Dipartimento di Ingegneria e Scienza dellInformazione

– KnowDive Group –

Spatial modelling extension for protected areas within OSM

Document Data:	Reference Persons:
14 December 2018	Nicolò Alessandro Girardini - 203265
	Diego Lobba - 198331

Index:

1 Scenario **1**

2 Model Formalisation **2**

2.1 Standards Followed 2

2.2 Data Analysis & Preparation 2

2.3 Ontology Contruction 3

2.3.1 Classes Definition 3

2.3.2 Object Properties Definition 4

2.3.3 Data Properties Definition 5

3 Lexical Information **7**

3.1 Data Properties 7

3.2 Object Properties 8

3.3 Classes 8

4 TopLevel Grounding **9**

5 Model Visualisation **11**

6 Generalised query **13**

7 Final considerations **15**

A Informal modelling **16**

B Lexical information mapping **17**

C Evaluation Model **20**

1 Scenario

Our service aims at providing users with a tool for a smoother tourist experience targeting protected areas within Italy. The system is designed to provide answers to common user needs that can be faced during a vacation. Namely, it is focused on transports' travelling timings by exploiting vicinity relationships between areas of interest that can be reached by public means of transportation. This features can be helpful in scheduling decisions when tourists have limited resources available in terms of time and the number of places they can visit must be weighted by factors like the quality of the transportation service and the vicinity to the place of interest.

A series of sample scenarios is now given, to put emphasis on the set of tasks the system is going to help the user with. In the first scenario, we consider Mark, a teenager who is travelling across Europe by train, who arrives in Trento for a break of two days before continuing with its trip. Mark, who loves walking in the nature, wants to visit the national parks around Trento and the set of lakes hosted in the city neighbourhood. By being without private vehicles, Mark wants to reach these destinations by using public means of transportation.

In the second scenario it is considered Pascal, who is a scuba diver and wants to schedule a trip on the most famous maritime sites in Italy. Due to diving recovery time, Pascal has to wait a certain amount of time before being able to travel by airplane again. Therefore he has to take small breaks that allow him to also visit neighbouring areas. Thus Pascal would like to have an advice that allows him to plan the trip in such a way to reduce the means of transportation used to reach each site, considering airports, seaports and cableways in order to reach mountain lakes too. Additionally he would like to have suggestions on local areas that he can visit during the recovery periods.

In the final scenario modelled, it is considered Grazia, an employee within the public administration of the municipality of Agrigento, who according to a regional plan wants to enforce the public transportation service to reach the regional park and maritime areas in the region of Sicily. Her needs are than linked to the availability of transportation around those protected areas: the region is historically weak in the sector and so improving lines that lead to hardly reachable areas will improve tourists' experience and make them willing to come back again to visit the remaining areas of the region. Grazia's job would be really simplified if she already had a service able to identify those links between protected areas and public transportation.

By considering the scenarios described we can summarise the operations the system is going to cover as:

- Given an administrative area (region, nation, province), lists protected areas and water spots within the given boundary.
- By defining protected areas and water spots as general point of interests, it is desirable to list all the means of transportation departing in the neighbourhood of the point of interest.
- Given a geographic reference a good service would be to provide a schedule for the user in order for him to visit a series of protected areas within a range of distance from the reference point.

2 Model Formalisation

2.1 Standards Followed

According to the methodology after the definition of the scenarios of our task we proceeded by examining what the available data and specifications were.

Regarding protected sites we found the INSPIRE documentation [4] providing basic information about a global standard to model protected areas according to the different set of specifications adopted at both national and international level. INSPIRE's implementation is the WDPA (World Database on Protected Areas), which can be found on the website [11]. For our scenario we focused on the Natura2000 directive[1] targeting European areas, which will be additional information integrated with WDPA.

For the building of the ontology the only good standard we found was SmOD INSPIRE Vocabularies [10], which was good, but very focused on protected sites. We decided then to keep some of the definitions, like *SpatialObject* and obviously *Species* and *ProtectedSite*, while discarding others that were not compliant to the other specifications we had (Open Street Map).

Transportation information instead has been obtained through the freely accessible OSM[7] database. In particular we restricted our focus on any transportation service, either public or private, available within an administrative division.

2.2 Data Analysis & Preparation

The building of the ontology had to be compliant to all the standards we decided to follow. Datasets available have been inspected in order to have a glimpse on what the available information were, that would be relevant for our task and that it would be possible for them to be integrated according to some common attribute.

We started first by considering data available related to protected areas (hence data coming from WDPA and Natura2000 databases) and delineated a first set of entities that it was possible to model out of this data, that were suitable for answering our queries. We distinguished among the following basic entities:

- Protected Site: representing the area which is pivoting for the resolution of our queries;
- Managing Authority: the authority, in general, supervising the protected site. We decided to include this information for our application so that it would be possible for the user to obtain some contact information and therefore increasing the possible use of the information she can gain from the query;
- Species: though not required from the task was an information that was directly available from the Natura2000 dataset and that could provide not only a valuable additional information, but also a crucial feature that could make our work unique in its class.

Data coming from Open Street Map has been examined indipendently. The variety and the loose structure of data is the main issue during the modelling phase. We finally decided to use this source only for means of transportation, while Natura2000 and WDPA datasets will be used for what concerns protected areas.

OSM extracts have passed through a pre-processing phase, where only nodes matching a predetermined set of tags related to transportation have been kept. This phase leads to data for which it was not possible to obtain the context surrounding nodes. For instance, given a node defined by a longitude and a latitude value, it was not possible to determine where the node lied with regards to an administrative division.

For this reason a *Geographic Information System* (GIS) tool, namely QGIS[8], has been used to load nodes together with the shapefile related to administrative divisions, both provinces and regions, of Italy. A spatial join has been applied in order to join information of nodes intersecting a given administrative division. Out of this process it was possible to have a *csv* file containing the province and the region where a node lies.

Finally, results obtained out of the previous process have been through an additional processing phase, where OSM tags have been manipulated to merge similar information and to clearly define the scope of each node in terms of the concepts required for our application. In particular, each node's tags have been analysed in order to distinguish to which particular transportation service the node referred to.

From the processed file related to OSM data we could have a clearer idea on what different transportation services were, and which attributes could have been of interest for our purposes. It is clear that the main concept underlying this information is related to the notion of mean of transportation, but indeed it would be wrong to assign such a title, since data are not describing neither vehicles nor stations (in particular). Data represent *nodes*, with a specific geographical location, where a specific transportation service is provided. We therefore felt more convinced to describe these as *Transportation Points*. A transportation point, as mentioned before, has a location expressed by a longitude-latitude pair and a name which identifies it within the type of service provided through it (for instance the name of a certain bus stop). A transportation point is not a self-standing entity in general, but it is a node of a more complex transportation network. We therefore defined the *operator*, and some contact information in order for users of our application to have all the information required to plan a trip.

2.3 Ontology Contruction

Having processed data in such a way allowed us to have a better understanding of all the information we could gather through all these sources. We built an informal model (see appendix A) merging knowledge encoded in our datasets. Thanks to the ER model we were able to identify all the classes needed to design our *domain ontology*.

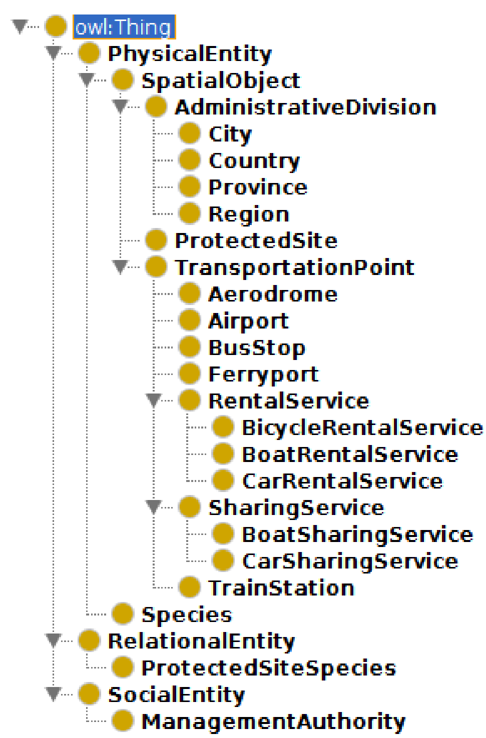

Figure 1: Entities hierarchy identified for our application

2.3.1 Classes Definition

The first classification needed was the distinction between *PhysicalEntity*, *RelationalEntity* and *SocialEntity*. The last one having just the leaf *ManagementAuthority*. A PhysicalEntity denotes any entity that is a concrete element in the environment, either by being a place or by being an animated being. A SocialEntity, instead, denotes legal entities: the one needed for our task is *ManagementAuthority*, which is the organisation in charge of the administration of protected sites.

PhisicalEntity is then split into two broad concepts, namely *Species* and *SpatialObject*. A SpatialObject is any entity for which a location information is defined. Therefore *ProtectedSite*, *AdministrativeDivision* and *TransportationPoint* are all subclasses of SpatialObject. On the other hand, *Species*, representing animals' species hosted within a protected site, is on the same level of the hierarchy, child of PhisicalEntity. Animals described by this entity are representative both with regards to the *Habitats directive*[3] and to the *Birds directive*[2], for which Natura2000 keeps status information.

The *ProtectedSite* class represents the core concept in our task. It merges information related to WDPA and Natura2000 sites. The merge has not been an easy task, since there is no direct mapping between WDPA entries and the corresponding Natura2000 information. Therefore we were able to obtain a single entity by expecting the merge to work both on the site designation and on the local name of the site, which is data available in both datasets. Due to the fact that we based our class mainly on Natura2000, it was not possible to perform a complete integration with the whole WDPA data. Still, this is not an issue for our task, since the more datasets related to different site designations are considered, the better the integration result one can obtain through the use of the WDPA standard. This means one could add knowledge to our result by merging datasets following standards compliant to INSPIRE.

AdministrativeDivision is used to represent geopolitical boundaries. This comprises entities such as *Country*, *Region*, *Province* and *City*. The Region entity is the most relevant for our task, due to the fact that it is used as reference administrative division in which a protected site lies in. Province instead is used as reference to the location for a particular transportation point.

The final entity, for what concerns this level in the hierarchy, is *TransportationPoint*. A TransportationPoint is any geographical location, described by means of coordinate values, that identifies a place where it is possible to have a transportation service. Indeed, this entity was the hardest to map due to the enormous variety of the possible configurations a transportation point can manifest. For instance, we chose to perform a first split basing the distinction from transportation services that are commonly agreed to be ”private” or ”public”. This first attempt of classification was wrong, due to the dynamic behaviour of ownership in the transportation sector. We then kept a more coarse-grained classification by just dividing the services between *RentalService* and *SharingService* in addition to classical transport vehicles used by commuters, such as *BusStop*, *TrainStation*, *FerryPort* and also *Aerodrome* and *Airport*.

Finally the RelationalEntity represents entities that are just used to describe relationships which possess attributes, and therefore cannot be represented directly with a simple object property. This group contains only the *ProtectedSiteSpecies* class, which describes properties related to the relationship between species housed within a protected site.

The overall classification of the classes resulting from this phase is depicted in figure fig. 1.

2.3.2 Object Properties Definition

In the object property hierarchy, we defined properties that are exploited to bind class entities together by means of a syntactical meaning. We therefore reproduced an overall structure resembling the class hierarchy previously described.

In *TransportationProperty* it is found the *Connected* property, which is used to map a TransportationPoint to any SpatialObject, this is indeed a very broad object property that we though was convenient to add, but we specialised this relation to be specifically

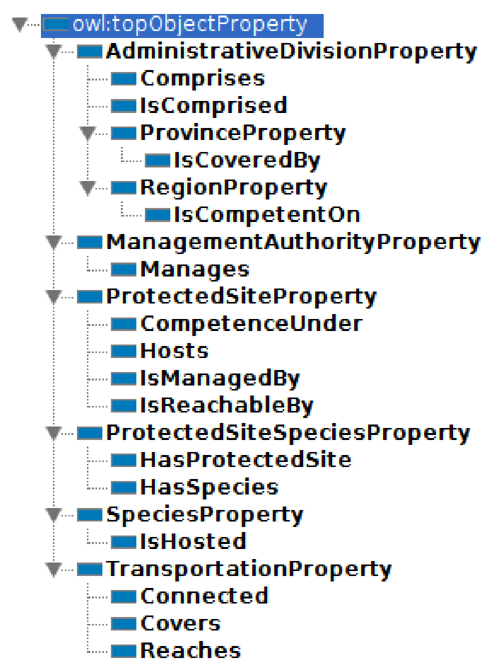

Figure 2: Object properties hierarchy identified for our application

related to protected sites through the *Reaches* property. We then added the *IsCoveredBy* property to map the transportation point to the province it lies in. This two properties allow to connect a TrasportationPoint to a ProtectedSite, enabling us to answer customer's queries about the nearby sites, and to connect it to the territory it lies within (namely the province).

SpeciesProperty contains the *IsHosted* property which is used to associate the protected site where the species can be found. To map this concept, due to the fact that the relationship has a property, namely the abundance category of the species within a particular protected site, this object property ranges to a *ProtectedSiteSpecies* class, which describes the relatioship itself. This pattern was required since it is not possible to add information to an object property directly, and it is described in [5].

The *ManagementAuthority* property contains the *Manages* property, used to identify the management authority entitled for the administration of a protected site.

The *AdministrativeDivisionProperty* property contains the mapping required to bind the various administrative divisions one another, making up the various sub-national divisions. The mapping is given using the *Comprises* and *IsComprised* properties which generally map an AdministrativeDivision to another. A more fine-grained control is then realised in the AdministrativeDivision's subclasses by means of constraints. We have that: a City `IsComprised` within a Province; a Province `IsComprised` within a Region and `Comprises` some City; a Region `IsComprised` within a Country and `Comprises` some Province; a Country `Comprises` some Region. In addition to these properties, there is a subgroup for the *RegionProperty* which includes the *IsCompetentOn* property used to map the Region to a ProtectedSite. On the same level there is the *ProvinceProperty* subgroup including the *IsCoveredBy* property, used to map TransportationPoints instances belonging to a specific Province.

The *ProtectedSite* property comprises the largest set of object properties, since the ProtectedSite is bound directly to several entities, namely: Species, ManagementAuthority, TransportationPoint and Region.

Respectively, the *Hosts* property is used to define the Species present within a protected site. In a similar way to the IsHosted property for Species, the object property ranges to ProtectedSiteSpecies instances. The *IsManagedBy* property defines the ManagementAuthority managing the area; *IsReachableBy* denotes the TransportationPoint that can be used to reach the site; finally the *CompetenceUnder* property maps the ProtectedSite to the Region it lies in.

Eventually we have the *ProtectedSiteSpecies* property comprising the *HasSpecies* and *HasProtectedSite* object properties that are used to establish the binds between the relationship entity and the corresponding arguments.

All the hierarchy can be seen in figure fig. 2.

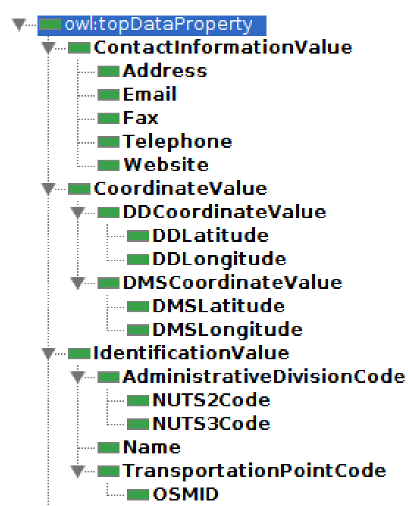

Figure 3: Data properties hierarchy identified for our application

2.3.3 Data Properties Definition

Data properties have been used to provide a more limited, hence precise, range of attributes defining the class entities.

IdentificationValue property is used to generally identify the entity, by means of some name. Each class has a *Name* which identifies the entity. To provide this broad range to the Name attribute, we decided to set this constraint to the *owl:Thing* entity, from which all others derive. Within the IdentificationValue, we found the *AdministrationDivisonCode* which comprises the *NUTS2Code* and *NUTS3Code* properties, which are used to uniquely identify Region and Province entities, respectively. The *TransportationPoint-*

Code contains *OSMID* identifying the node id of the corresponding object within the OpenStreetMap database.

The *ContactInformationValue* properties comprise the *Address*, *Email*, *Fax* and *Telephone* and *Website* values used both for ManagementAuthority and TransportationPoint, since it contains information about the service operator.

For the Species we defined the *SpeciesValue* properties. A species is characterised by an *Natura2000SpeciesID* which is the ID used within the Natura2000 to uniquely identify the species, the *Supergroup* which is the animated beings class the species belongs to, and *AbundanceCategory* which represents the rarity of the species with respect to the protected site considered. To represent this we created a new dataype, namely *psso:abundance*, which can assume the following values: `common`, `present`, `rare`, `unknown`, `very rare`. The AbundanceCategory is contained within the relationship entity ProtectedSiteSpecies.

Regarding the location of the SpatialObjects we defined *CoordinateValue*, distinguishing between *DDCoordinateValue* and *DMSCoordinateValue*. The first is the latitude and longitude representation with decimal degrees, while the latter is represented with degrees, minutes, seconds format.

Richer is the *ProtectedSiteValue* node. Here we needed to model: *AreaCovered* as the total area of the site; *Designation* as the official designation of the site; *LocalName* as the site's name in local language; *MarineAreaCovered* as area covered by the sea in case of marine or coastal sites; *PSType* which is the type of the site (with values `Terrestrial`, `Coastal`, `Marine`); *Status* and *StatusYear* model the ProtectedSite status, so wether it is `Proposed`, `Inscribed`, `Adopted`, `Designated`, `Established`. Regarding the ids we get from the two datasets we decided to keep them both (*Natura2000PSID* and *WDPAID*), making easier future extensions or mapping to the data.

Finally the *TransportationPointValue* properties are used by TransportationPoint instances and contains the *Operator* related to the transportation service. RentalService instances have the *OpeningHours* property under the *RentalServiceValue*. In particular, for busses there is the properties group *BusStopValue*, describing the bus lines crossing the transportation point (*RouteLine*, given by OSM), the availability of a shelter (*Shelter*) and the possibility to pick up people having motor dysfunction by the *Wheelchair* data property.

The whole hierarchy can be seen in figures fig. 3 and fig. 4.

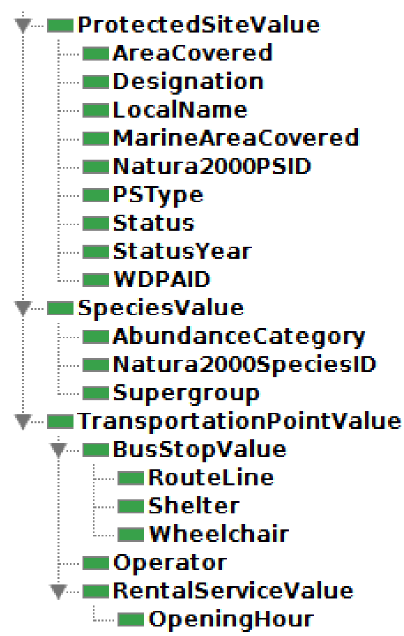

Figure 4: Data properties hierarchy identified for our application

3 Lexical Information

Here it is described the mapping process between our domain entities and properties to Wordnet *synsets* which allows to extend the syntactical knowledge and reusability of our model. To perform the mapping, we searched for our domain entity within Wordnet, we identified the synset ID and we put this ID as *synsetID* as annotation to the corresponding entity in the model (see appendix B). When we couldn't be able to identify a proper mapping we created a new synset, setting the value to *-1*. We also specified the part-of-speech tagging (*posTag*), the meaning of the word (*generalConcept*) and the ID of the parent if the concept had to be added in Wordnet tree (*parentID*).

3.1 Data Properties

We will start from the data properties. We didn't mapped the roots related to classes because it was not really useful for the lexical understanding, they are there just for the logical structure of the properties.

The *CoordinateValue* has been assigned to the synset related to the term *"coordinate"*, since it generally represents a value within a coordinate system. In Wordnet we didn't find any reference neither to the notion of the Degree-Minute-Second representation nor to the Decimal-Degree representation, therefore we created two new synsets. Both *DDLatitude* and *DMSLatitude* have been mapped to the same latitude definition. The same goes for the *DDLongitude* and *DMSLongitude*.

We then moved to the *IdentificationValue* which we mapped to the term *"identifier"*, since all its subproperties identify some entity. Subproperties of IdentificationValue are quite domain-specific and we were not able to define any synset for *AdministrativeDivisionCode*, *NUTS2Code*, *NUTS3Code*, *TransportationPointCode* and *OSMID*. For these properties, the best equivalent term is still *"identifier"*, so we created new synsets. Finally, *Name* has been mapped to the meaning of *"name"* defined as the language unit for which an entity is known.

We then proceeded by mapping the *ContactInformation* to the *"contact"* term: here this term is referred to any means of communication or access. Within this data property there are *Address* that we mapped to the term *"address"*, describing the location a entity can be found. For both *Email* and *Fax* we were not able to find any synset describing the fact that these are addresses: they only describe these terms as the phenomenon and the object. We therefore decided not to map these terms with any existent synset. We found the desired synset for *Telephone*, that we decided to map to *"phone number"*. For *Website* we decided to keep the definition of *"website"*.

With regards to properties related to protected sites, both *AreaCovered* and *MarineAreaCovered* have been mapped to the term *"surface area"*, which clearly express it as the extent of a 2D area enclosed within a boundary. For *Designation* it has been chosen the synset of *"designation"* comprising terms like "appellation" and "denomination": these should highlight the fact that the designation for the site has been given according to a directive. The *Status* property describes the status of the designation process the protected site currently is. Therefore we mapped the property to the *"status"* noun, which generally defines the condition of an entity. Since properties *PSType*, *StatusYear*, *WDPAID* and *Natura2000PSID* are very domain-specific, we were not able to find any meaningful mapping, so new synsets were needed. Regarding the *LocalName* the term chosen for the mapping is *"toponym"* which has the following definition: it is the name by which a geographical place is known.

For the properties related to species, the *Natura2000SpeciesID* property was impossible to map, since it was too tied to our domain. The *AbundanceCategory* property has been mapped to the noun *"rareness"*, which indeed can be used to denote the scarcity of a particular species within a given protected site. Finally, the *Supergroup* data property has been mapped to the *"family"* noun, namely to the synset related to the biological taxonomic group comprising one or more genera: examples are"fish family", "birds family", "mammal family" and so on.

Eventually we mapped *TransportationPointValue* properties. *Operator* has been mapped to *"operator"*, which is someone who owns a business. We had to create a new synset for *OpeningHours*, to describe intervals of time defined by hours in which the service is available. We then mapped *RouteLine* to *"itinerary"*, denoting an established line of travel, *Shelter* to its physical meaning *"shelter"* (protective covering) and we created a new synset for *Wheelchair* since the synset described just the object, which was a concept too far from the service provided by busses.

3.2 Object Properties

We start from the *AdministrativeDivisionProperty*. The *Comprises* and *IsComprised* properties have been mapped to the *"comprise"* term, in the synset describing the "part of" meaning of the noun.

The *IsCompetentOn* property and the *CompetenceUnder* property have been assigned the *"competent"* synset that represents a legally qualified entity. In out domain, a protected site can span different regions, but only one of the is competent for the protected site, meaning that administratively the site is assigned to that region.

The *Manages* and *IsManagedBy* properties are mapped to the *"manage"* synset describing the set of verbs whose meaning is to supervise on something: the synset comprises terms like "oversee", "supervise" and "manage".

For *Hosts* and *IsHosted*, we opted for the verb *"house"*, which best describes the act of accomodating species.

The *"accessible"* verb has been assigned to *IsReachableBy* property, in particular the synset picked is related to the property of a site to be capable of being reached by a given transportation point.

Finally, the *Connected* property has been mapped to the *"connected"* verb, to denote that any two spatial object entities can be connected in some ways.

3.3 Classes

Here the mapping of synsets to classes is discussed, starting from the SocialEntity The *ManagementAuthority* has been mapped to the term *"management"*, which describes the organisation responsible for running a business.

We decided to map *RelationalEntity* with *"relation"*, to simply highlight its abstract nature, also fit for *ProtectedSiteSpecies* (which has more lexical information looking at the object properties related to it).

ProtectedSite has been mapped to *"park"* which refers to an area preserved in its natural state as public property.

For *Species* we decided to map the term to the noun *"species"*, denoting a taxonomic group in general.

For the *AdministrativeDivision* class we used the noun *"division"*, defined as an administrative unit in government. Then the general rule used to map subclasses of AdministrativeDivision was to select the political meaning of the subarea. Therefore *Country* has been mapped with the concept of territory occupied by a nation, and *City* to the definition of an incorporated administrative district. The difficult part here was to choose the mapping for *Region* and *Province* since the difference can be subtle and very bound to the state (in our case Italy). To take into consideration these issues we have assigned to Region the term *"state"*, defined as the territory occupied by one constituent administrative district of a nation. In the same synset it is found the also noun "province". The Province class instead has not been assigned to a new synset, since we found no equivalent term.

For *TransportationPoint* the noun *"transportation system"* was used. For these classes, we have chosen to assign entities to the kind of service available in that point. *Aerodrome* and *Airport* have been assigned the same synset, defining generally an airfield where accomodations for passengers are provided. *BusStop*, *TrainStation* were mapped to the corresponding, unique, synsets. *Ferryport* has been mapped to the noun *"seaport"* instead.

Finally, we conclude the lexical mapping with the assignment of synsets to the *SharingService* and the *RentalService*. The latter has been mapped to the *"rental"* noun, denoting a property that could be rent. Same goes with SharingService, mapped to *"sharing"*.

4 TopLevel Grounding

In section it is described the process by which our domain specific ontology has been mapped to the CSK (Common Sense Knowledge) ontology. The CSK mapping grants an increased reusability over the concepts introduced by our model, in addition to check the foundations of our own model.

The mapping process has required a slight modification in our hierarchy, with respect to what shown in fig. 1. Namely, following the four core operations (subsumption, equivalence, addition, deletion), we had mainly to substitute the common entity types we created, using instead the concepts retrieved by the CSK. Consequently we needed also to find a new mapping with Wordnet for those involved entities. We felt that the process has been extremely easy to do, and we are therefore confident that the model we created originally was describing correctly the entity hierarchy.

Regarding the object properties we did not do a mapping, because of how specific our properties are: in CSK concepts are very general as it is a top level ontology, while our model is too domain specific to find a good mapping to the top level.

The *ManagementAuthority* has been mapped to the *Organisation* class of the CSK, which has been considered equivalent to our *SocialEntity*, describing any group defined by an organisational structure. With this definition, provided by the CSK, one could quite argue the decision of this choice, but by looking at the new mapping with Wordnet needed, we found a more convincing definition for the term *"organisation"*: *"the persons (or committees or departments etc.) who make up a body for the purpose of administering something"*.

The *Species* class has been mapped to the *LivingBeing* entity in the CSK with a subsumption: alongside the changes needed for the SpatialObject we decided here to delete the entity *PhysicalObject*, which was superfluous. This mapping is very intuitive, although we argue that our term could be used even as parent of *Person* (humans are after all an animal species) already present in CSK. We decided by the way to maintain the CSK structure and map Species as Person's sibling. The lexical concept of Wordnet chosen for LivingBeing is *being*: in fact its synset comprises also "organism", which is exactly a living thing that can act indipendently.

We proceed the mapping with the *SpatialObject*, which is equivalent to the *Location* concept (and so it has been substituted by it). Indeed every SpatialObject is a Location and therefore we were quite convinced with the mapping, but we still have that a TransportationPoint provides a *Service*. Thus, one could argue that the choice has not been the most appropriate, but the coverage we obtained by the single mapping with Location seems to support our thesis. Location has been mapped to the synset of *"location"* itself, denoting a point or extent in space.

Finally we have the *RelationalEntity*, for which we did not find any concept appropriate for the mapping. We are not quite sure whether our approach was different from the methodology expected by CSK or if it is actually rare to find this scenario. We built our solution following [5] and worked around this mapping problem by letting RelationalEntity as one of the top-level concepts.

The final mapping is depicted in fig. 5. Boxes in blue denotes domain-specific entity classes, concepts used by our model. Yellow boxes describe top-level entity types that were originally proposed in our hierarchy and persist also after the mapping. Green boxes are CSK concepts, mostly leaves with respect to the hierarchy, describing the new common types abstracting our domain concepts.

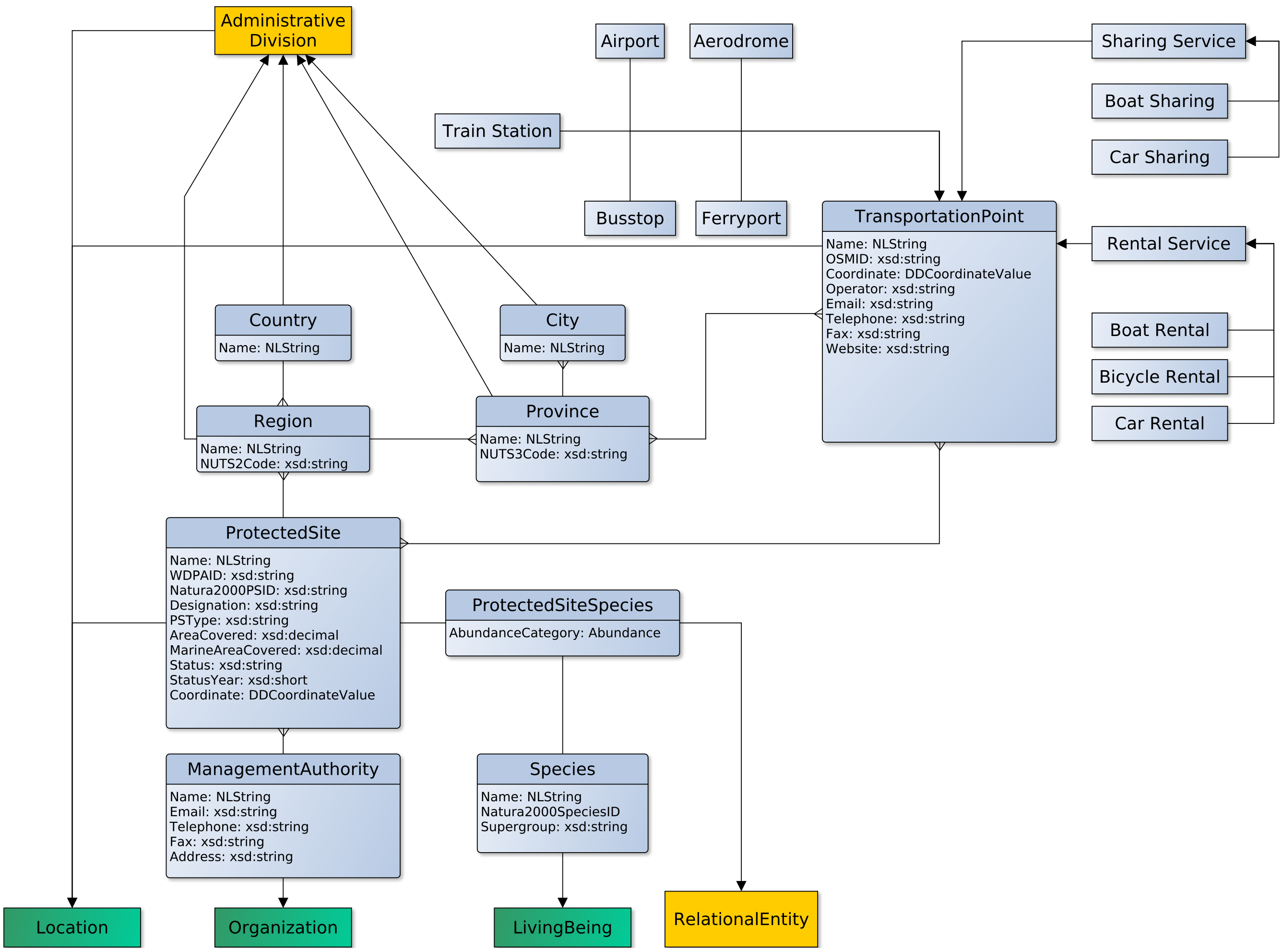

Figure 5: Mapping of the CSK leaves to the formal model roots.

5 Model Visualisation

In fig. 6 it is depicted the final formal model for our application, as it is given by the Ontograf plugin within Protegè [9]. As it is possible to see from the links, the most connected classes are the ProtectedSite and TransportationPoint entities, which are the core entities in our service.

This graphs is basically the Ontograf view of what has been already explained in 2 and 4.

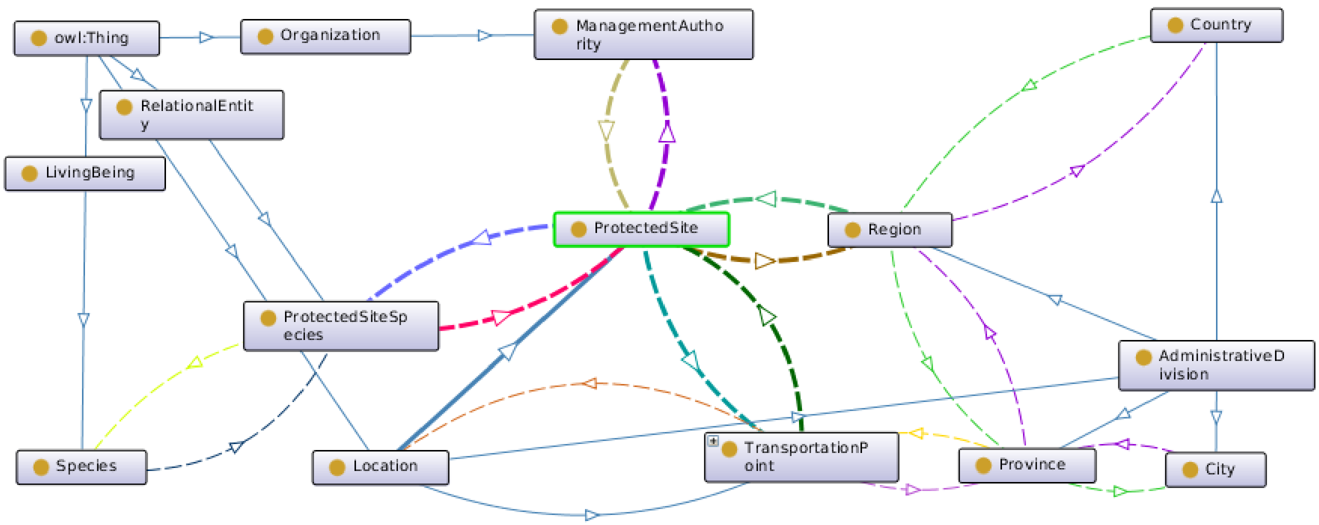

Figure 6: Ontology representation relative to our formal model.

By adding the data in Protegè we were also able to generate an *rdf* file and import it into GraphDB [6]. This allows us to display the links of real instances of our classes.

For the sake of an easier visualisation we imported just 10 sites from the province of Trento and 10 from Bolzano and the first 20 animals present at least in one of the sites taken in account. We imported also TransportationPoint instances, but we connected them randomly to the sites, just to display their connection. That's why we came up with the links highlighted in fig. 7.

We wanted then to display a graph containing at least one instance of the most important classes of our model. This helped us in checking whether the job we did was compliant to the goal of our work. What we can see in image fig. 8 are two ProtectedSite instances (*Biotopo Ontaneto di Sluderno* and *Biotopo Ontaneto di Cengles*) connected with various kind of TransportationPoint (purple and light green). We also have the ManagementAuthority which is violet (*Provincia Autonoma Bolzano - Ripartizione Natura*) and the Region (*Trentino Alto-Adige*): those are the organisations responsible for the management of the protected sites.

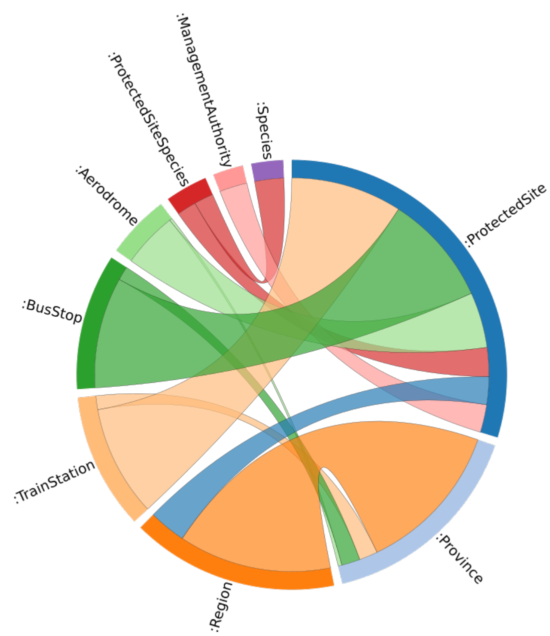

Figure 7: Links between classes distribution

Various species are present in these two sites, but in the graph only the name of one is present: the *Alauda Arvensis*, actually linked to both sites. Still we can see no direct connection between the species and the sites: this happens because of the introduction of the *ProtectedSiteSpecies* class and it is evident here why it was needed. A species cannot be linked to multiple sites and have multiple *Abundance* values without using another class. ProtectedSite-Species is not representented in the graph with its name, because there is none, but the class is identified with a unique number.

With this figure it is also possible to see how the built ontology could be used to answer the queries we assessed in section 1. All the connections between ProtectedSite and TransportationPoint can be queried to show how a tourist could visit some sites using available means of transportation. Also it is possible, starting from a Region or a Province to see the nearby sites and therefore also the species present.

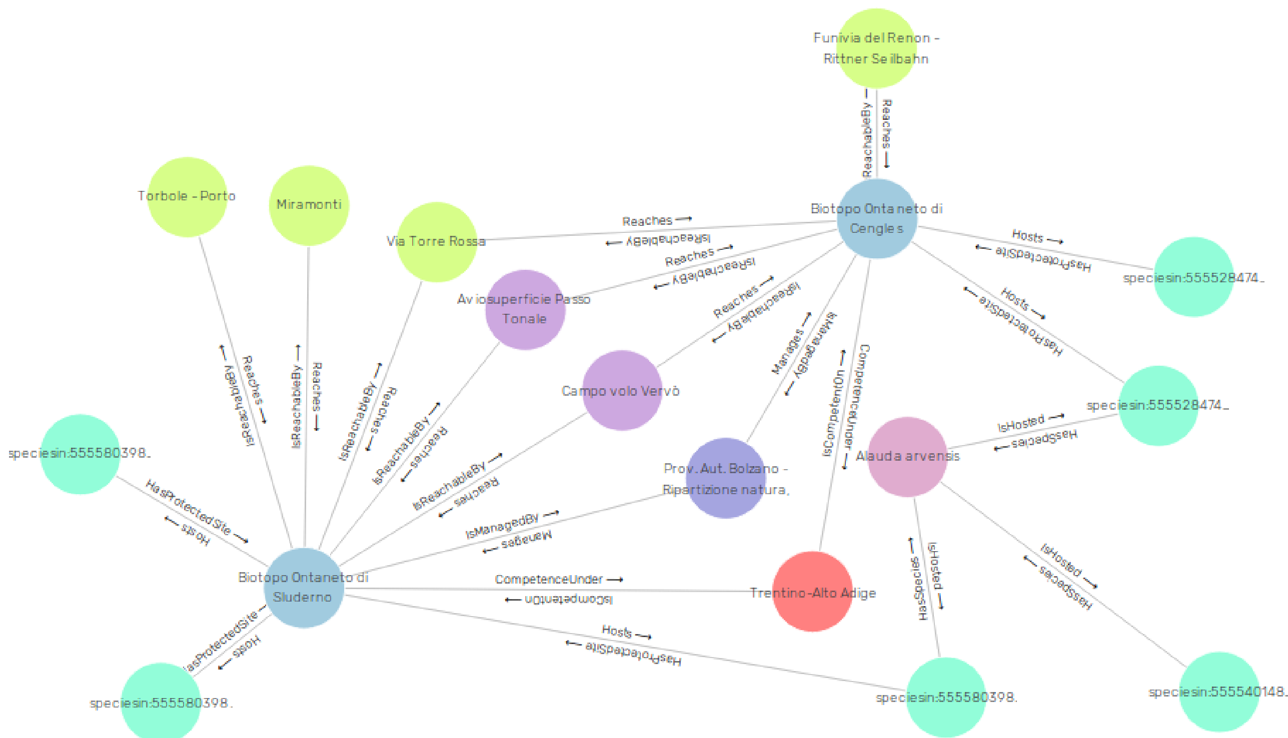

Figure 8: Connection between instances represented with a graph

6 Generalised query

In this section it is shown some of the queries that can be addressed using the formal model proposed in this report. For each query we provide a practical use case, a formalisation of the query in SPARQL syntax and a small set of results returned by the query.

To start with, one could be interested in knowing which are the protected areas in a given region, Trentino-Alto Adige for instance. This is indeed the first scenario described in section 1 and can be answered quite simply: *"Return every ProtectedSite within Trentino-Alto Adige"*. In this case the query should be the following:

```
SELECT ?pname
WHERE {
  ?site rdf:type psso:ProtectedSite  ;
        psso:CompetenceUnder ?region ;
        psso:LocalName ?pname .
  ?region psso:Name ?region_name .
FILTER
  (?region_name IN ("Trentino-Alto_Adige"))
}
```

```
pname
---
Biotopo Ontaneto di Cengles
Alta Val La Mare
Tre Cime Monte Bondone
Lago Nero
Palu Longa
```

Once we have the protected site, we are interested to know how it is possible to reach it. We are therefore interested in the transportation points that reach protected sites found in the previous query. *"Return the protected site and the nearest set of transportation points within Trentino-Alto Adige"*.

```
SELECT ?tpoint_name ?type_name ?psite_name
WHERE {
  ?psite rdf:type psso:ProtectedSite  ;
         psso:CompetenceUnder ?region ;
         psso:LocalName ?psite_name ;
         psso:IsReachableBy ?tpoint .
  ?tpoint psso:Name ?tpoint_name ;
          rdf:type ?type .
  ?type rdfs:label ?type_name .
  ?region psso:Name ?region_name .
  FILTER
  (?region_name IN ("Trentino-Alto_Adige"))
}
```

```
tpoint_name | type_name    | psite_name
------------+--------------+---------------------------+
Caldes      | TrainStation | Bio. Ontaneto di Oris
Intermodale | TrainStation | Bio. Ontaneto di Cengles
Dermulo     | TrainStation | Bio. Ontaneto di Sluderno
Miramonti   | BusStop      | Alta Val di Rabbi
Miramonti   | BusStop      | Alta Val La Mare
```

The next use case is given by considering the third scenario described in section 1. One could be interested in knowing which are the protected sites within a given region and check how they are connected to public transportation services, like bus stops or train stations. This would be indeed a filtering over the previous query. Instead we propose here an alternative approach, that could be helpful in detecting weak spots with regards to a given protected site. Namely we list all protected sites with the corresponding number of public transport services (bus stops and train stations are considered).

```
SELECT ?site (COUNT(?tpoint) as ?degree)
WHERE {
  ?tpoint rdf:type ?type;
          psso:Reaches ?site.
  FILTER
    (?type IN (psso:TrainStation,psso:BusStop))
}
GROUP BY ?site
```

```
site       | degree
-----------+-----
#555528475 | 8
#555540151 | 8
#555580398 | 8
#555580403 | 8
#555580413 | 8
```

From the result it is possible to see that every protected site has the same number of public transport services associated. This is given by the procedure we crafted to populate data for individuals.

A valuable information that could represent the core of a service is given by the species housed in protected sites. For example, one might be interested in listing all birds species that can be found in protected sites in the

Trentino-Alto Adige region. Informally, this can be stated as: *"Find all birds species housed in protected sites within Trentino-Alto Adige"*.

```
SELECT distinct ?psite_name ?species_name
WHERE {
  ?psite rdf:type psso:ProtectedSite ;
         psso:CompetenceUnder ?region;
         psso:LocalName ?psite_name;
         psso:Hosts ?relation .
  ?region rdfs:label ?name;
          psso:Name ?region_name .
  ?relation psso:HasSpecies ?species .
  ?species psso:Name ?species_name ;
           psso:Supergroup ?family .
  FILTER
    (?name IN ("Trentino-Alto Adige"@en) &&
    (?family IN ("Birds")))
}
```

```
psite_name           | species_name
- - - - - - - - - - -+- - - - - - - - - - -
Bio. On. di Sluderno | Anas platyrhynchos
Bio. On. di Cengles  | Gallinula chloropus
Bio. On. di Sluderno | Gallinula chloropus
Bio. On. di Cengles  | Streptopelia turtur
Bio. On. di Cengles  | Caprimulgus europaeus
Bio. On. di Cengles  | Alauda arvensis
Bio. On. di Sluderno | Alauda arvensis
Bio. On. di Cengles  | Lanius collurio
Bio. On. di Sluderno | Lanius collurio
```

The final query we present is quite particular. We thought that protected sites often host species whose beauty and rareness is often the motivation behind a trip. Therefore it is of interest to know, given a particular species, where it is possible to find it and how one could be able to reach that location. For the example proposed the species considered is the *Anas platyrhynchos*, which is a particular species of duck typical of northern region of Italy.

```
SELECT ?psite_name ?tpoint_name
       ?province_name ?region_name
WHERE {
  ?species rdf:type psso:Species;
           psso:Name ?species_name;
           psso:IsHosted ?relation.
  ?relation psso:HasProtectedSite ?psite.
  ?psite psso:LocalName ?psite_name;
         psso:IsReachableBy ?tpoint;
         psso:CompetenceUnder ?region.
  ?region psso:Name ?region_name.
  ?tpoint psso:Name ?tpoint_name;
          psso:Covers ?province.
  ?province psso:Name ?province_name.
  FILTER
    (?species_name IN ("Anas platyrhynchos"))
}
```

```
psite_name                  | tpoint_name                  | province_name | region_name
- - - - - - - - - - - - - - -+- - - - - - - - - - - - - - - + - - - - - - - + - - - - - - - - - -
Biotopo Ontaneto di Sluderno | Miramonti                    | Bolzano       | Trentino-Alto Adige
Biotopo Ontaneto di Sluderno | Serravalle all'Adige         | Trento        | Trentino-Alto Adige
Biotopo Ontaneto di Sluderno | Campo volo Vervo'            | Trento        | Trentino-Alto Adige
Biotopo Ontaneto di Sluderno | Caldes                       | Trento        | Trentino-Alto Adige
Biotopo Ontaneto di Sluderno | Aviosuperficie Passo Tonale  | Trento        | Trentino-Alto Adige
Biotopo Ontaneto di Sluderno | Torbole - Porto              | Trento        | Trentino-Alto Adige
Biotopo Ontaneto di Sluderno | Corso Verona/Ospedale        | Trento        | Trentino-Alto Adige
Biotopo Ontaneto di Sluderno | Waidbruck - Ponte Gardena    | Bolzano       | Trentino-Alto Adige
```

7 Final considerations

With this section, we conclude our presentation of a formal model to address protected site queries in the spatial context surrounding them. The model addresses most of the scenarios considered, in addition to provide an additional value given by the presence of species information, and we consider this a great achievement.

Unfortunately, we were not able to provide the fundamentals required to exploit also the last scenario presented, *i.e.* the trip suggesting feature constrained by time limits (to improve reachability of protected sites). The concept of *time* is totally missing in our model but we are quite convinced the model will be easy to extend by anyone willing to exploit this information.

Concepts introduced to describe the domain have been generalised and mapped to the Common Sense Knowledge ontology, in order to increase the reusability and interoperability of our work. A possible extension in this direction is given by mapping domain specific concepts to intermediary ontologies, to provide more uniformity directly to the leaf concepts in the hierarchy. For instance, we adopted a generalised concept of CoordinateValue, but if the widely deployed WGS84 ontology concepts were used, the impact on the query capabilities with regard the spatial domain would have been greatly increased. In particular, one would have been able to query for protected sites whithin a given, user defined, distance from a point.

From the point of view of the lexical information, our concepts have been mapped almost completely, and where a synset was not found the nearest concept is provided. This is of great value for a large number of applications that go also beyond the scope of the knowledge information and integration tasks.

In appendix C it is shown an evaluation of our domain formal model, according to the methodologies explained in class.

The table contains values matching both the model and the dataset. For the computation the generalised queries described in section 6 were used instead of the competence queries, as suggested by our project coordinator.

From the second method of evaluation it emerges that some issues afflict our model about the semantics of the concepts introduced. Fortunately, the hierarchy allows one easily to address this issue and to provide a solution in the next version of the ontology. Those issues are by the way marginal and isolated and do not touch many classes or properties, thus not affecting the model in a critical way.

A Informal modelling

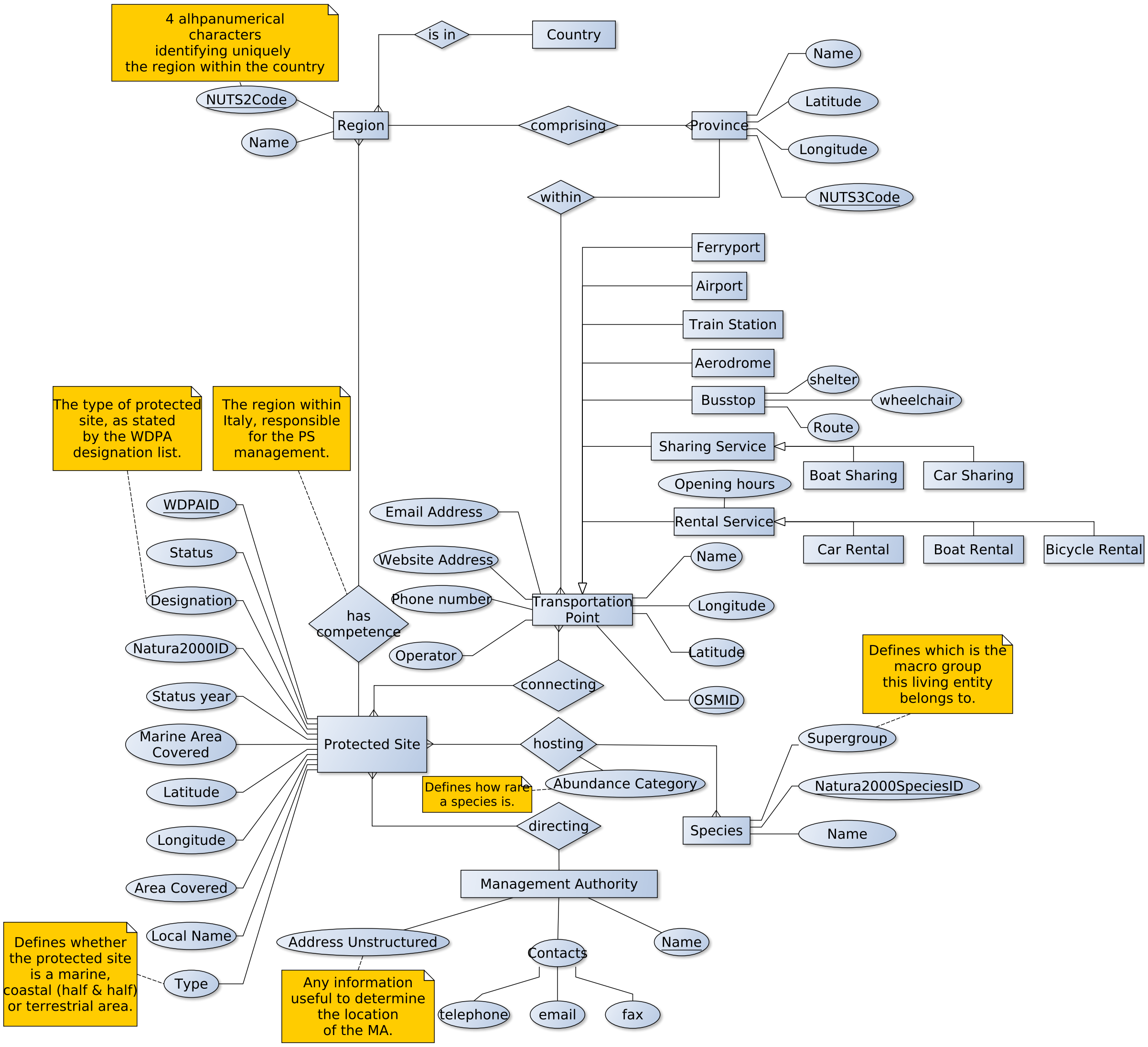

Figure 9: ER model representing the structure of the underlying datasets.

B Lexical information mapping

In this appendix some charts depicting the mapping between the domain entities and the synsets in Wordnet are given.

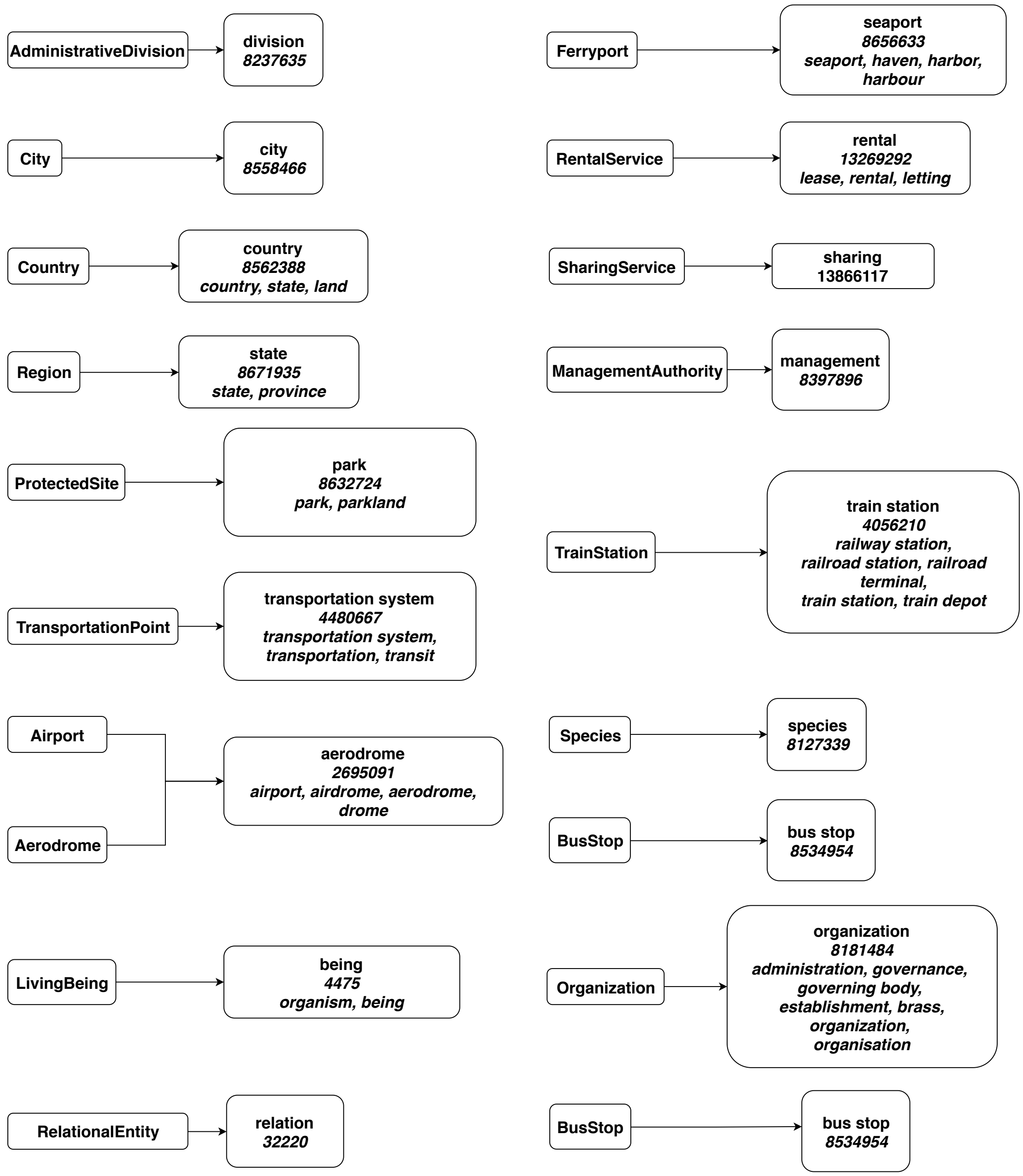

Figure 10: Diagram of the mapping related to class entities.

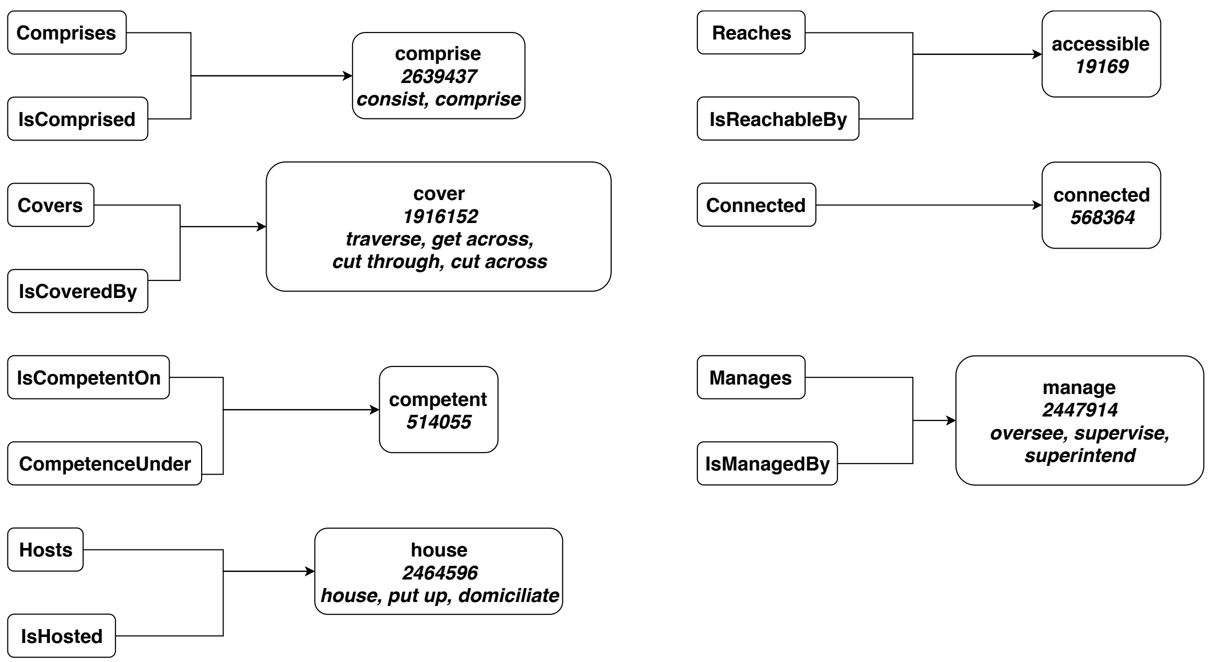

Figure 11: Diagram of the mapping related to object properties.

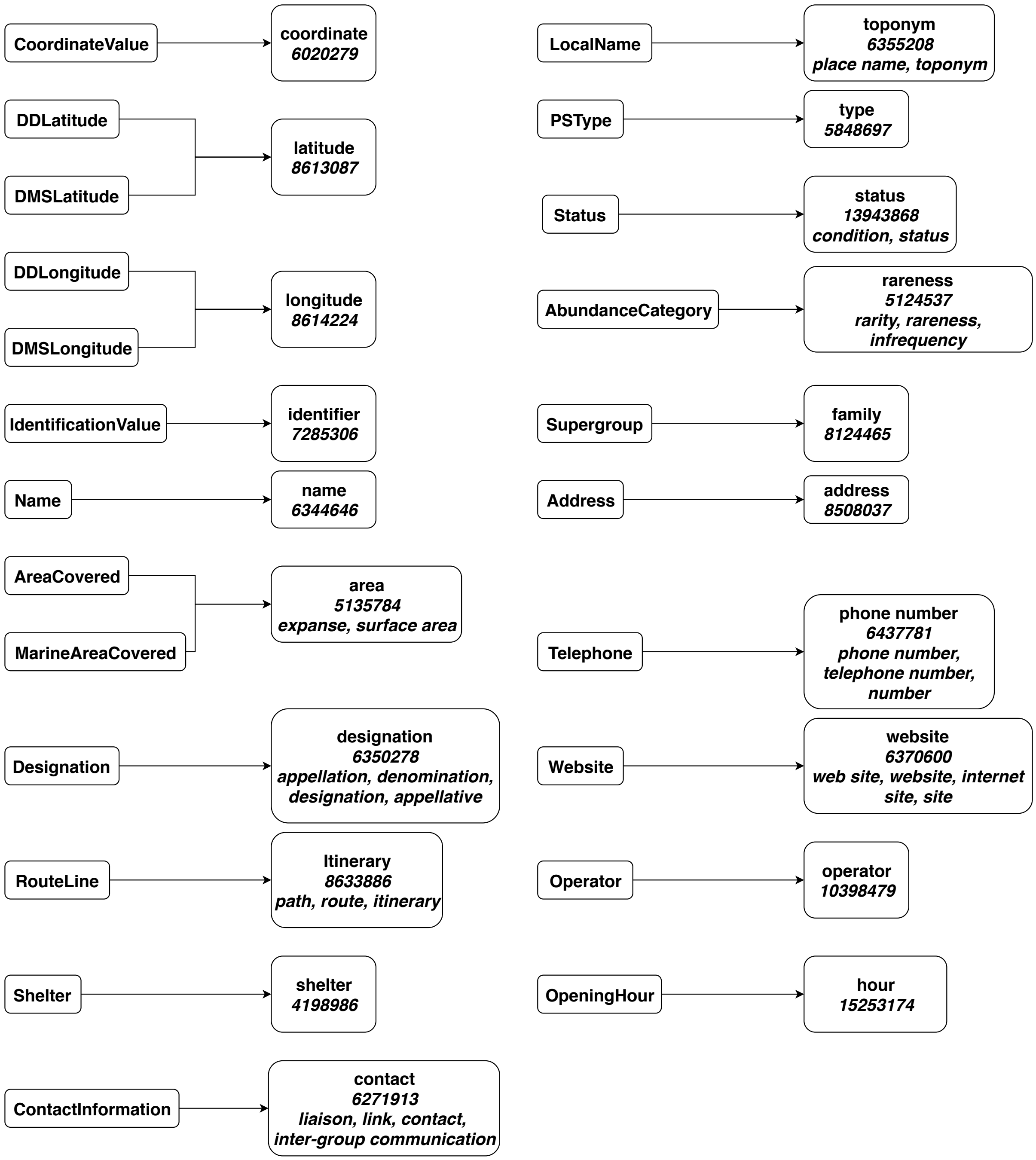

Figure 12: Diagram of the mapping related to data properties.

C Evaluation Model

	Class Coverage	Class Flexibility	Attribute Coverage	Attribute Flexibility
CQ-Model	$\simeq 0.8$ (Ideal) [1]	$\simeq 0.2$ (Ideal) [0.78]	$\simeq 0.8$ (Ideal) [1]	$\simeq 0.5$ (Ideal) [0.78]
DS-Model	$\simeq 0.8$ (Ideal) [1]	$\simeq 0.2$ (Ideal) [0.67]	$\simeq 0.8$ (Ideal) [1]	$\simeq 0.5$ (Ideal) [0.4]
CQ-DE	$\simeq 0.8$ (Ideal) [1]	$\simeq 0.2$ (Ideal) [0.33]	$\simeq 0.8$ (Ideal) [1]	$\simeq 0.5$ (Ideal) [0.63]

Schema Level

- Does the model including cycles in the class hierarchy ? No
- Does the Model uses any polysemous terms for its class or property name? Yes (Telephone, Fax, Email, Website)
- Is Multiple Domain / Range defined for any property ? No
- Does any class have more than one direct parent class ? No
- Does the Model include multiple classes which have same meaning ? Yes (Airport, Aerodrome)
- Is the class Hierarchy over specified? No
- Does the model use isA as a object Property or relation? No
- Does the model have any leaf class for which there is no relation with the rest of the model? No
- Did you use miscellaneous or others as one of the class name? No
- Does the model have any chain of Inheritance in class hierarchy? No
- Do all properties have explicit domain and range declarations? Yes
- Does the model have any classes or properties which are not used? Yes (the property: Connected)
- Are a collection of elements included as a group in a number of class/attribute ? No

Linguistic Level

- Does all elements of the model (i.e. class and property) have human readable annotations? Yes
- Do all elements of the model follow the same naming convention? Yes

Metadata Level

- Is provenance information (Creator, Version, Date) available for the final protege model? Yes
- Is provenance information available for any property or class which is taken from some reference standard or ontology? No

References

[1] European Commission, *Natura 2000*, http://ec.europa.eu/environment/nature/natura2000/, [Online; accessed 26-November-2018].

[2] European Commission, *The birds directive*, http://ec.europa.eu/environment/nature/legislation/birdsdirective/, [Online; accessed 26-November-2018].

[3] ———, *The habitats directive*, http://ec.europa.eu/environment/nature/legislation/habitatsdirective/, [Online; accessed 26-November-2018].

[4] ———, *Inspire*, https://inspire.ec.europa.eu/, [Online; accessed 26-November-2018].

[5] Natasha Noy, Alan Rector, Pat Hayes, Chris Welty, *Defining N-ary Relations on the Semantic Web*, https://www.w3.org/TR/2006/NOTE-swbp-n-aryRelations-20060412/, [Online; accessed 05-December-2018].

[6] Ontotext, *GraphDB*, http://graphdb.ontotext.com/, [Online; accessed 10-December-2018].

[7] OpenStreetMap contributors, *Planet dump retrieved from https://planet.osm.org* , https://www.openstreetmap.org, 2017.

[8] QGIS Development Team, *Qgis geographic information system*, Open Source Geospatial Foundation, 2009.

[9] Sean Falconer, *OntoGraf*, https://protegewiki.stanford.edu/wiki/OntoGraf, [Online; accessed 10-December-2018].

[10] Tatiana Tarasova, Jindřich Mynarz, Phil Archer, *SmOD INSPIRE*, https://www.w3.org/2015/03/inspire/, [Online; accessed 14-December-2018].

[11] United Nations, *Protected Planet*, https://www.protectedplanet.net/c/about, [Online; accessed 26-November-2018].