take into consideration the semantics of geospatial data to enable its full interoperability.

To this end, we designed a new conceptual framework for geospatial data interoperability and introduced the notion of *geosemantic proximity* based on human communication and cognition paradigms. This paper reviews this framework and the notion of *geosemantic proximity*. It also presents the *GsP Prototype*, which demonstrates the relevance of our framework and of the notion of *geosemantic proximity* for geospatial data interoperability. More specifically, we describe the architecture of the *GsP Prototype*, its implementation, and tests that have been conducted.

* Corresponding author. Fax: +1 819 564 5698.
E-mail address: brodeur@rncan.gc.ca (J. Brodeur).

[1] Also at Centre for Research in Geomatics (CRG).

0198-9715/$ - see front matter
doi:10.1016/j.compenvurbsys.2004.04.001

1. Introduction

Many geospatial databases have been set up during the last twenty years by different organizations to establish information bases corresponding to their specific needs. In this respect, the National Topographic Data Base (NTDB) (Natural Resources Canada, 1996) was elaborated for national mapping and GIS application purposes in Canada. Also, the VMap libraries (VMap, 1995) that also include topographic features of Canada were developed for military purposes. Moreover, Statistics Canada established the Street Network Files and the Digital Cartographic Files for socio-demographic and enumeration purposes (Statistics Canada, 1997). Additional topographic data sources produced at larger scales by provincial departments (e.g. OBM, 1996; Québec, 2000) are also other Canadian geospatial database examples. Each of these examples describes topographic features in different manners. To illustrate this, we have observed that forest-like phenomena are abstracted as *vegetation* ▣ in NTDB, *trees* ▣ in VMap, *woodedarea* ⊡▣ in Ontario Digital Topographic Data Base, and *milieu boisé* ▣ in the *Base de données topographiques du Québec* (where the pictograms point out the type of geometry used to map the feature geometry: point/⊡, line/⧄, or polygon/▣; see (Bédard & Proulx, 2002) for the description of spatial pictograms).

Since these organizations found that their respective databases are of general interest, they made them available to the public. Today, the Internet, the Web, and geospatial data infrastructures such as the Canadian Geospatial Data Infrastructure (CGDI) (GeoConnections, 2002) and the National Spatial Data Infrastructure (NSDI) (FGDC, 2002) facilitate the access to these geospatial databases.

Because users have access to several topographic databases, they expect to find, get and integrate the exact data they need from various databases according to their own perception and abstraction of the topographic reality. Hence, such a situation raises problems of syntactic, structural, semantic, geometric, and temporal heterogeneities between geospatial databases (Bishr, 1997; Charron, 1995; Laurini, 1998; Ouksel & Sheth, 1999; Sheth, 1999).

The idea of interoperability of geospatial databases has been promoted in the nineties to overcome the above mentioned heterogeneity problems and to allow the sharing and the integration of geospatial data and geospatial resources (Kottman, 1999). The current basis of geospatial data interoperability has been worked out by organizations such as the Open GIS Consortium Inc. (OGC), ISO/TC 211, governmental organizations, the geographic information industry and the geographic information academic community. They have made considerable progress particularly with regards to syntactic and structural heterogeneities (Egenhofer, 1999; Ouksel & Sheth, 1999; Rodriguez, 2000). Documents such as (ISO/TC 211, 2003a, 2003b; Open GIS Consortium Inc., 1999, 2001) define the content and the structure of geometric data as well as the syntactical description of geospatial data. But, to enable complete interoperability of geospatial data, it is essential to go beyond structural and syntactic heterogeneities and to address semantic heterogeneities as well as geometric and temporal heterogeneities (Egenhofer, 1999; Ouksel & Sheth, 1999).

Recently, we proposed a conceptual framework for geospatial data interoperability based on an analogy with human communication and have also introduced the notion of *geosemantic proximity* (*GsP*) (Brodeur, 2004; Brodeur & Bédard, 2001; Brodeur, Bédard, Edwards, & Moulin, 2003) as a solution to problems of semantic, spatial, and temporal heterogeneities of geospatial data. We also developed an experimental prototype, called *GsP Prototype*, to validate both our conceptual framework for geospatial data interoperability and the notion of *geosemantic proximity*. This paper specifically aims at presenting this prototype and the experiments we have conducted.

The remaining sections of this paper are structured as follows. The next section reviews geospatial data interoperability in the context of the communication process, the notion of *geosemantic proximity*, and the notion of a geospatial repository, which serve as the agent's application *ontologies* (Gruber, 1993; Guarino & Welty, 2000) in the prototype. In Section 3, we present the *GsP Prototype*, its architecture, its operation, and tests. We conclude and present future work in Section 4.

2. Geospatial data interoperability

During the last decade, interoperability has been developed within the OGC context as heterogeneous software components and Web services that adhere to common interface definitions. These components are assembled and operate together like a single system, transparently to users, even if they are located in a distributed environment (McKee & Buehler, 1998; Sondheim, Gardels, & Buehler, 1999). On the other hand, the Institute of Electrical and Electronics Engineers (IEEE) defined interoperability as the ability of two or more components to exchange information and to use the information that has been exchanged (Institute of Electrical & Electronics Engineers, 1990). This second definition emphasizes the ability of interpreting data across systems (Renner, 2001). Interoperability, and more specifically semantic interoperability, has to go beyond the delivery of data between systems that are interconnected to each other using common interfaces by providing reasoning capabilities to enable semantic interactions between systems (Kuhn, 2000). Accordingly, interoperability of geospatial data is presented hereafter in the spirit of the IEEE definition, in which geospatial concepts have reasoning capabilities that increase the ability of information systems and services to work in co-operation without human intervention (Kuhn, 2003). Our approach uses ontologies to enable agents to reason about information required by another agent and, as such, follows the idea of *Ontology-Driven Geographic Information Systems* (*ODGIS*) (Fonseca & Egenhofer, 1999; Fonseca, Egenhofer, Davis, & Borges, 2000).

2.1. A communication account of interoperability

Because people usually understand each other when communicating, we suggested that interoperability of geospatial data conforms to a human communication process (Brodeur et al., 2003). Harvey (2002), Xhu and Lee (2002), and Uitermark, van

Oosterom, Mars, and Molenaar (1999) also support this idea. As such, we developed a conceptual framework for geospatial data interoperability as a human-like communication process. In this section, we review our conceptual framework for geospatial data interoperability, which is the foundation of the architecture of the prototype presented in the next section.

According to Schramm (1971), a human communication process involves a source, a message, and a destination. When the source transmits information to the destination, he/she encodes a message, that is, to identify the information to be communicated and to transform it into physical signals. At this point, the message is still tied to the source's meaning. Afterwards, the source releases the message in the communication channel towards destination. Then, the message is released from the source's meaning. The message plays a mediating role between the source and the destination. When the message arrives at its destination, the destination begins to decode it. It recognizes the signals that compose it and assigns them a specific meaning. The communication process is working perfectly when the source's meaning and the destination's meaning of the message are the same. However, a possible source of noise can interfere with message's signals in the communication process and affect the transmission of the message. On the other hand, the communication process includes a feedback mechanism, which acts as a function to check how well the communication is performed. For instance, feedback may inform the source whether the destination has understood the message properly. As we can see, multiple representations of reality take place in the human communication process, namely the source's and destination's cognitive models, and the physical signals used for the message transmission. In the communication process, by definition the source and the destination succeed in exchanging information when they interoperate with each other.

The source's and destination's cognitive models result from the direct and the indirect observation (e.g. through sensors such as Earth observation satellite or aerial digital camera) of real-world phenomena and intentionally-produced signals received from other people. Human sensory systems capture signals and form so-called *perceptual states* (Barsalou, 1999). From *perceptual states*, the human selective attention collects the properties of interest and records them permanently as *perceptual symbols*, also known as *concepts* (Barsalou, 1999). As a cognitive element, a *concept* can never be accessed directly by another individual. It must be translated into physical signals, here called *conceptual representations*, in order to be communicated. A *concept* consists therefore of hidden-like data elements and a translation function that encapsulates these data elements. This translation function operates in two directions: (1) to generate *conceptual representations* when one wants to send a message and (2) to recognize *conceptual representations* when one wants to understand a received message.

Based on the human communication process, we developed a conceptual framework for geospatial data interoperability (Fig. 1) (Brodeur et al., 2003). Let us use the following situation to explain our framework. An individual (shown as a user agent in Fig. 1, A_u) wants information about the hydrologic network for flood analysis within a predefined area of the city of Sherbrooke. He/she encodes a query

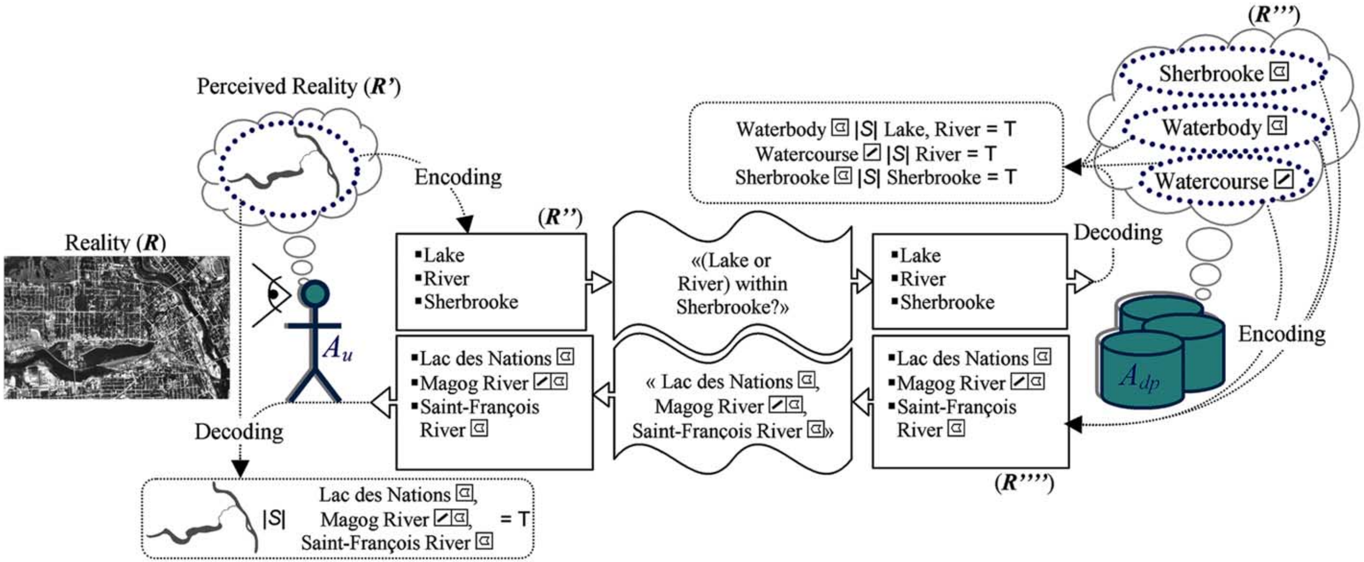

Fig. 1. A framework for spatial data interoperability (Brodeur et al., 2003).

to have information about lakes and rivers in the specified area—i.e. the conceptual representations—and sends the query to a geospatial database (shown as a data provider agent in Fig. 1, A_{dp}). When the database gets the request, it decodes it, that is, to find and assign concepts of the database that recognize the conceptual representations received, for instance *watercourses* and *waterbodies* in the neighbourhood of *Sherbrooke*. According to its interpretation, the database then gathers data, encodes and sends them—i.e. *Lac des Nations*, *Magog River*, and *Saint-François River*—to the individual, who evaluates that the received data answers his initial request. In this situation, the individual and the geospatial database use their respective vocabulary to communicate. They end up understanding each other because of their common set of conceptual representations and backgrounds, as well as reasoning capabilities, which enable them to recognize and generate messages as described in Section 3.5.

Our framework, illustrated in more detail in Fig. 1, encompasses five different expressions of the same topographic reality ($\boldsymbol{R}$, $\boldsymbol{R}'$, $\boldsymbol{R}''$, $\boldsymbol{R}'''$, and $\boldsymbol{R}''''$). These expressions, that we called the *five ontological phases of geospatial data interoperability* (Brodeur et al., 2003), are related because of the communication process. It is our interpretation that the *five ontological phases of geospatial data interoperability* are complementary to the (Frank, 2001)'s *five tiers of ontology* as they deal with the different abstraction levels of an agent in building its cognitive model. First, we have the topographic reality ($\boldsymbol{R}$) at a given time about which A_u wants information. This topographic reality is beyond description. Second, we have the A_u's set of properties ($\boldsymbol{R}'$) organized into concepts that represents $\boldsymbol{R}$. $\boldsymbol{R}'$ refers to the A_u's cognitive model. Third, we have the set of conceptual representations ($\boldsymbol{R}''$) that A_u generates to communicate data about $\boldsymbol{R}'$. These conceptual representations consist of relevant properties that describe $\boldsymbol{R}'$ concepts in a given context. They consist of the data employed for interoperability with A_{dp}. In Fig. 1, $\boldsymbol{R}''$ is illustrated by "*Lakes or Rivers within Sherbrooke.*" Fourth, we have the set of concepts ($\boldsymbol{R}'''$), which A_{dp} maintains. When A_{dp} receives $\boldsymbol{R}''$ conceptual representations, it uses $\boldsymbol{R}'''$ concepts to recognize and assign a meaning to $\boldsymbol{R}''$ conceptual representations and, afterwards, to collect the information that complies with A_u's initial request. In Fig. 1, these concepts correspond to *watercourses*, *waterbodies*, and *Sherbrooke*. In turn, A_{dp} uses again $\boldsymbol{R}'''$ concepts along with the corresponding information to encode conceptual representations ($\boldsymbol{R}''''$) to answer A_u. These conceptual representations consist of *Lac des Nations*, *Magog River*, and *Saint-François River* in Fig. 1. Finally, when A_u receives $\boldsymbol{R}''''$ conceptual representations, he/she decodes them, that is again to recognize and assign them a meaning, and validates them against $\boldsymbol{R}'$ concepts. If $\boldsymbol{R}''''$ conceptual representations correspond to the requested $\boldsymbol{R}'$concepts, then we can say that interoperability happens between A_u and A_{dp}. Accordingly, interoperability is a bidirectional communication process that includes a feedback mechanism in both directions, to control the proper reception of messages and ensure that they were understood properly. This description of interoperability is somewhat different from the usual way to query geospatial databases (e.g. SQL or WFS queries), as people have to know in advance the exact description and representation of concepts they are interested in before sending a query to a geospatial database otherwise they will not get any data. From a GIS and a database perspective, we believe that this communication-based conceptual framework, which takes into consideration the semantic issue, allows further progress towards the resolution of the geospatial data access issue.

As mentioned previously, $\boldsymbol{R}$, $\boldsymbol{R}'$, $\boldsymbol{R}''$, $\boldsymbol{R}'''$, and $\boldsymbol{R}''''$ consist in different facets of the reality which are concerned about ontology, even if they have similarities. In philosophy, ontology is a subject matter dealing with:

- the description of the world (Peuquet, Smith, & Brogaard, 1998);
- a model and an abstract theory of the world (Smith & Mark, 1999);
- the science of being (Bittner & Edwards, 2001; Peuquet et al., 1998);
- the science of the type of entities, of the objects, of the properties, of the categories, and of relationships, which constitute the world (Lehmann, 1992; Peuquet et al., 1998; Smith & Mark, 1999).

Ontology is also a subject of interest in artificial intelligence and database. It has been defined by (Gruber, 1993) as "an explicit specification of a conceptualization". However, (Guarino, 1998) refined Gruber's definition taking into account the philosophical meaning of ontology and defined ontology as "a logical theory accounting for the intended meaning of a formal vocabulary." Hence, in the scope of this paper, we consider an ontology as being "a formal representation of phenomena with an underlying vocabulary and axioms including definitions that make the intended meaning explicit and describe phenomena and their interrelationships" (Brodeur et al., 2003).

In the database realm, the representation of real-world phenomena is widely developed using conceptual models (e.g. E-R or UML models) and feature dictionaries. Together, these two components constitute a comprehensive set of metadata describing the content and the structure of databases, which are better known as *database repositories* (Brodeur, Bédard, & Proulx, 2000; Jones, 1991; Marco, 2000; Moriarty, 1990; Prabandham, Selfridge, & Mann, 1990). A conceptual model is a tool to capture abstract representations of real-world phenomena from a data-centred analysis perspective. It is also used to support the development of databases. It structures and stores features of interest using general categories, object classes, properties, relationships, generalizations, aggregations, roles, constraints, behaviours, and more specifically in the context of geospatial databases, geometry and temporality. The dictionary stores the intended meaning (in other words *the semantics*) of all elements that compose the conceptual model. In geographic information, *Perceptory* (Bédard & Proulx, 2002) is a tool specially developed to build, manage, and exploit geospatial data repositories. It consists of a UML-based conceptual modeling tool enhanced with the *Plug-in for Visual Language* (PVL) (Bédard, 1999; Bédard & Proulx, 2002) for spatial and temporal data modeling and an object class dictionary. As such, geospatial repositories developed with *Perceptory* can serve as application ontologies.

Practitioners of different backgrounds and professional experiences typically abstract identical phenomena and develop geospatial databases with their respective repository differently. The situation and the circumstances surrounding the perception of geospatial phenomena guide the manner with which these geospatial phenomena are abstracted. This refers more specifically to the *context*. The context is an abstract notion, which drives the definition of concepts and conceptual representations, and the choice of properties that are used for their description (Simsion, 2001). It is the context that provides the inherent semantics to concepts and conceptual representations (Kashyap & Sheth, 1996). Hence, the same part of the topographic reality is typically represented differently from one database to another because of their specific context. This causes interoperability problems when merging data from different geospatial databases. Notwithstanding this, context is a fundamental element for the assessment of the semantic, spatial, and temporal interoperability of geospatial data. Accordingly, the assessment of semantic, spatial, and temporal interoperability of geospatial data needs the addition of reasoning capabilities to geospatial concepts that take the context into consideration. Keeping this in mind, we developed the notion of *geosemantic proximity* (Brodeur, 2004) following a context-based orientation (Kashyap & Sheth, 1996).

We mentioned earlier that a concept has a translation function in order to generate and recognize conceptual representations. Hence, *geosemantic proximity* (*GsP*) consists of a basic component of this translation function, which specifically applies to geospatial concepts. *GsP* evaluates qualitatively the semantic similarity (Kashyap & Sheth, 1996; Sheth & Kashyap, 1992) of a geospatial concept (hereafter called a geoConcept) with a geospatial conceptual representation (hereafter called a geoConceptRep) by the comparison of their respective context. In *GsP*, the context (C) consists of the set of inherent properties of a geoConcept or a geoConceptRep. These properties are classified in two types: intrinsic and extrinsic. Intrinsic properties (C°) provide the literal meaning of the geoConcept or the geoConceptRep. They consist of the identification, attributes, attribute values, geometries, temporalities, and domain of a geoConcept or a geoConceptRep. Extrinsic properties (∂C) are properties that are subject to external factors. They give meaning by the action that these factors exercise on the geoConcept or the geoConceptRep. Behaviours as well as semantic, spatial, and temporal relationships are kinds of extrinsic properties. Although geoConcept or geoConceptRep may have an identical semantics, they may be depicted differently with different geometric and temporal primitives, which may own a specific semantic (geometric or temporal). For instance, the geometry of a geoConcept "road" could described in one case the position of the road way, in a second case the cadastral boundaries of the road, and in a third case the centreline of the road. The different manners a geoConcept or GeoConceptRep is depicted geometrically and temporally lead also to different spatial and temporal relationships and behaviours with other geoConcepts or GeoConceptReps. All these considerations are explicitly taken into consideration in a *geosemantic proximity* assessment. We use a segment (Fig. 2), which holds in a semantic space, to illustrate the context of a geoConcept or geoConceptRep. Intrinsic properties correspond to the interior of the segment, whereas extrinsic properties correspond to the boundary of the segment. Hence, the context (C) consists of the union of intrinsic and extrinsic properties: $C = C^\circ \cup \partial C$. Therefore, the *GsP* of a geoConcept (K) with a geoConceptRep

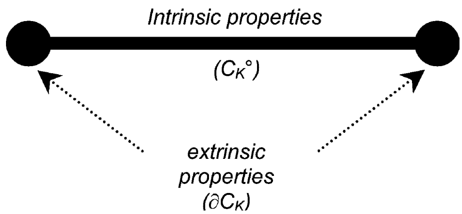

Fig. 2. Context of concepts or conceptual representations.

(L) can now be defined by the intersection of their respective context, $GsP(K,L) = C_K \cap C_L$, which becomes a four-intersection matrix when consolidated with intrinsic (C°) and extrinsic (∂C) properties.

$$GsP(K,L) = \begin{pmatrix} \partial C_K \cap \partial C_L & \partial C_K \cap C_L^{\circ} \\ C_K^{\circ} \cap \partial C_L & C_K^{\circ} \cap C_L^{\circ} \end{pmatrix}$$

Each component of the matrix can be evaluated empty (denoted by *f* or false) or not empty (denoted by *t* or true). Accordingly, we derived sixteen (2^4) predicates that are presented in the matrix row major form (i.e. row by row) with the prefix "*GsP_*": *GsP_ffff* (or *disjoint*), *GsP_ffft*, *GsP_fftt* (or *contains*), *GsP_tfft* (or *equal*), *GsP_ftft* (or *inside*), *GsP_tftt* (or *covers*), *GsP_ttft* (or *coveredBy*), *GsP_fttt* (or *overlap*), *GsP_tttt*, *GsP_tfff* (or *meet*), *GsP_tftf*, *GsP_tttf*, *GsP_ttff*, *GsP_fttf*, *GsP_fftf*, *GsP_ftff* (Brodeur, 2004). These predicates are used to qualify the *GsP* of a geoConcept with a geoConceptRep. As such, *GsP* are consistent with Allen's and Egenhofer's well known approaches of temporal and spatial topological relationships respectively (Allen, 1981, 1983; Egenhofer, 1993; Egenhofer & Franzosa, 1991). The extension of topology to geoConcept and geoConceptRep is possible because both geoConcept's and geoConceptRep's respective context agrees with Hausdorff's definition of *topological space* (Weisstein, 1999), whose description is beyond the scope of this paper and is presented in (Brodeur, 2004).

Let us use the following example to illustrate the relevance of *GsP*. According to our conceptual framework, a user agent, which is based on the *Base de données topographiques du Québec* (BDTQ) ontology, aims to update its road network information. It asks a data provider agent, which is based on the National Topographic Data Base (NTDB), for information about *street* ⧄—i.e. an encoded geoConceptRep. When the data provider agent receives the request, it looks through the geoConcepts it knows to find one that is *geosemantically* (i.e. semantically, spatially and temporally) similar to *street* ⧄. The data provider agent identifies that its geoConcept *road* ⧄ has an attribute *classification*, which can take the value *street* of similar definition to geoConceptRep *street* ⧄. Also, *road* ⧄ and *street* ⧄ have the same type of geometry. As such, they hold common intrinsic properties. As defined in BDTQ, *street* ⧄ possesses relationships with other road classes. But, these road classes are already included in the *road* ⧄ description. As such, *street* ⧄'s extrinsic properties intersect with *road* ⧄'s intrinsic properties. Accordingly, *GsP* of the *road* ⧄ geoConcept when compared to the *street* ⧄ geoConceptRep is *GsP_ttff* (or *contains*) and, as such, *road* ⧄ can be used to answer the request of the user agent about *street* ⧄.

As used in (Fowler, Perry, Nodine, & Bargmeyer, 1999; Payne, Paolucci, Singh, & Sycara, 2002; Sycara, Klusch, Widoff, & Lu, 1999), software agents appear well suited to develop user agents and data provider agents as illustrated in our conceptual framework to experience the *GsP* notion within a prototype. According to (Nwana, 1996), a software agent is defined as "a component of a software and/or hardware which is capable of acting exactingly in order to accomplish tasks on behalf of its user." In the specific context of the prototype on semantic, spatial, and temporal interoperability of geospatial data presented in the following section, user and data provider agents are deployed as software agents, which own a particular ontology to interoperate with other agents. However, the description of software agent is beyond the scope of this paper and can be obtained in (Nwana, 1996; Nwana & Wooldridge, 1996).

2.2. Related research work on geospatial data interoperability

In the last few years, a number of authors have also addressed semantic interoperability, but differently. Bishr (1997) and Benslimane (2001) both worked on mediation approaches. Uitermark et al. (1999) and Cruz, Rajendran, Sunna, and Wiegand (2002) worked on the geospatial data integration issue. Rodriguez (2000) developed a method for measuring the semantic similarity between geographic concepts.

The Bishr's (1997) approach, called *Semantic Formal Data Structure* (*SFDS*), was composed of three components: an export schema, a federated schema, and a proxy context mediator. The export schema serves to publish local database concepts to users. The federated schema collected domain specific concepts such as for transportation, soils, etc. Finally, the proxy context mediator consisted of a common ontology that mapped export schema concepts to federated schema concepts in order to enable the exchange of geospatial information from one system to another.

The Benslimane's (2001) approach, called the *Isis solution*, was developed in two layers: the data and the mediation layer. The data layer referred to heterogeneous databases with their respective local schemata whereas the mediation layer was described in terms of a universe of discourse, a global ontology, a context of reference, and database specific co-operation contexts. Co-operation contexts were made of co-operation classes, which served to map concepts from heterogeneous databases. They serve in a semantic translation process that takes place in the exchange of geospatial information.

In Uitermark et al. (1999), domain and application ontologies were used to describe concepts pertaining to a given discipline such as topographic mapping and concepts that were stored in individual geospatial databases, respectively. In this approach, a static semantic mapping was developed between concepts of application ontologies and concepts of domain ontology in Prolog clauses, which allowed the integration of geospatial data from heterogeneous databases when queried based on the domain ontology.

Similarly, Cruz et al. (2002) worked on a system to overcome semantic interoperability problems when integrating geospatial data from heterogeneous databases. In their approach, they, proposed the development of a common knowledge—i.e. the ontology, which was used to interact with local heterogeneous geospatial databases. Each local database was associated with a static *agreement* document encoded in XML that mapped its local concepts to *ontology*'s concepts. When the system received a query, the query was rewritten based on the ontology and the local *agreement* documents in a series of sub-queries that were executed on each local database. This approach resembles a loosely coupled federated database system (Sheth, 1999).

Rodriguez (2000) studied the similarity of geographic concepts. She proposed the Matching Distance (MD) model, based on the *ratio model* (Tversky, 1977), to measure a conceptual distance between two geographic concepts to quantify the similarity between two geographic concepts. This conceptual distance consisted in a weighted sum of the semantic proximity of parts, functions, and attributes of two geographic concepts, where the semantic proximity of these three facets corresponded to the ratio of common elements to the sum of their common and distinguishing elements.

The *GsP* approach differs from the above approaches as it proposes a dynamic computation of the semantic mapping between two geospatial abstractions (i.e. between a geoConcept and a geoConceptRep) instead of developing static mapping files between concepts of different ontologies. Also, as human beings seem to reason in more of a qualitative manner, the *GsP* approach expresses the semantic similarity between a geoConcept and a geoConceptRep qualitatively, using a set of mutually exclusive predicates to describe their commonness and difference as opposed to a quantitative measurement.

3. The GsP prototype

To evaluate the *GsP* notion, we built a software prototype, called *GsP Prototype*, which agrees to our interoperability conceptual framework illustrated in Fig. 1. With the *GsP Prototype*, software agents are instantiated and can interoperate with each other. This section presents successively a high level architecture of the prototype, the way the prototype operates, and the experimentations conducted so far.

3.1. Architecture

The architecture of *GsP Prototype* illustrated in Fig. 3 depicts a communication process, which takes place between two software agents (Agents A and B) interacting through a communication channel. It details more specifically one agent's internal structure and operations as well as the manner in which agents exchange information. However, this architecture is not limited to only two agents but can be expanded to multiple agents interacting in pairs.

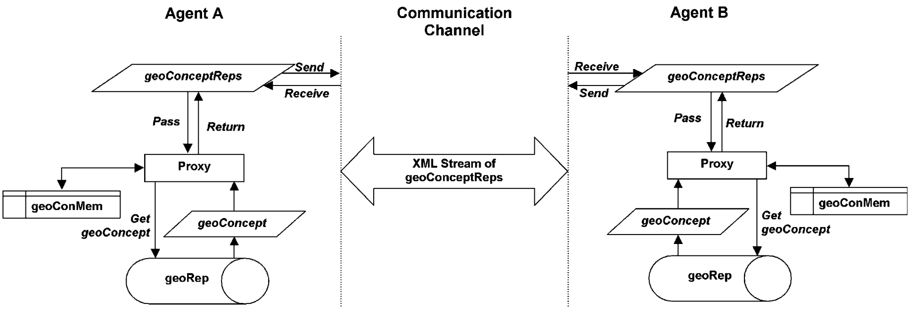

Fig. 3. Architecture of the *GsP Prototype* (from Brodeur and Bédard, 2002).

In this architecture, all agents have an identical internal structure and operate in the same manner. They communicate using messages composed of geoConceptReps encoded in XML streams. When an agent receives a message, it captures the inner XML geoConceptReps of the message and places them in a transitory internal data structure containing geoConceptReps. Each geoConceptRep stored in this data structure can be compared to a human perceptual state.

The geoConceptReps are then passed to a *Proxy*. The *Proxy* is a server responsible for finding geoConcepts that match the geoConceptReps in order to assign them a meaning. This is the recognition process. The *Proxy* has to mediate between two geoConcept storages: *geoConMem* and *geoRep*.

GeoConMem is a cache memory limited in size, which stores for a short period the most recent geoConcepts (the geoConcept structure is detailed further in this section) used by the agent. It may be compared to the short-term memory of a human being.

GeoRep consists of a geospatial data repository that holds the description of all geoConcepts that the agent knows; it is a direct access storage. In this case, the *geoRep* storage is implemented using *Perceptory* and consists of a graph representation of geoConcepts in a UML class diagram along with a dictionary which manages the description of semantic, spatial, and temporal properties of geoConcepts. *GeoRep* may be compared to the long-term memory of a human being.

When processing, the *Proxy* examines one by one all the geoConceptReps that the agent received in a message. For each geoConceptRep of the *geoConceptReps* data structure, the *Proxy* looks first in *geoConMem* to visit the geoConcepts it stores until a geoConcept that has a *GsP* of "*GsP_tfft*" (or equal) with a first geoConceptRep is located. It is the geoConcept that is responsible to evaluate its *GsP* with the geoConceptRep. As such, it compares all its intrinsic (i.e. identification, attributes, attribute values, geometries, temporalities, and domain) and extrinsic (i.e. relationships and behaviours) properties to the geoConceptRep's intrinsic and extrinsic properties as in the four-intersection matrix presented in the previous section. However, if no geoConcept shows a *GsP* of "*GsP_tfft*" (or equal) with the geoConceptRep, then the *Proxy* continues to search in *geoRep* to find the most similar geoConcept to the geoConceptRep, which consists in the geoConcept that has the highest *GsP* with the geoConceptRep. To this end, the *Proxy* visits *geoRep*'s geoConcepts to compute their respective *GsP* with the geoConceptRep. As such, it uses a graph traversal algorithm and begins with the geoConcept of the *geoConMem* cache memory that has the highest *GsP* with the geoConceptRep. *GeoRep* provides geoConcepts to the *Proxy* using a *geoConcept* data structure. When the *Proxy* gets a *geoConcept* from *geoRep*, it evaluates its *GsP* with the geoConceptRep and stores it. This process continues until a geoConcept that has a *GsP* of "*GsP_tfft*" (or equal) with the geoConceptRep is found or all concepts are visited. When the process is completed, geoConcepts having a *GsP* different from "*GsP_ffff*" (or disjoint) with the geoConceptRep are then sorted from the highest to the lowest *GsP*. The geoConcept with the highest *GsP* constitutes the most similar geoConcept to the geoConceptRep, which is then used to assign a meaning to the geoConceptRep.

It might happen that no geoConcept is found similar to the geoConceptRep and, accordingly, no meaning can be assigned to the geoConceptRep. Therefore, the agent will not be able to answer to the other agent on this geoConceptRep. The resulting set of geoConcepts matching the geoConceptReps of the message are then used by the agent to reply to the other agent. As such, the geoConcepts generate geoConceptReps that are then encoded in an XML stream and sent through the communication channel to the other agent.

Similarly to concepts that compose human cognitive models, geoConcepts obtained from either *geoConMem* or *geoRep* consist here of non-visible data elements (or *private* as in Java or C++), which are obviously inaccessible to other agents. These data elements are encapsulated by three functions: recognize, generate, and gspRelate (Fig. 4). The recognition and generate functions serve as the main geoConcept interfaces, which are supported by gspRelate that evaluates the *GsP*.

Fig. 5 draws a more detailed description of the geoConcept object structure in a UML class diagram. In this diagram, geoConcepts conform to the class *GEOCONCEPT*, which inherits its data structure from the class *GEOABSTRACTION*. *GEOABSTRACTION* aims at defining the properties used to identify and describe a geospatial phenomenon. These properties are divided into two types: intrinsic and extrinsic.

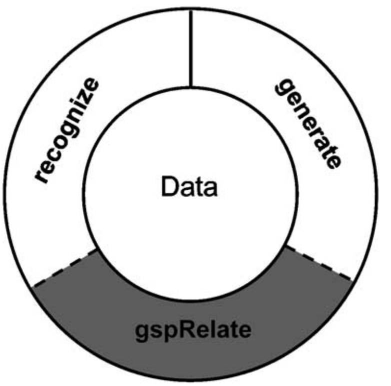

Fig. 4. Object structure of a concept (from Brodeur & Bédard, 2002).

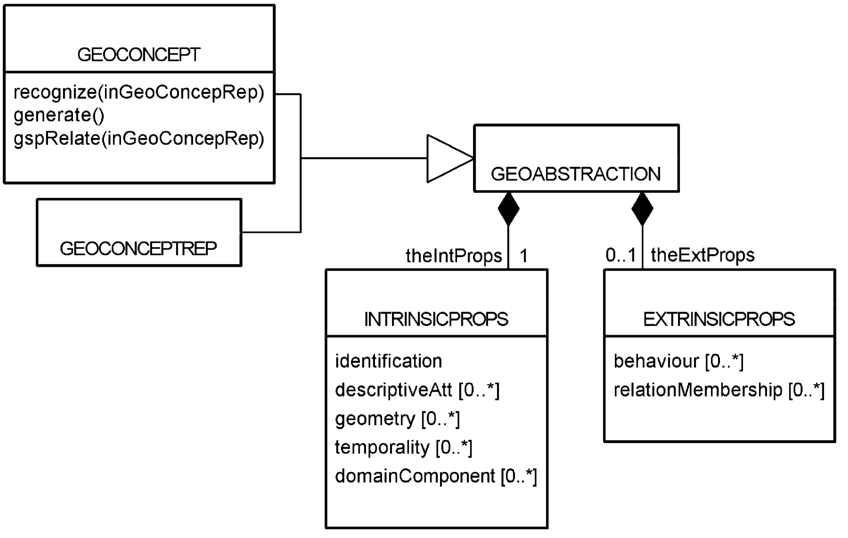

Fig. 5. UML class diagram of GEOABSTRACTION, GEOCONCEPT, and GEOCONCEPTREP.

On the one hand, the class *INTRINSICPROPS* accounts for intrinsic properties and captures the *identification*, *descriptiveAtts* (i.e. descriptive attributes), *geometries*, *temporalities*, and the *domainComponents* (i.e. various component of the domain) of a *GEOABSTRACTION*. Essentially, the *identification* refers to the name and the definition given to the *GEOABSTRACTION*. The *descriptiveAtts* report on the inherent characteristics of a phenomenon. A name, a definition, and a domain of values distinguish each descriptive attribute from another. *Geometries* refer to the various types of geometry such as simple geometry (e.g. point, line, or surface), geometric aggregate, complex geometry, and alternate geometry (Bédard, 1999) that are used to depict the phenomenon spatially along with its inherent semantics (e.g. building basement footprint). Similarly to *geometries*, *temporalities* refer to the various types of temporality such as instant, period, temporal aggregate, and alternate temporality (Bédard, 1999) that are used to depict the phenomenon temporally along with its inherent semantics (e.g. date when the building construction is completed). The domain consists of the numerous combinations of attribute values, geometry, and temporality that the *GEOABSTRACTION* can take where each combination refers to one *domainComponent*.

On the other hand, the class *EXTRINSICPROPS* provides the details of the extrinsic properties. Extrinsic properties are described in terms of *behaviours* and memberships in relationships (*relationMembership*). A *behaviour* refers to an operation that a phenomenon can accomplish. A name, a definition, a list of parameters, and a return type differentiate each behaviour of a phenomenon. A membership in a relationship expresses the participation of the phenomenon in a relationship with another phenomenon. It identifies the relationship (name of the relationship and the list of members), the role played by the *GEOABSTRACTION* and its minimum and maximum cardinalities.

The class *GEOCONCEPT* has the three functions mentioned above (*recognize*, *generate*, and *gspRelate*). The function *recognize* takes a geoConceptRep as an input. It identifies geoConcepts that are similar to one geoConceptRep prioritized by their *GsP*. The function *gspRelate* assists the function *recognize* by computing the *GsP* of the geoConcept with the geoConceptRep. It evaluates to what extent the geoConcept matches the geoConceptRep. The function *gspRelate* assists the function *recognize* by computing the *GsP* of the geoConcept with the geoConceptRep. Finally, the *generate* function produces a geoConceptRep of this geoConcept, which holds in a specific context. Again, the *gspRelate* function assists the *generate* function to ensure that the generated geoConceptRep is similar to the geoConcept.

Being a subtype of the class *GEOABSTRACTION*, GEOCONCEPTREP inherits also the data structure of *GEOABSTRACTION* (Fig. 5). Accordingly, GEOCONCEPTREP's data structure and GEOCONCEPT's data structure are identical. Because a geoConceptRep is essentially encoded data of a geoConcept, the class GEOCONCEPTREP does not possess any function. When an agent releases geoConceptReps in the communication channel, it transforms them in an XML stream and sends this XML stream to its destination. Accordingly, the XML encoding of geoConceptReps adheres to a predefined definition described either in a Document Type Definition (DTD) or an XML Schema. For the purpose of the prototype, the XML encoding of geoConceptReps satisfies the following DTD:

```
<?xml version = “1.0” encoding = “UTF-8”?>
<!– edited with XML Spy v4.0.1 U (http://www.xmlspy.com) by Jean Brodeur
   (Natural Resources Canada)–>
<!ELEMENT GsPmessage (conceptualRepresentation*)>
<!ATTLIST GsPmessage
   type CDATA #REQUIRED
   recognition (true | false) “true”>
<!ELEMENT conceptualRepresentation (intrinsicProperties,
   extrinsicProperties?)>
<!ELEMENT intrinsicProperties (identification, descriptiveAttribute*,
   geometry*, temporality*, domainElement*)>
<!ELEMENT identification (name, definition?)>
<!ELEMENT descriptiveAttribute (name, definition?, attributeValue*)>
<!ELEMENT attributeValue (name, definition?)>
<!ELEMENT geometry (#PCDATA)>
<!ELEMENT temporality (#PCDATA)>
<!ELEMENT domainElement (attValue+, geometry?, temporality?)>
<!ELEMENT attValue (descriptiveAttribute, attributeValue)>
<!ELEMENT extrinsicProperties (relationMembership*, behaviour*)>
<!ELEMENT relationMembership (relation, role?, cardMin?, cardMax?)>
<!ELEMENT relation (name, firstMember, secondMember?)>
<!ELEMENT behaviour (name, definition, parameter+, returnType)>
```

```
<!ELEMENT parameter (conceptualRepresentationName, defaultValue?)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT definition (#PCDATA)>
<!ELEMENT role (#PCDATA)>
<!ELEMENT cardMin (#PCDATA)>
<!ELEMENT cardMax (#PCDATA)>
<!ELEMENT firstMember (#PCDATA)>
<!ELEMENT secondMember (#PCDATA)>
<!ELEMENT conceptualRepresentationName (#PCDATA)>
<!ELEMENT defaultValue (#PCDATA)>
<!ELEMENT returnType (#PCDATA)>
```

3.2. Implementation

Based on the above architecture, *GsP Prototype* was implemented with Java™ and XML technologies in combination with *Perceptory*-based geospatial repositories. The reasons supporting this choice of technologies are:

1. XML is by far a widely recognized technology for the communication of information;
2. Availability of Java libraries to process XML documents (namely the Java API for XML Processing (JAXP)) (Sun Microsystems Inc., 2002) that includes the Xalan (The Appache Software Foundation, 2002) and the Xerces (The Appache Software Foundation, 2002) libraries for parsing and manipulating XML documents;
3. Portability of the development on the Web; and
4. *Perceptory* is a technology very well suited to develop geospatial repositories agreeing to ISO19103 *Geographic information—conceptual schema language* (ISO/TC 211, 2001b) and ISO19110 *Geographic information—methodology for feature cataloguing* (ISO/TC 211, 2001a), which can then serve as agent's ontologies.

This section presents in detail the implementation of the prototype and the way it operates.

The *GsP Prototype* uses interfaces of two kinds: software agent interfaces and an Agent Manager interface. A software agent appears as a window (Fig. 6). The window's title bar identifies the agent's name along with its ontology source name (e.g. agent1 (NTDB_RN)). The remaining part of the window is divided into two sections: the Console and the Communication Monitor.

The Console section consists of three components. The first component is a drop-down menu, which presents the list of geoConcepts that compose the agent's ontology. Each geoConcept is identified by a unique name. The next item is the **Send Query** button. When clicked, this button initiates a query about the geoConcept selected from the drop-down menu towards an external agent. The external agent is

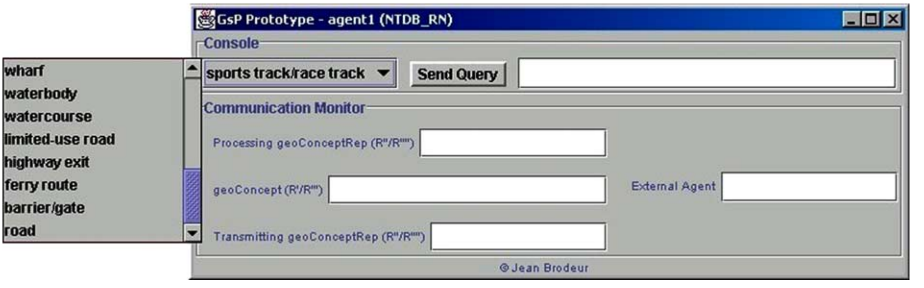

Fig. 6. The agent window.

identified by filling its name in the External Agent field of the Communication Monitor section. The last Console's component is a field in which the agent displays messages.

The Communication Monitor section shows the different steps of the communication process that are accomplished. When an agent receives a message from an external agent, the name of the external agent appears in the **External Agent** field. Following this, the agent extracts the geoConceptReps from the message and displays the name of the geoConceptRep being processed in the **Processing geoConceptRep** (***R''*/*R''''***) field one by one. Then, the agent initiates the recognition process of the geoConceptRep and, as such, visits the geoConcepts of the ontology until one is found similar to the geoConceptRep. Once a geoConcept is identified, its name is displayed in the **geoConcept** (***R'*/*R'''***) field. When a reply is expected by the external agent (e.g. answer to a query), the corresponding geoConcept generates a geoConceptRep of itself and the name of the transmitted geoConceptRep is displayed in the **Transmitting geoConceptRep** (***R''*/*R''''***) field.

The Agent Manager interface (Fig. 7) is used to instantiate software agents and displays one agent's state upon user request. The instantiation of an agent requires two elements: its identification and the name of an ontology source. The agent's identification is a unique identifier. The ontology source name consists of the source

Fig. 7. The Agent Manager interface.

name of a geospatial data repository. In our case, it corresponds to an ODBC data source name, which refers to the database containing the geospatial data repository. Once the name field and the ontology field are filled in, the agent is instantiated by clicking the **New** button. At this time, the agent is alive but not active. It becomes active by clicking the **Start** button. The agent's state can be set inactive (or sleeping) but still alive by clicking the **Stop** button. This is needed for management purposes. Even if the agent is inactive, it keeps all its properties and when it is re-started (by clicking the **Start** button again), it becomes active again. Finally, an agent is completely eliminated by clicking the **Kill** button. At any time, it is possible to look at an agent's state simply by filling in the agent's name in the name field and by pressing "return". The agent's state can be one of the following:

- **Null:** the agent does not exist;
- **Operating:** the agent is alive and active;
- **Sleeping:** the agent is alive but not active.

Fig. 8 illustrates the way the prototype operates. In this figure, *agent2* (*BDTQ_RN*) sends a query to *agent1* (*NTDB_RN*) for information about street. As such, it uses its geoConcept *street* to generate a geoConceptRep of the same name (i.e. *street*), encodes and sends it to *agent1* as in the following XML document.


```
<?xml version="1.0" encoding="UTF-8"?>
<GsPmessage type="query">
  <conceptualRepresentation>
    <intrinsicProperties>
      <identification>
        <name>street</name>
        <definition>rue : voie de communication généralement bordée
          de bâtiments dans une agglomération.</definition>
      </identification>
      <geometry>1</geometry>
    </intrinsicProperties>
    <extrinsicProperties>
      <relationMembership>
        <relation>
          <name>Inheritance</name>
          <firstMember>street</firstMember>
          <secondMember>communication route</secondMember>
        </relation>
        <role>subtype</role>
      </relationMembership>
    </extrinsicProperties>
  </conceptualRepresentation>
</GsPmessage>
```

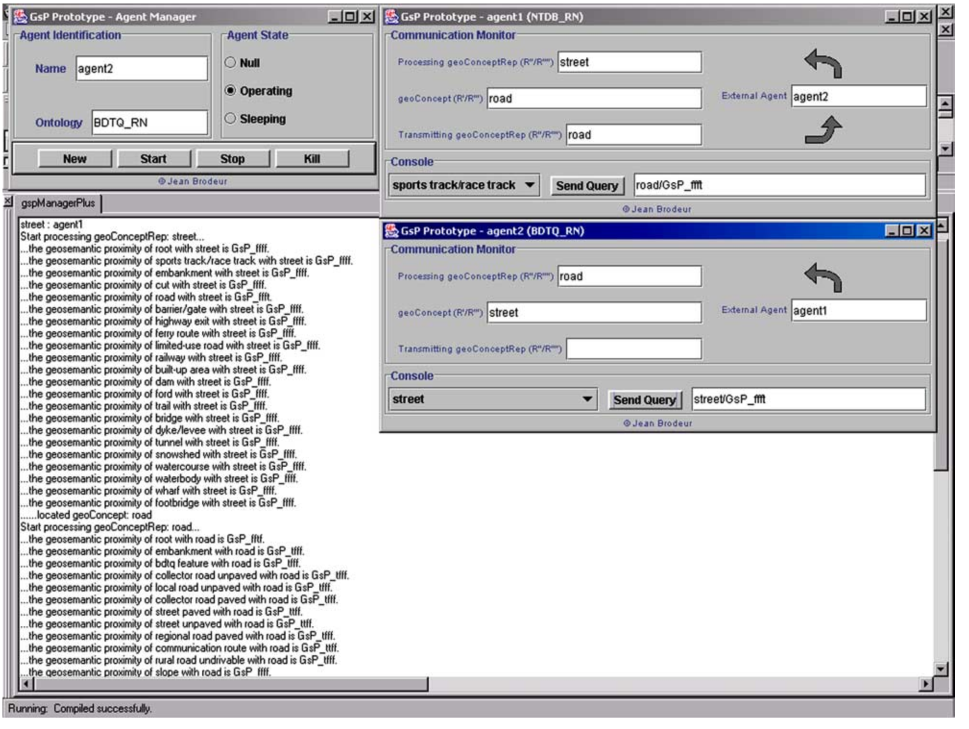

Fig. 8. Example of the prototype operation.

When *agent1* receives the XML document, it identifies its source and displays the name in the **External Agent** field—i.e. *agent2*. Following this, it extracts the message type—i.e. query—and the included geoConceptReps—i.e. *street*. Then, it processes the geoConceptReps one by one. In this example, there is only the geoConceptRep *street* to process. As such, *Agent1* displays the name *street* in the **Processing geoConceptRep** ($\boldsymbol{R''}/\boldsymbol{R''''}$) field.

To process the geoConceptRep *street*, *agent1* looks first for geoConcepts in its short-term memory. If no geoConcept has a *GsP* of *GsP_tfft* (or equal) to *street*, then it goes on searching its long-term memory until a geoConcept showing a *GsP* of *GsP_tfft* (or equal) with *street* is found or until all geoConcepts have been visited. As we can see in Fig. 8, *agent1* visited all geoConcepts. The computation of the *GsP* of a geoConcept with *street* takes into consideration their identification, their descriptive, geometric, and temporal properties (i.e. the intrinsic properties) as well as their behaviours and their memberships to relationships (i.e. the extrinsic properties), respectively. As the geoConcept *road* shows common intrinsic properties with *street* and has the most significant *GsP*—i.e. *GsP_ffft*—*agent1* displays its name in the **geoConcept** ($\boldsymbol{R'}/\boldsymbol{R''''}$) field and as such uses it to assign a meaning to the geoConceptRep *street*. Now with the geoConcept *road*, *agent2* can answer *agent1*'s request. It produces a geoConceptRep of the same name, displays the name in the **Transmit-**

ting geoConceptRep (***R″*/*R″″***) field (e.g. *road*), encodes the geoConceptRep, and sends it to *agent1* using an XML document similar to the one presented previously where the attribute type of the gspMessage element is set to "answer".

In turn, when *agent2* receives the XML document, it initiates a similar process as *agent1* did. It identifies the message originator and displays its name in the **External Agent** field—i.e. *agent1*—extracts the message type—i.e. answer—and the geoConceptRep—i.e. *road* —and processes it. Then, *agent2* displays the name *road* in the field **Processing geoConceptRep** (***R″*/*R″″***) and computes that its geoConcept *street* is similar (e.g. *GsP_ffft*) to the geoConceptRep *road*. As such, *agent2* displays the name *street* in the **geoConcept** (***R′*/*R‴***) field. Therefore, *agent2* acknowledges that *road* answers its initial query and thus interoperability happens. Because the message is an answer, no further action is required and the process stops at this point.

3.3. Experimentation

Using the above software agent-based prototype, we conducted experimentations on road and hydrographic networks to assess the strength of our approach. These two themes have been chosen because they are both candidates of the essential content and the desirable content, respectively, of the Canadian GeoBase (2001), which is currently being developed. Briefly, GeoBase consists of "the fundamental geographic information that describes Canadian landmass above and below water" (CCOG, 2001) that is established in co-operation with Canadian federal, provincial, and territorial mapping agencies. The experimentation aimed at assessing computer feasibility and strength of the *GsP* approach a priori. Tests were limited to the interaction of software agents using on the one hand identical ontologies and on the other hand different ontologies.

As such, we built UML-based geospatial data repositories on road and hydrographic networks with *Perceptory* using the following topographic data product specifications:

(a) National Topographic Data Base—standards and specifications of Canada (Natural Resources Canada, 1996) (NTDB);
(b) *User's Guide to Digital and Hardcopy property and Basemap Products* of Prince Edward Island (P.E.I. Geomatics Information Centre, n.d.) (PEIBP);
(c) Quebec Topographic Data Base 1:20 000—production standards (Québec, 2000) (QTDB);
(d) Ontario Digital Topographic Database—1:10 000, 1:20 000—a guide for user (OBM, 1996) (ODTDB);
(e) *Digital Baseline Mapping at 1:20 000* of the province of British Columbia (BC Ministry of Environment Lands and Parks (Geographic Data BC, 1992) (BCDBM).

For example, Figs. 9 and 10 show UML class diagrams corresponding to both themes of the NTDB data product specifications.

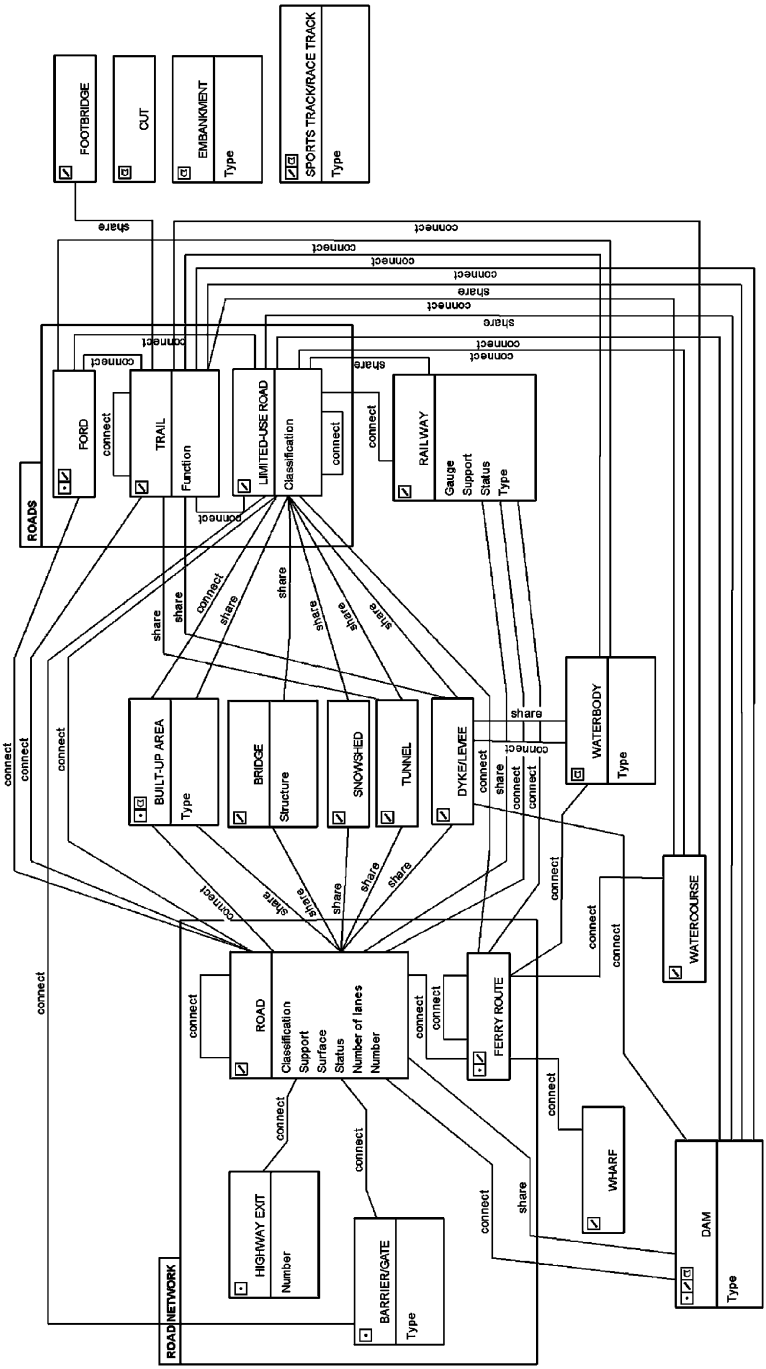

Fig. 9. NTDB road network UML class diagram.

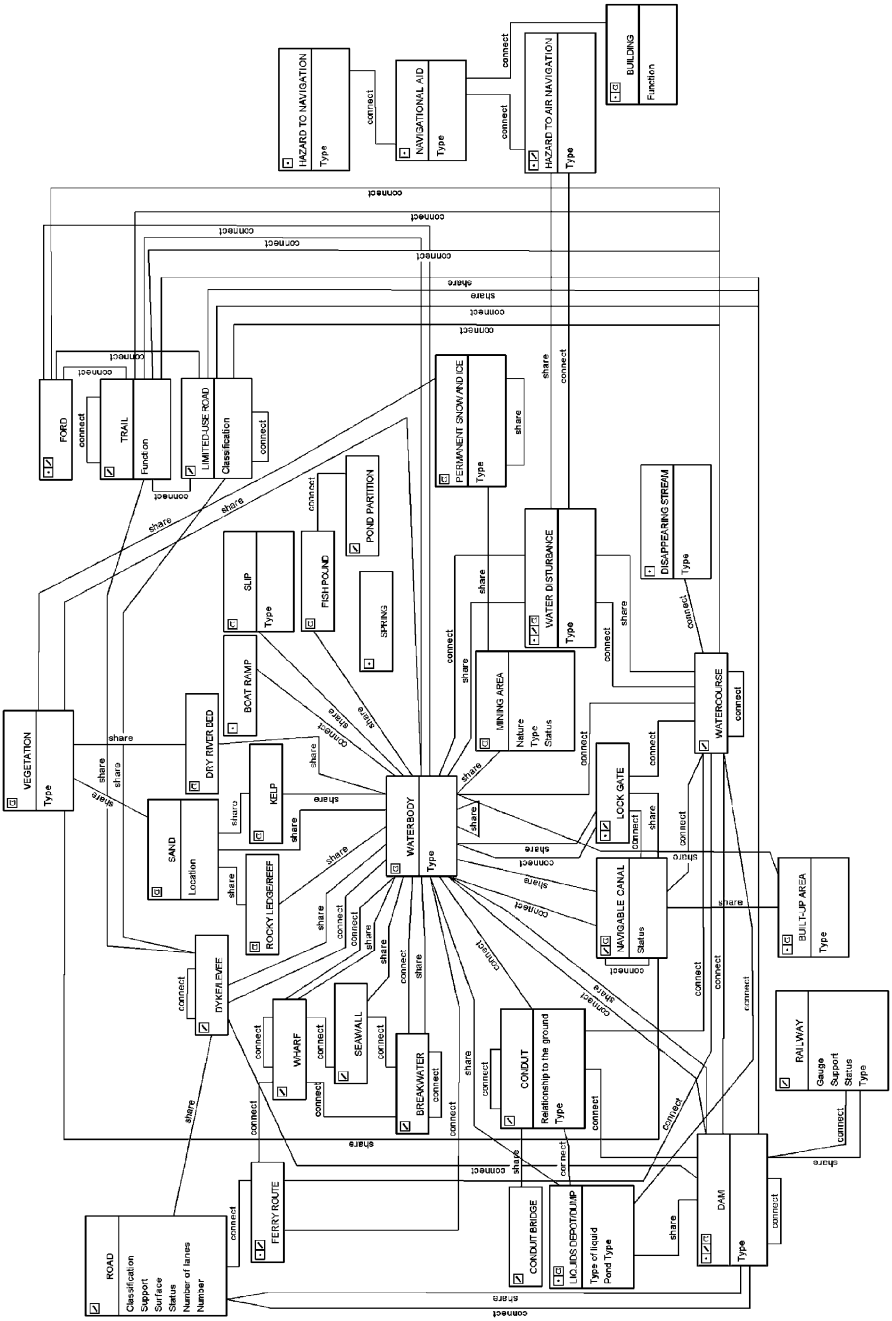

Fig. 10. NTDB hydrographic network UML class diagram.

Each object class and relationship is documented in a data dictionary, which provides its semantics and their inherent properties as shown in Fig. 11 for the class *road* of the NTDB road network (Fig. 9).

Software agents were instantiated using the above geospatial data repositories, which served as application ontologies. We used ten different software agents, one for each ontologies, and 46 road network related geoConcepts distributed among the five different road network ontologies and also 44 hydrographic network related geoConcepts distributed among the five different hydrographic network ontologies. We placed road network agents in interaction between themselves using the road network related geoConcepts, and did the same thing the hydrographic network agents. The results presented hereafter show the success rate where a data provider agent answered adequately to a query from a user agent. The data provider agent either answered it had not understood the query with its own ontology or used a similar geoConcept, which was recognized as such by the user agent, to answer the query. In the case where interacting software agents were of the same ontology, we observed that agents used in all cases the same geoConcept to generate and recognize the geoConceptRep of the message, which results in a success rate of 100% for both road and hydrographic networks (Figs. 12 and 13). For example, the message receives by an NTDB road network-based agent including a geoConceptRep generated from the geoConcept Highway exit of another NTDB road network-based agent was always recognized by the geoConcept *Highway exit* with a *GsP* of *GsP_tfft* (or equal) with the geoConceptRep.

When software agents of different ontologies but related to the same network (road or hydrographic) were interacting, a geoConcept of the destination agent succeeds in recognizing the geoConceptRep generated by the source agent when common intrinsic and extrinsic properties have been identified. We observed that software agents succeeded in recognizing messages received from another software agent of a different ontology in a success rate ranging from 30% to 100% depending on ontologies with a mean of 59% for the road network and 61% for the hydrographic network (Figs. 12 and 13). The difference between these results and 100% is explained because we used an artificial *root* geoConcept to link the sub-networks composing an ontology in order to use a graph traversal algorithm to navigate from one geoConcept to another within the ontology. This artificial *root* geoConcept has caused undesirable situations for instance when a geoConcept and a geoConceptRep had both a relationship with this *root* geoConcept, they showed a false *geosemantic proximity*. Table 1 illustrates a few examples of software agent's geoConcepts that automatically recognize geoConceptsReps encoded by another software agent where both agents were using different ontologies.

In all these examples, even if the *geosemantic proximity* between the geoConcept and the geoConceptRep seems obvious in certain cases (e.g. trail or coastline) because they appear to be identical abstractions, they are essentially different but similar because of all their inherent properties. It is because of their similarity that the geoConcept can be used to assign a meaning to the geoConceptRep.

However, it is still possible that an agent's geoConcept may not recognize a geoConceptRep encoded and transmitted by another agent. The agent's ability to

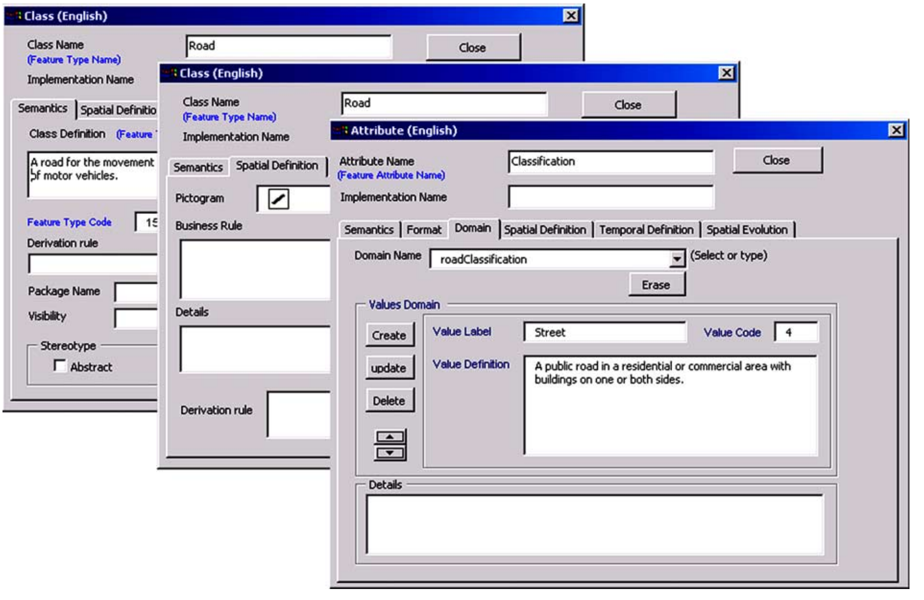

Fig. 11. Extract of the class *Road* of the data dictionary of the NTDB road network (made with *Perceptory*).

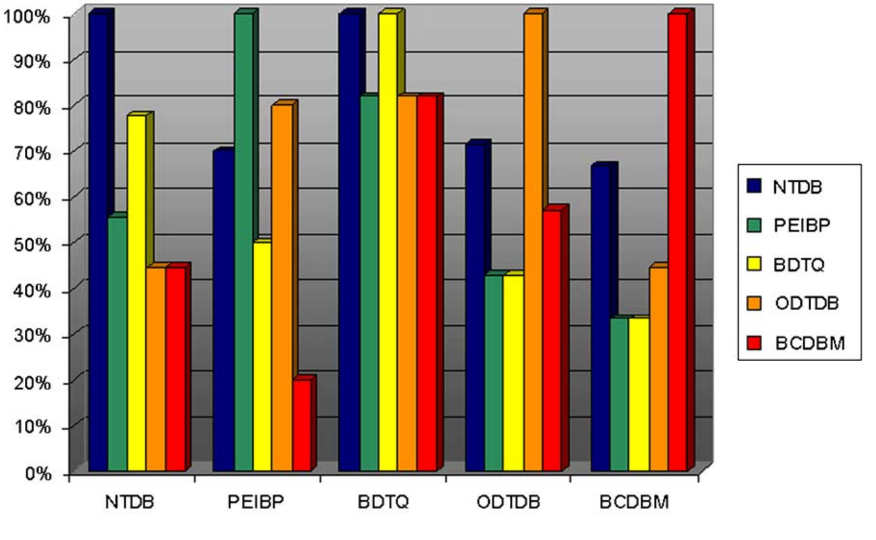

Fig. 12. Observed success rates—road network.

recognize a geoConceptRep resides in the richness of its ontology in terms of the geoConcepts it knows and the relationships between geoConcepts.

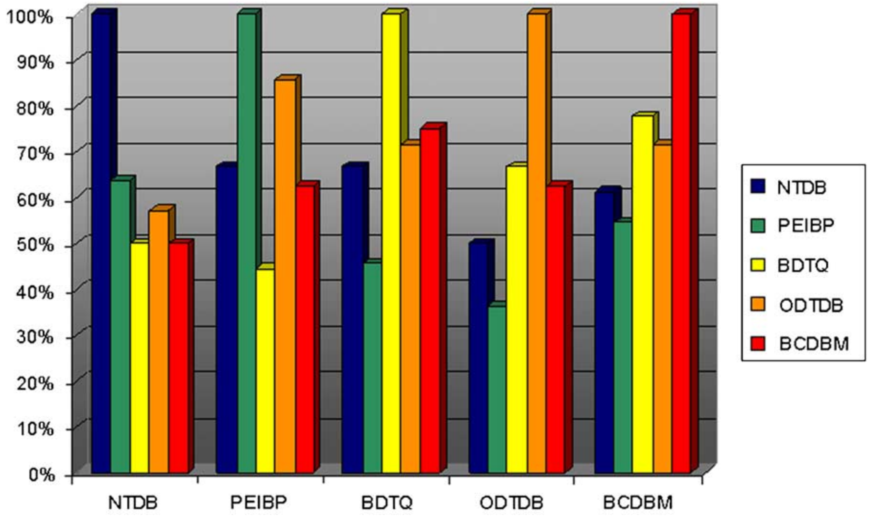

Fig. 13. Observed success rates—hydrographic network.

These results demonstrate that interoperability is possible between software agents of different ontologies although their respective data product specifications have not been developed explicitly for that purpose. To increase the level of interoperability of geospatial data, organizations involved in geospatial data acquisition, management, and dissemination should consider the development with meaningful geoConcepts in terms of content and relationships between each other, regardless the manner they are implemented in geospatial databases. The integration of software agents of domain and global ontologies in the prototype would also be an important improvement for geospatial data interoperability. Finally, the extraction of semantic information from definitions of geoConcepts, attributes, and attribute values typically stored in natural language in geospatial data repositories would further enhance the evaluation of *GsP*.

4. Conclusion

In this paper, we reviewed our conceptual framework for geospatial data interoperability, which has been derived from human communication and cognition theories. In this framework, user and data provider agents maintain in memory a set of geoConcepts, which constitute their respective ontologies. Agents communicate geoConcepts to others by generating and transmitting representations of the geoConcepts—i.e. geoConceptReps. When receiving a message, an agent goes through its geoConcepts to find those that recognize the message's geoConceptReps and then to give them a meaning. The notion of *geosemantic proximity* is here in support of these geoConcept's capabilities—i.e. to generate and recognize geoConceptReps. By the *geosemantic proximity*, a geoConcept assesses its semantic, spatial, and

Table 1
Examples of geoConcepts recognizing geoConceptReps, both of different ontologies

Agent		Recognizes (with the corresponding *GsP*)	Agent	
geoConcept	Ontology		geoConceptRep	Ontology
Road	NTDB	*GsP_tfft* (*equal*)	Road	PEIBP
Road	NTDB	*GsP_ffft*	Street	QTDB
Trail	ODTDB	*GsP_ffft*	Trail	NTDB
Lake	BCDBM	*GsP_tfft* (*equal*)	Lake	PEIBP
Coastline	PEIBP	*GsP_tfft* (*equal*)	Coastline	BCDBM
Rocky ledge/Reef	PEIBP	*GsP_ffft*	Rocky ledge/Reef	NTDB
Water disturbance	NTDB	*GsP_ffft*	Rapids	PEIBP
Disappearing stream	NTDB	*GsP_ffft*	Sinkhole	PEIBP

temporal similarity with a geoConceptRep. More specifically, the geoConcept evaluates the correspondence of its intrinsic and extrinsic properties with those of the geoConceptRep and expresses it in a four-intersection matrix.

Also, we presented the *GsP Prototype* that was developed to test the notion of *geosemantic proximity* within our framework. The *GsP Prototype* consists of software agents that communicate with each other by sending queries and replies. A software agent possesses its own ontology, which consists of a geospatial data repository. This agent accomplishes tasks on behalf of a user when it communicates with a data provider, and conversely, it accomplishes tasks on behalf of a data provider when it communicates with a user. As such, an agent communicates geoConcepts by generating and transmitting geoConceptReps in XML. An agent recognizes and assigns a meaning to a geoConceptRep by using the geoConcept of its ontology that has the most significant *geosemantic proximity* with the geoConceptRep. The *GsP Prototype* has been tested against agents using identical ontologies and agents using different ontologies. For this purpose, geospatial data repositories have been worked out to serve as agent's application ontologies from five different geospatial data product specifications. In the experimentation conducted, agents using identical ontologies always ended up understanding each other where the geoConcept that recognizes the geoConceptRep was always identical to the one that produces that geoConceptRep. Agents using different ontologies end up understanding each other when geoConcepts and geoConceptReps show sufficient commonalities. Limitations between agents of different ontologies come from the poverty of ontologies in terms of amount of geoConcepts and inherent structure as well as the difficulty to handle definitions in natural language.

Although we consider the prototype and the experimentation to be successful, a number of issues still need to be addressed, notably in (1) the development of more rigorous ontologies, (2) the extraction of intrinsic and extrinsic properties from natural language definitions of geoConcepts, attributes, attribute values, etc., (3) the integration of application, domain, and global ontologies and their interactions. Finally, we believe that this research takes a step forward in the achievement of the complete interoperability of geospatial data.

Acknowledgements

The authors wish to acknowledge the contribution of Natural Resources Canada—Centre for Topographic Information in supporting the first author for this research and of the GEOIDE Network of Centres of Excellence in geomatics, project DEC#2 (Designing the Technological Foundations of Spatial Decision-making with the World Wide Web).

References

Allen, J. F. (1981). An interval-based representation of temporal knowledge. In *Proceedings of international joint conference on artificial intelligence* (pp. 221–226).

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communication of the ACM, 26*, 832–843.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.

BC Ministry of Environment Lands and Parks (Geographic Data BC) (1992). *Digital baseline mapping at 1:20,000* (Release 2.0). Victoria, Province of British Columbia: BC Ministry of Environment, Lands and Parks.

Bédard, Y. (1999). Visual modelling of spatial database towards spatial PVL and UML. *Geomatica, 53*, 169–186.

Bédard, Y., Proulx, M.-J. (2002). Perceptory web site [web page document]. Available from: http://sirs.scg.ulaval.ca/Perceptory.

Benslimane, D. (2001). Interopérabilité de SIG: la solution Isis. *Revue internationale de géomatique, 11*, 7–42.

Bishr, Y. (1997). Semantics aspects of interoperable GIS, Ph.D. Dissertation. ITC Publication.

Bittner, T., & Edwards, G. (2001). Toward an ontology for geomatics. *Geomatica, 55*, 475–490.

Brodeur, J. (2004). Interopérabilité des données géospatiales: Élaboration du concept de proximité géosémantique, Thèse de doctorat, Université Laval.

Brodeur, J., & Bédard, Y. (2001). Geosemantic proximity, a component of spatial data interoperability. In *Proceedings of international workshop on semantics in enterprise integration* (*OOPSLA 2001*) (6 p).

Brodeur, J., & Bédard, Y. (2002). Extending geospatial repositories with geosemantic proximity functionalities to facilitate the interoperability of geospatial data. In *Proceedings of ISPRS, SDH2002 and CIG joint international symposium on geospatial theory, processing and application* (6 p).

Brodeur, J., Bédard, Y., Edwards, G., & Moulin, B. (2003). Revisiting the concept of geospatial data interoperability within the scope of a human communication process. *Transactions in GIS, 7*, 243–265.

Brodeur, J., Bédard, Y., & Proulx, M. J. (2000). Modelling geospatial application databases using UML-based repositories aligned with international standards in geomatics. In K. J. Li, K. Makki, N. Pissinou, & S. Ravada (Eds.), *Proceedings of eighth ACM symposium on advances in geographic information systems* (*ACMGIS*) (pp. 39–46). ACM Press.

CCOG—Working group on base data quality issue (2001). A new vision for quality base geographic information in Canada—GeoBase.

Charron, J. (1995). Développement d'un processus de sélection des meilleures sources de données cartographiques pour leur intégration à une base de données à référence spatiale, Mémoire de maîtrise. Université Laval.

Cruz, I. F., Rajendran, A., Sunna, W., Wiegand, N. (2002). Handling semantic heterogeneities using declarative agreements. In *Proceedings of ACM-GIS 2002, The 10th ACM international symposium on advances in geographic information systems* (pp. 168–174). McLean, VA.

Egenhofer, M. J. (1993). A model for detailed binary topological relationships. *Geomatica, 47*, 261–273.

Egenhofer, M. J. (1999). Introduction: theory and concepts. In M. Goodchild, M. Egenhofer, R. Fegeas, & C. Kottman (Eds.), *Interoperating geographic information systems* (pp. 1–4). Boston, MA: Kluwer Academic Publisher.

Egenhofer, M. J., & Franzosa, R. D. (1991). Point-set topological spatial relations. *International Journal of Geographic Information Science, 5*, 161–174.

FGDC (2002). NSDI [web page document]. Available from: http://www.fgdc.gov/nsdi/nsdi.html.

Fonseca, F., & Egenhofer, M. (1999). Ontology-driven geographic information systems. In *Proceedings of 7th International Symposium on Advances in Geographic Information Systems* (*ACMGIS*) (pp. 14–19). ACM Press.

Fonseca, F. T., Egenhofer, M. J., Davis, C. A., Jr., & Borges, K. A. V. (2000). Ontologies and knowledge sharing in urban GIS. *Computer Environment and Urban Systems, 24*, 251–271.

Fowler, J., Perry, B., Nodine, M., & Bargmeyer, B. (1999). Agent-based semantic interoperability in InfoSleuth. *Sigmod Record, 28*, 8.

Frank, A. U. (2001). Tiers of ontology and consistency constraints in geographic information systems. *International Journal of Geographic Information Science, 15*, 667–678.

GeoBase (2001). National GeoBase Definition.

GeoConnections (2002). Canadian geospatial data infrastructure (CGDI) architecture [electronic document]. Available from: http://www.geoconnections.org/architecture/index_e.html.

Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. Technical Report KSL 93–04 (22 p). Palo Alto, California: Knowledge Systems Laboratory.

Guarino, N. (1998). Formal ontology and information systems. In N. Guarino (Ed.), *Proceedings of formal ontology in information systems (FOIS '98)* (pp. 3–15). Amsterdam: IOS Press.

Guarino, N., & Welty, C. (2000). A formal ontology of properties. In R. Dieng & O. Corby (Eds.), *Proceedings of knowledge engineering and knowledge management: methods, models and tools* (*12th international conference, EKAW2000*)*: Vol. 1937. Lecture Notes in Computer Science* (pp. 97–112). Berlin: Springer-Verlag.

Harvey, F. J. (2002). Semantic interoperability and citizen/government interaction. In *Proceedings of spatial data handling 2002: joint international symposium on geospatial theory, processing, and applications* (13 p).

Institute of Electrical and Electronics Engineers (1990). *IEEE standard computer dictionary: a compilation of IEEE standard computer glossaries* New York, NY.

ISO/TC 211 (2001a). *ISO/DIS 19110 geographic information—feature cataloguing methodology*. Geneva, Switzerland: International Organization for Standardization.

ISO/TC 211 (2001b). *ISO/PDTS 19103 geographic information—conceptual schema language*. Geneva, Switzerland: International Organization for Standardization.

ISO/TC 211 (2003a). *ISO 19107:2003 geographic information—spatial schema*. Geneva, Switzerland: International Organization for Standardization.

ISO/TC 211 (2003b). *ISO 19115:2003 geographic information—metadata*. Geneva, Switzerland: International Organization for Standardization.

Jones, M. (1991). Brave new world: a vision of IRDS. *Database Programming and Design*, 43–49.

Kashyap, V., & Sheth, A. (1996). Semantic and schematic similarities between database objects: a context-based approach. *The VLDB Journal, 5*, 276–304.

Kottman, C. (1999). The open GIS consortium and progress toward interoperability in GIS. In M. Goodchild, M. Egenhofer, R. Fegeas, & C. Kottman (Eds.), *Interoperating geographic information systems* (pp. 39–54). Boston, MA: Kluwer Academic Publisher.

Kuhn, W. (2000). Ontologies from texts. In *Proceedings of first international conference on geographic information science* (p. 3). Savannah, Georgia.

Kuhn, W. (2003). Semantic reference systems. *International Journal of Geographical Information Science, 17*, 405–409.

Laurini, R. (1998). Spatial multi-database topological continuity and indexing: a step towards seamless GIS data interoperability. *International Journal of Geographic Information Science, 12*, 373–402.

Lehmann, F. (1992). Semantic networks. *Computers and Mathematics with Applications, 23*, 50.

Marco, D. (2000). *Building and managing the meta data repository*. Wiley.

McKee, L., & Buehler, K. (Eds.). (1998). *The OpenGIS guide* (third ed.). Wayland, MA: OpenGIS Consortium Inc.

Moriarty, T. (1990). Are you ready for a repository? *Database Programming and Design, March 1990*, 62–71.
Natural Resources Canada (1996). *National topographic data base—standards and specifications* (3.1 ed.). Sherbrooke, Quebec: Centre for Topographic Information—Sherbrooke.
Nwana, H. S. (1996). Software agents: an overview. *The Knowledge Engineering Review, 11*, 205–244.
Nwana, H. S., & Wooldridge, M. (1996). Software agent technologies. *BT Technology Journal, 14*, 68–78.
OBM (1996). *Ontario digital topographic database—1:10,000, 1:20,000—a guide for user*. Toronto, Ontario: Ministry of Natural Resources.
Open GIS Consortium Inc. (1999). *OpenGIS simple features specification for SQL*. Wayland, MA: OpenGIS Consortium Inc.
Open GIS Consortium Inc. (2001). *Geography markup language (GML) 2.0*. Wayland, MA: Open GIS Consortium Inc.
Ouksel, A. M., & Sheth, A. (1999). Semantic interoperability in global information systems: a brief introduction to the research area and the special section. *Sigmod Record, 28*, 5–12.
P.E.I. Geomatics Information Centre (n.d.). *User's guide to digital and hardcopy property and basemap products*. Charlottetown, P.E.I.: Provincial Treasury—Taxation and Property Records Division.
Payne, T. R., Paolucci, M., Singh, R., & Sycara, K. (2002). Communicating agents in open agent systems. In *Proceedings of first GSFC/JPL workshop on radical agent concepts* (*WRAC*) (p. 10).
Peuquet, D., Smith, B., & Brogaard, B. (1998). The ontology of fields. In *Proceedings of summer assembly of the university consortium for geographic information science*.
Prabandham, M., Selfridge, W. J., & Mann, D. D. (1990). The role of IRDS. *Database Programming and Design, April 1990*, 41–48.
Québec (2000). Base de données topographiques du Québec (BDTQ) à l'échelle de 1/20 000—Normes de production (Version 1.0) [CD document].
Renner, S. A. (2001). A "Community of Interest" approach to data interoperability. In *Proceedings of federal database colloquium 2001* (p. 7). San Diego.
Rodriguez, M. A. (2000). Assessing semantic similarity among entity classes, Ph.D. Thesis. University of Maine.
Schramm, W. (1971). The nature of communication between humans. In W. Schramm & D. F. Robert (Eds.), *The process and effects of mass communication* (pp. 3–53). Champaign-Urbana, IL: University of Illinois Press.
Sheth, A. (1999). Changing focus on interoperability in information systems: from systems, syntax, structure to semantics. In M. Goodchild, M. Egenhofer, R. Fegeas, & C. Kottman (Eds.), *Interoperating geographic information systems* (pp. 5–29). Boston, MA: Kluwer Academic Publisher.
Sheth, A., & Kashyap, V. (1992). So far (schematically) yet so near (semantically). In D. K. Hsiao, E. J. Neuhold, & R. Saks-Davis (Eds.), *Proceedings of IFIP WG2.6 database semantics conference on interoperable database systems* (*DS-5*)/IFIP transaction (A-25) (pp. 283–312). Elsevier Science Publishers B.V.
Simsion, G. C. (2001). *Data modeling essentials—analysis, design, and innovation* (2nd ed.). Scottsdale, Arizona: Coriolis.
Smith, B., & Mark, D. (1999). Ontology with human subjects testing: an empirical investigation of geographic categories. *American Journal of Economics and Sociology, 58*, 245–272.
Sondheim, M., Gardels, K., & Buehler, K. (1999). GIS interoperability. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical information systems: principles, techniques, applications and management* (pp. 347–358). New York: John Wiley and Sons, Inc.
Statistics Canada (1997). *Digital boundary file and digital cartographic file 1996 census* (*reference guide*). Ottawa: Minister of Industry.
Sun Microsystems Inc. (2002). Java™ API for XML Processing (JAXP) [web page document]. Available from: http://java.sun.com/xml/jaxp/index.html.
Sycara, K., Klusch, M., Widoff, S., & Lu, J. (1999). Dynamic service matchmaking among agents in open information environments. *Sigmod Record, 28*, 47–53.
The Appache Sofware Foundation (2002). Xalan-Java version 2.4.0 [web page document]. Available from: http://xml.apache.org/xalan-j/.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Uitermark, H. T., van Oosterom, P. J. M., Mars, N. J. I., & Molenaar, M. (1999). Ontology-based geographic data set integration. In M. H. Böhlen, C. S. Jensen, & M. O. Scholl (Eds.), *Proceedings of spatio-temporal database management, international workshop STDBM'99: vol. 1678. Lecture notes in computer science* (pp. 60–78). Berlin, Heidelberg: Springer-Verlag.

VMap (1995). Vector Map (VMap), Level 1. Mil-V-89033 (331 p). Bethesda, MD: US National Imagery and Mapping Agency.

Weisstein, E. W. (1999). Topological space [web page document]. Available from: http://mathworld.wolfram.com/TopologicalSpace.html.

Xhu, Z., & Lee, Y. C. (2002). Semantic heterogeneity of geodata. In *Proceedings of ISPRS commission IV symposium 2002: joint international symposium on geospatial theory, processing, and applications* (6 p).