SEMANTIC HETEROGENEITY OF GEODATA

Zhu Xu [a, *], Y.C. Lee [b]

[a] Department of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University, Hong Kong, China, zhu.xu@polyu.edu.hk

[b] Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N.B. E3B 6Y1, Canada, yclee@unb.ca

[b] National Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, China (LIESMARS), Wuhan University, Wuhan, China

Commission IV, WG IV/2

KEY WORDS: Semantic heterogeneity, Spatial database generalization, Geodata sharing, Semantic resolution

ABSTRACT: Semantic heterogeneity is a major barrier to geodata sharing. The aim of this paper is to improve understanding of semantic heterogeneity. It provides discussions on what semantic heterogeneity of geodata is, what its sources are, what semantic heterogeneity is irresolvable and how to avoid. The conclusion is semantic poverty and poor conceptual modeling are major forms of semantic heterogeneity. We also argue that spatial database generalization is a critical technique to facilitate geodata sharing.

1. INTRODUCTION

Semantic heterogeneity is a general term referring to disagreement about the meaning, interpretation or intended use of the same or related data. This problem is poorly understood, and there is not even an agreement regarding a clear definition of the problem [Sheth and Larsen, 1990]. The importance of being aware of semantic heterogeneity and doing semantic reconciliation is to guarantee meaningful data sharing, i.e. data exchanged is correctly interpreted and used. It is now well accepted that semantic heterogeneity of geodata is the major impediment to geodata sharing.

The aim of this article is to improve our understanding of semantic heterogeneity of geodata and discuss possible solutions to semantic reconciliation if they are resolvable and how to avoid or compensate irresolvable semantic heterogeneity. The article is organized as follows. In next section, we discuss what semantic heterogeneity is. After briefing the database perspective, we present our linguistic view on semantic problem in geodata sharing and identify sources and forms of semantic heterogeneity. Followed is an analysis sources and forms of conceptual heterogeneity in section 3, which is the major source of irresolvable semantic heterogeneity. We emphasise the special characters of geographic conceptual modelling. Section 4 is devoted to resolvability of semantic heterogeneity. We point out major irresolvable heterogeneity. Finally, we summarize our study in section 6.

2. MEANING, COMMUNICATION AND SOURCES OF SEMANTIC HETEROGENEITY

The understanding of semantic heterogeneity has evolved much during the past two decades. Semantic heterogeneity is perhaps most studied in the domain of information sharing in general and interoperating database in particular. At early stage, semantic heterogeneity mainly refers to the difference in database models and the alternative ways of implementing a certain conceptual model using a database model, usually a relational model and more recently an OO model.

Such heterogeneity is now called structural semantic heterogeneity by Colomb [1997] and schematic heterogeneity by Bishr [1998] and it is found that the more troublesome problem is what is called fundamental conceptual heterogeneity. According to [Colomb 1997], fundamental semantic heterogeneity occurs when terms in two different ontologies have meanings that are similar, but not quite the same. Furthermore, neither database contains sufficient information to resolve the differences.

Although the database perspective on semantic heterogeneity is still valid and helpful, it also limits our understanding of the problem and limits us in a from-within-database solution to this problem. We will discuss in more detail about this later. In the next subsection, we will compare communication by natural language, human-database interaction and database-database interaction to have a deeper understanding of semantic heterogeneity and its sources and forms.

2.1 A working definition of meaning

As mentioned above, semantic heterogeneity occurs when there is a disagreement about the meaning, interpretation or intended use of data. Therefore, it seems necessary to inspect what meaning is, how meaning is conveyed, the process of communication, and how disagreements upon meaning may happen.

According to the American Heritage Dictionary, meaning is interpreted as "something that is conveyed or signified", "something that one wishes to convey, especially by language" or "an interpreted goal, intent, or end." The above statements are largely from the viewpoint of communication and seem to distinguish three kinds of meaning: the intentional meaning (of the speaker), the explicated meaning (of the representation) and the interpreted meaning (of the audience). But it does not concern the nature of what that *something* is.

It is extremely difficult to give a general definition of that *something* which is being communicated. In the case of data sharing, we take meaning of data (that *something*) as the sum of a conceptualization of the universe of discourse and the facts observed according to this conceptualization. We call this meaning the *sensible meaning*. Our emphasis is to make this meaning independent of the language used to explicate and convey it. The counterpart of sensible meaning is *formal meaning*, which we define as the meaning of data in a certain

* Corresponding author.

language that can be understood and processed by the speakers of that language. When the language is a natural language and the speaker is a human being, formal meaning is just sensible meaning. When the language is some artificial language and the speaker is computer, formal meaning is the part of sensible meaning that is explicated or formalized and can be processed by the computer. One benefit of making the distinction is that we can tell the different between a database system and an information system, i.e. the meaning level of database system is that of that the database language and the meaning level of information system is promoted to sensible meaning due to the participation of human operator.

2.2 A communication model

From language studies [Jeffries 1998], we know that to communicate effectively by natural language we need to *a)* grasp the grammar and a (symbolic) vocabulary of the language, *b)* be equipped with necessary commonsense or domain knowledge to make sense of the vocabulary and *c)* have a correct perception of the context. To illustrate these notions, we give some simple examples. When reading the sentence *"Tom likes coffee"*, we are likely to interpret it as "Someone, named Tom, likes drinking coffer", if we know English to some extent. However, the sentence *"An algebra is a set together with operations defined in the set that obey specified laws"* doesn't make much sense to those who are totally new to algebra. It is not the new words if any that are hard to understand but rather that one lacks the knowledge to make sense of it. Yet, the sentence *"Tom is chasing Jerry"* cannot be unambiguously interpreted without referencing a certain context. Finally, while the statement *"This spatial database is of scale 1:5000"* might be used to mean "the spatial database is created by digitizing 1:5000 paper maps", and it is not a rigorous statement since the concept of scale of spatial database has no well-accepted definition. This shows domain knowledge needs to be built and shared to enable precise exchange of meaning.

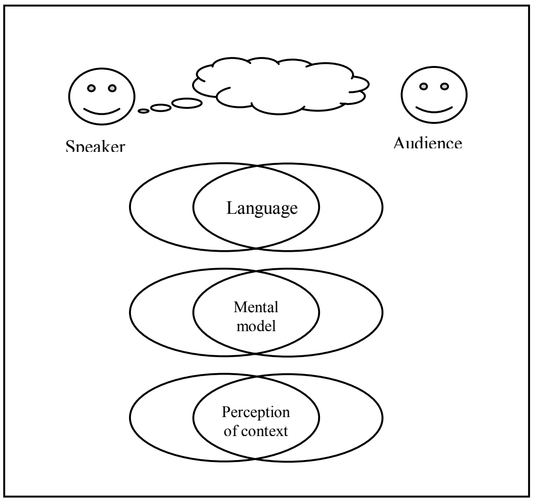


Figure 2**.** A communication model

To identify the sources and forms of semantic heterogeneity, we base further discussion on a communication model as shown in figure 3. In this model, the communication environment consists of a pair of speaker and audience and a conversation context. Both the speaker and the audience have their own mental models about the topic and may speak different languages. They may also have different perceptions of the context. A speaker has an opinion (the intentional meaning) and explicates it by first representing it using his language and then issuing a piece of utterance. The audience receives the utterance and tries to understand its meaning if he knows the language. First, the utterance is parsed according to the language grammar to obtain a structure of the language items. Then, the audience applies his/her vocabulary and perception of the context to interpret the utterance. During the process of interpretation, the audience may need to do some disambiguation on words, the structure of the utterance and the context. Finally, the audience applies his/her conceptualization to make sense of the interpretation to obtain the meaning. It should be noted that the audience should understanding the meaning of the speaker but does not necessarily agree with it. The disambiguation process may repeat sever times until the audience believes the "right" meaning is obtained.

Interactive information processing employing a database system is comparable to communication using a natural language. When we issue a query against the database, we get an answer, which is assumed to be semantically correct and complete. We now examine how meaning is exchanged between the user and the computer system and how semantic heterogeneity is avoided in the human-computer interaction, in particular human-database interaction. In this interaction, the common language used is the database language, which is free of grammatical ambiguity. Users need to understand the database language. Documents have to be provided to users for them to operate the views/tables in a meaningful way. All terms are defined including table names and filed names, i.e. the vocabulary is fixed. The database system can understand the programs and process data according to its own language and its own "mental model", i.e. the data schema. To the computer, data has only formal meaning in the sense of the database language and no intentional meaning in the sense of the specific application. In summary, in this interaction the user adapts the database language in the communication. The user must stick to the vocabulary and data schema, of which any change or deviation will produce errors unless the program is informed and/or updated. In fact, the computer does not care if the audience/speaker is a human user or a remote computer as in the case of homogeneous distributed database system. It should be stressed that the human operator is in charge of promoting the formal meaning of data to sensible meaning that makes sense to the application or decision-making.

	human-database interaction	Database-database communication
Speaker/ Audience	*Human operator*	*Database system*
Audience/ Speaker	*Database system*	*Database system*
Common Language	*Database language*	*A common database language*
Common Context	*That of the database system*	*Compromised context*
Common mental model	*Data schema*	*Integrated schema*
Common vocabulary	*Name of tables, fields, views etc.*	*Reconciled vocabulary*
Common Meaning level	*Formal meaning of database language*	*Formal meaning of database language*

Table 1. Communication setting of human-database interaction and a possible communication setting (in federated database system) for successful database-database communication.

Communication would fail if a common setting of the language, the context, the mental model can't be reached. Misunderstanding may occur if there is any disagreement upon the communication setting but not identified. In general, it's easy for human being to reach a common communication setting and intentionally change to another setting, given our intelligent capability while it is difficult for computer systems to reach a common communication setting. A possible communication setting for successful database interoperation is shown in table 1.

2.3 Sources of semantic heterogeneity

Based on the above discussions, we can identify sources of semantic heterogeneity of geodata as shown in figure 2, which include formalization difference, conceptualisation difference and difference in assumed/perceived context. By formalization heterogeneity, we mean the difference in formalizing a conceptualisation using data languages. It results from two factors. One factor is that we can use different data languages to formally represent a conceptualisation and these different languages may be of different semantic richness. For representing and processing geographic information, many GIS software packages with varying sophistication have been developed, of which all have their own data languages. There are also many data languages designed for exchanging geodata. The languages employ different spatial data model and language syntax. The spatial data models of some geodata languages are quite complex, providing plenty of geometric and topological data types while some are relatively simple. The other factor is that there are alternative ways to formalize a conceptualisation using one language or different languages. [Podczasy 1994] gives a comparison of some spatial data models and [Lee 1990] provides a framework for comparing syntax of different spatial data formats. For relational data model, [Kim and Seo 1991] gives a systematic analysis of representation difference.

Semantic heterogeneity
|→ *Formalization heterogeneity*
| |→ *Language heterogeneity*
| |→ *Representation difference*
|→ *Conceptualisation heterogeneity*
|→ *Context heterogeneity*

Figure 2. A classification of semantic heterogeneity

Conceptualisation heterogeneity is distinguished from formalization heterogeneity in the sense that it is the result of fundamental differences in modelling the real world. It usually means that the data is not readily suitable for a certain application. [Garcia-Solaco *etc.*1996] gave a systematic analysis of conceptualisation difference, in which O-O model is employed as the conceptual model. Regarding difference in conceptual spatial modelling, there have also been some studies. [OGC 1998] gave six cases of semantic relationship between terms defined in feature catalogues of different geodata standards. [Bishr 1998] discussed semantic heterogeneity basing on a language analogy. [Xu etc. 2000] applied the method of [Garcia-Solaco *etc.*1996] to analyse semantic heterogeneity of geodata. These studies helped us understand semantic heterogeneity of geodata. However, a common limitation of these studies is that geographic conceptual modelling is considered as an ordinary database modelling and the special problems of spatial modelling were ignored.

The third source of semantic heterogeneity is related to the context of geodata. In linguistic studies, context refers to the words around a word, phrase, statement, etc. often used to help explain (fix) the meaning. The more general meaning of context refers to the general conditions (circumstances) in which an event, action etc. occurs [Akman and Surav 1996]. For databases, [Sciore etc. 1994] used context to refer to the (implicit) assumptions made when an interoperating agent routinely represents or interprets data. When different agents try to cooperate, there often exists context heterogeneity that need to be resolved. For geodata, unit, reference system, and projection used when creating geodata are among the most significant contextual factors.

All the three kinds of semantic heterogeneity need to be tackled to achieve meaningful geodata sharing. In the rest of this article, we are particularly concerned with conceptualisation heterogeneity because it is relatively less studied and fundamental differences in conceptualisation may render geodata not sharable. We will examine conceptualisation difference in the course of geographic conceptual modelling in detail in the next section.

3. CONCEPTUALISATION HETEROGENEITY

In order to inspect conceptualization difference, we need to base further discussions on a meta model that is used for geographic modeling. In particular, we need to identify the overall idea used in geographic modeling and those modeling constructs. We base our discussion of geographic conceptual modelling on the well-accepted field-based model and feature-based model and for the latter, the Object-Oriented Model is taken as the conceptual formalism.

3.1 The field model

Space can be observed (i.e. applying a *function* to) from the viewpoint of a certain *Spatial Property*, which can be an *attribute* or *operation*. Spatial property can be natural, such as height, or artificial, such as "are there roads?" This observation happens in a certain spatial and property *resolution*. To be formal, this observation needs to be represented. Therefore, we need first to design a structure to sample the space. The *sampling structure* can be irregular like a TIN, or regular like a grid. The function need also be represented. If the function takes value of either *true* or *false*, this function is called an *assertion* about the property at a position. If the function takes continuous value in the sense of point set topology, it is called a *quantification* of property. If the function cannot be appropriately quantified, we may make it *quantified* with some fuzziness. In cases of quantification, the measurement scheme of the function is also associated with a resolution, which we call *thematic resolution* of the function. This is in fact the field-based model as known in GIS field.

3.2 The object model

Besides field-based modeling, feature-based modeling is the other essential way of spatial modeling, in which we see individual objects, i.e. geographic features. To form an identifiable object, we need not only a spatial property but also an identifying criterion. The process of identification is basically cognitive and the formal result will create an object and associate it with an identifier, which is often of sensible meaning that enables us to reference a geographic object without explicitly referencing its geometry. The criterion of identification is often complicated and involves a set of asserting rules. Individual objects are grouped into feature classes according to similarity in the criterion of identification.

As mentioned above, the Object Oriented Model is a well accepted and widely used meta model for object-based modeling, in which classification, specialization/generalization and aggregation are basic modeling methods. Here we are particularly interested in the special character of specialization and generalization when they are applied to geographic objects.

3.2.1 Spatial specialization and non-spatial specialization

After we have identified geographic features, we can attach other properties to them. Every time we introduce a new property to a class of objects, we are in fact making a further classification, i.e. specialization. For example, when adding *gender* property to *person* class, *person* is categorized into *male* and *female*. Here we make a distinction between spatial specialization and non-spatial specialization. If the criterion of specialization is spatial, it's called a spatial specialization. Otherwise, it is called an non-spatial specialization. The point is that a spatial specialization creates new spatial objects by dividing one object of superclass into component objects of subclasses while a non-spatial specialization creates new classes without creating new objects. For example, the specialization of *school* (a geographic feature class) into primary *school/secondary school/college etc*. is a non-spatial specialization. The specialization of *road* into *pedestrian path/carriageway/carriageway with walkway etc*. is a spatial specialization because this specialization creates new spatial objects, i.e. the starting and ending positions of which are generally not coincident with those of original *roads* identified using other criterion.

Therefore, the big difference between spatial and non-spatial specialization is that spatial specialization is accompanied by a process of feature identification. Further, only if the object identification of the superclass is respected, i.e. the spatial specialization is done by fragmenting objects of superclass into component objects, can we derive objects of superclass from objects of subclasses. For example, suppose we have a class called *named road* and then we introduce the spatial property of *usage* which takes value of *pedestrian path/carriageway/carriageway with walkway* and get a subclass called *road-usage-segment*, only when the secondary object identification process is applied to individual *named roads*, can we derive *named road* from *road-usage-segment*.

3.2.2 Spatial aggregation and non-spatial aggregation

The criterion for specialization is applied to a class of objects. Aggregation is different. Conceptually, aggregation is to group some objects according to some criteria. The objects grouped may or may not belong to the same class. Here we first distinguish non-spatial aggregation and spatial aggregation, of which the latter create new spatial objects while the former does not. Then, we distinguish spatial-dominant aggregation and thematic-dominant aggregation according to criterion of aggregation.

Non-spatial aggregation:
Although aggregation of geographic objects will result in new geographic objects, however, the location of the aggregates may not be of concern for a certain application. For example, a university may have several campuses in different locations. Generally, campus is a geographic object, so does a university. However, the location of the university might be not significant in some applications and in database the spatial object represent university is not explicated.

Spatial aggregation
A spatial aggregation is one that creates new spatial objects. For example, the buildings together with other accessory facilities might be aggregated and considered as a new feature, a campus, and the spatial object representing it is of significance. In this example, the criterion of aggregation is non-spatial. In fact, the aggregation is to group a set of features that functions as a whole to serve as a campus. Such an aggregation is called *thematic (spatial) aggregation* here. Some aggregations are based on spatial distribution pattern. For example, the buildings enclosed by major streets can be considered as a street block and an area with a high density of building and roads can be considered as a so-called built-up area. Such aggregations are called *pattern-based (spatial) aggregation*.

The importance of distinguishing pattern-based aggregation and thematic aggregation is that in the case of pattern-based aggregation, the processing can in theory be automated by developing spatial pattern recognition techniques while a thematic aggregation can not be automated because the criteria for aggregation is not derivable from spatial pattern. However, there is no distinct separation of thematic and pattern-based aggregation. Some thematic aggregation can be considered as pattern-based aggregation to some extent because features are related to nearby features more than faraway features. [Devogele *etc.* 1998] gave an example of spatial aggregation. They discussed the problem from the perspective of data integration and suggested to use location as an identifier to match one node to a set of *waysections*. Their basic idea is that the correspondence between the traffic circle node and a set of *waysection* arcs can be found by first forming a proper buffer of the node and then find out those *waysection* arcs falling into the buffer (see figure 3). However, it should be noted that such matching is not applicable if the case is spatial data generalization rather than data integration, in which we don't have a start point to do matching. In theory, we can apply pattern recognition technique to detect a traffic circle and do the generalization. But in practice, it is rather difficult and requires much cognitive intelligence.

Both spatial specialization / generalization and aggregation/segregation may be done recursively and thus form classification hierarchy and aggregation hierarchy. What we emphasize here is that the modelling constructs of specialization and aggregation are special to geographic features. Spatial specialization requires an object identification procedure and it is preferable that sub-identification process is applied to features of

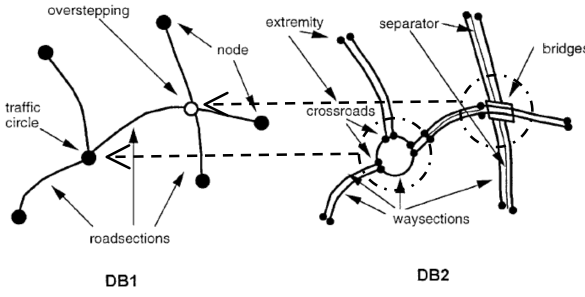


Figure 3 Thematic Spatial Aggregation (adapted from [Devogele *etc.* 1998])

superclass in order to allow the formation of a spatial classification hierarchy and allow spatial generalization be computable. In order that pattern-based spatial aggregation can be automated it is preferable that the ISPARTOF relation between component features and aggregate features be explicated. For thematic spatial aggregation, this explicitness is a necessity for automating the aggregation process.

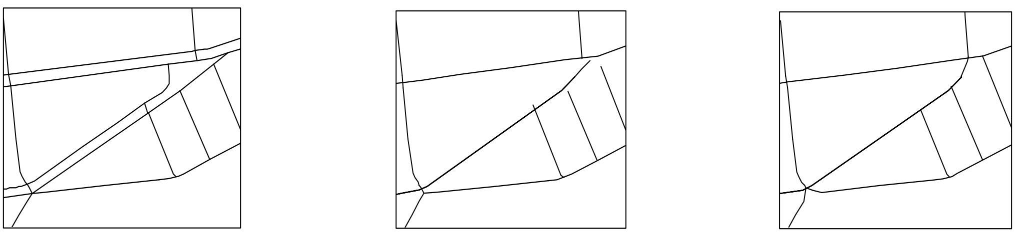
c) Spatial relationship and non-spatial relationship
In object-based modelling, relationship between objects is an important part of conceptualisation. GIS tool is well known for its capability of computing spatial relationship between features. However, attention should be paid to some non-spatial relationships that look like spatial relationships. For example, the traffic connectivity of road segments often deviates from physical connectivity of road segments. Also the *building-along-road* relationship is not fully computable only from position of buildings and roads. In most cases the *Road* to which a *Building* logically "belongs", will be the closest. However, this will not always be the case and one *Building* may belong to two or more different *Road Elements* if the *Building* has several entrances located at different *Road* [CEN TC 278, 1995]. Therefore, whether or not a relationship is a computable spatial relationship largely depends on whether additional information is available.

3.2.3 Representation dimension and semantic constraints

Although geographic features always occupy some portion of a 3-dimensional space, conceptually they are considered as point, line or area features in most cases to achieve computational efficiency. Different applications may represent the same class of features differently. Although it is possible to geometrically derive a lower dimensional representation from a high dimensional representation, result of such geometric generalization may not make sense for a certain application.

We have discussed the criterion of spatial aggregation above and we now discuss the derivation of spatial objects representing the aggregate feature (aggregate spatial object, in short) from spatial objects representing component features (component spatial objects in short). Generally, the aggregate spatial object can only be derived if the component objects cover the whole aggregate object. The above specialization of *road and road-usage-element* example is such a case. The derivation from lower order administrative area into higher order of administrative area is another example. However, in many cases, the component objects do not cover whole area of the aggregate object. For example, the represented road elements constituting a traffic circle do not cover the area of the traffic circle and the represented buildings and facilities do not cover the area of a building complex, such as a school. In these non-covering cases, there exists the problem of how to derive a suitable aggregate spatial object from its components.

What's more, features of the same class or features of different classes are often assumed to conform to some semantic constraints, especially topological constraints. Many spatial analyses rely on such assumptions. Semantic constraints are application dependent. For example, in order to analyse the accessibility of road-building, the application may require that buildings accessible from a certain road be represented as points lie on the road line or polygon have part of road lines as its boundary to facilitate evaluating accessibility by examining topological relation between geometric objects representing them. Some semantic constraint is somewhat betraying the real world facts such as the one mentioned above. While some semantic constraint respects real world facts. For example, the constraint that road lines should intersect or meet if the road they stand for intersect or meet. In such cases, often a spatial data model support topology is preferred. It should be stressed that even though the spatial semantic constraint respects real world fact and such constraint might be maintained by applying some spatial data model supporting such constraint, it can be generated and maintained only when the geometric objects representing them have enough high accuracy. Otherwise, the computer can't generate false spatial relationship due to poor accuracy. Such an observation although is widely known but worth stressing in cases in which the geometric objects are obtained by such kind of map generalization process, especially when it is an automated processing. It is often that we are able to get a generalized representation but it is hard for us evaluate its accuracy. Figure 4 illustrates the problem.



(a) (b) (c)

Figure 4. Topology can not be correctly built without accurate estimation of accuracy of representational generalization.
(a) the source representation of road network, (b) the generalized represented of (a), (c) the desired representation

when we have no accurate estimation of the accuracy of map generalization, we can't use topological cleaning and building process to obtain and maintain the connectivity constraint on road lines. If it is case, connectivity information should be incorporated from the original data if available, manually or automatically. In the example given above, the semantic constraint can be derived. However, it is not always the case because the representation dimension may change or the semantic constraint has changed.

4. RESOLVABILITY

The problem of geodata sharing can be considered as that of deriving an intended representation of real world from an available representation. For data created using field-based modelling, the problem of conversion among various sampling structures and deriving lower resolution representations from that of higher resolution has been widely studied. Generally, the conversion and derivation are possible but sometimes it is quite difficult and different methods are needed to deal with fields of different properties to produce useful information. We would not elaborate on this problem further. In fact, the less studied problems are those in feature-based modeling.

When sharing data created using feature-based modelling, the derivation of intended data from source data is in fact a problem of database transformation, which can generally be resolved by a series of (possibly complex and subtle) view definitions [Colomb 1996] or schema transformation [Xu and Lee 2000].[Xu etc. 2000] and [Devogele etc.1998] proposed candidate schema mapping languages for geodata translation and geodata integration respectively. For detailed discussion of this approach, please refer to foregoing literatures. However, the prerequisite of schema transformation approach is that the information needed in the target database is contained in the source database. Based on the discussion of featured-based modelling, it can be observed that semantically poor geodata set render the data not sharable. The sources of poor geodata set include:

- Non-general definition of feature classes
- Few thematic properties available
- Low thematic resolution
- Poor strategy of spatial object identification in spatial specialization
- Implicit object identification criterion in thematic aggregation
- Few relationships between features available

In fact, we have some examples of inappropriately identified road boundaries, semantically poor database of buildings and poorly modelled road network database, which can not be shown here due to limited space.

When some information is absent, such as the identifying property for thematic aggregation, thematic information need to be computed from spatial pattern of the features and human interpretation is often a necessity. Further, the computed or interpreted information need be confirmed and checked by field survey. When all necessary information for doing spatial database transformation are present, specialized techniques are still needed to deal with spatial objects derivation and semantic constraint maintenance when there are reduction in abstraction level. These issues are addressed traditionally under the title of map generalization and more recently spatial database generalization.

5. CONCLUSION

Geodata sharing is highly desirable because geographic information is public domain information which is used extremely wide and very costly to collect. In this paper, we inspect the greatest barrier to geodata sharing, the semantic heterogeneity. Our emphasis is the special characters of spatial conceptual modelling. We have shown that spatial specialization and aggregation create new spatial objects, difference in representation dimension needs representational generalization of spatial objects while at the same time relationship between generalized features need to be derived from component features. It is found that semantic poorness is the major source of semantic heterogeneity, which includes narrowly defined feature classes when creating data, few feature properties and poor thematic resolution. Another significant problem is geodata is poorly modelled. This includes using unmeaning identifiers for spatial objects, ignoring the specialization and aggregation hierarchy of feature classes. This can bring great trouble for spatial database generalization, which heavily rely on information that identify ISA and ISPARTOF relationships. We believe the cause of semantic poorness and poor conceptual modelling is that when collecting geodata we have borne in mind only traditional application of geographic information, i.e. map production, and largely ignored to take into account of GIS-based applications. It should also be stressed that spatial database generalization is a critical technique in geodata sharing as abstraction level is a very common semantic heterogeneity when geodata is used for unintended purposes.

ACKNOWLEDGEMENTS Zhu Xu thanks The Hong Kong Polytechnic University for funding his Ph.D. study.

REFERENCES

Akman V. and Surav M., 1996, Steps Toward Formalizing Context, AI Magazine 17(3):55-72.

Bishr, Y.A., 1998, *Overcoming the semantic and other barriers to GIS interoperability*, Int. J. Geographic Information Science, Vol. 12, No.4, pp299-314.

Burrough P.A. and McDonnell R.A., 1998, Principles of Geographic Information Systems, Oxford University Press, 333p.

CEN TC 278, 1995, Geographic Data Files, Version: 3.0 FIRST DRAFT.

Colomb, R. M., 1997, Impact of Semantic Heterogeneity on Federating Databases, THE COMPUTER JOURNAL, Vol. 40, No. 5, p235-244

Devogele T., Parent C. and Spaccapietra S., 1998, On spatial database integration, , IJGIS vol.12, no.4, p335-352.

Garcia-Solaco, M, Saltor, F. and Castellanos, M., 1996, *Semantic heterogeneity in multidatabase systems*, In Object-oriented Multidatabase System---A Solution for Advanced Applications (Bukhres, O.A. and Elmagarmid, A.K., Eds.), Prentice-Hall Inc., pp129-202

Jeffries L., 1998, Meaning in English: an introduction to language study, MACMILLAN PRESS LTD, 267p.

Kim, W. and Seo, J., 1991, *Classifying schematic and data heterogeneity in multidatabase systems*, IEEE Computer, 24:12, pp. 12-18.

Lee, Y.C., 1990, A Framework for Evaluating Interchange Standards, CISM Journal Vol. 44 No. 4 pp391-402.

OGC, 1998, *The OpenGIS Guide* (Third Edition, Eds. Buehler, K. and McKee, L.), http://www.ogis.org

Podczasy K. A., 1994, Spatial Data Exchange Standards Harmonization for the Defense Mapping Agency, GIS/LIS'94 Proceedings, Phoenix, Arizona, USA, 25-27 Oct. 1994, p629-638.

Sciore E., Siegel M. and Rosenthal A., 1994, Using semantic values to facilitate interoperability among heterogeneous information systems, ACM Transaction on Database Systems, Vol. 19 No.2 , June 1994, Page 254-290

Sheth, A. P. and Larsen, J., 1990, Federated database systems for managing distributed, heterogeneous and autonomous databases. *ACM Comput. Surveys*, **22**, 183.236 (special issue on heterogeneous databases).

Xu Z., Lee Y.C. and Chen Y.Q., 2000, Schema transformation for semantic geodata translation, International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B6. Amsterdam 2000, p195-202