A RDF Vocabulary for Spatiotemporal Observation Data Sources

Karine Reis Ferreira [1], Diego Benincasa F. C. Almeida[1], Antônio Miguel Vieira Monteiro[1]

[1]DPI – Instituto Nacional de Pesquisas Espaciais (INPE)

{karine, benincasa, miguel}@dpi.inpe.br

Abstract. *Observations are our means to assess spatiotemporal phenomena in the real world. They are basic units for spatiotemporal data representation and are distributed by data providers using different formats and standards. In this work, we propose an approach to discover, access and integrate spatiotemporal observations from different kinds of data sources using RDF framework and SPARQL language. This paper presents an ongoing work on defining a RDF vocabulary for describing spatiotemporal observation data sources.*

1. Introduction

The recent technological advances in geospatial data collection, such as Earth observation and GPS satellites, have created massive data sets with better spatial and temporal resolution than ever. This scenario has motivated a challenge for Geoinformatics. We need geographical information systems (GIS) that can access spatiotemporal data sets from different kinds of data sources and analyze them in an integrated way.

Observations are our means to assess spatiotemporal phenomena in the real world. Although most spatiotemporal phenomena are continuous over time and space, they are often measured through discrete observations. Observations link information to reality and provide the building blocks of conceptualizations [Kuhn 2009]. Recent research draws attention to the importance of using observations as a basis for designing geospatial applications [Kuhn 2005]. Following this trend, we proposed a data model for spatiotemporal data representation grounded on observations [Ferreira et al. 2014] and implemented it in the TerraLib GIS library [Camara et al. 2008]. Taking observations as basic units for spatiotemporal data representation, this work focuses on *how to access and combine observations from different kinds of data sources*.

RDF (Resource Description Framework) is a data model for describing and connecting resources and SPARQL is a query language for RDF data sets. Both are World Wide Web Consortium (W3C) standards and are key techniques in *Linked Data* and *Semantic Web* [Berners-Lee et al. 2001] [Wood et al. 2014]. The term Linked Data refers to a set of best practices for publishing and connecting structured data on the Web using international standards of W3C. Semantic Web provides technologies that allow data to be shared with explicit meaning and processed by machines. Publishing information as linked data is the first step towards the world of semantic web [Herman 2012]. In this work, we propose an approach that uses RDF to describe data sources that provide spatiotemporal observations and SPARQL to discover information about these data sources.

RDF describes resources using the concepts of classes, properties, and values. The term *vocabulary* refers to a set of classes and properties that are defined specifically for a certain application. RDF Schema (RDFS) is a framework to define such vocabularies, that is, to describe application-specific classes and properties. Examples of vocabularies are Dublin Core Vocabulary[1] that defines a set of predefined properties for describing documents and FOAF (Friend Of A Friend) Vocabulary[2] that describe relationships among people. This paper presents an ongoing work on defining a RDF vocabulary for describing spatiotemporal observation data sources.

2. Related Work

In a previous work, we propose an approach to access trajectories of moving objects from different kinds of data sources based on XML metadata files [Ferreira et al. 2013]. Differently, this paper presents a new approach to access spatiotemporal observations from different kinds of data sources, using RDF framework and SPARQL language. The new approach is more comprehensive than the previous one because it is for spatiotemporal observation data sources and not only for trajectory data sources. Besides that, the new approach uses RDF files to describe data sources. It allows the use of SPARQL language on these files to discover information about the data sources.

There are many RDF vocabularies for different application domains, such as Dublin Core for describing documents and FOAF for describing relationships among people. Examples of RDF vocabularies for describing geospatial data are W3C Basic Geo vocabulary, GeoOWL ontology, NeoGeo Vocabulary and GeoSPARQL [Battle and Kolas 2012]. GeoSPARQL is an Open Geospatial Consortium (OGC) standard that defines a vocabulary for representing geospatial data in RDF and an extension to the SPARQL query language for processing geospatial data [OGC 2012a].

RDF framework is the key technique in Linked Data and Semantic Web. Many initiatives and projects focus on transforming geospatial data into linked RDF files, such as the LinkedGeoData and Geonames.org[3] [Stadler et al. 2012] [Janowicz et al. 2012]. The LinkedGeoData project provides a RDF serialization of Points Of Interest from Open Street Map. Geonames.org provides a database that contains over 10 million geographical names and an interface to generate a RDF-dump of this database. In our proposal, we use RDF as linked metadata files, that is, files that describe *how* data sources represent spatiotemporal observations and *links* among these data sources. In this first step, we will not transform the spatiotemporal observations from their original data sources and formats into RDF files.

3. An Observation-Based Model for Spatiotemporal Data

We proposed a data model for spatiotemporal information grounded on observations and specified it using an algebraic formalism [Ferreira et al. 2014]. Algebras describe data types and their operations in a formal way, independently of programming languages. By separating specification from implementation, they help to develop interoperable, reliable and expressive applications [Frank 1999]. The proposed algebra

[1] http://dublincore.org/
[2] http://xmlns.com/foaf/spec/
[3] http://www.geonames.org/

is extensible, defining data types as *building blocks* for other types, as shown in Figure 1.

The proposed model defines three spatiotemporal data types are defined as abstractions built on observations: *time series*, *trajectory*, and *coverage* [Ferreira et al. 2014]. A *time series* represents the variation of a property over time in a fixed location. A *trajectory* represents how locations or boundaries of an object change over time. A *coverage* represents the variation of a property in a spatial extent at a time. We also define an auxiliary type called *coverage series* that represents a time-ordered set of coverages that have the same boundary. Using these types, we can construct *objects* and *events*.

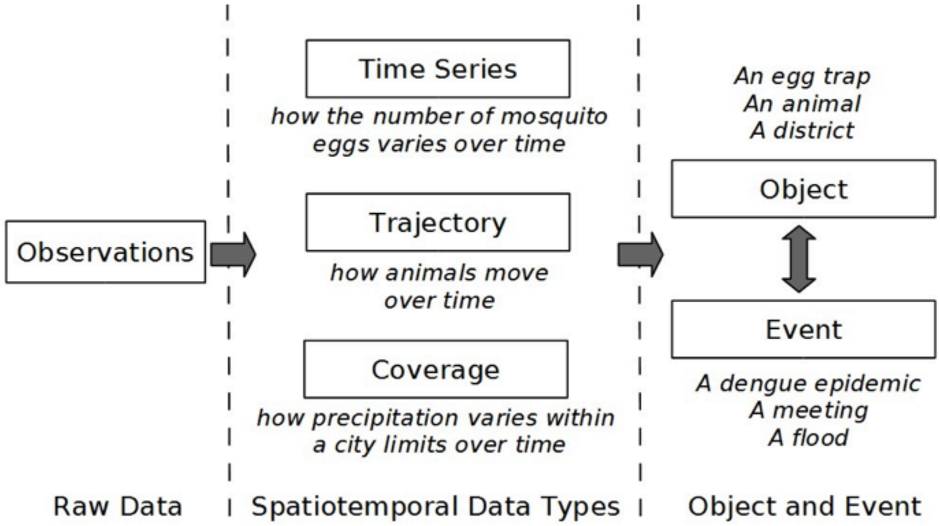

Figure 1. An Algebra for Spatiotemporal Data: From Observations To Events. Source [Ferreira et al. 2014]

Using these three spatiotemporal data types, we can create different views on the same observation set, meeting the application needs. Consider a set of cars equipped with GPS and air pollution sensors. Figure 2 shows tracks of three cars in a city during one day. These cars produce an observation set, where each one contains a car identity, a time instant, a location and an air pollution value. The observations are taken at each hour. From this observation set, we can extract information using different kinds of queries, such as *how the average air pollution varies over time in the city*; *how the cars move over time and space*; and *how pollution varies within the city limits*.

Each application needs different queries and each kind of query is suited to a specific data type. Taking the whole city as a fixed reference, we can get a *time series* that represents the variation of the average air pollution in the city per hour. Considering each car an individual object, we can get a set of *trajectories*. Making the whole day as a time reference and taking all observations at that day, we can create a *coverage* to represent the air pollution variation within the city limits during that day.

Algebraic specifications are language-independent. Programmers can translate them into software using programming languages of their choice. We implemented the proposed algebra in the TerraLib GIS library [Camara et al. 2008]. We developed a new module called "TerraLib ST module" to deal with spatiotemporal information that contains all data types and operations described in the proposed algebra.

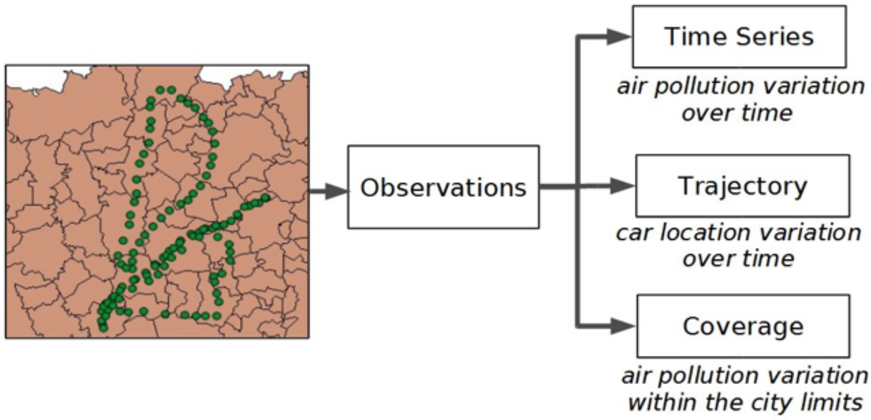

Figure 2. Different views on observations produced by moving cars. Source [Ferreira et al. 2014]

4. The Proposal

Spatiotemporal observations can be stored and disseminated by data providers in different ways. Recently, most data providers throughout the world have organized their geospatial data sets and made them available on the Internet via files, spatial databases and web services, following well-established standards[4] defined by ISO and OGC. Geography Markup Language (GML) and Keyhole Markup Language (KML) are examples of OGC's file formats for spatial data interchange. Spatial extensions of traditional database management systems, such as PostGIS and Oracle Spatial, deal with vector spatial information in compliance with the OGC Simple Feature Access (SFA) specification. Besides that, there are standards for serving spatial data and processes via web services such as Web Feature Service (WFS) and WCS (Web Coverage Service).

OGC proposes a standard called Sensor Observation Service (SOS) that defines a web service interface for disseminating and querying observations, sensor metadata and observed features [OGC 2012b], based on the OGC Observations and Measurements (O&M) specification [OGC 2010]. However, many data providers store and disseminate spatiotemporal information using other formats and standards, not only SOS, as shown in Figure 3. GIS tools must be able to access different types of spatiotemporal data sources, without forcing the use of a specific format or standard.

Therefore, we propose an approach to access spatiotemporal observations from different kinds of data sources using RDF framework and SPARQL language. The central idea is to use RDF files for describing how spatiotemporal observations are represented in data sources and SPARQL language for discovering information about these data sources. Each data source has an associated RDF file and all RDF files are based in the same vocabulary, as presented in Figure 3. We are working on defining and implementing a RDF vocabulary for spatiotemporal observation data sources. The main features of this vocabulary are:

1. It will be based on the observation concept defined in the OGC O&M specification [OGC 2010]. So, it must contain classes and properties to describe

---

[4] http://www.opengeospatial.org/standards/is

how phenomenon time, result time, valid time, and observed attributes are represented in different kinds of data sources.

2. It will use the OGC GeoSPARQL schema [OGC 2012a] to represent spatial information in RDF files. This is important, for example, to encode observation spatial extents in RDF files.
3. It must support link among data sources. Two RDF files that describe data sources can be linked. For example, the observations can be stored in a data source and their spatial extent can be stored in another data source. In this case, the RDF file that describes the first data source must have a link to the RDF file that describes the second one.

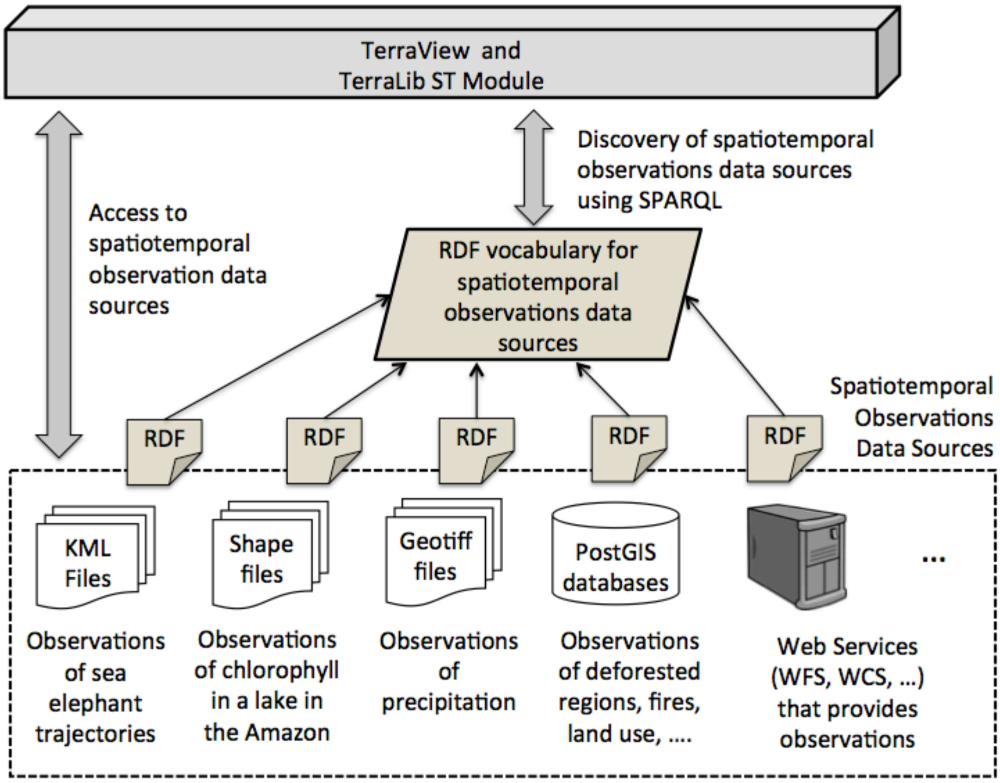

Figure 3. The proposal

The TerraLib ST Module contains the data types and operations defined in the model presented in Section 3. It is responsible for discovering information about the spatiotemporal observation data sources using SPARQL on the RDF files and accessing these observations. After loading observations from data sources, users can create the most suitable views on them, as shown in Figure 2, in the application level. Observations are basic units without strong semantics. TerraView/TerraLib users decide what abstraction, *time series*, *trajectory* or *coverage*, will be created on these observations to fit the application needs.

5. Final Remarks

This paper presents an approach to discover, access and integrate spatiotemporal observations from different kinds of data sources using RDF framework and SPARQL language. This is an ongoing work. In this approach, we are working on defining a RDF vocabulary for describing spatiotemporal observation data sources based on the OGC O&M and GeoSPARQL specifications.

Our proposal considers that data sources provide observations which are basic units for spatiotemporal phenomenon representation. This allows application users to create different views on observations, according to the application needs. This work is crucial in extending TerraLib GIS library and TerraView to deal with spatiotemporal data.

References

Battle, R. and Kolas D. (2012) "Enabling the geospatial semantic web with parliament and GeoSPARQL". *Semantic Web Journal* 3(4).

Berners-Lee, T.; Hendler, J.; and Lassila, O. (2001). The semantic web. *Scientific american*, 284(5), 28-37.

Camara, G.; Vinhas, L.; Queiroz, G. R.; Ferreira, K. R.; Monteiro, A. M. V.; Carvalho, M. T. M.; Casanova, M. A. (2008) "TerraLib: An open-source GIS library for large-scale environmental and sócio-economic applications". *Open Source Approaches to Spatial Data Handling*. Berlin: Springer-Verlag.

Ferreira, K. R.; Camara, G.; Monteiro, A. M. V. (2014) "An algebra for spatiotemporal data: From observations to events". *Transactions in GIS*, 18(2), 253-269.

Ferreira, K. R.; Vinhas, L.; Monteiro, A. M. V.; Camara, G. (2013) "Moving Objects and Spatial Data Sources". *Revista Brasileira de Cartografia*, (64/4).

Frank, A. U. (1999) "One step up the abstraction ladder: Combining algebras – from functional pieces to a whole". In Freksa C and Mark D (eds) *COSIT: Conference on Spatial Information Theory*. Berlin, Springer Lecture Notes in Computer Science, 1661, 95-108

Herman, I. (2012) "Tutorial on Semantic Web". Available at: http://www.w3.org/People/Ivan/CorePresentations/SWTutorial/

Janowicz, K.; Scheider, S.; Pehle, T.; Hart, G. (2012) "Geospatial semantics and linked spatiotemporal data–Past, present, and future". *Semantic Web Journal*, 3(4), 321-332.

Kuhn, W. (2005) "Geospatial Semantics: Why, of What, and How?". *Journal of Data Semantics*, 3(1).

Kuhn, W. (2009) "A functional ontology of observation and measurement". In Janowicz K, Raubal M, and Levashkin S (eds) *International Conference on GeoSpatial Semantics (GeoS 2009)*. Berlin, Springer Lecture Notes in Computer Science, 5892, 26-43

Open Geospatial Consortium – OGC (2012a) "OGC GeoSPARQL - A Geographic Query Language for RDF Data". Available at: http://www.opengeospatial.org/

Open Geospatial Consortium – OGC (2012b) "OGC Sensor Observation Service Interface Standard". Available at: http://www.opengeospatial.org/

Open Geospatial Consortium – OGC. (2010) "Geographic Information: Observations and Measurements". Available at: http://www.opengeospatial.org/

Stadler, C.; Lehmann, J.; Höffner K.; Auer S. (2012) "Linkedgeodata: A core for a web of spatial open data". *Semantic Web Journal* 3(4).

Wood, D.; Zaidman, M.; Ruth, L. (2014) *Linked Data – Structured Data on the Web*. Manning Publications.