Huerta, Schade, Granell (Eds): Connecting a Digital Europe through Location and Place. Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Castellón, June, 3-6, 2014. ISBN: 978-90-816960-4-3

Towards Spatio-temporal Data Modeling of Geo-tagged Shipping Information

Amin Mobasheri *
GIScience research group, Heidelberg university
Berliner str.48, 69120
and
Cluster of Excellence Asia & Europe in a Global Context, Voss str. 2, 69115
Heidelberg, Germany
Amin.mobasheri@geog.uni-heidelberg.de

Mohamed Bakillah
GIScience research group, Heidelberg university
Berliner str.48, 69120, Heidelberg, Germany
and
Department of Geomatics Engineering
University of Calgary, Alberta, Canada
Mohamed.Bakillah@geog.uni-heidelberg.de

Abstract

Spatio-temporal data models deal with capturing information characterized by both spatial and temporal semantics. In this paper we review current approaches for spatio-temporal data modelling and present out initial results for selecting the most relevant approach: Object-Oriented modeling for means of modeling geo-tagged shipping information. The shipping information is provided by the well-known LLLOYD's lists dataset. We have introduced the case study and dataset characteristics used in the research project and presented our data model in Unified Modeling Language (UML). The model focuses on spatio-temporal events where characteristics are categorized as thematic, spatial and temporal attributes. The paper follows up with discussion on the selected approach and results, and finally ends with presenting the future outlook.

Keywords: Spatio-temporal, data modeling, Lloyd's lists

1 Introduction

Spatio-temporal data models deal with capturing information characterized by both spatial and temporal semantics. Research in this area started decades ago when collection and management of data, related to both spatial and temporal changes was recognized as an essential task. Before that, earlier works in this area began separately in spatial [1] and temporal [2] data models. These efforts later became the basis for generation of data models which can handle spatio-temporal context as a whole.

Nowadays, several data models have been presented regarding research and practice on spatio-temporal modelling. Each model has its own approach for dealing with spatio-temporal data and facing the challenges concerned with the integration of spatial and temporal concepts. Therefore, each model would be useful to be applied for a specific application (depending on the requirements of application as well as data characteristics) and not for another.

In this research, we aim to study the available spatio-temporal data models and select the best candidate for designing a spatio-temporal data model for geo-tagged shipping information. In the next section we provide the basic definitions and provide a brief review of the literature regarding spatio-temporal data models. In section 3, our case study and data is presented and based on discussion and interpretations, the relevant approach for modelling this data is selected. This section also provides our initial results of applying the approach, and modelling our data. Finally, in section 5, the paper ends with providing the discussion, conclusion and our plan for future work.

2 Spatio-temporal data modeling

Spatio-temporal data models define object data types, relationships, rules and constraints that maintain the integrity of a database [8]. A well-defined data model must anticipate spatio-temporal queries, geo-analytical and geo-statistical methods to be performed in a Geographical Information System (GIS). Their purpose is to deal with real world applications where spatial changes occur over time. The following four categories define the main requirements that need to be addressed by a spatio-temporal data model [8]:

- **Spatial semantics:** this category contains several criterions dealing with pure spatial aspects such as *structure of space*, *orientation/direction*, *measurement* and *topology*.
- **Temporal semantics:** this category deals with the nature of time and the basic features which are used to describe it such as *granularity*, *time density*, *time order*, *transaction*, etc. The issue of whether time should be modelled as continuous elements or as discrete elements are covered with these criterions.
- **Spatio-temporal semantics:** this category contains the most important and challenging criterions that do not exists in single spatial or temporal data models. *Data types*, *primitive notions*, *type of change*, *evolution in time and space*, *space-time topology*, *object identities*, and *dimensionality* are the seven factors/criterions which for instance, makes the data model able of capturing changes in shape and size of the object features and/or whether

* Corresponding author: Amin Mobasheri, E-mail: amin.mobasheri@geog.uni-heidelberg.de

a model supports spatio-temporal real world objects that change continuously or just objects that are subject to discrete changes.

- **Query capabilities:** this category is devoted to classification of existing spatio-temporal data models based on their query capabilities. Such queries may contain:
    - Queries about locations, spatial properties, and spatial relationships
    - Queries about time, temporal properties, and temporal relationships, and
    - Queries about spatio-temporal behaviours and relationships

However, the above queries form a minimum functionality that a spatio-temporal system should provide [3].

Several models exist for spatio-temporal data modelling. These models range from simple approaches such as The snapshot model [10], Simple Time-stamping [5], and Space-Time Composite (STC) [11] to more sophisticated models such as Entity-Relationship (STER) model [12], and Spatio-temporal Object-Oriented data models [13, 14].

The Snapshot model is one of the simplest spatio-temporal data models where temporal information has been incorporated into the model by time-stamping layers [10]. The model considers every layer as collection of temporally homogeneous units of one theme and shows the states of a geographic distribution at different times without considering temporal relations among layers, explicitly. This is the simplest way to capture spatio-temporal information and therefore has limitation in means of being capable to support complex queries.

Another approach is to tag every object with a pair of timestamps. One tag could be for the time of creation and the other for the time of cessation. This approach is also known as data models based on simple time-stamping. Previous implementation of such an approach shows that time slices can easily be retrieved by simple queries [5]. The authors argue that storing layers of geo-information for different time periods is impractical. Therefore, they develop a system that keeps a graphic file of current parcels for day to day use, but archive historical spatial data into a separate file and keep the references alive.

In addition to these approaches, there are other data models that follow the conceptual database models. The Spatio-Temporal Entity-Relationship (STER) model [6] is one of the earliest and best known models among them. The STER model is able to deal with complex geo-entity sets and interrelations of spatial and temporal semantics allows description of attributes and relationships among entity sets. The model is universal in terms of reusability because of its flexible notation [7, 8].

As another approach, spatio-temporal Object-Oriented data models incorporate the features of object oriented technology such as classes, instances, abstract data types, encapsulation, aggregation, inheritance, polymorphism, etc. In a research study, Wachowicz and Healy [9] present an object-oriented spatio-temporal model of real-world phenomena and events. The phenomena are represented as complex versioned objects with geometric, topological and thematic properties. In this approach, for every version of the object which establishes a hierarchical structure for the past, present and future of the object, a new instance of the object with a different identifier is created. In addition, "events are manifestations of actions, which invoke update procedures on one or more objects" [8]. In this approach, time is represented as an independent, linear dimension. The time reference is absolute and the time order is linear. Last but not least, space is conceptualised in three linear dimensions [9].

In practice, there is no "complete" model of any kind for any application domain. The decision for selecting the relevant approach of modelling should be made based on the requirements for the specific application as well as the characteristics of information aimed to be modelled. Therefore, in the next section we present our case study and thorough discussion select the most relevant approach to be applied for spatio-temporal modelling of geo-tagged shipping information in our research.

3 Case study: The LLOYD's Lists

Since the late seventeenth century, the shipping newspaper Lloyd's List and its direct predecessors contain weekly and later daily information on global shipping. The core of the Lists' mostly tabular contents is formed by the categories *"Shipping Intelligence"*, *"Speakings"*, *"Foreign Mail"*, *"Casualities"*, and *"War"*. Specifically, the first two categories are essential in our research. The "*Shipping Intelligence*" consists of exhaustive lists of the arrivals, departures and other nautical activities of civilian ships in practically all important ports of the world. The *"Speakings"* list sightings of ships at the high seas and give both the sighted and the reporting ship with name and geographical coordinates.

The *"Speakings"* and *"Intelligence"* hold much information that with a primarily quantitative approach, will allow us to analyze, for instance, the shifting patterns of shipping routes; the time it took to get information, goods, and humans from one harbor to another depending on the year and season; the "black spots" and interruptions of service due to natural disasters or wars; the shifts in trade intensity between specific regions; the constantly changing patterns of transcontinental/international trade and migration; or, in short, the transformation of a variety of "global spaces" during times of rapid globalization [15].

As an initial step, spatio-temporal modeling of Lloyd's data is essential in order to capture the information in a coherent and flexible manner which would later allow spatio-temporal analysis and querying of this information. The *"Speakings"* information contains latitudes and longitudes that could be best captured by a point feature as well as two timestamps. The first timestamp is for the date that the actual *speaking* has occurred and the other timestamp contains the actual date when this information has been reported.

In order to deal with modeling of Lloyd's data, some approaches cannot be used due to their specific drawbacks. For example, the Snapshot model is not appropriate for means of describing changes in space thorough time. Each snapshot is relevant to a specific time, but in order to understand how T_i differs from T_j, two snapshots should be compared exhaustively. As an example of disadvantages of the Space-

Time Composite (STC) model, the fact that it is very difficult to define rules of internal logic and/or integrity constraints is a big problem. This is because the model does not provide understanding of the constraints upon the temporal structure [8]. Employing the STER model on the other hand has several benefits but the main problem according to a research study [8], which makes this model un-suitable for our research too, is the fact that it lacks the ability to capture the actual motion of the process of change and does not indicate if a spatial object is dynamic or static.

Among several approaches of spatio-temporal data modeling discussed in the previous section, we select the object oriented modeling because of its four main advantages in spatio-temporal modeling [8, 13]:

- A single object represents the whole history of an entity
- Efficient temporal data handling
- Uniform treatment of spatial and temporal data handling
- Simple queries due to its capability of dealing with each single object of an entity

As an initial attempt towards digitization, a large amount of published information for specific time periods which was of interest for historians; specifically years 1851 and 1871 were read and transferred into Excel spreadsheets in a formal tabular format. In the next step, in order to design a geo-database for the Lloyd's data, a geo-data modeling task using the object oriented modeling was performed. As an extra data source, we will also use the digitized maps of shipping routes downloaded from David Rumsey's map collection [17] and the CLIWOC database [16] which provide weather data (e.g. wind speed and direction, air temperature, etc.). The aim would be to integrate these datasets with the Lloyd's data in order to improve the normal trajectories (digitized from the maps) to a higher abstraction level, leading to semantic trajectories. From all the necessary information that needed to be recorded, a total amount of 9 tables were designed, each of which carry special attributes.

Figure 1 illustrates the data model for each table in a class diagram fashion using Unified Modeling Language (UML). Note that the first four tables are normal tables containing several necessary information (e.g. Ship name, Captain name, Event date, Journal Date, etc.), yet the two last tables are considered as feature types since they have spatial components, thus can be treated as point features. As it can be seen two main tables are *tbl_Speakings* and *tbl_Intelligence* which contain the shipping information. Furthermore, *tbl_Speakings* is the most special table in this model which contains coordinate values in forms of latitude and longitudes where the actual speaking has happened (somewhere in the sea) as well as information about the plan of the other ship (the sighted ship) such as its port of origin and destination. *tbl_weather* captures weather information collected from CLIWOC database [16] for the relevant temporal periods of LLOYD's shipping data. Last but not the least, *tbl_Routes* captures the data model behind shipping routes that are digitized based on famous maps provided by David Rumsey's map collection [17] (note that additional data are stored in *tbl_Route_Metadata*).

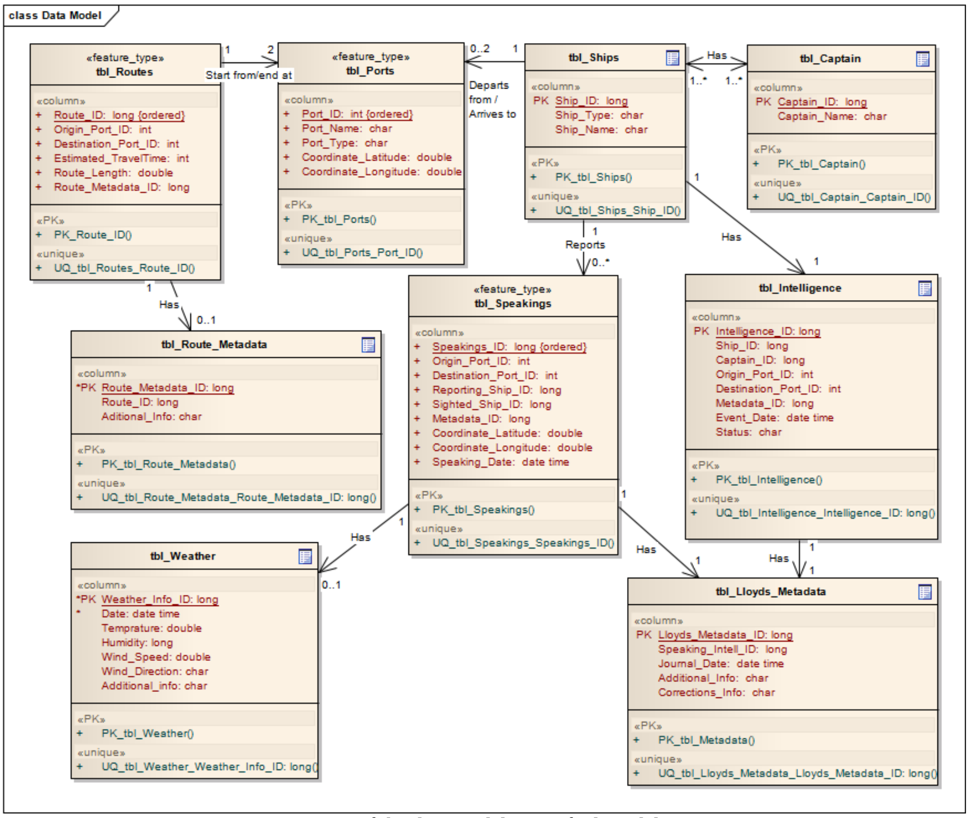

Figure1: Main part of the data model in Unified Modeling Language

The model focuses on spatio-temporal events where characteristics are categorized as thematic, spatial and temporal attributes. In this approach, for every version of the object which establishes a hierarchical structure for the past, present and future of the object, a new instance of the object with a different identifier is created. In addition, “events are manifestations of actions, which invoke update procedures on one or more objects”. In our model, speakings of ships are the special events that occur in the sea and need to be recorded in an efficient and effective manner. In this approach, time is represented as an independent, linear dimension. The time reference is absolute and the time order is linear. Last but not least, space is conceptualized in two linear dimensions.

4 Discussion and Conclusion

In this paper we presented a brief overview of available approaches for spatio-temporal data modeling and elaborated on some of the advantages and disadvantages of them. In the next step, we introduced the data of our case study: Lloyd’s lists, and through a discussion selected the object oriented modeling as the best candidate for spatio-temporal modeling of geo-tagged shipping information. The paper also presents our initial attempt to design our model using Unified Modeling Language (UML). We concluded that our model focuses on spatio-temporal events and for every version of the object which establishes a hierarchical structure for the past, present and future of the object, a new instance of the object with a different identifier is created. *Speakings* of ships are seen as specific events happening in the sea with additional semantic information and temporal attributes that need to modelled effectively and efficiently in order to be ready for spatio-temporal analysis and reasoning tasks.

Furthermore, using simple sample points would not be the only case in this research, we will also use a higher level of abstraction in order to introduce semantic trajectories of data. The reason is that more application-oriented ways of analyzing segments of movement suitable for specific purposes of the application domain is needed. Therefore, in the future, the model provided here would be extended in order to be capable of capturing trajectories (e.g. lines) in addition to point features. Another aim for future work is design and populate a geo-database based on the data model and to review and apply relevant geo-statistical methods for means of analyzing and discovering spatio-temporal patterns of shipping information and detecting hot-spots of shipping *speakings* considering different temporal timestamps.

5 Acknowledgments

This research is made possible by a grant from the Cluster of Excellence “Asia and Europe in a Global Context” as well as GIScience research group of University of Heidelberg. Also, the authors would like to thank the hard work of students in the cluster as well as the department of history (HGIS Club) for extracting the data samples of the Lloyd’s lists. The authors would also like to thank Alexander Zipf, Kilian Schultes and Roland Wenzlhuemer from Heidelberg University for their comments and support.

References

[1] R.H. Guting, “An Introduction to Spatial Database Systems”, VLDB Journal 4 (1994), 357-399.

[2] A.U. Tansel, J. Clifford, S. Gadia, S. Jajodia, A. Segev and R. Snodgrass, “Temporal Databases: Theory, Design and Implementation”, Benjamin/Cummings Publishing Company, 1993.

[3] M. Koubarakis, T. Sellis et al. (eds.), Spatio-Temporal Databases: The Chorochronos Approach, 2003. Springer-Verlag LNCS 2520.

[4] G. Langran, “A Framework for Temporal Geographic Information Systems”, Cartographica, 25 (3), 1988.

[5] G. J. Hunter and I. P. Williamson, “The Development of a Historical Digital Cadastral Database”, Int. Journal of Geographic Information Systems, 4(2), 1990.

[6] N. Tryfona and C. S. Jensen, “Conceptual Data Modeling for Spatiotemporal Applications”, GeoInformatica, Vol. 3: 245-268, 1999.

[7] N. Tryfona and C. S. Jensen, “Using Abstractions for Spatio-Temporal Conceptual Modeling”, Proceedings of the 2000 ACM Symposium on Applied Computing, Como, Italy, 2000.

[8] P. Nikos, B. Theodoulidis, I. Kopanakis, and Y. Theodoridis. "Literature review of spatio-temporal database models." The Knowledge Engineering Review 19, no. 03 (2004): 235-274.

[9] M. Wachowicz and R. G. Healey, “Towards Temporality in GIS”, In Worboys M. F. editor, Innovations in GIS, Taylor & Francis, 1994.

[10] G. Langran, “A Framework for Temporal Geographic Information Systems”, Cartographica, 25 (3), 1988.

[11] G. Langran, “Time in Geographical Information Systems”, ed. Taylor & Francis, London, 1992.

[12] I. Theodoulidis and P. Loucopoulos, “The Time Dimension in Conceptual Modeling”, Information Systems, vol. 16, no. 3, pp. 273-300, 1991.

[13] M. F. Worboys, H. M. Hearhshow, D. J. Maguire, "Object-Oriented Modeling for Spatial Databases", Int. Journal of GIS ol. 4, No. 4, 1990.

[14] M. Wachowicz and R. G. Healey, "Towards Temporality in GIS", In Worboys M. F. editor, Innovations in GIS, Taylor & Francis, 1994.

[15] A. Mobasheri, M. Bakillah, K. Schultes, A. Zipf. Towards Integrated Web Processing Services for Spatio-temporal Analysis of Geo-data, The Case of Lloyd's Lists. Scientific Computing and Cultural Heritage Conference (SCCH), Heidelberg, Germany, 2013.

[16] Wheeler, D., Garcia-Herrera, R., Koek, F.B., Wilkinson, C., Können, G.P., Prieto, M.R., Jones, P.D. and R. Casale. 2006: CLIWOC, Climatological database for the world's oceans: 1750 to 1850; Results of a research project EVK1-CT-2000-00090. European Commission, Brussels, ISBN 92-894-8279-6.

[17] David Rumsey Map Collection, available at http://www.davidrumsey.com/view, accessed in May 2014.