EXTENDING GEOSPATIAL REPOSITORIES WITH GEOSEMANTIC PROXIMITY FUNCTIONALITIES TO FACILITATE THE INTEROPERABILITY OF GEOSPATIAL DATA

Jean Brodeur [a, b], Yvan Bédard [a]

[a] Centre for Research in Geomatics, Laval University, Cité universitaire, Québec QC Canada G1K 7P4
[b] Natural Resources Canada, Centre for Topographic Information, 2144 King Street West, Sherbrooke QC J1J 2E8
Brodeur@RNCan.gc.ca, Yvan.Bedard@scg.ulaval.ca

Commission IV, WG IV/1

KEY WORDS: Geospatial data interoperability, Geospatial repository, Geosemantic proximity, Communication, Semantic similarity

ABSTRACT:

Today, with the common availability of Internet technologies, interoperability of geospatial data has become a necessity for sharing and integrating geospatial data. More specifically, it is seen as a solution to solve syntactic, structural, semantic, geometric and temporal heterogeneities between data sources. In Canada, we observe such heterogeneities from existing geospatial databases. For example, *Vegetation*, *Trees*, *Wooded area*, *Wooded area*, *Milieu boisé* and *Zone boisée* (unknown geometry), found in different geospatial data specifications, describe the same type of phenomena. Recently, we have proposed a conceptual framework for geospatial data interoperability based on human communication concepts. This framework introduces the idea of *geosemantic proximity*, which provides reasoning capabilities to assess the semantic, geometric, and temporal similarities between geospatial concepts and geospatial conceptual representations. In the present paper, we review the conceptual framework and present an architecture of a system based on this framework. In fact, the architecture uses a geospatial repository, namely *Perceptory*, as a data source's ontology upon which we add *geosemantic proximity* functionalities. These functionalities evaluate the similarity of the information stored in the data source with the information required by another one in order to facilitate the interoperability of geospatial data.

1. INTRODUCTION

Many geospatial data sources are today publicly available to end users, especially by the way of Internet and geospatial data infrastructure (e.g. CGDI in Canada and NSDI in the United States). Typically, these data sources have been produced for specific purposes of people and organizations. For example, in Canada, there is the National Topographic Data Base (NTDB) (Natural Resources Canada, 1996) produced for national mapping and GIS application purposes, the VMap libraries (VMap, 1995) for military purposes, the Street Network Files and the Digital Cartographic Files (Statistics Canada, 1997) for enumeration purposes, and various topographic data sources produced at larger scales by provincial departments (e.g. (OBM, 1996; Québec, 2000)). These geospatial data sources abstract the topographic reality in various ways, which causes problems of sharing and integration when users join data from many sources. To illustrate this, we observe that forest-like phenomena are abstracted as *Vegetation* in NTDB, *Trees* in VMap, *Wooded area* in Ontario Digital Topographic Data Base, and *Milieu boisé* in the *Base de données topographiques du Québec* (where the pictograms mean polygonal, linear or point geometry; see (Bédard and Proulx, 2002) for the description of spatial pictograms). This rises up syntactic, structural, and, moreover, semantic, geometric, and temporal heterogeneities between the various data sources (Bishr, 1997; Charron, 1995; Laurini, 1998).

Interoperability of geospatial data has been introduced early in the 1990's as a solution for the sharing and the integration of geospatial data and geoprocessing resources (Kottman, 1999). The Open GIS Consortium Inc., the ISO/TC 211-Geographic Information/Geomatics, governmental organizations, the geographic information industry, and the research community have worked in co-operation to achieve the current foundation of geospatial data interoperability. Major progresses are noted especially for syntactic and structural heterogeneities (Rodriguez, 2000). As such, documents like (ISO/TC 211, 2001a; ISO/TC 211, 2001b; Open GIS Consortium Inc., 1999; Open GIS Consortium Inc., 2001) provide definitions of the content and the structure of geometric data as well as syntactical descriptions of geospatial data. However, interoperability of geospatial data must go beyond this fact to include solutions for semantic heterogeneity (Egenhofer, 1999). Following this line of thought, we have recently proposed a conceptual framework for geospatial data interoperability in order to position a new approach called *geosemantic proximity* for the assessment of semantic interoperability of geospatial data (Brodeur and Bédard, 2001). It consists in a human communication-like process between two agents, which exchange in order to share and integrate geospatial data from each other.

The elaboration of geospatial repositories is recognised as a good practice in the development of geospatial databases. A geospatial repository constitutes a comprehensive source of knowledge that captures the semantics and the structure of the data being stored in a geospatial database (Brodeur et al., 2000). Then, it can be considered as an application *ontology* (Gruber, 1993) to support geospatial data interoperability.

In this paper, we present an architecture of a system that integrates the idea of *geosemantic proximity* with geospatial repositories to assess the semantic interoperability of geospatial data.

The rest of the paper is subdivided as follows. In the next section, we review related notions to geospatial data interoperability. The third section summarizes our conceptual framework of geospatial data interoperability and the idea of *geosemantic proximity*. In section 4, we present a system architecture that implements the idea of *geosemantic proximity* upon *Perceptory*, a typical geospatial repository (Bédard and Proulx, 2002), based on our framework for geospatial data interoperability. Finally, we conclude in section 5 and present future work.

2. INTEROPERABILITY-RELATED NOTIONS

Our conceptual framework for geospatial data interoperability and the *geosemantic proximity* approach are based on a number of fields such as philosophy, human communication, cognition, computer science, and geographic information. More specifically, we have considered works related to ontology, context, semantic proximity, topology, semantic interoperability, and mapping specifications.

As people end up understanding each other when communicating, we think that interoperability of geospatial data obeys a *human communication process* (as described by (Schramm, 1971)). A human communication process consists in an individual transmitting details depicting real-world phenomena he/she has in mind to someone else. Typically, the communication process is made up of a human source, a human destination, physical signals, a communication channel, a source of noise, and a feedback component. Multiple representations of reality are involved in a communication process, namely the source and destination cognitive models and the physical signals that are used to transmit a message between the source and the destination.

The source and the destination cognitive models are built from the direct observation of phenomena and from the observation of intentional semantic signals from someone else. These signals are captured by our sensory systems to form *perceptual states* (Barsalou, 1999). The human selective attention gathers properties of interest and stores them as *perceptual symbols* (hereafter called *concepts*) (Barsalou, 1999) permanently in memory. A concept consists of both cognitive elements–i.e. hidden data-like elements–and a translation function that encapsulates these elements. This translation function recognizes and produces physical signals (hereafter called *conceptual representations*) about that concept (Barsalou, 1999). In memory, concepts are aggregated in clusters, which express a kind of similarity between each others (Krech and Crutchfield, 1971).

A conceptual representation acts as an intermediary between the source and the destination. It transmits a source's concept adapted to a specific context and a particular use towards destination.

To illustrate the communication process in the context of geospatial data interoperability, let's use the following example inspired from (Kottman, 1999). A person interesting to rent a house or the like asks an agent of a local service for information about *houses for rent* in the area of *Sherbrooke*. Once the agent receives the request, he/she interprets it, and provides a list of available *dwellings for rent* in the area of *Sherbrooke* to the person, which answers completely the person's request. In such a communication process, interoperability happens between the person and the agent.

The communication process as a model of interoperability for geospatial data includes multiple representations and descriptions of real-world phenomena. The representations and descriptions of real-world phenomena is still a subject matter studied in ontology and database modeling.

In its philosophical meaning, ontology is concerned with the description of the world in itself, with a model and an abstract theory of the world, and with the science of being, of the type of entities, of properties, of categories, and of relationships (Peuquet et al., 1998; Smith and Mark, 1999). However, in artificial intelligence, it corresponds to "an explicit specification of a conceptualisation" (Gruber, 1993) and "a logical theory accounting for the intended meaning of a formal vocabulary" (Guarino, 1998). As mentioned previously, one phenomenon could be described in multiple ways. Thus, following Gruber's and Guarino's definitions, we assume an ontology to be a formal representation of phenomena supported by a vocabulary and definitions that explicit the intended meaning, and represents phenomena with their interrelationships.

In database modeling, a conceptual model consists of an abstract description of a portion of reality from a data-centered point of view. A conceptual model is a tool to think about, to document, to communicate, and to develop data sources about parts of reality (Bédard, 1999). It captures, structures, and catalogues selected features in general categories, classes, properties, relationships, generalizations, aggregations, roles, constraints, behaviours, geometric properties, temporal properties, and so on, using a given formalism (e.g. UML). A good practice in developing conceptual model is to support it with a data dictionary that specifies the intended semantics of features. The conceptual model along with the data dictionary constitutes the repository's essential components (Brodeur et al., 2000). A repository is defined as a collection of metadata that is structured in such a way to provide the semantics and the structure of the objects stored in a database (Brodeur et al., 2000). It consists of information to assess fitness for use of data, to support data integration from multiple sources, and to support data interoperability. However, databases with their corresponding repositories (conceptual models and data dictionaries) are usually developed to serve specific needs and specific uses by database practitioner of various backgrounds and experiences. Consequently, the same part of reality is abstracted differently from one conceptual model to another. This drives to problems of interoperability when merging conceptual models and geospatial data to elaborate a more comprehensive set of data.

The abstraction of real-world phenomena is basically driven by the situation or the circumstances in which phenomena are perceived and used. This refers to the *context*, which influences the definition of concepts and conceptual representations with specific intrinsic and extrinsic properties (described later in this section). Context is recognized as a basic element for the assessment of semantic interoperability, which provides abstractions with their fundamental semantics (Kashyap and Sheth, 1996). A main requirement for semantic interoperability of geospatial data is to have reasoning methods that take the context into consideration. Semantic proximity in a context-based perspective is seen as such a reasoning method, which expresses qualitatively the semantic similarity between abstractions (e.g. semantic resemblance, semantic relevance, semantic relation, semantic equivalence and semantic incompatibility) (Kashyap and Sheth, 1996).

Conceptual representations are used to transmit concepts within a given context. However, to be interoperable, concepts and conceptual representations have to refer to the same phenomenon and, thus, the same identity of the phenomenon must be recognized from these various abstractions. Therefore, concepts and conceptual representations are not as important as the phenomena they represent. Identity is then considered to be a closely related notion to geospatial data interoperability. It consists in a meta-property from which we distinguish and individualize phenomena (Guarino and Welty, 2000), which allows the recognition of real-world phenomena representations.

Concepts and conceptual representations typically circumscribe a particular set of phenomena. We can imagine they follow a geometric-like metaphor, such as a segment on a semantic axis, with an interior and boundaries. The interior of the segment consists of the set of intrinsic properties and boundaries of the segment, of the set of extrinsic properties. Intrinsic properties are those providing the literal meaning. They describe the essential nature of a phenomenon. They are not dependent of any external factors. Basically, the identity of a phenomenon can be recognized from the intrinsic properties. Identification, attributes, attribute values, geometries, temporalities, and domain are good intrinsic property candidates for geospatial concepts and geospatial conceptual representations. Extrinsic properties are those influenced by external factors. They provide meaning based on their interaction with other concepts or conceptual representations and as such, set the limit of the literal meaning of a concept or a conceptual representation. Behaviours and relationships (semantic, spatial, and temporal) are good extrinsic property candidates for geospatial concepts and geospatial conceptual representations. Thus, extrinsic properties are associated to the notion of boundary. The notion of boundary has been discussed in (Casati et al., 1998); they have identified two types of boundaries: *bona fide* and *fiat* boundaries. *Bona fide* boundaries are associated to genuine or physical demarcation as it is the case for *buildings*, *runways*, and *the body of a person*. *Fiat* boundaries correspond to human driven demarcations, which are more theoretical, mathematical, or virtual and have no direct relationship with physical objects. This is the case for *administrative boundaries* or the boundary between waterbodies such as the *St.Lawrence Gulf* and the *Atlantic Ocean*. Concepts and conceptual representations are essentially defined by humans and, thus, can be associated with *fiat* boundary (Smith and Mark, 1999). According to (Casati et al., 1998), objects of *fiat* boundaries follow a topology that includes an interior and a boundary. This kind of topology has been a subject of interest in geospatial information (Egenhofer et al., 1994). We propose to extend the use of topology for the assessment of *geosemantic proximity*.

3. INTEROPERABILITY OF GEOSPATIAL DATA AND GEOSEMANTIC PROXIMITY

This section reviews our conceptual framework for spatial data interoperability and the idea of *geosemantic proximity*. They constitute the theoretical foundation of the architecture presented in the next section.

3.1 Interoperability of Geospatial Data

As introduced in the previous section, geospatial data interoperability follows a human communication-like process. To illustrate this, let's assume the situation in which a user agent (A_u) of geospatial data wishes to have information about the hydrologic network for flood analysis in the region of Sherbrooke. He/she sends a query to a geospatial data source, called a data provider agent (A_{dp}), to get information about *lakes* and *rivers* within *Sherbrooke*. When A_{dp} receives the A_u's query, it interprets it–i.e. to find and associate concepts it knows with the received conceptual representations (e.g. *watercourses* and *waterbodies* in the proximity of *Sherbrooke*). Once the query has been interpreted, A_{dp} gathers and sends to A_u the information that fulfils totally his/her original request (e.g. *Lac des Nations*, *Magog River*, and *Saint-François River*). In this process, A_u and A_{dp} use their own vocabulary to communicate their respective abstractions of real-world phenomena. Because of their common set of symbols and backgrounds, they can end up understanding each other (Bédard, 1986).

From this situation, we elaborated a conceptual framework for geospatial data interoperability that uses five expressions of reality: ***R***, ***R′***, ***R″***, ***R‴***, and ***R′′′′*** (Figure 1). These expressions are five separate ontologies that are linked together within a communication process to form what we call the *five ontological phases of geospatial data interoperability*. ***R*** represents the topographic reality as observed by A_u at a given time about which he needs information. ***R*** is beyond description. ***R′*** is the A_u's abstraction of ***R***. It corresponds to the set of properties of ***R*** selected by A_u and arranged into concepts. These properties constitute the A_u's cognitive model. ***R″*** is the set of conceptual representations encoded by A_u, which express relevant properties of ***R′***s concepts to describe the A_u's specific need (e.g. *Lakes* or *Rivers* within *Sherbrooke*). These conceptual representations are the data used for interoperability that are placed in the communication channel towards destination A_{dp}. ***R‴*** refers to the set of concepts that A_{dp} has in memory. These concepts are used to decode ***R″***s conceptual representations and to assign them an explicit meaning, for instance *watercourses*, *waterbodies* and *Sherbrooke*. ***R′′′′*** gathers the conceptual representations encoded by ***R‴***s concepts, which comply with ***R′***–i.e. the A_u's initial request–(e.g. *Lac des Nations*, *Magog River*, and *Saint-François River*). These conceptual representations are finally decoded and validated by A_u to identify if they infer the needed information. We can say that interoperability occurs only once ***R′′′′*** is validated. As we can observe from this conceptual framework, geospatial data interoperability consists in a bi-directional process with a feedback mechanism in both directions, which ensures that messages get destination and are satisfactorily understood by the recipient.

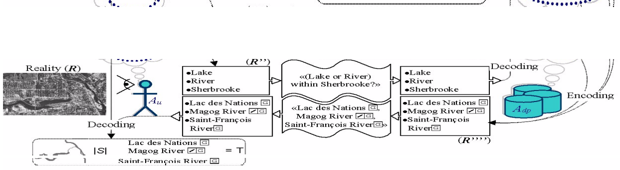

Figure 1: A Framework for Spatial Data Interoperability (Brodeur and Bédard, 2001)

Translation processes (encoding and decoding) are typically viewed as middleware components. However, in this conceptual framework, they are tied to concepts that are found in ***R′*** and ***R‴***. They are used to recognize and produce conceptual representations that fit the concept.

3.2 Geosemantic Proximity

As mentioned, a concept must be able to recognize and to generate conceptual representations. Therefore, concept's reasoning capabilities are considered a key element of the conceptual framework. As such, we suggest the idea of *geosemantic proximity* (*GsP*). It consists in a context-based approach that assesses the semantic similarity between a geospatial concept and a geospatial conceptual representation.

The context is basic in the abstraction process. Even if it is a fictitious and an imaginary notion, it is omnipresent and it guides the abstraction of phenomena. As such, the context makes some properties of phenomena more interesting than others do. Accordingly, the context (C) of a concept or a conceptual representation is described by the set of built-in properties, which are of two types of properties: intrinsic properties ($C°$) and extrinsic properties (∂C). Consequently, we define the context as in equation 1.

$$C_K = C_K° \cup \partial C_K \qquad (1)$$

where:
- C_K: Context of K
- $C_K°$: Intrinsic properties of C_K
- ∂C_K: Extrinsic properties of C_K

GsP is a component of the concept's translation process. Analogously to human reasoning, it qualifies the likeness of a geospatial concept with a geospatial conceptual representation by comparing their respective intrinsic and extrinsic properties. As such, *GsP* consists of the intersection of the context of the geospatial concept K with the context of the geospatial conceptual representation L (equation 2). It is further expanded to a four-intersection matrix (equation 3) to work out the different intersections between intrinsic and extrinsic properties. Each different intersections of the matrix can be tested empty (denoted by Φ or *f*) or non-empty (denoted by ¬Φ or *t*). The sixteen derived predicates (2^4) that are presented in row major form (i.e. row by row) according to the four-intersection matrix characterize the different *geosemantic proximity* cases: *GsP_ffff* (*disjoint*), *GsP_ffft*, *GsP_fftt* (*contains*), *GsP_tfft* (*equal*), *GsP_ftft* (*inside*), *GsP_tftt* (*covers*), *GsP_ttft* (*coveredBy*), *GsP_fttt* (*overlap*), *GsP_tttt*, *GsP_tfff* (*meet*), *GsP_tftf*, *GsP_tttf*, *GsP_ttff*, *GsP_fttf*, *GsP_fftf*, *GsP_ftff*.

$$GsP\ (\mathrm{K,L}) = C_K \cap C_L \qquad (2)$$

where:
- C_K: Context of the geospatial concept K
- C_L: Context of geospatial conceptual representation L
- *GsP* (K,L): *Geosemantic proximity* between K and L

To illustrate the *GsP* approach, let's look about the following example. Consider that we have two data sources as different agents like in Figure 1. First, we have the National Topographic Data Base (NTDB) and second, the *Base de données topographiques du Québec* (BDTQ). For some reason (e.g. for update purposes), BDTQ requests to NTDB information about *street* ◩. When NTDB receives the request, it searches in its content to find a concept that is semantically similar to *street* ◩. It finds that the concept *road* ◩ has an attribute called *classification*, which can take the value *street* of similar definition to *street* ◩. They have as well the same type of geometry. As such, they have common intrinsic properties. Also, as part of the BDTQ's description of *street* ◩, *street* ◩ shows relationships with other classes of road that are part of the NTDB's description of *road* ◩ and, as such, *street* ◩ extrinsic properties are related to *road* ◩ intrinsic properties. Consequently, the *geosemantic proximity* of the concept *road* ◩ from NTDB with the conceptual representation *street* ◩ requested by BDTQ is *GsP_ttff* or *contains*.

$$GsP\ (\mathrm{K,L}) = \begin{pmatrix} \partial C_K \cap \partial C_L & \partial C_K \cap C_L° \\ C_K° \cap \partial C_L & C_K° \cap C_L° \end{pmatrix} \qquad (3)$$

4. GEOSPATIAL REPOSITORIES AND *GsP* FUNCTIONALITIES

To validate the above conceptual framework of geospatial data interoperability and the *GsP* approach, a prototype is currently under development. It uses geospatial repositories that serve as agents' ontologies on top of which *GsP* functionalities are added. This section presents a system architecture, which describes how *GsP* functionalities are integrated to *geospatial repository*.

4.1 Geospatial Repository

A geospatial repository consists of a collection of metadata structured in such a way to provide the semantics and the structure of objects stored in a database. A geospatial repository can then be used as the ontology of the database that it describes.

In the architecture presented below, we are using *Perceptory*, which is a typical geospatial repository. *Perceptory* consists of a UML-based conceptual model building tool and an object database dictionary. It captures and manages representations of user's perceptions, which support the development of geospatial database. Accordingly, *Perceptory* allows definitions of *classes*, characteristics such as *descriptive attributes*, *geometries*, *temporalities*, and *visual information*, relationships including *associations*, *aggregations*, *compositions*, *dependencies*, and *generalizations*, *operations* (i.e. class behaviours), and more (Brodeur et al., 2000). It consists of a graph-like structure where relationships link classes together. However, the addition of *geosemantic proximity* functionalities would enhance *Perceptory* in order to serve for geospatial data interoperability.

4.2 An Architecture

The architecture, described below and illustrated in Figure 2, consists of three distinct components: two agents and a communication channel. They are describes below.

The two agents (A and B) are identical in this architecture. An agent can <u>receive</u> conceptual representations that are transmitted in the communication channel. Each of these conceptual representations is transformed in a data structure called conceptual representation, which takes place in the agent memory. This data structure is like a *perceptual state* of a cognitive agent. In order to be recognized, a conceptual representation is passed to a proxy. This proxy is a process that acts as an intermediary to locate a concept corresponding to the passed conceptual representation. A concept can be located either in the concepts data storage or in *Perceptory*. Concepts is an internal data storage where the most recent concepts used by the agent are placed temporarily. The object structure of a concept is described later in the section. This concepts storage is like the short-term memory of a cognitive agent. *Perceptory* is a direct access storage that includes a description of all classes, relationships, and so on that define the content of a geospatial database. *Perceptory* is like the long-term memory of a cognitive agent. The proxy looks first in the concepts storage to locate a concept that is similar to the conceptual representation. If a concept of this storage recognizes (i.e. is similar to) the conceptual representation, then it is used to answer the other agent. If not, then the proxy gets concepts from *Perceptory*. A graph traversal algorithm is used to navigate in *Perceptory*. It begins with a concept of the concepts storage that appears to be the most related one with the conceptual representation in order to access its associated concepts in *Perceptory*. The accessed concepts are returned in a concept (with no "s") object structure, which is identical to the structure of the concepts placed in the concepts data storage. Each concept obtained from *Perceptory* evaluates its similarity with the conceptual representation. This process is repeated recursively with concepts associated to the previous concepts until a concept that is *GsP_tfft* or *equal* is found or *Perceptory* is traversed completely. If an equal concept is located, then it is used to answer the other agent. Otherwise, concepts showing other kind of similarity with the conceptual representation are sorted by their *GsP* and the most similar one is used to answer the other agent. It might happen that no concept is found similar to the conceptual representation. Finally, the proxy returns the answer in term of encoded conceptual representations, which are sent towards destination the other agent in the communication channel.

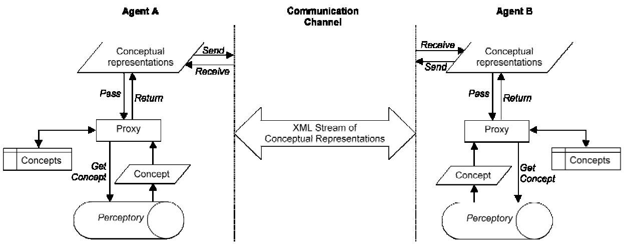

Figure 2: Architecture of the system

A concept is an object structure that consists of data elements that are hidden to other agents (Figure 3). These data elements are composed of a set of intrinsic properties and a set of extrinsic properties. On the one hand, the set of intrinsic properties is made of the definition of a *class,* its *characteristics* including *geometries* and *temporalities*, and its domain that are obtained from *Perceptory*. On the other hand, the set of extrinsic properties is made of the set of *operations* along with the membership of the class to *associations* and *generalizations* that are also obtained from *Perceptory*. As these data elements cannot be directly accessible by other agents, they are encapsulated by three functions: *recognize*, *generate*, and *gspRelate*. The *recognize* function evaluates if this concept can used to assign a meaning to the conceptual representation. The *gspRelate* function supports the *recognize* function by assessing the *geosemantic proximity* of the concept with the conceptual representation. The *generate* function produces conceptual representations that represents the concept within a specific context. Again, the *generate* function uses the *gspRelate* function to ensure that the generated conceptual representation is similar to the concept. Consequently, these three functions add reasoning functionalities on top of *Perceptory*, which provides the data. Thus, *Perceptory* and the *GsP* approach can be used together in order to assess automatically the interoperability of geospatial data.

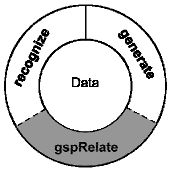

Figure 3: Object structure of a concept

Like a concept, a conceptual representation is formed of a set of intrinsic properties and a set of extrinsic properties, which encode selected intrinsic properties and extrinsic properties of a concept according to a specific context. Conceptual representations are transformed in an XML stream when placed in the communication channel and sent towards destination. An XML stream consists of a set of conceptual representation descriptions according to a predefined structure described by a DTD or an XML Schema. Instead of providing basic information of the concept such as the name of the concept, the attribute names and values, an XML conceptual representation provides also the definition of the concept, of the attributes, and of the attribute values. It provides also the domain, the geometry, and the temporality of the concept as well as the domain of attributes. Encoded likewise, a conceptual representation provides the necessary data to assess the semantic interoperability of geospatial data.

5. CONCLUSION

Supported by works on communication, cognition, ontology, geographic information, context, and semantic similarity, we have developed a conceptual framework for geospatial data interoperability. This framework corresponds to a human communication-like process that takes place between two agents. In this framework, we have differentiated two types of abstraction: *concept* (geospatial) and *conceptual representation* (geospatial). Concepts are stored in the agent memory and conceptual representations are used to communicate information about concepts. Concepts have reasoning capabilities, namely to recognize a conceptual representation and to generate a conceptual representation. Additionally, they are supported by the notion of *geosemantic proximity*, which is basically a context-based approach that qualifies the similarity of a concept with a conceptual representation. *Geosemantic proximity* corresponds to a four-intersection matrix between intrinsic and extrinsic properties of a concept and a conceptual representation. Finally, this conceptual framework serves to develop an architecture of a system that adds *geosemantic proximity* functionalities upon a geospatial repository, namely *Perceptory*. These functionalities extend geospatial repositories to facilitate the interoperability of geospatial data.

A prototype that is based on this architecture is currently under development. It aims at validating the conceptual framework and the *geosemantic proximity* approach. We expect that this research will conduct to important progress for semantic interoperability of geospatial data.

6. REFERENCES

Barsalou, L. W., 1999. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), pp. 577-609.

Bédard, Y., 1986. A Study of the Nature of Data Using a Communication-based Conceptual Framework of Land Information. Ph.D. Dissertation, University of Maine, Orono, Maine.

Bédard, Y., 1999. Visual Modelling of Spatial Database Towards Spatial PVL and UML. *Geomatica*, 53(2), pp. 169-186.

Bédard, Y., and Proulx, M.-J., 2002. Perceptory Web Site, http://sirs.scg.ulaval.ca/Perceptory

Bishr, Y., 1997. Semantics Aspects of Interoperable GIS. Ph.D. Dissertation, ITC Publication.

Brodeur, J., and Bédard, Y., 2001. Geosemantic Proximity, a Component of Spatial Data Interoperability. In: *Proceedings of International Workshop on "Semantics in Enterprise Integration" (OOPSLA 2001)*. Tampa, Florida, pp. 6.

Brodeur, J. et al., 2000. Modelling Geospatial Application Databases using UML-based Repositories Aligned with International Standards in Geomatics. In: *Proceedings of Eighth ACM Symposium on Advances in Geographic Information Systems (ACMGIS)*. Washington, D.C, ACM Press, pp. 39-46.

Casati, R. et al., 1998. Ontological Tools for Geographic Representation. In: *Formal Ontology in Information Systems*, Guarino, N. ed. IOS Press, Amsterdam pp. 77-85.

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.

Egenhofer, M. et al., 1994. The 9-Intersection: Formalism and Its Use for Natural-Language Spatial Predicates, Technical Report, 94-1, NCGIA.

Egenhofer, M. J., 1999. Introduction: Theory and Concepts. In: *Interoperating Geographic Information Systems*, Goodchild, M. et al. eds. Kluwer Academic Publisher pp. 1-4.

Gruber, T. R., 1993. A Translation Approach to Portable Ontology Specification, Technical Report, KSL 92-71, Knowledge Systems Laboratory.

Guarino, N., 1998. Formal Ontology and Information Systems. In: *Proceedings of Formal Ontology in Information Systems (FOIS '98)*. Trento, Italy, IOS Press, pp. 3-15.

Guarino, N., and Welty, C., 2000. A Formal Ontology of Properties (Preliminary Version). In: *Proceedings of ECAI-2000 Workshop on Applications of Ontologies and Problem-Solving Methods*, pp. 15.

ISO/TC 211, 2001a. *ISO/DIS 19107 Geographic Information - Spatial Schema*.

ISO/TC 211, 2001b. *ISO/DIS 19115 Geographic Information - Metadata*.

Kashyap, V., and Sheth, A., 1996. Semantic and Schematic Similarities Between Database Objects: a Context-Based Approach: a Context-Based Approach. *The VLDB Journal*, 5(pp. 276-304.

Kottman, C., 1999. The Open GIS Consortium and Progress Toward Interoperability in GIS. In: *Interoperating Geographic Information Systems*, Goodchild, M. et al. eds. Kluwer Academic Publisher pp. 39-54.

Krech, D., and Crutchfield, R. S., 1971. Perceiving the World. In: *The Process and Effects of Mass Communication*, Schramm, W., and D. F. Robert eds. University of Illinois Press pp. 235-264.

Laurini, R., 1998. Spatial Multi-Database Topological Continuity and Indexing: a Step Towards Seamless GIS Data Interoperability. *IJGIS*, 12(4), pp. 373-402.

Natural Resources Canada, 1996. *National Topographic Data Base - Standards and Specifications*. CTIS, Sherbrooke.

OBM, 1996. *Ontario Digital Topographic Database - 1:10,000, 1:20,000 - A Guide for User*.

Open GIS Consortium Inc., 1999. *OpenGIS Simple Features Specification for SQL*. OpenGIS Consortium Inc., Wayland, Mass.

Open GIS Consortium Inc., 2001. *Geography Markup Language (GML) 2.0*. Open GIS Consortium Inc., Wayland, Mass.

Peuquet, D. et al., 1998. The Ontology of Fields, Report of a specialist meeting held under the auspices of the Varenius Project, Bar Harbor, Maine.

Québec, 2000. Base de données topographiques du Québec (BDTQ) à l'échelle de 1/20 000 - Normes de production (V 1.0)

Rodriguez, M. A., 2000. Assessing Semantic Similarity Among Entity Classes. PhD. Thesis, University of Maine, Orono.

Schramm, W., 1971. How Communication Works. In: *Communication: Concepts and Processes*, DeVito, J. A. ed. Prentice-Hall Inc, New Jersey pp. 12-21.

Smith, B., and Mark, D., 1999. Ontology with Human Subjects Testing: An Empirical Investigation of Geographic Categories. *American Journal of Economics and Sociology*, 58(2), pp. 245-272.

Statistics Canada, 1997. *Digital Boundary File and Digital Cartographic File 1996 Census (Reference Guide)*. Minister of Industry, Ottawa.

VMap, 1995. Vector Map (VMap), Level 1, Military Specifications, Mil-V-89033.

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).