Towards a semantic framework for exploiting heterogeneous environmental data

Ba-Huy Tran* and Alain Bouju

L3i Laboratory,
Université de La Rochelle,
La Rochelle, France
Email: ba-huy.tran@univ-lr.fr
Email: alain.bouju@univ-lr.fr
*Corresponding author

Christine Plumejeaud-Perreau

LIENSs Laboratory – UMR 7266,
CNRS & Université de La Rochelle,
La Rochelle, France
Email: christine.plumejeaud-perreau@univ-lr.fr

Vincent Bretagnolle

Centre d'Etudes Biologiques de Chizé – UMR 7372,
CNRS & Université de La Rochelle,
F-79360 Villiers-en-Bois, France
Email: breta@cebc.cnrs.fr

Abstract: This paper presents the 'Environment and landscape geo-knowledge' (GEMINAT) project which aims to build an infrastructure favouring the cross-analysis of spatio-temporal heterogeneous data sources recorded at the Chizé environmental observatory since 1994. From a case study, we summarise the difficulties encountered by biologists and ecologists when maintaining and analysing collected environmental data, essentially the spatial organisation of the landscape, crop rotation, and wildlife data. We show how a framework which uses a spatio-temporal ontology as a semantic mediator can solve challenges related to the analysis and maintenance of these heterogeneous data.

Keywords: data integration; spatio-temporal ontology; environment; ecology.

Reference to this paper should be made as follows: Tran, B-H., Bouju, A., Plumejeaud-Perreau, C. and Bretagnolle, V. (2016) 'Towards a semantic framework for exploiting heterogeneous environmental data', *Int. J. Metadata, Semantics and Ontologies*, Vol. 11, No. 3, pp.191–205.

Biographical notes: Ba-Huy Tran is a PhD student at the University of La Rochelle. He received his MS degree in data mining in 2009 from the University of Nantes, France. His research interests include applied artificial intelligence, applied information systems, data mining, and knowledge discovery.

Alain Bouju received an Engineering degree in Computer Science in 1990, his MS degree in Computer Science and his PhD degree in Artificial Intelligence from National Higher School of Aeronautics and Space, France, and a habilitation to direct research (HDR) from the University of La Rochelle, France, in 2004. He is currently Associate Professor in L3i Laboratory. His current research interests lay in the area of the spatio-temporal data, mobility, trajectory, geographic information system, intelligent transportation systems and smart cities.

Christine Plumejeaud-Perreau received her Master's degree in applied mathematics in 2000, a research master in 2006 at LIG laboratory at Grenoble, France, followed by a PhD Thesis defended in 2011. She got a reward from the geomatics community through the GDR MAGIS price honoring her PhD's work on spatio-temporal databases for making long-term and sustainable information systems for socio-economic statistics. She was recruited by CNRS to work in an interdisciplinary laboratory, mixing ecology, biology, geography, history and geosciences disciplines. Her competencies address the field of computer science and geomatics, as well as the edge of data science analysis through spatio-temporal analysis and statistics.

Vincent Bretagnolle received his PhD at the University of Rennes, France, and his HDR at the University of Dijon, France. He is a senior scientist at CEBC-CNRS, Beauvoir sur Niort, France, head of CEBC-CNRS, of AGRIPOP Team and of the 'Plaine & Val de Sèvre' LTER Workshop Zone. His research interests are related to ecology, zoology, and marine biology.

This paper is a revised and expanded version of a paper entitled 'A semantic mediator for handling heterogeneity of spatio-temporal environment data' presented at the '9th International Conference on Metadata and Semantics Research - MTSR 2015', Manchester, UK, 9–11 September 2015.

1 Introduction

In rural areas with the predominance of agricultural activities, the study of environmental issues such as biodiversity preservation, soil erosion by water and tillage, erosive run-off, water pollution, and gene fluxes may benefit from the long-term analysis of the crop mosaic resulting from farming practices. In fact, agricultural landscapes are primarily designed by farmer decisions dealing with crop choices and crop allocation at the farm scale. The arrangement, shape and nature of crops compose the spatial organisation of the landscape which impacts ecological processes at various scales.

This spatial organisation evolves throughout time, because farmers can change the land use of their parcel (in cereal landscapes of western France for instance, the main rotation includes wheat, rapeseed, and sunflower, but other crops may include alfalfa, corn, hay, peas), as well as its shape (by splitting, merging or redistributing between each other's). This information is relevant for studying the links between socio-economic environment and agricultural practices and the subsequent spatial organisation of the landscape. Recognising the benefits of the long-term observation of agricultural practices for research on environmental issues, the UMR Chizé has established an observatory for crop rotation on the 'Plaine & Val de Sèvre' workshop area. Since 1994, a Geographic Information System for the Environment (GIS-E) has been deployed in order to monitor crop rotation of agricultural parcels (Figure 1).

Since 2013, an effort has been made to integrate semantic technologies within the GIS-E, joining thus the current move of the research towards infrastructures mixing heterogeneous and interdisciplinary data sets such as ones promoted by the Research Object[1] and Data One[2] initiatives, the Research Data Alliance organisation[3] or the ENVRIplus project.[4] Those movements have acknowledged that the use of fully qualified and documented data sets is fundamental for any reproducible research, but also they promote the use of ontologies to accelerate the bindings of new data sets. In the ecological field, various ontologies have been developed to address the need for interoperability between data sets, among which we can cite ENVO[5] (Buttigieg et al., 2013) and OBOE[6] (Madin et al., 2007). However, one should notice that they do not fully address the need for spatio-temporal analysis, hence giving an operational and effective support for spatio-temporal queries is still a challenge.

This paper presents first the local context of this interdisciplinary research along with the challenges related to the cross-analysis of spatio-temporal environmental data. In the next sections, a spatio-temporal ontology as well as a framework are proposed to fulfil the need for spatio-temporal data analysis. Finally, the conclusion summarises the progress achieved with the system while highlighting our future work.

Figure 1 Architecture overview of the GIS-E, with a central server serving content of various databases for analysis to scientists

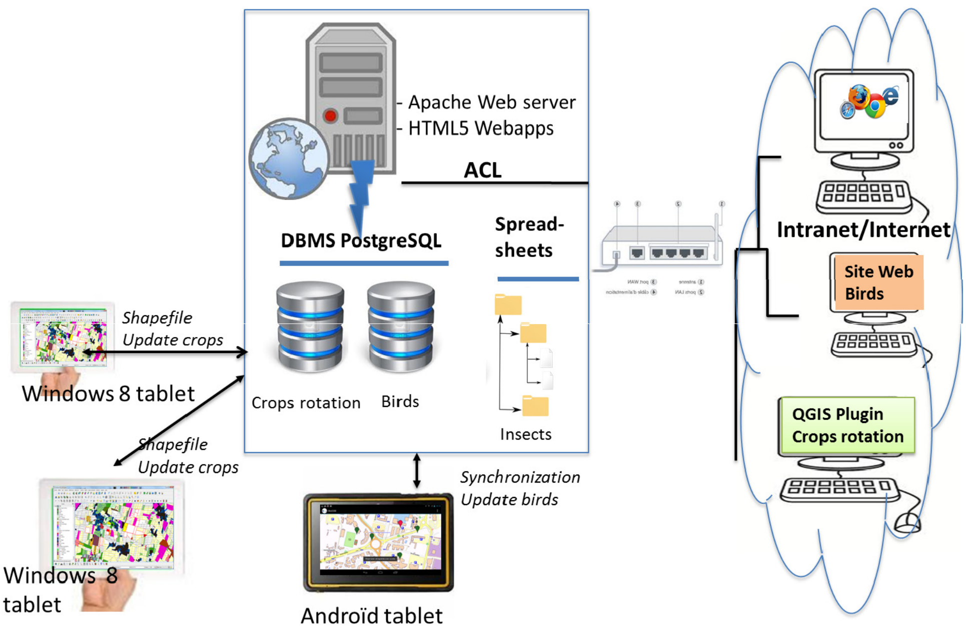
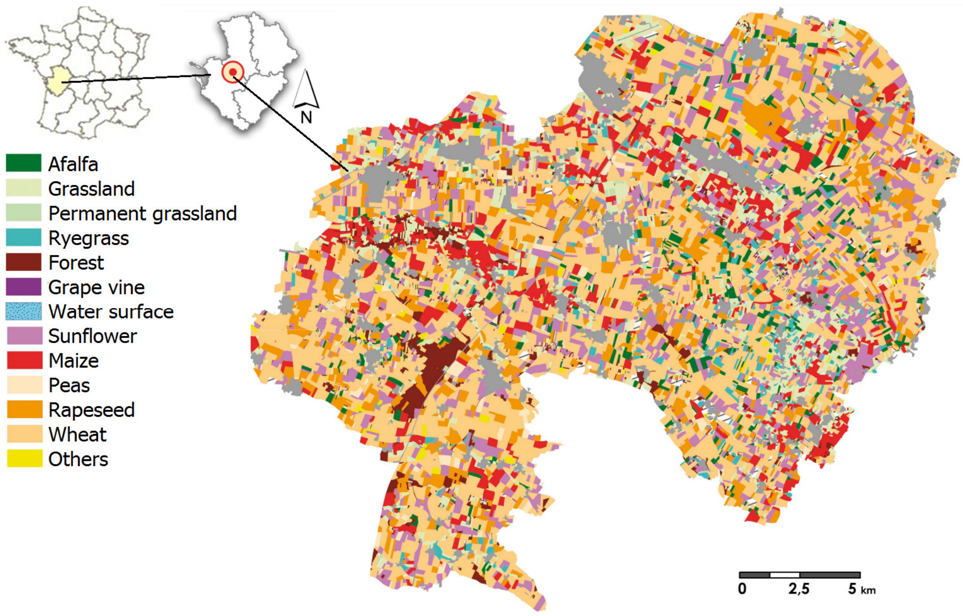


Figure 2 Map of the crop rotation covering the 450 km^2 of 'Zone Atelier Plaine & Val de Sèvre'



1.1 *The observatory*

The need for a long-term observatory of crop rotation and associated biodiversity indicators has been recognised in a research program led by the CNRS[7]) since 1994 called 'Zone Ateliers'[8] (workshop zones). The observation area of 450 km^2 is designated as 'Zone atelier Plaine et Val de Sèvres', located south of Niort, in the department of Deux-Sèvres, France. It is an intensive cereal plain: wheat, corn, sunflower, peas and rapeseed essentially where livestock (cattle, goats) is still present but in steep decline. Agricultural parcels are still medium-sized (4–5 ha), 15% of them are occupied by grasslands (artificial, permanent or temporary). The main research addresses the relation between farmer practices and ecological processes, through the study of the spatial organisation of the landscape. Since 1994, the land use of 19,000 agricultural parcels is recorded from the field each year by AGRIPOP team (CNRS Chizé) and centralised in a database called 'Assolement'. A parcel is defined as a management unit, a polygon surrounded by entities having different land uses in successive years (typically four). It is bounded by physical limits such as a road, river, field path or single field boundary. It contains only one type of crop. The parcel differs from the cadastral parcel and also the block stored in the RPG[9] which is updated every two years and distributed by IGN.[10] The RPG contains relevant information for monitoring parcels owners, but block boundaries do not match the observed parcels from the field scale. In fact, farmers often cut each of their blocks into several parcels but only the surface of these blocks is declared with the nature of the land use in percentage.

1.2 *Spatio-temporal environment data*

For over 20 years, several databases have been collected by the AGRIPOP team. These data can be categorised as follows (more details can be found at the workshop website[11]).

1.2.1 *Land use data*

The CEBC[12] carries out, every year, two land use surveys (in April and June) that take into account both early harvested and late planted crops. In each survey, the occupancy of parcels is recorded distinguishing 47 types of land use including 42 agricultural, three urban, and two forest lands. Since 1994, the land use and spatial organisation of over 19,000 agricultural parcels is recorded from the field and centralised in a database that is initially modelled based on the Space-Time Composite paradigm (Langran and Chrisman, 1998). The paradigm introduces a small geometry, here called *microparcel*, which is obtained by finding the intersection of all parcels during an observation period. The geometry of a parcel can be rebuilt by the spatial union of all *microparcels* belonging to it. The database contains over 600,000 records managed by the PostgreSQL DBMS extended with the PostGIS extension.

1.2.2 *Biodiversity data*

Meanwhile, wildlife data are collected in the field for several years by another AGRIPOP team. These data, timely and dated, come from different researchers who report their observations on over 600 species, mostly birds and plants, via their mobile devices. For birds, the behaviour of observed species, their nests, and the contexts such as vegetation height, date, time, location, and weather condition are recorded. Over 26,000 observations are also managed by the PostgreSQL DBMS with the PostGIS extension.

Another database that could be cited is the Micromammal data set. It is a set of 10,000 observations related to the three micromammal species that are good indicators of agricultural practices: Wood mouse, Greater white-toothed shrew, and Common vole. In each observation, the catch rates and individual characteristics are recorded.

There also exist numerous sets of structured data about different insects, often in spreadsheets or Microsoft Access databases. These data concern mostly observations of ground beetles and small beetles which are auxiliaries of the fields and very sensitive to the quality of the environment. These insects have been monitored for over nine years.

1.3 The need for spatio-temporal analysis

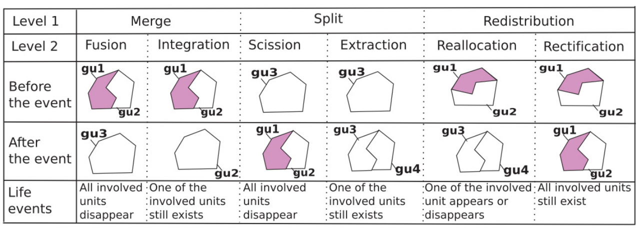
As a first step, three main types of spatio-temporal analysis can be conducted to exploit the available data sets presented above. These analyses, described as follows, require queries accomplished with spatio-temporal reasoning.

1 The analysis can be used first to verify the collected data sets. On crop rotation, experts can describe a set of successions rules in order to eliminate or correct questionable values. For example, the unlikely crop succession like *Sunflower-Sunflower* or *Sunflower-Rapeseed*, as well as the disappearance of wood in the workshop area, can be represented as a query to be detected and examined. This type of analysis needs primarily the reasoning about temporal relationships between intervals of recorded land use statements.

2 On another hand, territorial events, such as *fusion*, *integration*, *scission*, *extraction*, *reallocation and rectification* (Figure 3), are desired to be pointed out. Analysing these events allows discovering the correlation between crop rotation and the spatial organisation of the landscape. These events can be detected through queries with spatio-temporal reasoning.

3 Finally, experts also wish to seek the correlation between species observations and the land use of parcels. They could concern such animals preferences by the type and form of land use. Or they can verify the co-appearance of species according to the food chain or season. Cross-database queries with spatio-temporal reasoning are required to select observations occurring in the period of recorded land use statements.

Figure 3 A classification of territorial events (Plumejeaud et al., 2011) (see online version for colours)



2 Heterogeneous data integration

In order to exploit these heterogeneous data, first and foremost, data integration must be accomplished. Data integration is the process of combining data from different sources to provide users a unified view of the data. A data integration system often uses a global schema containing mappings from global definitions to the local schema of each data sources. In this context, the data query problem can be reduced to the problem of answering queries using materialised views. In the traditional data integration approach, data sources are queried through a global data model (Lenzerini, 2002). However, in general, there is not a common data model that is capable of accessing all available data with the granularity required by all users. One of the benefits often stated for RDF is the ease with which data can be integrated from distributed RDF sources (Gray et al., 2009). At the heart of semantic data integration is the concept of ontology where a conceptual representation of data and their relationships is used to eliminate heterogeneities, which then acts as a mediator for re-conciliating the heterogeneities between different data sources (Cruz and Xiao, 2005). Because the global schema is defined using an ontology, views are not just the mapping from data sources on the global schema but also include inference rules to obtain object properties (Zhao et al., 2008). By defining entities and their relations, ontologies are considered as a feasible solution to the semantic heterogeneity problem (Wache et al, 2001; Roussey et al., 2011) thus become the heart of semantic data integration systems (Cruz and Xiao, 2005).

Cruz and Xiao (2005, 2009) have identified five main roles of ontology in data integration process.

- Metadata representation: Metadata in each data source can be explicitly represented by a local ontology. These ontologies are homogeneous since they use the same representation language.
- Global conceptualisation and support for high-level queries: The global ontology provides a conceptual view of the heterogeneous source schemas. It provides a high-level view of sources. Therefore, a query can be formulated without specific knowledge of data sources. Using an ontology as a query model, the structure of query model should be more intuitive for users because it corresponds more to their appreciation of the domain.
- Declarative mediation: The global ontology can be used as a mediator for query rewriting across peers.
- Mapping support: A common vocabulary, which can be formalised as an ontology, can be used to facilitate the mapping process. Since ontologies contain a complete specification of the conceptualisation, the mappings can be validated with respect to ontologies to facilitate its automation.

In general, there exist three ontology-based approaches to form the global schema: single ontology approach, multiple ontologies approach, and hybrid approach. The integration based on a single ontology seems to be the simplest approach because it can be simulated by the others (Wache et al, 2001).

2.1 Single ontology approach

Single ontology approaches use a global ontology providing a shared vocabulary for the specification of semantics. All source schemas are directly related to the shared global ontology, which provides a uniform interface to users. However, the approach requires that all data sources provide nearly the same view on a domain, for example the same level of granularity.

2.2 Multiple ontology approach

Instead of using a global ontology, each data source is described by its own local ontology. This architecture can simplify the change, i.e. modifications in one data source or the adding and removing of sources. However, the lack of a common vocabulary makes it extremely difficult to compare local ontologies.

2.3 Hybrid ontology approach

A combination of these approaches is used to overcome the drawbacks of the single and multiple ontology approaches. First, a local ontology is built for each source schema which is mapped to the global ontology. New sources can be easily added with no need to modify existing mappings.

The hybrid ontology approach is used in our framework due to its advantages and the support of building central data integration systems. As a result, a global spatio-temporal ontology is introduced to which our data sources' schema are mapped.

3 A spatio-temporal ontology for environment

With the purpose of integrating and exploiting the presented data set, a global spatio-temporal ontology is introduced. The ontology is formed based on a temporal ontology and spatial ontology. A reasoning mechanism for spatio-temporal relations between entities is also proposed.

3.1 Temporal ontology

Time can be represented by time instances or intervals. Several ontologies have been proposed in the literature to represent time, including OWL-Time[13] (Hobbs and Pan, 2004) and SWRL Temporal Ontology[14] (O'Connor and Das, 2011).

The OWL-Time ontology, dedicated to concepts and temporal relationships as defined in the theory of Allen (1983), is certainly the best candidate. The ontology is used first to describe the temporal content of web pages and the temporal properties of web services. It is recommended by the W3C for modelling temporal concepts due to its vocabulary for expressing topological relations between instants and intervals.

Another ontology used to represent temporal information is the SWRL Temporal Ontology. The ontology defines OWL entities that represent an interval-based temporal information in OWL. It does not allow to establish a topological relationship between instants or time intervals but offers some temporal *built-ins* that allow checking of Allen relationships between different temporal entities when querying.

However, apart from language constructs for representing time in ontologies, there is still a need for mechanisms for representing the evolution of objects through time.

3.2 Representation of time in ontologies

There is a fundamental philosophical controversy concerning the evolution of objects throughout time, namely perdurantism and endurantism. Endurantism assumes that objects have three dimensions and are available in full at every moment of their lives. Thus, these objects do not have the temporal dimension. In contrast, perdurantist approach considers objects to have four dimensions. As a consequence, these objects have several time slices in their lives constituting the temporal dimension. This approach represents various properties of an entity over time as fluents that are only validated during certain intervals or instants. Therefore, perdurantist approach enables richer representation of real-world phenomenon through its flexibility and expressiveness (Al-Debei et al., 2012).

Ontologies contain definitions of concepts and their properties by means of binary relations, as a result, temporal relationships between objects are neglected. In order to represent dynamic objects and their properties, several methodologies have been proposed such as temporal description logic (Artale and Franconi, 2001), temporal RDF (Gutierrez et al., 2007), versioning (Klein and Fensel, 2001), reification, N-ary relationships (Noy and Rector, 2006) and 4D-Fluent model (Welty and Fikes, 2006). In the literature, 4D-Fluent, based on perdurantist view, is the most well-known approach to handle dynamic properties of temporal objects. It has a simple structure allowing transforming a static ontology into a dynamic one (Harbelot et al., 2013). The authors introduced the *TimeSlice* class to represent the temporal part of entities which is linked to the *TimeInterval* class, a class of the time domain. Each entity is associated with an instance of the *TimeSlice* class by the *tsTimeSliceOf* object property. This latter is connected to an instance of the *TimeInterval* by the *tsTimeInterval* property.

Several approaches based on the 4D-Fluent model have been introduced. tOWL (Frasincar et al., 2010) extends OWL with a temporal dimension in order to represent complex temporal aspects, such as process state transitions. SOWL (Batsakis et al., 2011) extends OWL-Time by enabling the representation of static and dynamic information. Recently, the Continuum model (Harbelot et al., 2013) was presented allowing tracking the identity of spatio-temporal entities through time. The model represents dynamic entities as constituted by timeslices with semantic, spatial, temporal and identity components. Therefore, it is able to link the diverse representations of an entity and allows the inference of qualitative information from quantitative one. Inferred data are later added to the ontology to enrich knowledge about the phenomenon. The model has been successfully applied in studies of the urban evolution (Harbelot et al., 2013) or decolonisation process (Harbelot et al., 2014).

Owing to the capacity of presentation real-world dynamic objects, the 4D-Fluent model along with the OWL-Time ontology are chosen to develop our spatio-temporal ontology that will be presented in a later section.

3.3 Spatial ontology

Spatial entities can be represented by points, lines (polygonal, lines) and their relationships. There is a broad range of spatial ontologies defined by several communities for different applications. Ressler et al. (2010) have studied 45 geospatial and temporal ontologies. Seven full spatio-temporal ontologies among them are recommended for reuse. The top of this list are the two version of GeoRSS[15] (Open Geospatial Consortium, 2006): GeoRSS Simple and GeoRSS GML.

GeoRSS has been designed as a lightweight and community-driven way to extend existing feeds with geographic information, since the location is described in an interoperable manner enabling processing, aggregating, sharing, and mapping of geographically tagged feeds. The GeoRSS XML schema contains four simple types of geographic locations: point, line, polygon, and box. These locations may be encoded in either a literal string of latitude and longitude, called GeoRSS Simple, or a more robust representation using GML, called GeoRSS GML, which is formally defined as a GML application profile and supports a greater range of features, notably coordinate reference systems other than the WGS84 latitude and longitude.

The GeoRSS Simple model is reused for our ontology development to represent the spatial dimension of environmental entities since the main concepts of the model are enough for representing the spatial dimension of environmental entities.

3.4 Spatio-temporal reasoning

As presented, perdurantist objects evolve through time. As a result, these objects may change their positions or their occupancies during their lives. The spatio-temporal reasoning is used to point out spatio-temporal relations between them. The reasoning is performed by a dynamic combination of spatial and temporal reasoning mechanism which are described bellow.

3.4.1 Temporal reasoning

Basic temporal formalisms can only be used for reasoning about objects of a single type, for instance Vilain's point algebra (Vilain, 1982) deals with three elementary relations between points and Allen's interval algebra (Allen, 1983) allows reasoning about 13 elementary relations between two intervals. Such restricted languages may not be sufficient for modelling real-world problems. Meiri's qualitative algebra (Meiri, 1996), originated from Vilain's point and interval algebra (Vilain et al., 1986), which is a temporal formalism, ables to represent both time points and time intervals. The algebra helps to relate points with points, points with intervals and intervals with intervals using an expressive set of qualitative relations.

In the qualitative algebra, there is a total of 26 relations categorised into three types of relations:

1. Point–point relations that can hold between a pair of points: *before*, *equals* and *after*.
2. Point–interval and interval–point relations that can hold between a point and an interval and vice-versa: *before*, *starts*, *during*, *finishes* and *after*.
3. Interval–interval relations that can hold between a pair of intervals: *precedes*, *meets*, *overlaps*, *during*, *starts*, *finishes*, *equals* and their reverse. These relations correspond to the 13 relations introduced in Allen's interval algebra.

At the present time, in order to reuse the OWL-Time ontology, we consider only the 13 relations between intervals and the *during* relation between a point and an interval which matches the *inside* relation of OWL-Time. In this manner, these intervals can be viewed as instances of the *ProperInterval* class of OWL-Time. An interval is linked to two instants by the *hasBeginning* and *hasEnd* attribute that determine their boundaries.

Let $I = [a, b]$, $J = [c, d]$ be two intervals with starting and ending points a, c and b, d, respectively; and $P = x$ a point at x, 14 relations, detailed in Table 1, are considered.

Table 1 Temporal relations used to develop the spatio-temporal ontology

Relation	Constraint	Schema	Inverse
before [b]	b<c	a I b c J d	after [a]
meets [m]	b=c	a I b c J d	met-by [mi]
equals [eq]	a=c ∧ b=d	a I b c J d	
overlaps [o]	d ∧ b	a I b c J d	overlapped-by [oi]
starts [s]	a=c ∧ b<d	a I b c J d	started-by [si]
during [d]	c ∧ d	a I b c J d	contains [di]
finishes [f]	b=d ∧ c<a	a I b c J d	finished-by [fi]
inside [i]	c<x<d	c I ● x d	

3.4.2 Temporal reasoning through SWRL rules and SPARQL

To discover temporal relationships between objects, these above relations can be expressed by a set of rules. The Semantic Web Rule Language (SWRL[16]) is commonly chosen owing to its available libraries, called *built-ins*, that provide several predicates, mostly for date-time and duration processing. In this way, qualitative temporal relationships between objects can be derived by the Pellet engine[17] through a set of SWRL rules. For example, the *inside* relation between an instant and an interval can be expressed by the following SWRL rule:

```
Instant(?x), ProperInterval(?a),
hasBeginning(?a,?bg),
inXSDDateTime(?bg,?dt1),
hasEnd(?a,?end),
inXSDDateTime(?end,?dt2),
inXSDDateTime(?x,?dt),
lessThan(?dt,?dt2),
greaterThan(?dt,?dt1) → inside(?x,?a)
```

This reasoning mechanism was applied in the SOWL ontology (Batsakis et al., 2011) which was afterward improved by the CHRONOS system (Anagnostopoulos, 2013). O'Connor and Das (2011) have developed SWRLTemporalBuiltIns, a set of *built-ins* that can be used in SWRL rules to perform temporal reasoning with their temporal ontology.

The application of SPARQL as a rule language is another approach. Angles and Gutierrez (2008) has proved that SPARQL is equivalent from an expressiveness point of view to Relational Algebra. Polleres et al. (2007) develop SPARQL++ which uses SPARQL as a rule language to define mappings between RDF vocabularies, allowing CONSTRUCT queries extended with *built-in* and aggregate functions. Schenk and Staab (2008) applied SPARQL as rule language to define new RDF data from existing data source. Knublauch et al. (2011) have introduced SPIN[18] that become the de-facto industry standard to represent SPARQL rules and constraints on Semantic Web models. Corby et al. (2012) have proposed Corese/KGRAM implementing SPARQL 1.1[19] to query data sources and apply inference rules to them.

In our case, the INSERT queries introduced in the SPARQL 1.1 query language is used instead of SWRL rules to infer new statements. This approach is chosen as a reasoning mechanism for non-spatial data in our framework. Query 1 is an example of how to infer the *inside* relation between an instant and an interval.

Query 1: Infercence of the *inside* relation between an instant and an interval

```
INSERT
{?x time:inside ?a.}
WHERE
{
   ?x rdf:type time:Instant.
   ?x time:inXSDDateTime ?dt.
   ?a rdf:type time:Interval.
   ?a time:hasBeginning ?bg.
   ?a time:hasEnd ?end.
   ?bg time:inXSDDateTime ?dt1.
   ?end time:inXSDDateTime ?dt2.
FILTER(?dt>?dt1 && ?dt<?dt2)
}
```

3.4.3 *Spatial reasoning*

Relations between spatial entities can be topological, orientation or distance-based. Furthermore, spatial relations can be partitioned into qualitative and quantitative ones. The qualitative approach is considered to be closer to the way humans represent spatial knowledge. The concept of neighbourhood is implicit and leads to the study of topology, which is a set of perceived relationships that allow situating an object in relation to others. In the literature, the topological analysis between spatial objects is often performed by the Nine-Intersection Model (Egenhofer and Herring, 1990) or the RCC8 model (Randell et al., 1992). In both cases, we obtain an equivalent set of eight basic pairwise disjoint topological relations which are mutually exhaustive: *equals*, *disjoint*, *intersects*, *touches*, *within*, *contains* and *overlaps*.

Unfortunately, these relations cannot be inferred with simple rules. In several studies (Karmacharya et al., 2010; Vandecasteele and Napoli, 2012), the authors have introduced SWRL *built-ins* to represent and infer spatial relationships between spatio-temporal objects but there are still limitations with regard mainly both to the system performance and reuse capability.

On the other hand, reasoning systems, such as Pellet-spatial (Stocker and Sirin, 2009), CHOROS (Christodoulou et al., 2012) or its next version CHOROS 2.0 (Mainas et al., 2014) improving the run-time performance, extract spatial relations from a knowledge base and reason over both these topological and directional relations. However, reasoning through specialised software also complicates the system re-usability and data sharing (Batsakis and Antoniou, 2014). Specifically, whenever the definitions of spatial relations and semantics in RDF/OWL format are modified, the spatial reasoners must be modified as well. In addition, the above-mentioned specialised software is needed for querying spatial information. Using specialised reasoners such as CHOROS, the reasoning performance increases but this approach is less flexible. Nevertheless, dealing with Big Data is currently beyond the capabilities of both approaches.

Therefore, in our project, the reasoning on complex spatial information is performed by a geospatial triplestore that is described in the next section.

3.4.4 *Spatial reasoning through a Geospatial RDF Store*

Triplestores are DBMS for data modelled in RDF. Currently, several triplestores support storing and querying spatial data using GeoSPARQL[20] or stSPARQL[21] that are two extensions of the SPARQL language. Those open-sources that manage the best are uSeekM,[22] Parliament[23] and Strabon.[24] Other triplestores support only a few type of geometries and geospatial functions. Strabon (Kyzirakos et al., 2012) is chosen since the triplestore has a good overall performance. This advantage can be explained by particular optimisation techniques allowing spatial operations to take advantage of PostGIS functionalities instead of relying on external libraries (Patroumpas et al., 2014) and the push of the evaluation of SPARQL queries to the underlying spatially enabled DBMS which has been enhanced recently with selectivity estimation capabilities (Garbis et al., 2013). Thus, spatial joins are also efficiently handled by the underlying PostgreSQL/PostGIS optimiser.

Strabon extends the Sesame triplestore, allowing spatial RDF data stored in the PostgreSQL DBMS enhanced with PostGIS. The triplestore works over the stRDF data model (Kyzirakos et al., 2012), a spatio-temporal extension of RDF in which the *strdf:WKT* and *strdf:GML* datatype are introduced to represent geometries serialised using the two OGC standards: WKT and GML. The *strdf:geometry* which is the union of *strdf:WKT* and *strdf:GML* and the appropriate relationships are also introduced to represent the serialisation of a geometry independently of the standard used serialisation. Among two geometries, Strabon supports 13 topological relations and four directional relations. Table 2 lists eight Strabon's spatial functions and their corresponding relation in the RCC8 model and the OGC standard.

The stSPARQL language used in Strabon has certain limitations, as it does not enable binary topological relations to be used as RDF properties. However, stSPARQL extends SPARQL 1.1 and overtakes GeoSPARQL by offering spatial aggregate functions and triple update commands. SQL functions can be utilised in SPARQL queries by defining a URI for each of them. Similarly, a boolean SPARQL extension function has been defined for each topological relation in three models: the OGC Simple Features, Nine-Intersection Model, and RCC8 model. Thus, stSPARQL supports multiple families of topological relations and can express spatial selections and spatial joins. Another benefit of stSPARQL is the support of federated queries and update statements that facilitates updating knowledge bases.

Table 2 Strabon's spatial functions and the corresponding relation in the RCC8 model and the OGC standard

RCC8 relation	OGC relation	Schema	Strabon function
disconnected [DC]	disjoint	DC(a,b)	strdf:disjoint(a,b)
externally connected [EC]	touches	EC(a,b)	strdf:touches(a,b)
partially overlapping [PO]	overlaps	PO (a,b)	strdf:overlaps(a,b)
equal [EQ]	equal	EQ(a,b)	strdf:equals(a,b)
tangential proper part [TPP]	within	TPP(a,b)	strdf:within(a,b)
non-tangential proper part [nTPP]	within	nTPP(a,b)	strdf:within(a,b)
tangential proper part inverse [TPPi]	contains	TPPi(a,b)	strdf:contains(a,b)
non-tangential proper part inverse [nTPPi]	contains	nTPPi(a,b)	strdf:contains(a,b)

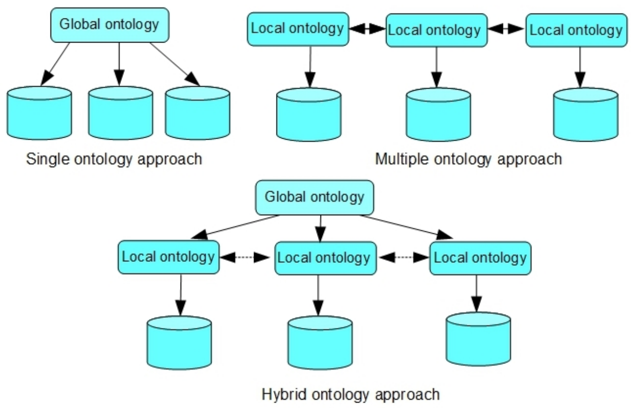
3.5 *A spatio-temporal ontology for environment*

We propose a spatio-temporal ontology (Figure 5 (a)) based on the 4-D Fluent approach that serves as a semantic mediator to integrate the presented data sets. The ontology is inspired by the Continuum model that examines the evolution of objects in both the temporal and spatial dimension. The core concept in our ontology is the *Spatio-temporal Element* which represents a species observation or a parcel timeslice.

3.5.1 *Spatio-temporal entities and Semantic component*

The main entities in our studies are parcels (*gem:Parcel*), animal nests (*gem:Nest*) and individuals (*gem:Animal*) belonging to different species (*gem:Species*). They have one or more spatio-temporal elements (*gem:STElement*) that correspond to their different characteristics and spatial occupations through their lives. The latter has two subclasses *gem:Obsv* and *gem:TimeSlice* representing, respectively, animal observations and parcel statements. Each spatio-temporal element has three components: the semantic (*gem:Description*), temporal (*time:TemporalEntity*) and spatial component (*georss:_geometry*). The semantic part represents the land use of parcels or description of animal observations such as the behaviour or context.

Figure 4 Ontology-based approaches for data integration (Wache et al, 2001) (see online version for colours)



3.5.2 *Temporal component*

While the land use or boundary changes of parcels are periodically archived by predetermined intervals of temporal validity, species observations are collected at will. For this reason, the 4D-Fluent model is extended by generalising the *Interval* class to the *TemporalEntity* class of OWL-Time that has two subclasses, *Interval* and *Instant*. The similar generalisation is also presented in the tOWL temporal language (Frasincar et al., 2010) to support both time points and time intervals and adopt an internal (perdurantist) view on the world.

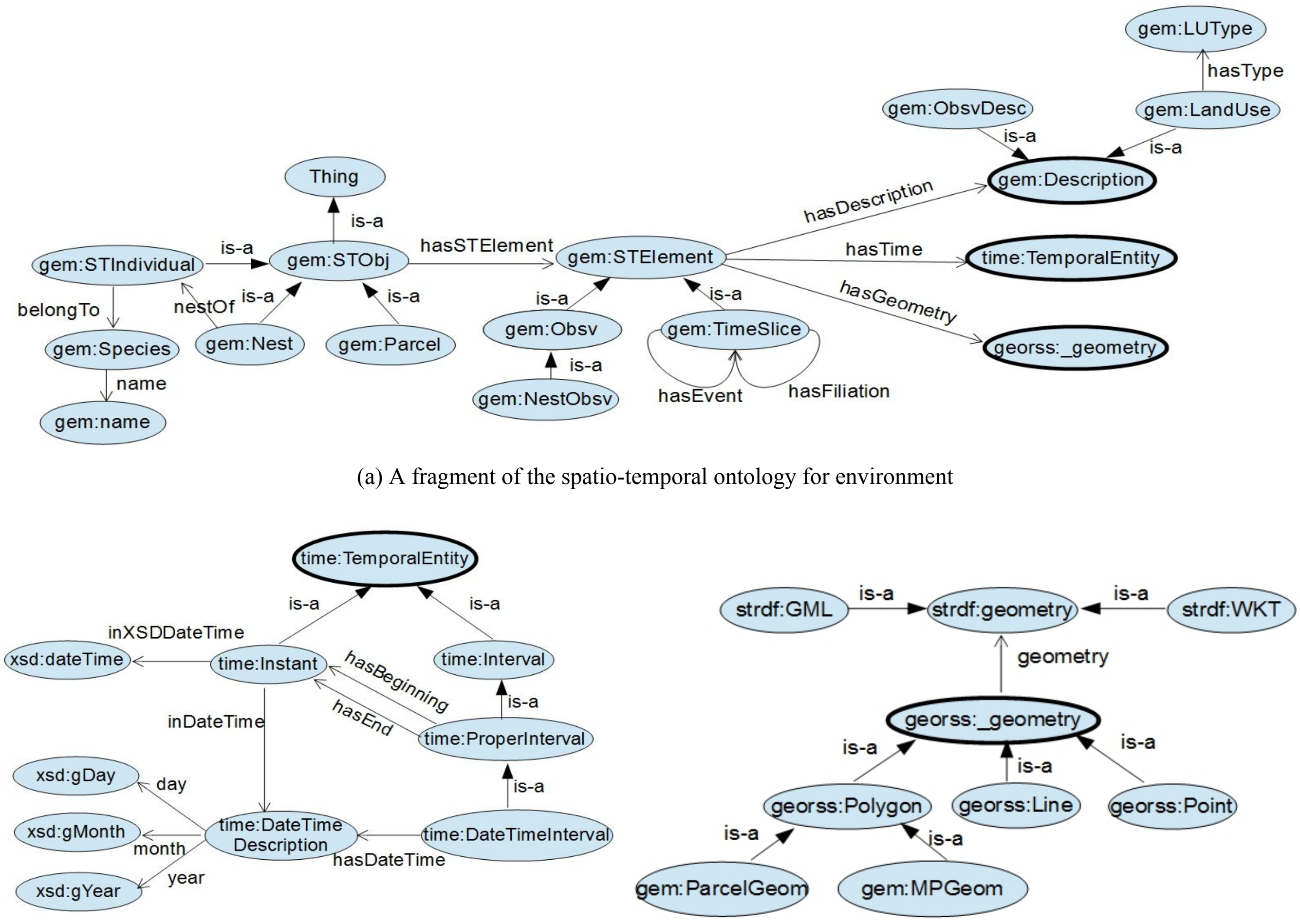
Figure 5 (b) shows the reused fragment of the OWL-Time ontology that has three main concepts: *Instant*, *Interval* for date-time representation and *DateTimeDescription* for date-time description facilitating date-time comparison.

3.5.3 *Spatial component*

In this component (Figure 5 (c)), the *_geometry* class of the GeoRSS model and its three subclasses are reused: *Polygon*, *Line*, *Point* to represent parcel geometries and observation points. The geometry of their entities is represented by the *strdf:geometry* datatype proposed by the Strabon triplestore that is the union of *strdf:WKT* and *strdf:GML*.

As presented, the land use database is built based on the *Space-Time Composite* paradigm which uses *microparcel* as a management unit. In consequence, we introduce the *MicroparcelGeometry* class as a subclass of the *Polygon* class that specialises the *_geometry* class. Since the spatial reference system used in these heterogeneous databases are different, the conversion of geometry data into the same spatial reference is performed via the mapping process. This approach improves the performance of spatial reasoning compared with maintaining the spatial reference for each spatial data.

Figure 5 A spatio-temporal ontology for environment (see online version for colours)

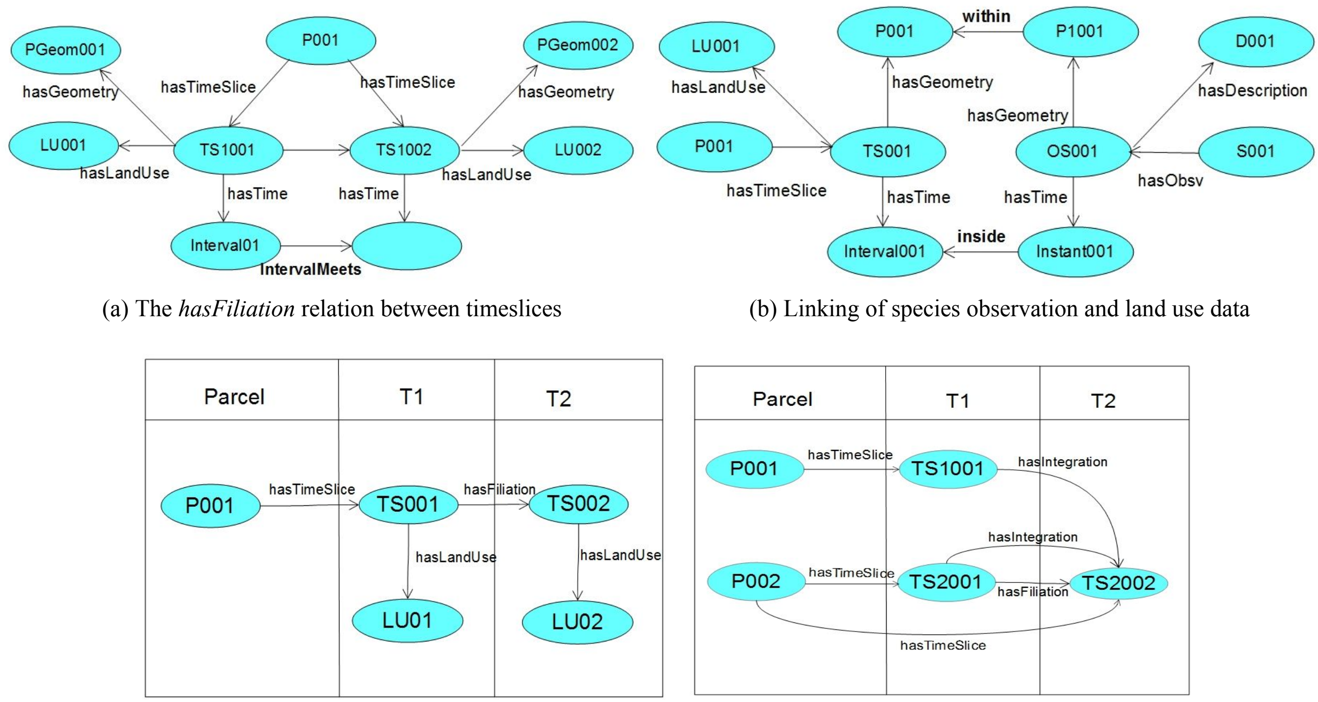


(a) A fragment of the spatio-temporal ontology for environment

(b) The temporal component

(c) The spatial component

Figure 6 Examples of spatio-temporal analysis (see online version for colours)



(a) The *hasFiliation* relation between timeslices

(b) Linking of species observation and land use data

(c) A two-year crop rotation

(d) An integration event between two parcels

3.5.4 Objects Identity

The *hasFiliation* relationship (Figure 6 (a)) between two timeslices is introduced to track and define the evolution of an object in the case of a continuation, i.e. object may change its semantic or its position while keeping its identity. Territorial events occurring in parcels can be represented and examined by a combination of the *hasFiliation* relation and one of six relations between timeslices that are subproprieties of the *hasEvent* propriety: *hasFusion*, *hasIntegration*, *hasScission*, *hasExtraction*, *hasReallocation* and *hasRectification*.

Through the inferred spatio-temporal relations, the three major needs of data analysis can be fulfilled. Let's examine three simple corresponding cases below:

1 Species observation and land use data can be linked by combining the *inside* temporal relation between an instant of observation and an interval of land use statement and the *within* spatial relation between an observation point and a parcel polygon (Figure 6 (b)).

2 Crop rotation can be verified by the *hasFiliation* relation between parcel timeslices or the *meets* temporal relation between timeslices' intervals of the same parcel (Figure 6 (c)).

3 Territorial events can be detected by incorporating the *intervalmeets* temporal relation between interval of different timeslices and the *within* spatial relation between parcel geometries. Or, they can be discovered in a simple manner with the *hasEvent* and *hasFiliation* relation between *timeslice*. Figure 6 (d) shows how to identify the two parcels that participate in an integration event. While the P001 parcel disappears after this event, the P002 parcel continue to exist so that the TS2002 timeslice is generated and associated to the earlier timeslice by the *hasFiliation* relation.

4 Towards a semantic framework for exploiting heterogeneous environmental data

A layered framework (Figure 7) is proposed in order to integrate and exploit heterogeneous spatio-temporal data. Data in different data sets are filtered and converted to RDF triples which are next imported to the GEMINAT knowledge base managed by the Strabon triplestore. The reasoning and data analysis are performed through the semantic query service.

4.1 The layer-based framework

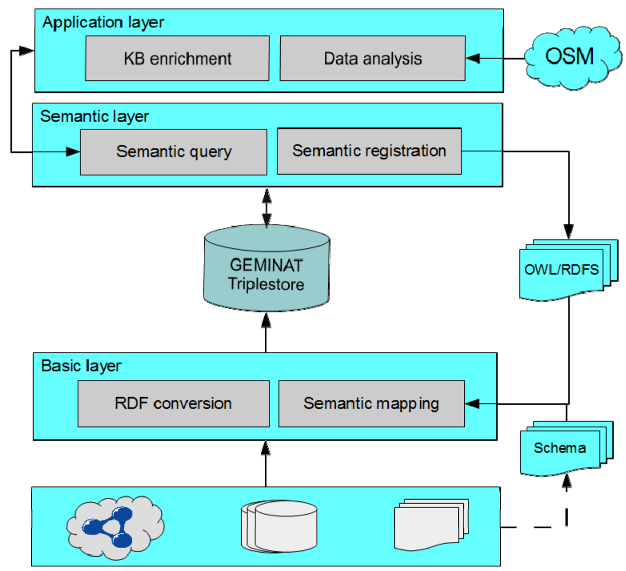
The framework consists of three principal layers: the basic, semantic, and application layer.

Basic layer

In order to populate ontologies with existing data sources, we rely on translation techniques that transform relational data into RDF graph.

The semantic mapping module establishes mappings from relational databases schema to ontologies. The module manages mapping files that define how to connect to databases and match ontologies to databases schema. Being converted from relational data through the RDF conversion module, RDF triples can be next imported to the knowledge base.

Figure 7 A framework for exploiting heterogeneous environmental data (see online version for colours)



In this layer, the D2RQ framework[25] (Bizer, 2004) is chosen as a translation tool owing to the support of different DBMS. Information extraction from Linked Data will be also supported in the next development of the framework.

Semantic layer

Two modules are developed. The semantic query module is used to process SPARQL semantic queries that are next sent to the triplestore. The semantic registration module is used to expose the shared ontologies that mediate heterogeneous relational databases. Domain and spatio-temporal ontologies can be registered and edited by this module. The essential tool used in this layer is the Jena framework.[26]

Application layer

There exist two modules: the reasoning and data analysis module, which are both referred to the semantic query module. In the first one, new statements can be inferred from spatio-temporal and business rules, while in the latter one, a web server is hosted to receive stSPARQL queries in the form of HTTP requests. The returned result is visualised by the OpenLayers library[27] with geographical data from OpenStreetMap.[28] The result is stored in several different layers to facilitate the presentation and analysis.

4.2 Reasoning and data analysis

The proposed framework along with the use of a spatio-temporal ontology as a semantic mediator can fulfil the needs of spatio-temporal analysis. Indeed, the data model in the form of subjacent RDF graph facilitates the integration of different databases. In addition, thanks to the Strabon triplestore, spatio-temporal relationships between objects can be inferred. At the first time, only land use and wildlife data are selected for experiments.

1 New statements can be inferred from knowledge base through spatial-temporal and business rules expressed by SPARQL Update queries. Query 2 is an example of how to detect and insert integration events applied on farmland into the triplestore.

2 To analyse correlations between crop rotation and biodiversity, experts can visualise the appearance of species by type and form of crop rotation. For example, they can check out the nesting preference of Montagu's harriers (*Circus pygargus*) for different types of grassland parcels (Figure 8 (a)). In this analysis, the *inside* temporal relation and the *within* spatial relation between entities are used (Query 3). In addition, observations of different types of species can be queried to discover their relation in the food chain. Figure 8 (b) visualises the nesting of Montagu's harrier and other species observed in the same period within the 100 m radius, provided by Query 4.

3 Territorial events applied on farmland can be discovered by combining qualitative spatio-temporal relationships. For example, integration events (parcels are absorbed by others) in 2009 can be retrieved and displayed on the map like (Figure 8 (c)). The analyses of these events along with the land use of involved parcels can reveal farmer practice preferences. Soil data and market price data could be used to complement these analyses.

Query 2: Detection of the integration events

```
INSERT
{
  ?tsa gem:hasIntegration ?tsc.
  ?tsb gem:hasIntegration ?tsc.
}
WHERE
{
  ?pc gem:hasTimeSlice ?tsa.
  ?pc2 gem:hasTimeSlice ?tsb.
  ?pc gem:hasTimeSlice ?tsc.
FILTER(?pc!=?pc2)
  ?tsa gem:hasTime ?intva.
  ?tsb gem:hasTime ?intvb.
  ?tsc gem:hasTime ?intvc.
  ?intva time:intervalMeets
?intvc.
  ?intvb time:intervalMeets
?intvc.
  ?tsa gem:hasPGeometry ?geoa.
  ?geoa gem:geometry ?geoma.
  ?tsb gem:hasPGeometry ?geob.
  ?geob gem:geometry ?geomb.
FILTER(strdf:intersects
(?geoma,?geomb))
    ?tsc gem:hasPGeometry ?geoc.
    ?geoc gem:geometry ?geomc.
FILTER(strdf:within(?geoma,?geomc)
    && strdf:within(?geomb,?geomc))
}
```

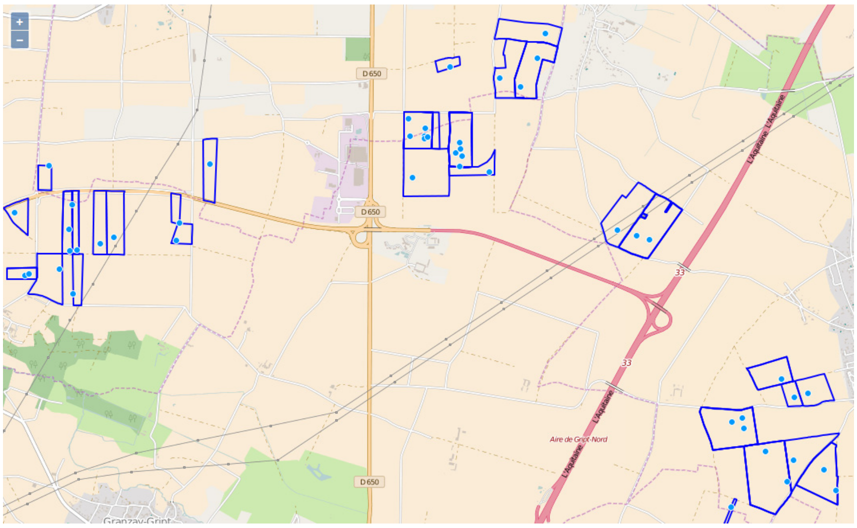
Query 3: Correlation between the nesting of Montagu's harrier and the land use of parcels

```
SELECT *
WHERE
{
  ?nest gem:nestOf ?indv.
  ?indv gem:belongTo ?sp.
  ?sp gem:name ?spname.
  ?nest gem:hasNestObsv ?nobsv.
  ?nobsv sig:hasTime ?inst.
  ?inst time:inside ?intv.
  ?ts gem:hasTime ?intv.
  ?ts gem:hasLandUse ?lu.
  ?nobsv gem:hasGeometry ?geom.
  ?geon gem:geometry ?geomn.
  ?ts gem:hasGeometry ?geo.
  ?geo gem:geometry ?geom.
FILTER (strdf:within(?geomn,?geom)
     && ?spname="Busard cendre")
}
```

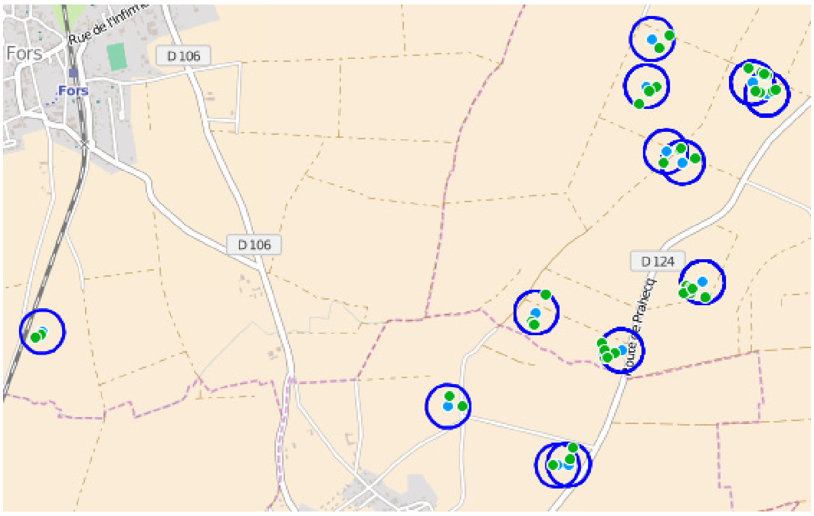
Query 4: Correlation between the nesting of Montagu's harrier and other species observed in the radius of 100 m and in the same period

```
SELECT (strdf:buffer(?geom1, 100,
ogc:metre)
       as ?geomBuf) ?geom1 ?name2
?geom2
WHERE
{
  ?nest gem:nestOf ?indv.
  ?indv gem:belongTo ?sp1.
  ?sp1 gem:name ?spname1.
  ?nest gem:hasNestObsv ?nobsv.
  ?nobsv sig:hasTime ?inst1.
  ?nobsv gem:hasGeometry ?geo1.
  ?geo1 gem:geometry ?geom1.
  ?inst1 time:month ?m1.
  ?inst1 time:year ?y1.
  ?indv gem:hasObsv ?obsv.
  ?obsv gem:hasTime ?inst2.
  ?obsv gem:hasGeometry ?geo2.
  ?geo2 gem:geometry ?geom2.
  ?inst2 time:month ?m2.
  ?inst2 time:year ?y2.
  ?indv gem:belongTo ?sp2.
  ?sp2 gem:name ?spname2.
FILTER(?spname1="Busard cendre"
&&  ?m1=?m2
   &&  ?y1=?y2  (strdf:distance
(?geom1,
    ?geom2, ogc:metre)<="100"^^
xsd:double))
}
```

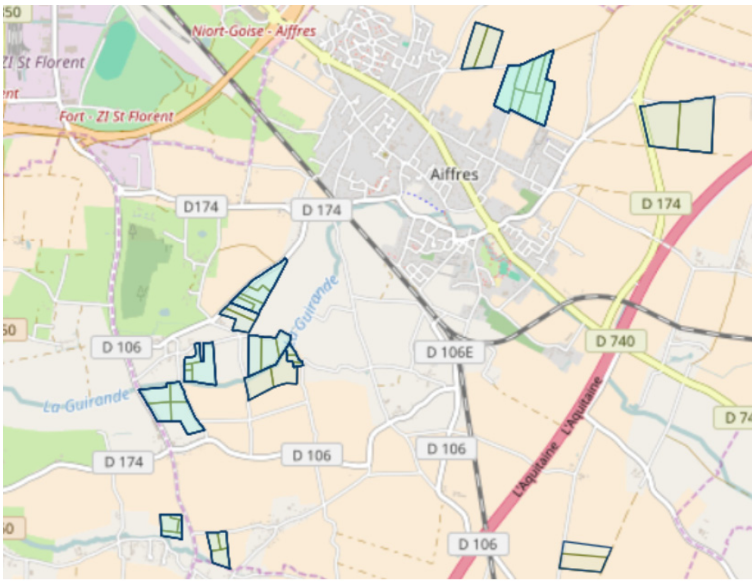
Figure 8 Application of the framework in spatio-temporal analysis (see online version for colours)



(a) Correlation between the nesting of Montagu's harrier and the land use of parcels



(b) Correlation between the nesting of Montagu's harrier (blue point, circle's centre) and other species observed (red point) in the radius of 100 m



(c) Detection of the integration events

5 Related work

Several frameworks were introduced to integrate heterogeneous data by means of ontologies. In the Geon project (Lin and Lud Ascher, 2003), a prototype system has been developed for registering data sets through ontologies to assist in integrating and querying heterogeneous geologic data. Each data set must be registered before it becomes available, and the registration semi-automatically generates a mapping from data sets to ontologies. The mapping is afterward used by applications to explore and extract information from the data set.

Dartgrid (Chen and Wu, 2005) is an application development framework with a set of practical semantic tools to facilitate the integration of heterogeneous relational databases using semantic web technologies. It allows developers to interconnect distributed legacy databases using richer semantics, to provide ontology-based query, search and navigation services as one huge distributed database, and to add additional deductive capabilities to increase the usability and re-usability of data.

Green et al. (2008) have investigated the use of semantic tools as an aid to data integration and identify the need to modify these tools to meet the needs of spatial data. The Demonstrator prototype is built to enable queries that are expressed in the vocabulary of the application domain rather than using the terminology of the database schemas. The system comprises a series of layered ontologies and a translator for converting data sources into a virtual RDF graph.

Durbha et al. (2009) introduced the MOB ontology-driven middleware (Middleware for Ontology-driven Brokering) that provides functionalities for ontology management, storage, query, and inference services. It also enables resource discovery and mediation through ontologies built on top of the metadata. The translation of metadata to semantic metadata through ontology-based approaches is also handled by MOB. The middleware was used to solve the problems of semantic heterogeneity in coastal zone data.

Recently, in the AgriNepalData project (Pokharel et al., 2014), the authors demonstrate how to convert several available, although not easily accessible, data sets to RDF to lower the barrier for data re-usage and integration. They describe the conversion, linking, and publication process as well as a use case of five different databases concerning the agriculture field in Nepal. An ontology is designated to represent and align heterogeneous data sources so that new knowledge can be inferred. In addition, after conversion, RDF is linked to Agrovoc DBPedia and to facilitate reusability and integrability.

In our framework, data integration is applied as an intermediate process followed by the knowledge base enrichment to enable the exploitation of environmental data, while the first four above systems do not focus on both the spatial and temporal dimension of data. Furthermore, it's unclear which tools are used for RDF translation, semantic querying and which ontologies are reused. While sharing the same approach, the last project uses the Virtuoso geospatial triplestore along with the GeoSPARQL language to exploit data in the domain of agriculture. On the other hand, belong to the larger GeoKnow European project, the authors apply other research of the group that is also the most recent work in the field such as Sparqlify for extraction, LIMES for data set linking, GeoLift for data set enrichment and Facete for data visualisation. Besides, since their main studied objects are administrative units that can hardly evolve through time, the incorporation of time to represent the evolution of entities turns out to be ignored.

6 Conclusion and future work

The presented work is part of the 'Environment and landscape geo-knowledge' interdisciplinary project which sets out to improve the use of collected environment data sets on the 'Plaine & Val de Sèvre' workshop observatory since 1994. We seek to develop an open-source framework to exploit heterogeneous environmental data through semantic web technologies. A spatio-temporal ontology is used as a mediator to resolve semantic heterogeneity of different data sets. The results show that the proposed framework along with the global spatio-temporal ontology can fulfil the need of spatio-temporal analysis. The introduced approach could be reused to perform management and analysis of long-term environmental data for others observatories.

In our perspectives, we consider integrating other data sets of the workshop area, such as the insect and botanical data, or satellite data. It will be then possible to use the system to enrich and qualify our data sources. The application of association rule mining on the knowledge base is also considered to discover new business rules and examine already known ones (for example the relation between species and land uses and between species themselves in their food chain). Finally, we plan to publish a portion of these data over the web as linked data in order to facilitate interchanges with other available data sets, especially the weather and soil data concerning the workshop area.

References

Al-Debei, M., Al Asswad, M., de Cesare, S. and Lycett, M. (2012) 'Conceptual modelling and the quality of ontologies: endurantism vs. perdurantism', *International Journal of Database Management Systems*, Vol. 4, pp.1–19.

Allen, J.F. (1983) 'Maintaining knowledge about temporal intervals', *Communications of the ACM*, Vol. 26, No. 11, pp.832–843.

Anagnostopoulos, E., Batsakis, S. and Petrakis, E.G.M. (2013) 'CHRONOS: a reasoning engine for qualitative temporal information in OWL', *Procedia Computer Science*, Vol. 22, pp.70–77.

Angles, R. and Gutierrez, C. (2008) 'The expressive power of SPARQL', *Proceedings of the 7th International SemanticWeb Conference, ISWC 2008*, Springer, Karlsruhe, Germany, pp.114–129.

Artale, A. and Franconi, E. (2001) 'A survey of temporal extensions of description logics', *Annals of Mathematics and Artificial Intelligence*, Vol. 30, pp.1–4.

Batsakis, S. and Antoniou, G. (2014) 'Representing and reasoning over spatial relations in OWL: a rule-based approach', *W3C/OGC Workshop on Linking Geospatial Data (LGD'14)*, Lightning Talks, London.

Batsakis, S., Euripides, G. and Petrakis, M. (2011) 'SOWL: a framework for handling spatio-temporal information in OWL 2.0', *Proceedings of the 5th International Conference on Rule-based Reasoning, Programming, and Applications, RuleML 2011*, Barcelona, Spain.

Bizer, C. (2004) 'D2RQ – treating non-RDF databases as virtual RDF graphs', *Proceedings of the 3rd International Semantic Web Conference (ISWC2004)*, Hiroshima, Japan.

Buttigieg, P., Morrison, N., Smith, B., Mungall, C., Lewis, S. and Consortium, E. (2013) 'The environment ontology: contextualising biological and biomedical entities', *Journal of Biomedical Semantics*, Vol. 4, p.43.

Chen, H. and Wu, Z. (2005) 'DartGrid III: a semantic grid toolkit for data integration', *SKG '05. First International Conference on Semantics, Knowledge and Grid*, IEEE, Beijing, China.

Christodoulou, G., Petrakis, E.G.M. and Batsakis, S. (2012) 'Qualitative spatial reasoning using topological and directional information in OWL', *Proceedings of 24th International Conference on Tools with Artificial Intelligence (ICTAI 2012)*, 7–9 November, IEEE Computer Society, Washington, DC, pp.596–602.

Corby, O., Gaignard, A., Faron-Zucker, C.M. and Montagnat, J. (2012) 'KGRAM versatile data graphs querying and inference engine', *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence, WI 2012*, IEEE Computer Society, Macau, China.

Cruz, I.F. and Xiao, H. (2005) 'The role of ontologies in data integration', *Journal of Engineering Intelligent Systems*, Vol. 13, pp.245–252.

Cruz, I.F. and Xiao, H. (2009) 'Ontology driven data integration in heterogeneous networks', *Complex Systems in Knowledge-based Environments: Theory, Models and Applications*, Springer, Berlin, Heidelberg, pp.75–98.

Durbha, S.S., King, R.L., Shah, V.P. and Younan, N.H. (2009) 'A framework for semantic reconciliation of disparate earth observation thematic data', *Computers & Geosciences*, Vol. 35, No. 4, pp.761–773.

Egenhofer, M.J. and Herring, J.R. (1990) *Categorizing Binary Topological Relations between Regions, Lines, and Points in Geographic Databases*, Technical Report, Department of Surveying Engineering, University of Maine.

Frasincar, F., Milea, V. and Kaymak, U. (2010) 'tOWL: integrating time in OWL', in de Virgilio, R., Giunchiglia, F. and Tanca, L. (Eds): *Semantic Web Information Management*, Springer, Berlin, Heidelberg, pp.225–246.

Garbis, G., Kyzirakos, K. and Koubarakis, M. (2013) 'Geographica: a benchmark for geospatial rdf stores', *CoRR*, Vol. 8219, pp.343–359.

Gray, A.J., Gray, N. and Ounis, I. (2009) 'Can RDB2RDF tools feasibily expose large science archives for data integration?', *Proceedings of the 6th European Semantic Web Conference on the Semantic Web: Research and Applications*, SpringerVerlag, Heidelberg, Berlin, pp.491–505.

Green, J., Hart, G., Dolbear, C., Engelbrecht, P.C. and Goodwin, J. (2008) 'Creating a semantic integration system using spatial data', in Bizer, C. and Joshi, A. (Eds): *International Semantic Web Conference*, Karlsruhe, Germany.

Gutierrez, C., Hurtado, A. and Vaisman, A. (2007) 'Introducing time into RDF', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19, pp.207–218.

Harbelot, B., Arenas, H. and Cruz, C. (2013) 'Continuum: a spatio-temporal data model to represent and qualify filiation relationships', *Proceedings of the 4th ACM Sigspatial International Workshop on Geostreaming*, ACM, Orlando, FL, pp.76–85.

Harbelot, B., Arenas, H. and Cruz, C. (2014) 'un modèle sémantique spatio-temporel pour capturer la dynamique des environnements', *14 ème conférence Extraction et Gestion des Connaissances*, Rennes, France. [In French]

Hobbs, J.R. and Pan, F. (2004) 'An ontology of time for the semantic web', *ACM Transactions on Asian Language Information Processing*, Vol. 3, pp.66–85.

Karmacharya, A., Cruz, C., Boochs, F. and Marzani, F. (2010) 'Use of geospatial analyses for semantic reasoning', in Setchi, R., Jordanov, I., Howlett, R. and Jain, L. (Eds): *Knowledge-based and Intelligent Information and Engineering Systems*, Vol. 6276, pp.576–586.

Klein, M. and Fensel, D. (2001) 'Ontology versioning on the semantic web', *Proceedings of the First International Semantic Web Working Symposium SWWS'01*, Standford University, CA, pp.75–91.

Knublauch, H., James, A.H. and Kingsley, I. (2011) *Spin – Overview and Motivation*. Available online at: http://www.w3.org/Submission/spin-overview/

Kyzirakos, K., Karpathiotakis, M. and Koubarakis, M. (2012) 'Strabon: a semantic geospatial DBMS', *The Semantic Web – ISWC 2012*, Springer Berlin Heidelberg, Boston, MA, pp.295–311.

Langran, G.E. and Chrisman, N.R. (1998) 'A framework for temporal geographic information. Cartographica', *The International Journal for Geographic Information and Geovisualization*, Vol. 25, No. 3, pp.1–14.

Lenzerini, M. (2002) 'Data integration: a theoretical perspective', *PODS'02: Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, Madison, WI, pp.233–246.

Lin, K. and Lud Ascher, B. (2003) 'A system for semantic integration of geologic maps via ontologies', *Proceedings of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data (SCISW)*, 20 October, Sanibel Island, FL.

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D. and Villa, F. (2007) 'An ontology for describing and synthesizing ecological observation data', *Ecological Informatics*, Vol. 2, No. 3, pp.279–296.

Mainas, N. and Petrakis, E.G.M. (2014) 'CHOROS 2: improving the performance of qualitative spatial reasoning in OWL', *2014 IEEE 26th International Conference on Tools with Artificial intelligence (ICTAI)*, Limassol, pp.283–290.

Meiri, I. (1996) 'Combining qualitative and quantitative constraints in temporal reasoning', *Artificial Intelligence*, Vol. 87, Nos. 1–2, pp.343–385.

Noy, N. and Rector, B. (2006) *Defining N-ary Relations on the Semantic Web*, W3C Working Group Note 12. Available online at: http://www.w3.org/TR/swbp-n-aryRelations/

O'Connor, M. and Das, A. (2011) 'A method for representing and querying temporal information in owl', in Fred, A., Filipe, J. and Gamboa, H. (Eds): *Biomedical Engineering Systems and Technologies: Third International Joint Conference, BIOSTEC 2010*, Valencia, Spain, No. 127 in Communications in Computer and Information Science, Springer, Berlin, pp.97–110.

Open Geospatial Consortium (2006) *An Introduction to GeoRSS*, Whitepaper, Wayland, MA.

Patroumpas, K., Giannopoulos, G. and Athanasiou, S. (2014) 'Towards GeoSpatial semantic data management: strengths, weaknesses, and challenges ahead', *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '14)*, ACM, New York, pp.301–310.

Plumejeaud, C., Mathian, H., Gensel, J. and Grasland, C. (2011) 'Spatio-temporal analysis of territorial changes from a multi-scale perspective', *International Journal of Geographical Information Science*, Vol. 10, pp.1597–1612.

Pokharel, S., Sherif, M.A. and Lehmann, J. (2014) 'Ontology based data access and integration for improving the effectiveness of farming in Nepal', *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Warsaw, pp.319–326.

Polleres, A., Schareffe, F. and Schindlauer, R. (2007) 'SPARQL++ for mapping between RDF vocabularies', *Proceedings of the 6th International Conference on Ontologies, DataBases, and Applications of Semantics, ODBASE 2007*, Vilamoura, Portugal, Vol. 4803 of LNCS, Springer, pp.878–896.

Randell, D.A., Cui, Z. and Cohn, A.G. (1992) 'A spatial logic based on regions and connection', *Proceedings 3rd International Conference on Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, CA, pp.165–176.

Ressler, J., Dean, M. and Kolas, D. (2010) 'Geospatial ontology trade study', *Proceedings of the 2010 conference on Ontologies and Semantic Technologies for Intelligence*, IOS Press, Amsterdam, The Netherlands, pp.179–211.

Roussey, C., Pinet, F., Kang, M.A. and Corcho, O. (2011) 'Ontologies for interoperability', *Ontologies in Urban Development Projects*, Springer, London, pp.39–53.

Schenk, S. and Staab, S. (2008) 'Networked graphs : a declarative mechanism for SPARQL rules, SPARQL views and RDF data integration on the web', *Proceedings of the 17th International Conference on World Wide Web, WWW 2008*, ACM, Beijing, China, pp.585–594.

Stocker, M. and Sirin, E. (2009) 'PelletSpatial: a hybrid RCC-8 and RDF/OWL reasoning and query engine', *CEUR Workshop Proceedings*, Vol. 529, pp.2–31.

Vandecasteele, A. and Napoli, A. (2012) 'Spatial ontologies for detecting abnormal maritime behaviour', *OCEANS 2012 MTS/IEEE Yeosu Conference: The Living Ocean and Coast – Diversity of Resources and Sustainable Activities*, Yeosu, South Korea.

Vilain, M.B. (1982) 'A system for reasoning about time', *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, Los Altos, CA, pp.197–201.

Vilain, M.B., Kautz, H. and Beek, P. (1986) 'Constraint propagation algorithms for temporal reasoning', *Readings in Qualitative Reasoning about Physical Systems*, Morgan Kaufmann, San Mateo, CA, pp.377–382.

Wache, H., Vogele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H. and Hubner, S. (2001) 'Ontology-based integration of information – a survey of existing approaches', *IJCAI-01 Workshop: Ontologies and Information*, Washington, DC, pp.108–117.

Welty, C. and Fikes, R. (2006) 'A reusable ontology for fluents in owl', *Proceedings of the Conference on Formal Ontology in Information Systems*, IOS Press, Amsterdam, The Netherlands, pp.226–236.

Zhao, T., Zhang, C., Wei, M. and Peng, Z-R. (2008) 'Ontology-based geospatial data query and integration', *GIScience '08 Proceedings of the 5th international conference on Geographic Information Science*, Park City, UT, pp.370–392.

Notes

1 Research Object: http://www.researchobject.org/initiative/

2 Data One: https://www.dataone.org/

3 Research Data Alliance: https://rd-alliance.org/

4 ENVRIplus project: http://www.envriplus.eu/

5 ENVO ontology: http://environmentontology.org/

6 OBEO ontology: https://semtools.ecoinformatics.org/oboe

7 The National Centre for Scientific Research: http://www.cnrs.fr/

8 Zone Ateliers program: http://www.za-inee.org

9 Database recording the identification of agricultural parcels for the French government

10 RPG in 2012: http://www.geoportail.gouv.fr/donnee/251/registreparcellaire-graphique-rpg-2012

11 Workshop website: http://www.za.plainevalsevre.cnrs.fr/

12 Centre for Biological Studies of Chizé: http://www.cebc.cnrs.fr/

13 OWL-Time ontology: http://www.w3.org/2006/time

14 SWRL Temporal Ontology: http://swrl.stanford.edu/ontologies/built-ins/3.3/temporal.owl

15 GeoRSS: http://www.georss.org/

16 SWRL language: http://www.w3.org/Submission/SWRL/

17 Pellet engine: http://clarkparsia.com/pellet/

18 SPIN – SPARQL Inferencing Notation: http://spinrdf.org/

19 SPARQL Update: https://www.w3.org/TR/sparql11-update/

20 GeoSPARQL: http://www.opengeospatial.org/standards/geosparql

21 stSPARQL: http://www.strabon.di.uoa.gr/stSPARQL

22 uSeekM: http://dev.opensahara.com/projects/useekm/

23 Parliament: http://parliament.semwebcentral.org/

24 Strabon: http://strabon.di.uoa.gr/

25 D2RQ framework: http://d2rq.org/

26 Jena framework: http://jena.apache.org/

27 OpenLayers library: http://openlayers.org/

28 OpenStreetMap website: http://www.openstreetmap.org