Linkable Geographic Ontologies

Francisco J.
Lopez-Pellicer
Universidad de Zaragoza
Zaragoza, Spain
fjlopez@unizar.es

Mário J. Silva
Universidade de Lisboa
Lisboa, Portugal
mjs@di.fc.ul.pt

Marcirio Chaves
Universidade de Lisboa
Lisboa, Portugal
mschaves@gmail.com

ABSTRACT

The performance of some tasks in Information Retrieval is strongly related to the extent and quality of the geographic knowledge about named places. This paper presents a conceptualization of the geographic knowledge, the Geo-Net vocabulary, and a tool for building large knowledge bases of named places, the GKB management system, developed in the GREASE-II project. The Geo-Net vocabulary is a conceptual model for describing geographic places, including their names, types, relationships and footprints. It uses URIs and the RDF data model to expose, share and connect pieces of geographic knowledge each other and to related data on the Web. The GKB system is a multi-paradigm knowledge management system that enables the development of geographic ontologies with the Geo-Net vocabulary. This paper also presents a geographic ontology of Portugal, Geo-Net-PT 02, created with the Geo-Net vocabulary and the GKB system.

Categories and Subject Descriptors

H.3.2 [**Information Storage and Retrieval**]: Information Storage; H.2.8 [**Database Management**]: Database applications—*Spatial databases and GIS*

General Terms

Design

Keywords

geo-ontologies, geographic information retrieval, linked data, geographic knowledge base

1. INTRODUCTION

Whatever occurs, occurs in a geographic location, physical or mental [25], and hence can be described, filed, visualized, queried, retrieved, and understood using geographic terms [13]. Some tasks in Information Retrieval are designed for searching on collections of resources rich in geographic information. These tasks retrieve and rank relevant documents by applying methods based on spatial reasoning algorithms.

For that purpose, we need to support the description and discovery of geographic features. The description entails the identification of place names and the geographic features that they designate. The discovery entails the identification of geographic features given some of their attributes. We use the term geographic feature instead of place or location to emphasize the relevance of the spatial information in these tasks. The term geographic feature comes from Geographic Information, where a geographic feature is an entity with a geographic locacion.

Our concern in this work is the provision of support for simple geospatial data identification. That is, the assembly of large datasets that can provide the answer to questions like: given a set of place names, which geographic features best match the set of place names and what are their locations on Earth?

The answer to the above question might require to explore the Web of data, that is, to look for resources in data Web sites, such as DBpedia [4]. Linked Data (`http://linkeddata.org/`) is one of the main communities of the Web of Data. It promotes best practices for exposing, sharing, and connecting pieces of knowledge via dereferenceable URIs making use of the architecture of the Web and the RDF data model.

This paper is structured as follows. Section 2 presents the Geo-Net vocabulary, a conceptual model for the description and discovery of geographic features in the Web of Data and proposes its RDF encoding. Section 3 describes the Geographic Knowledge Base (GKB) system, a manager of geographic ontologies based on the Geo-Net vocabulary. Section 4 presents the application of both in the development of the geographic ontology of Portugal Geo-Net-PT 02. Section 5 describes related work. Finally, Section 6 outlines future lines of work.

2. GEO-NET VOCABULARY

This section presents the Geo-Net conceptual model and its implementation as a RDF vocabulary. The conceptual model of Geo-Net extends a previous version [6], based on conceptualizations of Hill [12], Manov et al. [16], and Fu et al. [8]. Practitioners of Geographic Information Retrieval, Linked Data and Geographic Information are the intended audience of the Geo-Net vocabulary.
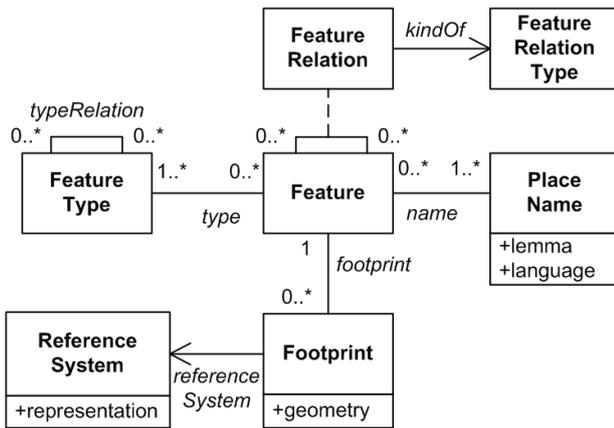
Figure 1: Core metamodel.

2.1 Conceptual Model

The Geo-Net vocabulary is intended to describe and discover toponymic datasets. By description, we mean the identification of relevant place names and geographic features that they refer to. By discovery, we mean the identification of geographic features that best match a query. This requires a set of assumptions about the world that are presented below.

The conceptual model of the Geo-Net vocabulary (Figure 1) defines the following concepts:

- Geographic features (class `Feature`).

- Place names (class `PlaceName`). The linguistic content of the place name is captured in `lemma`, a canonical form of the name, and `language`.

- Types (class `FeatureType`). They classify geographic features.

- Relations between features (class `FeatureRelation`). They are classified by the class `FeatureRelationType`.

- Locations (class `Footprint`). The spatial description of a location is captured in the field `geometry` and complemented by a `ReferenceSystem`.

We now examine the above concepts in detail. The relations `type` and `name` relate a `Feature` with feature types and place names, respectively. These relations have a cardinality of 1..* in the `FeatureType` and `PlaceName` ends. The cardinality reflects the assumption that each geographic feature must be partially specified in terms of declarative knowledge, that is, a feature needs to have at least a place name and a feature type. The metamodel allows the specification of place names and feature types not explicitly related to features.

The relation `typeRelation` describes binary semantic relationships among instances of the `FeatureType` class. The relation `typeRelation` provides the minimum machinery for supporting taxonomies of feature types.

The association class `FeatureRelation` describes binary relationships among features. Relations must be qualified by an instance of the class `FeatureRelationType`.

The class `Footprint` is related with the class `Feature` by means of the relation `footprint`. This relation has a cardinality of 1 in the `Feature` end, and enforces the conceptual

restriction that footprints are not shared among features. The data type of the field `geometry` is the type `Geometry`, a complex structure with the properties `coordinates` and `type`. The property `coordinates` values are either a list of coordinate points, or a set of instances of the type `Geometry`. The property `type` is an identifier that specifies the kind of geometric shape whose boundary is described in the coordinates field; possible values include point, line and polyline.

The `ReferenceSystem` class provides a description of the reference system encoded in the field `representation`. The value of the field `representation` provides directly or indirectly the geodetic datum, and, if needed, the projection. To be fully specified, an instance of the class `Footprint` requires the specification of a unique reference system for all the points stored in the geometry field. The list of coordinate points takes the form (x, y) if the georeferencing system is based on a planar, Cartesian or two-dimensional reference system, or $(latitude, longitude)$ if the georeferencing system is based on a three-dimensional reference system. A planar reference system requires a map projection that describes how to transform the planar coordinates onto the Earth's surface and vice versa. Both planar and three-dimensional reference systems need the definition of a geodetic datum.

The description of time is not explicitly considered in the Geo-Net vocabulary. It is only a framework to identify the above concepts and properties. The description of its temporal dimension should be done with appropiate vocabularies, such as the temporal vocabulary described in [10].

2.2 Vocabulary

The Geo-Net vocabulary is composed by a core set of classes and properties and three specialized modules: spatial reference systems, information domains, and provenance. Table 1 presents the terms of the Geo-Net vocabulary. The Geo-Net vocabulary, encoded in OWL 1, is available from `http://xldb.di.fc.ul.pt/wiki/Geo-Net-PT_02` together with vocabularies and datasets based on it, such as the Geo-Net-PT vocabulary and the Geo-Net-PT 02 ontology. The URI of the namespace of the Geo-Net vocabulary is `http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net#` with the associated prefix `gn:`.

OWL is a family of knowledge representation languages for exchanging ontologies that has formally defined semantics and well understood computational properties [24]. Consequently, a dataset described by the Geo-Net vocabulary is interpreted as a collection of individuals and a collection of relations between these individuals, where both collections are constrained by the axioms of the Geo-Net vocabulary.

Core Classes. The following classes define the core concepts of the Geo-Net vocabulary:

- `gn:GeographicConcept` represents identifiable things related with geographic features. The definition of this class is intentionally broad. Application vocabularies can extend the conceptual model using this class. For example, they can define new subclasses or use this class as a restriction in new properties. This class is the superclass of the classes `gn:Feature`, `gn:PlaceName`, `gn:FeatureType` and `gn:Footprint`, which are pairwise disjoint.

- `gn:Feature` represents any meaningful object that can be grounded directly or by reference. A feature can in-

Table 1: Classes and properties of the Geo-Net vocabulary.

	Term	Description	Specializes
Class	gn:GeographicConcept	root concept	-
	gn:Feature	feature	gn:GeographicConcept
	gn:PlaceName	proper name	gn:GeographicConcept
	gn:FeatureType	classifier	gn:GeographicConcept
	gn:Footprint	footprint	gn:GeographicConcept
	gn:InformationDomain	collection	-
	gn:Source	provenance metadata	-
	gn:ReferenceSystem	spatial reference system	-
Object property	gn:name	has name	-
	gn:type	is classified by	-
	gn:relation	is related with	-
	gn:inDomain	is member of	-
	gn:lineage	has provenance metadata	-
	gn:referenceSystem	has reference system	-
Datatype property	gn:lemma	has a canonical form	-
	gn:languageCode	belongs to language identified by	-
	gn:geometry	has a canonical encoding of its shape	-
	gn:representation	has literal description of reference system	-

clude additional annotations. For example, we can add multilingual labels and a descriptive statement using Dublin Core metadata terms, or links to linked data resources using the property `rdfs:seeAlso`.

- `gn:PlaceName` represents proper names of one or more features. The definition of proper name is not restricted. For example, authorities can define schemas of codes for identifying features. These codes, known as geographic codes, are considered proper names in this vocabulary.

- `gn:FeatureType` represents classifiers for features. The Geo-Net vocabulary does not intend to enforce a typing schema: the set of features classified by a type is defined by explicit enumeration. A set defined by enumeration allows to describe that these features share a set of properties without stating which or how fuzzy they are.

- `gn:Footprint` is a location on a surface. The location data can be described with well known vocabularies, such as Basic Geo Vocabulary [3].

Core Properties. The core Geo-Net vocabulary includes the following properties:

- `gn:name` asserts that a resource of type `gn:Feature` has as proper name a resource of type `gn:PlaceName`. A place name can play several roles along its existence as communication tool. These roles often reveal significant patterns of environment, settlement, colonization, organization, historical facts and folk etymology. Application vocabularies derived from the Geo-Net vocabulary can create specialized relations to represent these roles.

- `gn:type` declares that a resource of type `gn:Feature` is classified by a resource of type `gn:FeatureType`. The intended meaning of this relation is to assert the feature shares characteristics with other features classified by the same feature type.

- `gn:footprint` states that the location of a resource of type `gn:Feature` is described by a resource of type `gn:Footprint`. The footprint of features might be as complex as a survey description of the boundaries of the feature, or as simple as a pinpoint.

- `gn:relation` asserts that two instances of the class `gn:Feature` are related. Relationships record declarative geographic knowledge, such as *capital of, administrative division of* and *former part of*, and configurational geographic knowledge, such as *part of, adjoint to* and *connect to*.

- `gn:lemma` relates a resource of type `gn:PlaceName` to a unique lexical form considered the canonical representation of the denoted place name. The range is a typed literal with type `xsd:string`. The canonical form must be a Unicode string, used in communication, indivisible, and without additional context information. That is, the lemma identifies the lexical form used as place name, such as *Paris*, not the lexical form of the proper name of a specific feature, such as *Paris, Texas*. Application vocabularies can represent other lexical forms, such as the hierarchical form *Paris, Texas*, by extending the class `gn:PlaceName`.

- `gn:languageCode` is a functional property. It relates a resource of type `gn:PlaceName` to a well-known string value that identifies the language of the denoted place name. The range is an `xsd:string` literal. The allowed values are the three-letters identifiers from ISO 639-3.

- `gn:geometry` relates a resource of type `gn:Footprint` with a canonical geometrical description of the footprint shape. It must be encoded in one of the following encoding schemes for representing geometry data: Well-Know Text (WKT), and Geography Markup Language (GML). Both are GIS industrial standards specified by Open Geospatial Consortium (OGC). WKT syntax for 2D and 3D geometries is described in the OGC SFA specification [11], and GML is described in [19]. GML is also an ISO standard [23].

Spatial Referencing Systems. The property `gn:geometry` can hold a geometry with a reference to its coordinate reference system. We introduce here a class and two properties to explicitly assert the reference system used in the description of a footprint:

- `gn:ReferenceSystem` represents spatial reference systems. A reference system describes how to transform coordinates onto the Earth's surface and vice versa. this class is disjoint with other Geo-Net concepts.

- `gn:representation` relates a `gn:ReferenceSystem` instance with a literal description of the spatial reference systems encoded in WKT or GML.

- `gn:referenceSystem` relates a `gn:Footprint` individual with a resource that describes how to translate the description of the footprint into a location on the surface. A `gn:Footprint` individual can hold only one reference system.

Information Domains. The Geo-Net vocabulary provides support to organise data into collections. These collections are called information domains, and are represented by the class `gn:InformationDomain`. This concept disjoints with other concepts of the ontology. The property `gn:inDomain` marks the assignment of a resource to an information domain. Information domains allow the partitioning of data in coherent collections. Having said that, it is also possible the assigment of a geographic concept to multiple information domains. This enables the organisation of the geographic information in overlapping collections if necessary.

Provenance. The potential users of a dataset cannot evaluate the authority of the claimed facts without a proper description of their provenance. The class `gn:Source` represents sources whose contents have been added to a dataset described with the Geo-Net vocabulary. The information model applied to the description of a source should be similar to the applied to describe sources in the geographic information domain, such as the detailed in [18, 14]. The class `gn:Source` disjoints with other Geo-Net concepts. The provenance of items of information is stated through the property `gn:lineage` that relates a concept or statement with its source, a resource of type `gn:Source`.

3. GKB SYSTEM

The GKB system is a knowledge management system that implements a domain-independent geographic meta-model and integrates geographic knowledge collected from multiple sources. The core of the GKB system is an abstract data model, whose purpose is to formally describe objects grounded in some way to locations. The GKB system has components for loading, querying and exporting knowledge.

3.1 Architecture

The architecture of the GKB system is represented in Figure 2. We can analyzane its structure taking into account how a client application can interact with the managed content. It is possible to interact with an ontology instance managed by the GKB sistem in three different paradigms:

- A relational database schema accessible through SQL interfaces. This interaction mode entails that the client
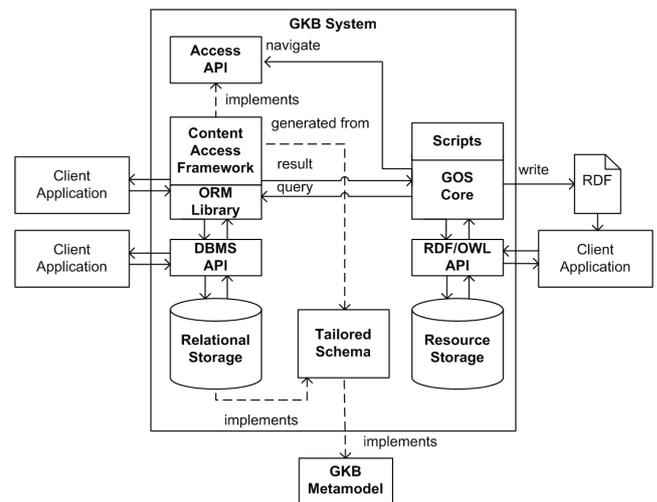

Figure 2: The architecture of the GKB System

application must query and reason following all the rules and assumptions formalized in the Geo-Net vocabulary.

- An object-oriented data-structure accessible through an API. The API works as a contract between the application and the GKB by exposing the domain model.

- A knowledge representation described by an OWL 1 ontology, which uses the Geo-Net vocabulary, accessible through a semantic aware interface, such as a SPARQL end-point [21], or serialized in a semantic format, such as RDF/XML [2].

Following these criteria, the GKB architecture can be divided in relational, object-based and resource-based persistence systems.

Relational Persistence System. The components of the relational persistence system are the tailored schema, the storage that implements it, and the API that provides interaction with the repository. Both the relational storage and its API must support the OGC Simple Feature specification for spatial datatypes [11]. The tailored schema is a platform-dependent model of the GKB metamodel that instances one or more information domains. These information domains can be modified to implement application requirements. The client applications can use the GKB instance by querying the relational persistence system.

Object-Oriented Persistence System. The components of the object-oriented persistence system are the content access framework, the access API, and the object-relational mapping (ORM) library. The ORM library must support code generation from schemas and standard spatial datatypes. The back-end of the object-oriented persistence system is the relational persistence system. The content access framework is an object-oriented library, which has been semi-automatically generated from the tailored schema of the relational-based persistence system. The access API is a lightweight set of interfaces that partially implements the GKB metamodel and the additional classes, and provides

graph navigation methods. The process that generates the content access framework is instructed to implement the access API in the generated code and wire the access API methods to fields of the tailored schema. The advantage of the object-oriented persistence system over the relational is that the generated code structure is closer to the GKB metamodel, and the ORM library can provide extra functionalities, such as caching and transparent fetching of data. The main disadvantage is the potential complexity of wiring the code and the additional overload of the ORM library to some tasks.

Resource-Oriented Persistence System. The resource oriented persistence system is composed by the Geographic Ontology Serializer (GOS), the resource-based storage, and the RDF/OWL API that manages the resource storage. The GOS is described in detail below. The resource-oriented persistence system employs the object-oriented persistence system as back-end database. Applications can access the resource-oriented persistence system using a SPARQL endpoint or the RDF/OWL API. In addition, the resource-oriented system can be used to create RDF serializations of the ontologies managed by GKB.

3.2 Geographic Ontology Serializer

GOS is a GKB module that takes an ontology hosted in a GKB instance and creates its representation in one of the available RDF serialization formats. The main components of GOS are:

- **Access API** implements the current GKB metamodel based on the Geo-Net vocabulary. This API is the facade for any serializable content.

- **Content Access Framework** provides an ORM to relational storage instances of a specific GKB schema. The content access framework is generated using reverse engineering techniques and implements the Access API.

- **Application Script** selects the contents to be serialized using the Access API and provides specific mappings between a GKB tailored schema and an application vocabulary describing its semantics.

- **GOS Core** orchestrates the creation of serializations of GKB instances. The GOS Core creates and manages an RDF staging area associated to the execution of an application script. This component is also responsible for writing out in the desired RDF format.

Users can run application scripts in the GOS Core. The application script can use GOS core to query the Content Access Framework using its Access API, or use the default Geo-Net mappings implemented in the GOS Core.

4. GEO-NET-PT 02 ONTOLOGY

The Geo-Net-PT 02, a geospatial ontology of Portugal, is an authoritative geographic knowledge dataset created with GKB [15]. Geo-Net-PT 02 is the evolution of the Geo-Net-PT-01 ontology [5]. The development of Geo-Net-PT 02 began with a GKB instance containing the data of Geo-Net-PT 01, mainly administrative features, which was enriched with data from the physical domain later [20].

4.1 Content Organization

The Geo-Net-PT 02 GKB repository consists of three instances: geo-administrative, geo-physical and network. The geo-administrative instance includes human geography features, such as administrative regions. The geo-physical instance includes physical geography features, such as natural regions and man-made spots. The network instance stores data about Web sites. Each of these instances contains only data about Portugal. The content of the geo-administrative and network domains is the same as in Geo-Net-PT 01, now mapped to fit the Geo-Net vocabulary.

4.2 Geo-Net-PT Vocabulary

The Geo-Net-PT vocabulary is an extension of the Geo-Net vocabulary. The goal of this extension is to annotate specific characteristics of the Geo-Net-PT data and the tailored instance of the GKB system (version 2.1) that stores this data. The prefix of this vocabulary in this paper is `gnpt:`. The terms of the Geo-Net-PT vocabulary are summarized in Table 2.

Name Relationships. The Geo-Net-PT vocabulary defines three subproperties derived from `gn:name`: `gnpt:preferred`, `gnpt:alternative` and `gnpt:identifier`. The properties that represent these roles are subproperties of the property `gn:name`. For example, `gnpt:preferred` is applied to mark the main label for the geographical feature in an information system. The notion of preferred name implies that a geographical feature can only have one such name per language. `gnpt:alternative` is applied to relate a feature with a resource of type `gn:PlaceName` not considered preferred. Finally, `gnpt:identifier` asserts that the place name uniquely identifies that feature.

Geographic Codes. A place name is not a safe feature identifier for management or administrative purposes. Authorities can define schemas of codes for identifying features. These codes are known as geographic codes. In the context of Geo-Net-PT, there are some administrative features, such as administrative divisions, and physical features, such as rivers, that have known geographic codes. These codes can be found in databases and specialized documents. The class `gnpt:GeographicCode` represents these codes. The controlled vocabulary where the geographic code is defined is identified by the functional property `gnpt:inSchema`, subproperty of `gn:lineage`.

Additional Relations. One of the goals of the Geo-Net-PT vocabulary is to describe conceptual, hierarchical and topological relations between features. These relations are subproperties of the property `gn:relation`. This vocabulary includes the following properties:

- `gnpt:scope` declares that the subject of the relation is a resource that identifies a real world object, such as a book, or a digital resource, such as a Web page, whose content has a spatial extent described by the object of the relation.

- `gnpt:isLocatedOn` asserts that the subject is a feature located on the land surface of the target feature.

- `gnpt:isAdjacentTo`, which is symmetric, states that its subject is a feature touching another feature.

Table 2: Classes and properties of the Geo-Net-PT vocabulary.

	Term	Description	Specializes
Class	gnpt:GeographicCode	alphanumeric identifier	gn:PlaceName
Object property	gnpt:preferred	has as preferred name	gn:name
	gnpt:alternative	has as alternative name	gn:name
	gnpt:identifier	is identified in some Schema	gn:name
	gnpt:identifies	identifies in some Schema	-
	gnpt:scope	spatial content about	gn:relation
	gnpt:isLocatedOn	is on the land surface of	gn:relation
	gnpt:isAdjacentTo	is adjacent to	gn:relation
	gnpt:isConnectedTo	is connected to	gnpt:isAdjacentTo
	gnpt:isPartOf	is part of	gn:relation
	gnpt:hasPart	has as part	gn:relation
	gnpt:inSchema	specified in	gn:lineage
Datatype property	gnpt:population	is inhabited by *literal value* people	-

- **gnpt:isConnectedTo**, which is symmetric, asserts that its subject is a feature attached to another feature such that objects may flow between them. This property is a subproperty of **gnpt:isAdjacentTo**.

- **gnpt:isPartOf**, which is transitive, declares that its subject is a feature that is a physical or logical component of the object feature.

- **gnpt:hasPart**, which is transitive and inverse of **gnpt:isPartOf**, asserts that its subject is a feature has as physical or logical component of the feature designated by the object.

The Geo-Net-PT vocabulary also has the datatype property **gnpt:population,** which asserts the inhabitants of the location. The value of this property is used to rank features.

4.3 Serialization for the Web

The Geo-Net-PT 02 serialization for the Web uses the Geo-Net and Geo-Net-PT vocabularies. The serialization procedure uses the property **rdfs:label** to provide a human readable name for the serialized resources. The property **rdfs:comment** and the Dublin Core metadata terms are also used to provide a human-readable description of the serialized resources. The serialization also uses properties and classes from known vocabularies including SIOC Core (prefix **sioc:**), FOAF (prefix **foaf:**) and Basic Geo (prefix **geo:**) when they can provide a better description of the semantics of the resource. The use of these vocabularies eases the reuse of the Geo-Net-PT 02 by other communities.

Features. Features are serialized as **gn:Feature** instances, classified by a **gn:FeatureType**, with a human readable **rdfs:label**, members of a **gn:InformationDomain**, and information about provenance captured with a **gn:lineage** property.

Features from the geo-administrative and the geo-physical information domains contain as many labels as place names. These labels are derived from the relations with place names asserted in **gnpt:preferred** and **gpnt:alternative**. Also, the features can be reused as spatial references. For that reason, we use the class **geo:SpatialThing** from the Basic Geo vocabulary to describe these resources as having spatial extent. Finally, if a detailed description of the spatial extent is available as a footprint, the property **gn:footprint** points to it.

Features from the network domain represent Web sites and internet domains. Each network feature is annotated as a resource of type **sioc:Space**. Its label is based on the URL of the Web page or on the registered name of the internet domain. If the network feature describes a Web page, the property **rdfs:seeAlso** links the feature to it.

Place Names. Place names are serialized as **gn:PlaceName** instances, with its lemma, its language and a human readable label. Also, they are tagged as members of one of the **gn:InformationDomain**. Only place names from the geo-administrative and the geo-physical information domains are serialized. Some geographic codes are explicitly serialized as resources of type **gnpt:GeographicCode**. These names are related to features by **gn:identifies** and **gn:identifier**. Also, the property **gnpt:inSchema** links the name with the schema where the geographic code is defined.

Feature Types. Feature types are serialized as instances of the **gn:FeatureType** class, with a human readable label, and members of a **gn:InformationDomain** instance. Feature types are also serialized using the SKOS Core Vocabulary [17] in a separate RDF dataset.

Footprints. Footprints are serialized as **gn:Footprint** instances in a **gn:InformationDomain** with a human readable label, based on the label of the feature that locates, and a lineage description. The geometry description is encoded in GML [19]. Footprints that have spatial extent can be pinpointed with a geodetic point. These characteristics are made explicit using the class **geo:SpatialThing** and the properties **geo:lat** and **geo:long** from the Basic Geo vocabulary.

Reference Systems. Reference systems are serialized as instances of the **gn:ReferenceSystem** class. The description of the reference system can be retrieved from a registry, such as the EPSG Geodetic Parameter Registry (**http://www.epsg-registry.org**), if the URI has the form **urn:ogc:def:crs:{authority}::{id}** where *authority* is the name of a known authority and *id* is the identifier of the coordinate reference system in its registry. A property **rdfs:seeAlso** links the resource with a GML description of it in that registry. If the registry does not contain the description of the reference system used, it is explicitly described in the ontology. The geodetic datum and, if needed, the map pro-

Table 3: Descriptive statistics of Geo-Net-PT 02 concepts; this table only considers direct assertions.

Concept	Adm.	(%)	Phy.	(%)	Net.	(%)
Features	386,067	92.9	5,676	1.4	23,666	5.7
Names	265,044	97.0	8,266	3.0	-	0.0
Footprints	4,597	100.0	-	0.0	-	0.0
Types	62	69.7	25	28.1	2	2.2
other	4,597	58.9	3,207	41.1	-	0.0
Total	660,637	94.2	17,174	2.4	23,668	3.3

jection, are encoded in GML as the value of the property `gn:representation`.

Sources. They are serialized as resources of type `gn:Source` with a rich metadata description. A property `rdfs:seeAlso` links, when available, with an online resource where the original data can be found or is described in detail.

Relationships. The serialization only asserts direct relationships. The serialized relationships include: administrative hierarchy containment, spatial proper containment, administrative and spatial adjacency, spatial connectivity, relative location, geographic scope web resources and location of the owner of a Web resource. The geographic scope and the location are special relationships. The geographic scope of a Web resource is derived from its content [22]. The scope is described by a `gnpt:scope` property that links a feature resource in the network domain to a feature resource in the geo-administrative domain. The owner of a Web resource is described using the concept `sioc:User`, and the ownership by the property `sioc:has_owner`. The location of the user is described using the property `foaf:based_near`.

4.4 Descriptive Statistics

Geo-Net-PT 02 defines 701,209 concepts, most of them administrative features and place names (see Table 3). Some of these concepts have additional types to ease the reuse in the Web of Data: 390,664 administrative and physical features and footprints are classified as `geo:SpatialThing` and 23,666 network features are classified as `sioc:Space`. Geo-Net-PT 02 identifies 22,980 owners of domains, which are classified as `sioc:User` instances. The administrative and physical features are classified by 81 feature types. Postal code, street layout and settlement are the most common feature types found in the geo-administrative domain. Hydrography and touristic resources, such as museums and hotels, are the most common feature types found in the geo-physical domain. The geographic descriptions are in 5 different coordinate reference systems, and there are two different coordinate reference systems for footprints located in Portugal's mainland (ETRS 1989 TM06-Portugal, Lisboa Hayford Gauss IGeoE). Geo-Net-PT 02 has 21 different sources. The main source is CTT (`http://www.ctt.pt`), Portugal's mail services. It provides mainly addresses which outnumber other kinds of features. Excluding CTT, about 5% of the data is derived using rules. Other relevant local sources are Fundação para a Computação Científica Nacional (FCCN, `http://www.fccn.pt`), Agência Portuguesa do Ambiente (APA, `http://www.apambiente.pt`) and Instituto Geográfico Português (IGP, `http://www.igeo.pt`). FCCN provides data about Web sites. APA provides data about features in the geo-physical domain. IGP provides official data about administrative features. Finally, Wikipedia is a complementary source that provides ancillary data about the administrative structure.

4.5 Production Environment

The RDBMS of choice for GKB is PostgreSQL 8.3.6. The support for geographic objects to the PostgreSQL is provided by PostGIS 1.3.6. The language for the development of the object-oriented and resource-oriented persistence systems is Java SE 1.6. The ORM library of choice for the object-oriented persistence system is Hibernate Core 3.3. The support for geographic objects is provided by the extension Hibernate Spatial. The storage system of choice for the resource-oriented persistence system is TBD, a large-scale non-transactional file-based RDF storage.

5. RELATED WORK

The use of Semantic Web languages and concepts of the Web of Data has generated interest in the literature. Abdelmoty et al. [1] review the limitations of the Semantic Web languages for the representation of named places, and proposes the combination of semantic web languages and rules for building and manage place ontologies. Our approach is similar but more flexible, because the ontologies based in Geo-Net can describe detailed footprints and names that are proper names of several places. The GKB system does not support rules, although applications built on it can use them for the derivation of spatial information.

On the side of publication and sharing of these datasets, Goodwin et al. [9] propose the use of RDFS to describe the contents of existing place name databases, and detects some issues, such as the need for partitioning the data in coherent collections and including data provenance information. These aspects are included in Geo-Net.

Linked open data practitioners have created large collections of named places on the Web. The GeoNames Ontology (`http://www.geonames.org/ontology`) contains about 6.2 million named places in RDF. The DBPedia Ontology (`http://dbpedia.org/About`) contains 339,000 places out of 1.17 million instances derived from Wikipedia. The above datasets are interlinked with other datasets, such as the CIA Word Factbook, US Census Data and WordNet. Geo-Net-PT 02 is focused on a limited geographic area (Portugal), but its level of detail is higher than other linked geographic datasets.

6. CONCLUSIONS AND FUTURE WORK

Geo-Net and GKB provide a flexible model and access interface to geographic ontologies. Both also allow the development of improved geographic ontologies, such as Geo-Net-PT 02, which was designed as an answer to the limitations that we found as users of Geo-Net-PT 01. Geo-Net-PT 02 now includes comprehensive information of the physical geography of Portugal, and semantic associations among the features in the physical and administrative domains.

Geographic ontologies truly demonstrate their usefulness if they are reused. The choice of RDF as the abstract data model of the Geo-Net vocabulary is a step in this approach. Future work will follow this path: its publication as Linked Data, and the interlinking of Geo-Net-PT 02 with other datasets of the Web of Data, such as GeoNames and DBPedia, using ontology alignment techniques [7].

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] A. I. Abdelmoty, P. Smart, and C. B. Jones. Building place ontologies for the semantic web: issues and approaches. In *GIR '07: Proceedings of the 4th ACM workshop on Geographical information retrieval*, pages 7–12, New York, NY, USA, 2007. ACM.

[2] D. Beckett. RDF/XML syntax specification (revised) [online]. W3C recommendation, W3, 2004. Available from: http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/.

[3] D. Brickley. Basic Geo (WGS84 lat/long) Vocabulary [online]. Technical report, W3C Semantic Web Interest Group, 2004. Available from: http://www.w3.org/2003/01/geo/.

[4] N. Cardoso, D. Batista, F. J. Lopez-Pellicer, and M. J. Silva. Where in the Wikipedia is that answer? The XLDB at the GikiCLEF 2009 task. In C. P. et al, editor, *Working Notes of CLEF 2009*, Corfu, Greece, September 2009.

[5] M. Chaves. Criação e expansão de geo-ontologias, dimensionamento de informação geográfica e reconhecimento de locais e seus relacionamentos em textos. In *Encontro Linguateca: 10 anos*, Aveiro, Portugal, Sep 2008. In Portuguese.

[6] M. Chaves. *Uma Metodologia para Constru cão de Geo-Ontologias.* PhD thesis, Faculty of Sciences, University of Lisbon, December 2008. In Portuguese.

[7] J. Euzenat and P. Shvaiko. *Ontology matching.* Springer-Verlag, Heidelberg, Germany, 2007. ISBN 3-540-49611-4.

[8] G. Fu, A. Abdelmoty, and C. Jones. Design of a geographical ontology. Technical Report WP D5 3101, SPIRIT project, 2003.

[9] J. Goodwin, C. Dolbear, and G. Hart. Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web. *Transactions in GIS*, 12(s1):19–30, 2009.

[10] C. Gutierrez, C. A. Hurtado, and A. Vaisman. Introducing time into RDF. *IEEE Transactions on Knowledge and Data Engineering*, 19:207–218, 2007.

[11] J. R. Herring. OpenGIS Implementation Specification for Geographic information - Simple feature access - Part 1: Common architecture. OpenGIS Implementation Specification OGC 06-103r3, Open GIS Consortium Inc., October 2006. Version 1.2.0.

[12] L. L. Hill. Core elements of digital gazetteers: Placenames, categories, and footprints. In *ECDL '00: Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries*, pages 280–290, London, UK, 2000. Springer-Verlag.

[13] L. L. Hill. *Georeferencing: The Geographic Associations of Information.* The MIT Press, March 2006. ISBN 978-0-262-08354-6.

[14] W. Kresse and K. Fadaie. *ISO Standards for Geographic Information.* Springer, Berlin, 2004. ISBN 978-3-54020-130-4.

[15] F. J. Lopez-Pellicer, M. Chaves, C. Rodrigues, and M. J. Silva. Geographic Ontologies Production in GREASE-II. Technical Report TR 09-18, University of Lisbon, Faculty of Sciences, LaSIGE, November 2009.

[16] D. Manov, A. Kiryakov, B. Popov, K. Bontcheva, D. Maynard, and H. Cunningham. Experiments with geographic knowledge for information extraction. In *Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references, May 31 2003, Edmonton, Alberta*, pages 1–9, Morristown, NJ, USA, 2003. Association for Computational Linguistics.

[17] A. Miles and S. Bechhofer. SKOS simple knowledge organization system reference [online]. W3C proposed recommendation, W3C, June 2009. Available from: http://www.w3.org/TR/2009/PR-skos-reference-20090615/.

[18] J. Nogueras-Iso, F. J. Zarazaga-Soria, J. Lacasta, R. Béjar, and P. R. Muro-Medrano. Metadata standard interoperability: application in the geographic information domain. *CEUS*, 28(6):611––634, 2004.

[19] C. Portele. OpenGIS Geography Markup Language (GML) Encoding Standard. OpenGIS Standard OGC 07-036, Open Geospatial Consortium Inc., July 2007. Version 3.2.1.

[20] C. Rodrigues. An ontology of the physical geography of portugal. Master's thesis, University of Lisbon, Faculty of Sciences, June 2009.

[21] A. Seaborne and E. Prud'hommeaux. SPARQL query language for RDF [online]. W3C recommendation, W3C, Jan. 2008. Available from: http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/.

[22] M. J. Silva, B. Martins, M. Chaves, N. Cardoso, and A. P. Afonso. Adding geographic scopes to web resources. *CEUS – Computers, Environment and Urban Systems*, 30(4):378–399, July 2006.

[23] Technical Committee ISO/TC 211, Geographic Information/Geomatics. Geographic information – Geography Markup Language (GML). Technical Report ISO 19136:2007, International Organization for Standardization, Geneva, Switzerland., 2007.

[24] F. van Harmelen and D. L. McGuinness. OWL Web Ontology Language Overview [online]. W3C recommendation, W3C, Feb. 2004. Available from: http://www.w3.org/TR/2004/REC-owl-features-20040210/.

[25] M. Wegener. *Spatial Models and GIS: New Potential and New Models*, chapter Spatial Models and GIS, pages 3–20. Taylor and Francis, London, 2000. ISBN 978-0-74840-846-7.