Geographic Ontologies Production in GREASE-II

Francisco J. Lopez-Pellicer
Marcirio Chaves
Catarina Rodrigues
Mário J. Silva

DI–FCUL TR–09–18

November 2009

Departamento de Informática
Faculdade de Ciências da Universidade de Lisboa
Campo Grande, 1749–016 Lisboa
Portugal

Geographic Ontologies Production in GREASE-II

Francisco J. Lopez-Pellicer [†], *Marcirio Chaves* [‡], *Catarina Rodrigues* [‡] *and Mário J. Silva* [‡]

[†] Departamento de Informática e Ingeniería de Sistemas, Centro Politécnico Superior, Universidad de Zaragoza, Spain
[‡] Departamento de Informática, Faculdade de Ciências da Universidade da Lisboa, Portugal

Abstract

This report presents the development of authoritative geographic knowledge and related conceptualization tools in the project GREASE-II. The feature concept, an abstraction for geographical entities, is the centre of the conceptual model, which is oriented towards a representation of geographic data as interlinked web resources. The conceptualization tools are the Geographic Knowledge Base (GKB) system (version 2.1), a system for geographic knowledge management, and the Geo-Net vocabulary for describing geographic knowledge. Both contribute to the production of the new version of the geospatial ontology of Portugal, Geo-Net-PT 02. The innovations introduced since version 1.0 of the GKB system and Geo-Net-PT 01 include a formalization of the model, the support for generic property sets, detailed information provenance, better geographic descriptions and the definition of Geo-Net vocabulary, a domain vocabulary.

Contents

List of Figures

List of Tables

Acronyms

APA	Agência Portuguesa do Ambiente
API	Application Programming Interface
CC-BY	Creative Commons Attribution 3.0 License
CRS	Coordinate Reference System
CTS	Coordinate Transformation Service
CURIE	Compact URI
DCMI	Dublin Core Metadata Initiative
EPSG	European Petroleum Survey Group
ESRI	Environmental Systems Research Institute
ETRS	European Terrestrial Reference System
FCCN	Fundação para a Computação Científica Nacional
FOAF	Friend of a Friend project
GIS	Geographic Information System
GKB	Geographic Knowledge Base management system
GML	Geography Markup Language
GOS	GKB Ontology Serializer
HENDRIX	HENDRIX is an Entity Name Desambiguator and Recognizer for Information Extraction
IETF	Internet Engineering Task Force
IGeoE	Instituto Geográfico do Exército
IGP	Instituto Geográfico Português
INE	Instituto Nacional de Estatística
ISO	International Organization for Standardization
ITRF	International Terrestrial Reference System
KOS	Knowledge Organization System
OGC	Open Geospatial Consortium

ORM	Object-relational mapping
OWL	Web Ontology Language
RCF	Request for Comments
RDF	Resource Description Framework
RDFa	RDF in annotations
SFA	Simple Feature Access
SIOC	Semantically-Interlinked Online Communities Project
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
Turtle	Terse RDF Triple Language
UML	Unified Modelling Language
URI	Uniform Resource Identifier
UTM	Universal Transverse Mercator
WGS	World Geodetic System
WKT	Well-known text markup language

Conventions

Fonts and formatting

In this report, we use *italics* for emphasis and for identifying examples embedded in paragraphs. For URIs and machine-readable examples, we use a `typewriter` font.

Terms and definitions

For this report, the following terms and definitions apply:

Coordinate system The set of mathematical rules for specifying how to assign points from coordinates (ISO 19111:2007).

Datatype property A property that relates instances of a class and literals.

Datum A parameter or set of parameters that define the position of the origin, the scale, and the orientation of a coordinate system (ISO 19111:2007).

Feature An object grounded in some way by one or several locations. The term *feature* is widely used in the domain of geographic information to denote an element of a geographic dataset. This term is part of the terminology of ISO 19100 standards for geographic information and is applied to abstractions, as type or instance, of located phenomena (Kresse and Fadaie, 2004).

Feature type A classifier for feature types. This term is part of the terminology of ISO 19100 standards for geographic information and is applied to denote types of located entities (Kresse and Fadaie, 2004).

Functional property A property that can only hold a single value for a given individual.

Geodetic datum A datum describing the relationship of a 2- or 3-dimensional coordinate system to the Earth (ISO 19111:2007).

Inverse functional property A property that that can only hold a single individual for a given value.

Map projection Any method of representing the surface of a sphere or other shape on a plane.

Object property A property that relates instances of two classes.

Symmetric property A property where if the pair (a, b) is an instance of the property then the pair (b, a) is also instance of the property.

Transitive property A property where if the pairs (a, b) and (b, c) are instances of the property then the pair (a, c) is also instance of the property.

Examples

The examples of descriptions of geographic knowledge are about *Lisboa*, the capital of Portugal (`http://en.wikipedia.org/wiki/Lisboa`). They are often written in Turtle language. The examples use human readable identifiers, such as `:Lisboa` for the feature and `:LisboaName` for its place name.

Prefixes and namespaces

Throughout this document, vocabularies and examples about geographic knowledge use the pairs of prefixes and namespaces of the table below:

Prefix	Namespace
gn:	http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net#
gnpt:	http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net-pt#
gnpt02:	http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net-pt-02#
rdfs:	http://www.w3.org/2000/01/rdf-schema#
rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
xsd:	http://www.w3.org/2001/XMLSchema#
dc:	http://purl.org/dc/elements/1.1/
geo:	http://www.w3.org/2003/01/geo/wgs84_pos#
sioc:	http://rdfs.org/sioc/ns#
foaf:	http://xmlns.com/foaf/0.1/
ex:	http://www.example.com/simple#
:	#

Vocabularies

This document uses concepts and relations defined in the vocabularies of the following table:

Vocabulary	Prefix	Description
DCMI Metadata Terms	dc:	A standard vocabulary for cross-domain information resource description (see http://dublincore.org/documents/dcmi-terms/).
Basic Geo vocabulary	geo:	A basic RDF vocabulary for representing latitude, longitude using WGS84 as reference datum (see http://www.w3.org/2003/01/geo/).
SIOC ontology	sioc:	An ontology for describing the information contained in online communities (see http://rdfs.org/sioc/spec/).
FOAF ontology	foaf:	An ontology for describing people, their activities and their relations (http://xmlns.com/foaf/spec/).

Notation in diagrams

The diagrams that appear in this report describing structure are presented using the UML static structure diagram. The diagrams that describe graphs are presented using RDF graph diagrams (see Miller and Manola, 2004).

This page intentionally left blank.

1 Introduction

Whatever occurs, occurs in a physical or mental geographic location (Wegener, 2000) and hence can be described, filed, visualized, queried, retrieved, and understood using geographic terms (Hill, 2006). These activities require the use of search engines designed for searching on collections of resources rich in geographic information, which retrieve and rank relevant documents applying methods based on spatial reasoning algorithms.

For that purpose, we need to support the description and discovery of geographic data. Describing geographic data entails indentifying place names and the features that they designate. By discovery, we mean the identification of geographic features given (some of) their attributes. Our concern is the provision of support for simple geospatial data identification. That this, providing the answer to questions like given a set of place names, which available features best match the set of place names and what are their locations on Earth?

In this report, we present the conceptualization and methods for authoritative geographic knowledge management developed in project GREASE-II, *Geographic Reasoning for Search Engines* (Silva et al., 2006). These include providing a representational machinery in the form of (i) definitions, that is, a specification of domain concepts represented in a formal language, and (ii) tools, that is, software for managing the classes and instances of the specification. For this purpose, we introduce in this report a revised conceptual model for geographic knowledge, including a vocabulary for sharing geographic knowledge, a geographic knowledge management system updated to this model, and the software for serializing this knowledge as Web ontologies:

- **Geo-Net**, geographic knowledge vocabulary, is intended to support the data description and discovery in contents with geospatial data.
- **GKB**, Geographic Knowledge Base management system, is a knowledge management system that implements a domain-independent meta-model for integrating geographic knowledge collected from multiple sources. The core of the GKB system is an abstract data model, whose purpose is to describe formally objects grounded in some way to locations. The GKB system has components for loading, querying and exporting knowledge. Chaves et al. (2005b) describes the original version of the GKB system. This report presents a reengineered GKB with new representation features including multilingual support, description of feature types and their relationships, features and types from the geo-physical domain and shapes representing footprints.
- **GOS**, Geographic Ontology Serializer, is a new component of the GKB system for the serialization of the content represented following the GKB abstract data model and the Geo-Net vocabulary as OWL 1 (Dean and Schreibe, 2004) and SKOS (Miles and Bechhofer, 2009) in the desired formats such as RDF/XML Turtle (Beckett and Berners-Lee, 2008) and Notation 3 (Berners-Lee, 1998).

With the above machinery, we have implemented an authoritative geographic knowledge dataset:

- **Geo-Net-PT 02** is a geospatial ontology of Portugal, contains more than 700.000 concepts stored in a GKB system (version 2.1). The Geo-Net-PT 02 is an extension to the Geo-Net-PT 01 ontology presented in Chaves et al. (2005b). The serialization of Geo-Net-PT 02 with the GOS is released as a public resource.

Geo-Net-PT 02 is built from *gratis* information sources, that is, sources which are distributed by an authority and available free of charges. This concept is opposite to *libre* sources, that is, sources with very few limitations on distribution or modification, which are common in the Web. Authoritative is a property related to trustiness, a characteristic that libre sources often miss. An example of libre source is a dataset without metadata or license description that anybody could have edited. The distinction between *gratis* and *libre* sources available on the Web is relevant in the context of the GREASE-II project. The evaluation of tools, methods and interfaces developed in GREASE-II depends on the degree of trust on the geographic knowledge dataset used.

Geo-Net-PT 02 is available in the XLDB Node of Linguateca (`http://xldb.di.fc.ul.pt/wiki/Geo-Net-PT_02`). This ontology is licensed under a Creative Commons Attribution 3.0 License (CC-BY). To view a copy of this license, visit `http://creativecommons.org/licenses/by/3.0/`.

1.1 Report organization

This report is organized as follows. The conceptual model and its formalization are presented in Section 2. The Geo-Net vocabulary is detailed in Section 3. The architecture of the GKB system (version 2.1) and its latest improvements can be found in Section 4. Details of the construction, sources and lessons learned in the development of Geo-Net-PT 02 are discussed in Section 5. Finally, Section 6 presents final remarks.

2 Conceptual model for geographic knowledge

This section presents the conceptual model that we use for reasoning about geographic knowledge, which extends the conceptual model presented in Chaves et al. (2005b). This model is based on ideas presented in Hill (2000), Manov et al. (2003), and Fu et al. (2003). The conceptual model for geographic knowledge is materialized in the Geo-Net vocabulary and the GKB system (version 2.1) to be described in the next two sections.

This section is organized as follows: characterization of the conceptual model, formalization of the conceptual model as a metamodel, and lessons learned during the implementation of the metamodel in different scenarios.

2.1 Characterization

The characterization includes the description of the base knowledge organization systems (KOS) that serve as framework, the characteristics of the geographic knowledge that the model manages, and its concepts, relations and constraints. The main characteristics of the conceptual model are summarized in Table 1.

KOS base models KOS is a term introduced by Hodge (2000) intended to encompass all types of schemas for organizing information and promoting knowledge management. The KOS base model provides the semantic structure and specifies features for the conceptual model, such as the vocabulary for describing concepts, their meaning and the relations among them.

Our context is the model of concepts such as cities, like *Lisboa* and related Web resources, such as the Web page about *Lisboa* in the Wikipedia, and relations among these concepts. The creation

Characteristic	Summary
Base models	- Geographic dictionary and Ontology.
Conceptualizing	- Declarative and configurational geographic knowledge.
Concepts	- Features, place names, feature types, footprints and relation types.
Relations	- Features are known by place names, classified by feature types, grounded by footprints, and related each other.
Mandatory constraints	- Features are classified by enumeration, footprints grounds only one feature, and relations between features are typed.
Conditional constraints	- Feature instantiation.

Tab. 1: Summary of the characteristics of the conceptual model.

of a knowledge representation of the facts that a text about a geographic resource, such as the the following quote:

> Lisboa is the geographic scope of the Web page `http://en.wikipedia.org/wiki/Lisboa`

involves the combination of a geographic dictionary that provides geographic features, such as *Lisboa*, and an ontology defining concepts, such as *geographic scope*.

Geographic dictionary Contains information about places, their names, types, locations, relationships, and other descriptive information. This dictionary is also known as gazetteer. This KOS relates coordinates, names and categorization schemas to conceptualizations of geographic locations. We can ask to this kind of KOS *where* and *what* queries, such as:

> Where is located Lisboa?
>
> What are the names of streets in Lisboa are near of something named Carmo?

Ontology Tools and products defined by their use as providers of representational machinery (Gruber, 2009) for an explicit specification of a conceptualization (Gruber, 1993), that is, a specification of domain concepts representing complex relationships, rules and axioms that takes the form of formal definitions.

Conceptualized knowledge Mark (1993) proposes a classification of geographic knowledge in three broad categories:

- *Declarative geographic knowledge*: facts that may or may not be associated with a clear and crisp idea of where are these named places are located on Earth. In the phrase *Lisbon is the capital of Portugal,* we express a set of facts that are independent of where on Earth *Lisbon* and *Portugal* are located.
- *Configurational geographic knowledge*: facts that range from basic topological relations to complete coordinate descriptions. In the phrase *Lisbon is part of Portugal and is located at 38° 42′ N, 9° 11′,* we express a basic configuration as a topological relation and a location described with coordinates.

- *Procedural geographic knowledge*: rules and clues that allow a person to perform a spatial task using geographic knowledge.

We use both declarative and configurational knowledge in the conceptual model. The distinction between declarative and configurational knowledge is relevant because, when using geographic knowledge in everyday texts, we are likely to say *Lisbon, Portugal* rather than *Lisbon (38° 42′ N, 9° 11′)* or similar configurational description. We do not use *procedural geographic knowledge* in the conceptual model. Procedural geographic tasks, for example grounding a document, are performed by other tools, such as the HENDRIX text-mining tool (Batista, 2009), also developed in the GREASE-II project.

Concepts and relations The description of a geographic location can be complex and there are many points of view involved (see López-Pellicer et al., 2007). The concepts and relations of the conceptual model are a subset, although widely used in information retrieval tasks, of all the possible information about a geographic location. The concepts are (see Figure 1):

- *Feature*, a geographic feature,
- *Place name*, the proper names given to geographic features,
- *Feature type*, a classifier of geographic features,
- *Footprint*, location information that grounds geographic features, and
- *Relation type*, a classifier of relations between geographic features.

Figure 1 also contains information represented as relations among objects:

- Features are *known by* place names,
- Features are *grounded by* footprints,
- Features are *classified by* feature types, and
- Features are *related with* other features with typed relations.

Mandatory constraints The constraints are about the footprints, the relationships and how features are classified.

The footprint of physical and mental features might be as complex as a survey description of the boundaries of the feature or as simple as a pinpoint. This configurational knowledge is restricted by the following rules:

1. a footprint can be instantiated in the conceptual model if it grounds at least a feature, and
2. a footprint cannot ground more than one feature.

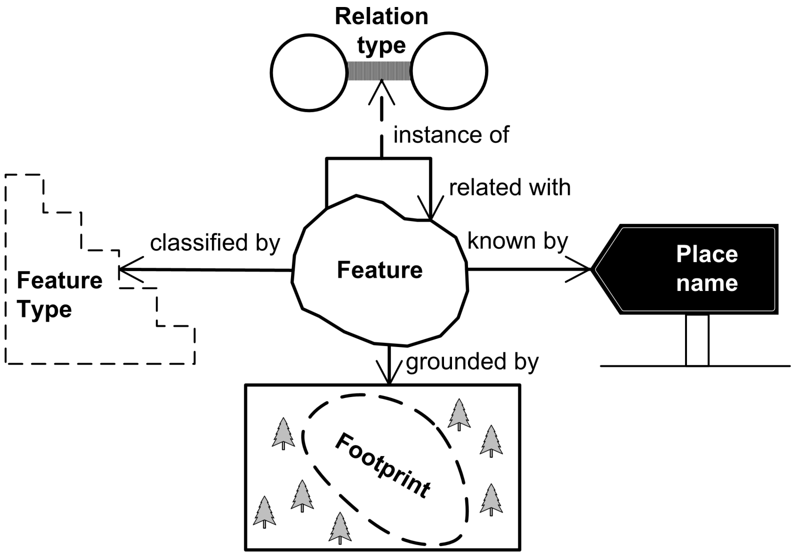


Fig. 1: Conceptual model for Geographic Knowledge.

This constraint assumes that footprints are tightly bounded to only one feature.

Relationships document the configurational geographic knowledge derived from the footprints. These include relations such as *part of*, *adjoint to* and *connect to.* Relationships also describe declarative geographic knowledge, such as *capital of*, *administrative division of* and *former part of.* At least, relationships should be documented to determine if they are pre-calculated from spatial relationship or they extend beyond the configurational knowledge.

Classification is the process of assigning elements or units to classes carrying some kind of geographic meaning according to some criteria. Our conceptual model does not intend to enforce a typing schema. We assume that the set of features classified by a type is defined by explicit enumeration, that is, a set of statements that assert membership. A set defined by enumeration allows to describe that these features share a set of properties without stating which or how fuzzy they are. The main drawback of this assumption is that the system cannot use the properties associated to each feature type to infer or verify knowledge.

Conditional constraints Each geographic feature is required to be partially specified in terms of the declarative geographic knowledge, that is:

1. a feature can be instantiated in the conceptual model if it has at least a place name or a label and a type, and
2. two different features can have the same name and type signature.

This represents the fact that people can recognize place names in the context of a communication with little context. Place names are inherently ambiguous, and they do not need to be georeferenced to be acknowledged.

In some cases, the name that identifies the feature is just a label. For example, the name of a digital resource that describes the city of Lisboa can have as name *Fernando Pessoa: Lisbon,*

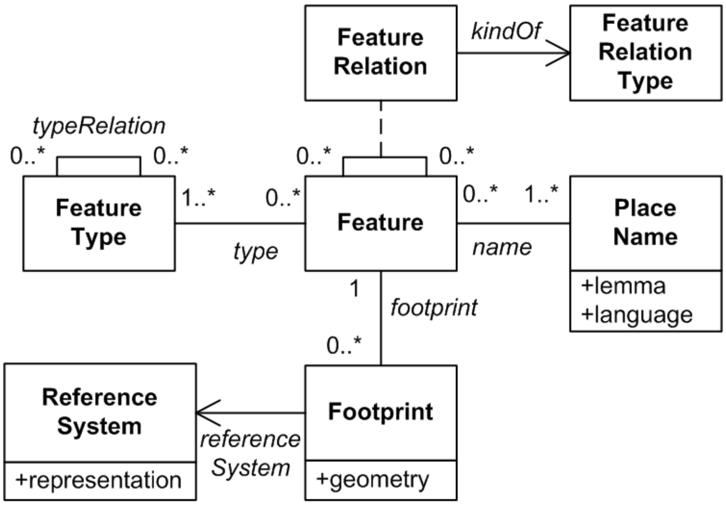


Fig. 2: Core metamodel.

what the tourist should see (http://www.shearsman.com/pages/books/catalog/2008/pessoa_lisbon.html). Only *real* place names deserve the qualification of place names, and, hence, if a kind of feature does not have real place names, the place name is not needed to the instantiation.

2.2 Formal model

The description presented above introduces the core elements of our geographic abstract model: features, place names, feature types, footprints and relation types. The model is backwards compatible with previous versions of the GKB system. It is divided in a core metamodel and a set of rules for implementing and extending the metamodel. An example application is the GKB instance that manages the Geo-Net-PT 02 ontology of Portugal to be described in section 5.

Core metamodel The core metamodel (Figure 2) includes the concepts described as first class citizens in the conceptual model, that is, having an identifier:

- Geographic features (instances of the 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 that classify geographic features are represented by the class `FeatureType`.
- Locations that are referred by geographic features are represented by the class `Footprint`. The spatial description of the location is captured in the field `geometry`.
- Relations between features are represented by the association class `FeatureRelation`, which is classified by the class `FeatureRelationType`.

The relations `type` and `name` relate a `Feature` with the classes `FeatureType` and `PlaceName,` respectively. These relations have a cardinality of 1..* in the `FeatureType` and

PlaceName ends. This reflects the requirement in the conceptual model that each geographic feature requires to be partially specified in terms of declarative knowledge, that is, 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 association class FeatureRelation describes binary relationships among features. These relations must be qualified by an instance of the meta class FeatureRelationType.

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

The class Footprint is related with the class Feature by 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 following sub-properties:

- coordinates, a set of coordinate points or a set of instances of the type Geometry.
- type, 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 data type Geometry describes how to document a geometric shape as a set of points that conform to a rule, and then how to project them into a surface. These coordinate points take 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. 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 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.

General rules The additional classes (see Figure 3) and general rules for implementing and extending the metamodel are enumerated below:

- The classes Feature, PlaceName, FeatureType, Footprint and ReferenceSystem are instances of the metaclass Concept.
- The association class FeatureRelation and the associations typeRelation, name, type, footprint and referenceSystem are instances of the metaclass Relation.
- The class FeatureRelationType is instance of the metaclass RelationType.
- The class DataItem represents an item of knowledge. The classes InformationDomain, Source and Metadata provide meta information about the items of knowledge.
- Data items can be organized in collections, named information domains, represented by instances of the class InformationDomain. Information domains are useful when a group of features shares something and it is convenient to group them under a common label for management purposes.

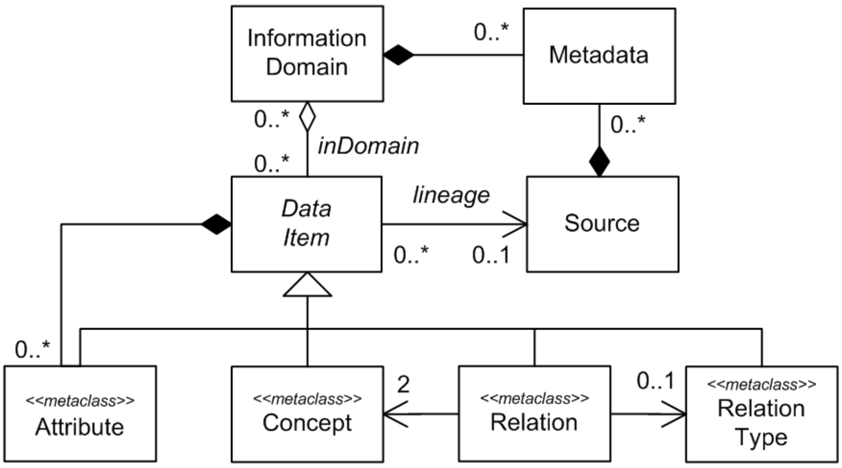


Fig. 3: Extended metamodel: additional classes for implementing and extending the core metamodel.

- Each data item may have a provenance description, instance of the class `Source`, which indicates that this item was originally sourced from another resource. The provenance description can be shared among data items. Each source has its own descriptive metadata.
- Each data item may have as many attributes as needed, represented by instances of the class `Attribute`. These attributes should be modelled as fields if they do not have lineage.
- Each instance of the class `Relation` defines a binary relation between concepts and may have an additional description in a relation type. This description is mandatory when the relation is among instances of the class `Feature`. Relations must be modelled as association classes if they have attributes, lineages or a description as a relation type.
- Relations between instances of two different information domains and their corresponding relation types, if any, are known as inter-domain relationships. This kind of relationships represents links between information domains.
- Each of the elements can be subclassed in the implementations. The semantics and the constraints of the original elements must be maintained. Properties can be replaced by other more appropriate, while maintaining the original semantics.
- The cardinalities expressed in the metamodel can be modified when subclassing. The minimum cardinality can be increased, but never above the original maximum cardinality. For example, an optional cardinality 0..1, such as `lineage,` can be transformed in mandatory 1..1. The maximum cardinality can be decreased, but never below the original minimum cardinality. For example, the cardinality of the `PlaceName` end of the relationship name of 1..* can be decreased to 1..1.

2.3 Best practices

The following best practices reflect lessons learned during the implementation of the conceptual model in different scenarios.

Temporal scope The conceptual model does not include an explicit mention to the description of the temporal scope of the facts. If the fact has a temporal scope, the recommended practice is to document the temporal scope in the descriptive metadata of the fact.

The canonical form is internationalized The recommended practice is to state that the values of the property `lemma` are Unicode strings. This allows storing names such as Лиссабон, a variant name of Lisboa in non-roman characters. If the implementation requires the persistence of values in other character encodings, such as ASCII, these never must be stored in the property `lemma`.

Natural identifier for place names The recommended practice is no two Unicode-equivalent string values of the property `lemma` should appear in separate place name descriptions with the same value of the property `language`. The implementation may create an unique identifier for each place name derived from the values of the properties `lemma` and `language`.

The canonical form does not encode any relationship Place names are context dependent. They often require for their understanding an additional context in the form of additional details. Consider for example *Lisbon, Maine. Lisbon* is a place name and the additional context is *Maine*, a place name too. The conceptual model does not consider such additional context as part of the place name. The recommended practice for applications is to assume that the property lemma is indivisible and does not contain additional context information.

The canonical form is indivisible Place names can contain a proper name part and a generic part that may hint the feature type. However, the inference of the generic part is ambiguous. For example, the place name with the lemma *Rio Douro* can refer to the *Douro* River or an administrative division with that name. The recommended best practice for applications is to avoid practices such as splitting the lemma in parts to infer possible feature types.

Matching the canonical form Names are vague and ambiguous, not only because language users can use them to refer to vague places, but also because they can use different case, spellings and abbreviatures. For example, *Olissippo*, *Olisipo*, *Ulisipo*, *Olisponna* but also *OLISIPO* and *olisipo* are different strings, but they all are versions of the historical name of *Lisboa* during the Roman Empire. The decision of how a string is matched to a name is dependent of how the conceptual model is implemented. The recommended practice is to formalize the method that transforms a string to a form considered canonical. This method can be as simple as the lower case function, and as complex as a complete lemmatization.

Controlled values for language The recommended practice for values of the property language is to encoding these values using the language tags defined in IETF RCF 4646 and 4647 (Phillips and Davis, 2006a,b). These tags allow identifying regional variants of the language.

Qualified relations between features and place names A place name can play several roles along its existence as a communication tool that often reveals significant patterns of environment, settlement, colonization, exploration, organization, historical facts and folk etymology. The recommended practice is to add a relation type to the relation between a feature and a place name describing the kind of role played by the name, and create specialized relations with a default role.

Representing ternary relations between features and place names The conceptual model only considers binary relations between features and place names. Some relations, such as those that involve translated place names, for example *New Lisbon is the translation to English of the Portuguese name Nova Lisboa*, are ternary relations. The recommended practice for ternary relations that involve an additional name is to model them as a binary relationship between the feature and the place name and add to the metadata of the relation the additional name.

Features without a real place name Implementations can deal with features, such as a Web resource, that do not have a *real* place name but a label. In this circumstance, the recommended practice is twofold: to subclass the class `PlaceName`, and to consider the new subclass as a simple label store.

Features with geographic codes Implementations may need to store geographic codes, that is, alphanumerical representations in a schema that uniquely identifies a feature. The recommended practice for geographic codes is to define a class named `GeographicCode`, subclass the class `PlaceName`. In this new class, the `language` field is ignored and replaced by a field that identifies the schema. The combination of the latter with the `lemma` uniquely identifies instances of this new class.

Syntax encoding scheme for geometries and coordinate reference systems Powell et al. (2007) defines *syntax encoding scheme* as the set of strings and the associated set of rules that describe a mapping between that set of strings and a set of resources. The encoding rules may define the syntax to serialize the resource within a text string. The recommended practice for the serialization as plain text of crisp and fuzzy 2D geometries, from the `geometry` field of the class `Footprint`, and coordinate reference systems, from the `representation` field of the class `ReferenceSystem`, is to use WKT, a text markup language for representing geometry data, as syntax encoding scheme. WKT is a GIS industrial standard and is supported by many spatial databases. The WKT syntax for 2D and 3D geometries is described in the OGC SFA specification (Herring, 2006). Crisp 2D geometries, such as points, lines, polygons and combinations of these, are represented as WKT geometries with coordinates in 2D. Fuzzy 2D geometries are represented as WKT geometries with coordinates in 3D (see Jones et al., 2008). The WKT syntax for coordinate reference systems is described in the OGC CTS specification (OGC, 2001).

The recommended practice for the serialization as XML is to encode the geometry data as GML, an OGC and ISO standard XML Schema grammar used to describe geographic information (Portele, 2007, Technical Committee ISO/TC 211, Geographic Information/Geomatics, 2007b). The use of GML in plain text serialization is admissible only if it is serialized as a XML literal.

Use general purpose metadata standards The recommended practice for the descriptive metadata about sources is to use the Dublin Core metadata element set (DCMI, 2003) as vocabulary. The Dublin Core metadata element set is a standard for the description of cross-domain information sources regardless the media format, the area or specialization, the kind of resource or its cultural origin. When available, the description of a source should be Dublin Core metadata harvested from the source. It is common to find Dublin Core metadata about data sources in the geographic information domain (Nogueras-Iso et al., 2004).

	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	-

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

3 Geo-Net vocabulary

The Geo-Net vocabulary, based on the above described conceptual model for geographic knowledge, is intended to support the description and the discovery of toponymic datasets. By description, we mean the identification of the relevant toponyms and the features they refer. By discovery, we mean the identification of features that answer the question: given a set of names, which available features best match the set of names and what is their location on Earth?

Table 2 presents the terms of the Geo-Net vocabulary. The designated namespace for this vocabulary is:

```
http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net#
```

with the associated prefix `gn:`. The Geo-Net vocabulary, encoded in OWL 1, is available from `http://xldb.di.fc.ul.pt/wiki/Geo-Net-PT_02` .

The introduction of this vocabulary will be illustrated with a running example describing Lisboa, Portugal. The examples presented here will show how to map to this vocabulary the text in the quote:

> The city of Lisboa is the capital of Portugal. Lisboa is known in English as Lisbon. Lisboa is known in German as Lissabon. We can find additional information about Lisboa in Wikipedia. For example, its coordinates are 38° 42′ N, 9° 11′ W.

3.1 Core

Classes The Geo-Net vocabulary defines the following basic classes.

Class gn:GeographicConcept Represents the descriptive elements about geographic features introduced in the conceptual model. 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. The class `gn:GeographicConcept` is the superclass of the classes `gn:Feature`, `gn:PlaceName`, `gn:FeatureType` and `gn:Footprint`, which are disjoint.

Class gn:Feature Represents any meaningful object that can be grounded directly or by reference.

As an example, we give a description of *Lisboa*. We will first coin an identifier, and type it as a `gn:Feature`:

```
:Lisboa a gn:Feature .
```

A feature can include additional annotations. For example, we can add multilingual labels and a descriptive statement using DCMI metadata terms, or links to resources providing additional information, using the property `rdfs:seeAlso`:

```
:Lisboa dc:title "Lisboa"@pt , "Lisbon"@en , "Lissabon"@de ;
        dc:description "Lisbon is the capital of Portugal"@en ;
        rdfs:seeAlso <http://pt.wikipedia.org/wiki/Lisboa> .
```

Class gn:PlaceName Represents proper names of one or more features.

Following with the running example, we can define from the quoted text three place names for Lisboa:

```
:LisboaName a gn:PlaceName ;
            dc:title "Lisboa"@pt .
:LisbonName a gn:PlaceName ;
            dc:title "Lisbon"@en .
:LissabonName a gn:PlaceName ;
            dc:title "Lissabon"@de .
```

Class gn:FeatureType Represents classifiers for features.

From the example, we can define *city* as a feature type:

```
:City a gn:FeatureType ;
     dc:title "City"@en .
```

Class gn:Footprint Represents a description of a location on a surface that can ground a feature.

We can use specialized vocabularies for describing the footprint. For example, the Basic Geo vocabulary (Brickley, 2004) specifies a RDF vocabulary for describing Points with `lat`itude, `long`itude, and `alt`itude properties in the WGS84 reference datum specification. Hence, we can describe the footprint of *Lisboa* as a point using the properties `lat` and `long` of the Basic Geo vocabulary (prefix `geo:`):

```
:LisboaFootprint a gn:Footprint ;
                 geo:lat "38.7"^^xsd:double ;
                 geo:long "-9.183333"^^xsd:double .
```

Properties The Geo-Net vocabulary defines the following properties:

Object property gn:name Asserts that a resource of type `gn:Feature` has as a proper name a resource of type `gn:PlaceName`.

For example, we can specify that *Lisboa*, *Lisbon* and *Lissabon* are names for the feature *Lisboa:*

```
:Lisboa gn:name :LisboaName , :LisbonName , :LissabonName .
```

Object property gn:type Asserts that a resource of type `gn:Feature` is classified by a resource of type `gn:FeatureType`.

Then we specify that Lisboa is a city using this property in a triple:

```
:Lisboa gn:type :City .
```

Object property gn:footprint Asserts that the location of a resource of type `gn:Feature` is described by a resource of type `gn:Footprint`.

Thus, we can specify the location of Lisboa with:

```
:Lisboa gn:footprint :LisboaFootprint .
```

Object property gn:relation Asserts that two instances of the class `gn:Feature` are related.

In the example, Lisboa is related to Portugal. Then:

```
:Lisboa gn:relation :Portugal .
:Portugal a gn:Feature ;
          dc:title ‘‘Portugal’’@pt .
```

Functional data property 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`.

We can now add to the description of the names Lisbon and Lisboa the following assertions:

```
:LisboaName gn:lemma "Lisboa"^^xsd:string .
:LisbonName gn:lemma "Lisbon"^^xsd:string .
:LissabonName gn:lemma "Lissabon"^^xsd:string .
```

Functional data property gn:languageCode 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 a typed literal with type `xsd:string`.

The following assertions complete the description of the names:

```
:LisboaName gn:languageCode "pt"^^xsd:string .
:LisbonName gn:languageCode "en"^^xsd:string .
:LissabonName gn:languageCode "de"^^xsd:string .
```

Functional data property gn:geometry Relates a resource of type `gn:Footprint` with a canonical description of the footprint shape encoded as a literal.

For example, we can describe the location of Lisboa as a Point geometry encoded in GML:

```
:LisboaFootprint gn:geometry """
                      <gml:Point
                           xmlns:gml="http://www.opengis.net/gml">
                           <gml:pos>38.7, -9.183333</gml:pos>
                      </gml:Point>"""^^rdf:XMLLiteral .
```

3.2 Spatial referencing systems

The property `gn:geometry` can hold geometries 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.

Class gn:ReferenceSystem Represents spatial reference systems. This class is disjoint with other Geo-Net concepts.

Datatype property gn:representation Relates a gn:ReferenceSystem with a literal description of the spatial reference systems encoded in a well known syntax encoding scheme, such as WKT.

Functional object property 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.

For example, we can assert that the geometry of the resource `:LisboaFootprint` uses as reference system the geodetic reference system WGS 84, a common reference system, which is identified with the URI `urn:ogc:def:crs:EPSG::4326` in the EPSG Geodetic Parameter Registry (see `http://www.epsg-registry.org/`):

```
:LisboaFootprint gn:referenceSystem urn:ogc:def:crs:EPSG::4326 .
```

3.3 Collections

Data is organized in collections of concepts, named information domains. Information domains are useful when a group of features shares something and it is convenient to group them under a common label.

Class gn:InformationDomain Represents a collection of geo concepts. An information domain is a convenience concept. Hence, the type `gn:InformationDomain` is disjoint from the type `gn:GeographicConcept`.

An use case for information domains is grouping large collections of homogeneous concepts. For example, the Geo-Net-PT ontology defines three information domains: geo-administrative, geo-physical and network (Chaves et al., 2005b). The properties of a network resource, such as a Web site about the Tejo River, and those of a geo-physical resource, such as the Tejo River, can be quite different.

Object property gn:inDomain Marks resources of type `gn:GeographicConcept` as member of a information domain. A resource can be member of several information domains.

Information domain is a convenience concept. For example, we can tag all the concepts of the running example as members of the geo-administrative information domain:

```
:GeoAdministrative a gn:InformationDomain ;
                   dc:title "Geo-Administrative"@en.
:Lisboa gn:inDomain :GeoAdministrative .
:LisboaName gn:inDomain :GeoAdministrative .
:LisbonName gn:inDomain :GeoAdministrative .
:LissabonName gn:inDomain :GeoAdministrative .
:LisboaFootprint gn:inDomain :GeoAdministrative .
:City gn:inDomain :GeoAdministrative .
```

There is no restriction to assert inter-domain relationships using the `gn:relation` property. The follow example is valid:

```
<http://www.cm-lisboa.pt/> a gn:Feature ;
                           gn:inDomain :Network ;
                           dc:title """Webpage of the
                                       Lisbon City Council"""@en ;
                           gn:relation :Lisboa .
```

3.4 Provenance

Without a proper description of provenance, potential users of a dataset described with the Geo-Net vocabulary would not have criteria for evaluating the authority of assertions.

Class gn:Source Represents sources whose contents have been added to a toponymic dataset described with the Geo-Net vocabulary. `gn:Source` is disjoint with other Geo-Net concepts.

Functional object property gn:lineage Indicates that a concept or relationship was originally sourced from another resource relating a concept or statement with its source, a resource of type `gn:Source`.

For example, if we want to assert that the provenance of the resource `:LisbonName` is the resource `:Wikipedia,` we can represent this assertion as:

```
:LisbonName gn:lineage :Wikipedia .
```

3.5 Use case: geographic annotation of an XHTML document

In this example, we annotate a text sentence, *Lisbon is the capital of Portugal,* found in a simplified XHTML document:

```
<html xmlns="http://www.w3.org/1999/xhtml">
    <body>
        <p>Lisbon is the capital of Portugal</p>
    </body>
</html>
```

The first step is to provide an instance of the Geo-Net vocabulary, which is depicted graphically in Figure 4. This is its content:

```
ex:City a gn:FeatureType .
ex:Country a gn:FeatureType .
ex:PortugalName a gn:PlaceName ;
                gn:lemma "Portugal"^^xsd:string ;
                gn:language "en"^^xsd:string .
ex:LisboaName a gn:PlaceName ;
              gn:lemma "Lisboa"^^xsd:string ;
              gn:language "pt"^^xsd:string .
ex:LisbonName a gn:PlaceName ;
              gn:lemma "Lisbon"^^xsd:string ;
              gn:language "en"^^xsd:string .
ex:Portugal a gn:Feature ;
            gn:name ex:PortugalName ;
            gn:type ex:Country .
ex:Lisboa a gn:Feature ;
          gn:relation ex:Portugal ;
          gn:name ex:LisboaName , ex:LisbonName ;
          gn:type ex:City .
```

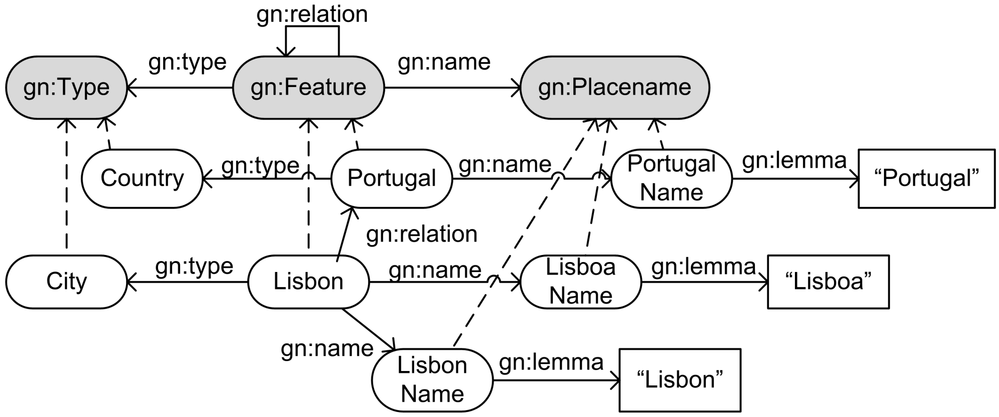


Fig. 4: Example ontology; shaded figures represent classes of the Geo-Net vocabulary; dotted arrows are statements with an `rdf:type` predicate.

Using the ontology a text processing application can discover that the tokens *Lisbon* and *Portugal* are referents of the entities identified as `ex:Lisboa` and `ex:Portugal`. Then, we can annotate the string with this information so the next application in the pipeline can use the assertions about `ex:Lisboa` and `ex:Portugal` defined above. The annotation can take the form of embedded RDFa annotations (Adida and Birbeck, 2008). The following is an example of RDFa annotations where the URIs are formatted in a compact notation named safe CURIEs (Birbeck and McCarron, 2008):

```
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:ex="http://www.example.com/simple#">
  <body>
    <p>
      <span about="[ex:Lisbon]">Lisbon</span>
      is the capital of
      <span about="[ex:Portugal]">Portugal</span>
    </p>
  </body>
</html>
```

4 GKB management system (version 2.1)

This section provides an overview of the current architecture of the GKB system (version 2.1). The GKB Ontology Serializer and the differences from GKB system (version 2.0) are also described in detail.

4.1 Architecture

The architecture of the GKB system, which implements the above described metamodel, is represented in Figure 5. We can analyze its structure taking into account how we can interact with the managed contents. Each instance of the GKB system manages an ontology instance. It is possible to interact with the ontology instance in three different ways:

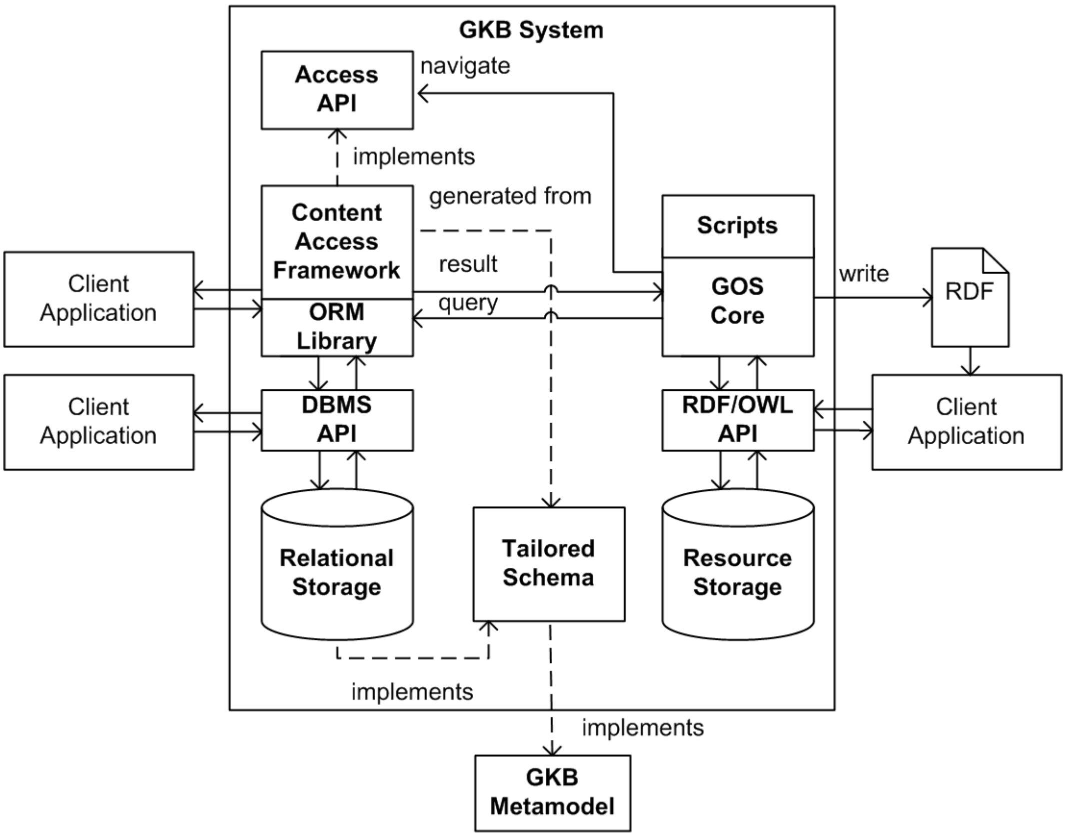


Fig. 5: Architecture of the GKB system.

- A relational database schema accessible through SQL interfaces. This entails that the client application must query and reason following all the rules and assumptions described in the conceptual model of the GKB system and 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 a OWL 1 ontology, which uses the Geo-Net vocabulary, accessible through a semantic aware interface such as a SPARQL end-point (Seaborne and Prud'hommeaux, 2008) or serialized using a syntax such as RDF/XML (Beckett, 2004).

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

Relational persistence system A tailored schema, a storage that implements it and the API that provides interaction with the repository, compose the relational persistence system. Both the relational storage and its API must support the OGC Simple Feature specification for spatial datatypes (Herring, 2006). The tailored schema is a platform-dependent model of the GKB metamodel that instances one or more information domains modified to implement the requirements of the client applications. The client applications can use a GKB instance by querying the relational persistence system.

Object-oriented persistence system A generated content access framework, the access API and an ORM library with support to code generation from schemas and standard spatial datatypes, compose the object-oriented persistence system. The backend of the object-oriented persistence system is the relational persistence system. The content access framework is an object-oriented library, semi-automatically generated from the tailored schema of the relational-based persistence system. The access API is a lightweight set of interfaces that partially implement the GKB metamodel and the additional classes, and provide 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 task.

Resource-oriented persistence system The resource oriented persistence system is composed by the GOS, a resource based storage, and a RDF/OWL API that manages the resource storage. The resource-oriented persistence system uses the object-oriented persistence system as back-end database. The GOS is described in Section 4.2. Applications can access the resource-oriented persistence system programmatically using the RDF/OWL API or using a SPARQL end-point. In addition, the resource-oriented system can be used to create RDF serializations of the ontologies managed by GKB.

The resource-oriented persistence system uses the Geo-Net vocabulary described in the Section 3 as the basic vocabulary to describe the contents of a GKB instance.

4.2 Geographic Ontology Serializer

Serializing the contents of a GKB instance for the Web means to deploy it on the Web as RDF and making it accessible by assigning an URL. 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.

Components The main components of GOS are (see Figure 6):

Access API Implements the current GKB metamodel. A complete description is provided below.

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. Any serializable content is accessible programmatically by implementation mappings between the ORM default generated API and the GKB Metamodel API.

Application Script Selects the contents to be serialized and provides specific mappings between a GKB tailored schema and an application vocabulary describing its semantics.

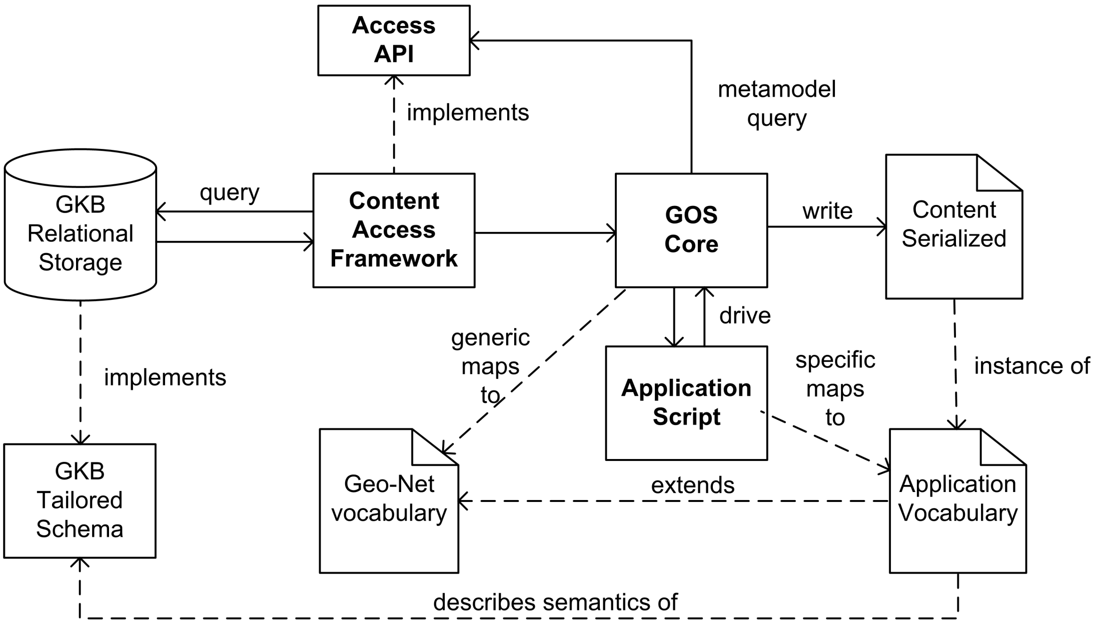
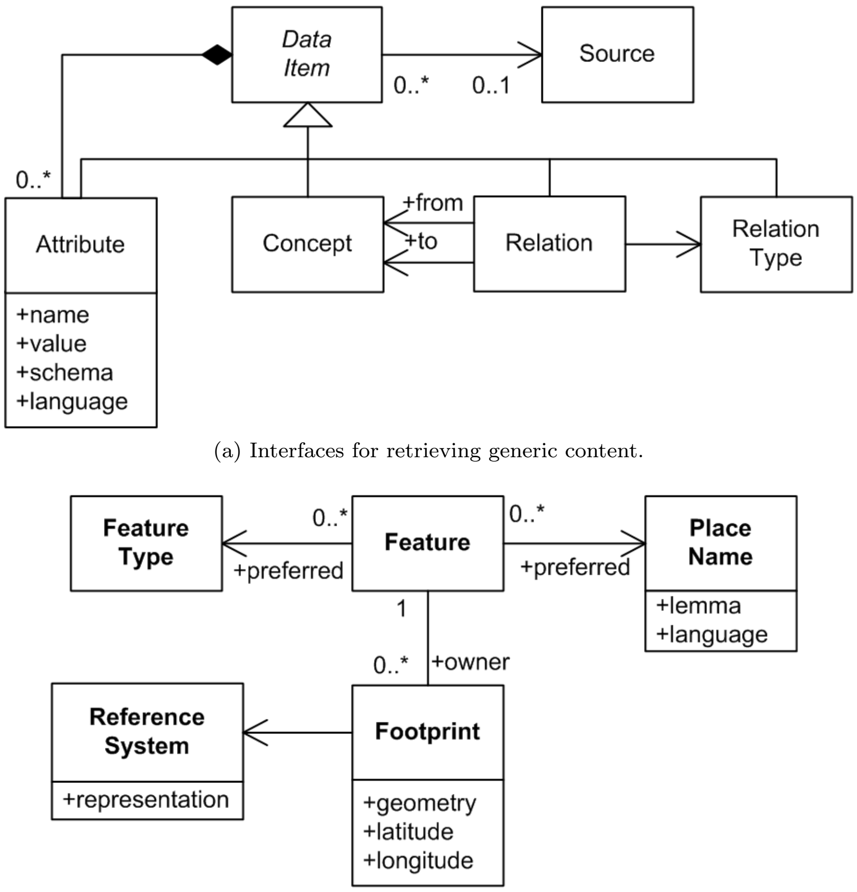


Fig. 6: Serialization of GKB instances: interaction of the main components of the GOS (in bold) with vocabularies, schemas and repositories.

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. It queries the Content Access Framework using its Access API to retrieve the contents of the relational storage. An application script that can be written to produce an intended serialization can then map the retrieved content to the application vocabularies or use the default Geo-Net mappings implemented in the GOS Core. This component is also responsible for writing out in the desired RDF format.

Access API The interfaces described in Figure 7a are based in the auxiliary classes for the conceptual model. These interfaces are `DataItem, Concept`, `Attribute, Relation`, `RelationType` and `Source`. The `DataItem` interface is the root of the GKB Metamodel API hierarchy and all the interfaces are subclasses of `DataItem`. It provides the identifier hierarchy. In the Access API, all the items can have an identifier. The `Attribute` interface provides access to typed attribute data owned by an object. The `Relation` interface gives access to any binary relation. A type qualifies this binary relation. The `RelationType` interface allows the identification of the kind-of relation. Finally, the `Source` interface allows the retrieval of basic metadata associated to any item.

The interfaces described in Figure 7b provide the retrieval of some of the core concepts described in the conceptual model: `Feature`, `FeatureType`, `PlaceName`, `Footprint` and `ReferenceSystem`. The `Feature` interface allows the retrieval of the preferred name and feature type, if they are explicitly defined. In the Access API, a feature can be the owner of a set of footprints. The `PlaceName` interface defines the properties `lemma` and `language`. The `Footprint` interface defines the properties `latitude`, `longitude` and `geometry`. The `latitude` and `longitude` properties represent a point inside the extension of the footprint. The `geometry` property represents any kind of object that contains a geographical description. The reference system of the extent can be



(a) Interfaces for retrieving generic content.

(b) Interfaces for retrieving geographic content.

Fig. 7: Interfaces of the Access API.

optionally described by a reference system object. Finally, the `ReferenceSystem` interface provides a representation of the specification of the reference system.

4.3 Changes from previous versions

This section describes the differences between the GKB system (version 2.0), which was described in Chaves et al. (2007), and the current version of the GKB system. The following list gives a brief account of the changes.

- Explicit inclusion of the concept information domain in the core metamodel using the class `InformationDomain`.
- Interdomain relationships and relation types are not different from other relations and relation types, but they are not members of any `InformationDomain` instance.
- The restriction that a `Feature` could not be associated to more than a `FeatureType` has been eliminated.

- A Footprint is no longer considered an attribute. However, it now requires the relation to a feature and a reference system to be fully specified.
- Explicit inclusion of the used reference systems in the core metamodel using the class ReferenceSystem.
- Any item in the model can have attached a set of attributes. Even attributes can have their own attributes if they are complex data structures. It is the responsibility of the instance design to define which concepts require that feature.
- Any item in the model can have a lineage attribute. It is the responsibility of the instance design to define which concepts demand this feature.
- Some of the names have been renamed to convey a clearer meaning, for example, the classes Name and Type in the GKB system (version 2.0) have been renamed to PlaceName and FeatureType, respectively.

5 Geo-Net-PT 02 ontology

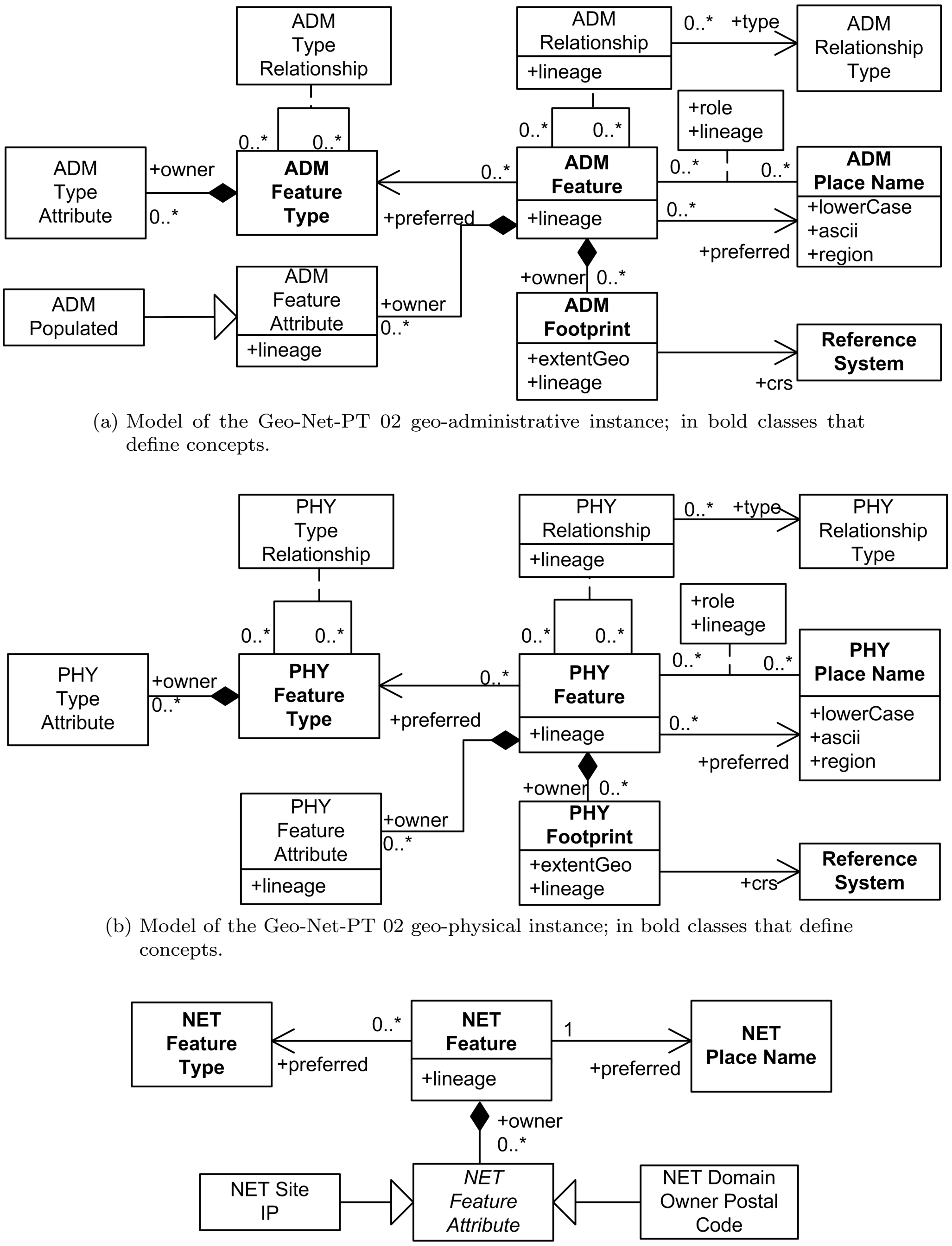
The Geo-Net-PT 02 geospatial ontology of Portugal is an authoritative geographic knowledge dataset, elaborated in GREASE-II. Geo-Net-PT 02 is the evolution of an ontology developed for GREASE named Geo-Net-PT-01 described in Chaves et al. (2005a). The development of Geo-Net-PT 02 began with a GKB instance containing the data of Geo-Net-PT 01, mainly administrative features, which was enriched with data from the physical domain (Rodrigues, 2009).

5.1 Content organization

The Geo-Net-PT 02 GKB repository consists of three GKB 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 GKB schema described in Section 4. The geo-physical domain content was not included in Geo-Net-PT 01.

Geo-Administrative domain The abstract model of the implementation of this domain is represented in Figure 8a. The classes prefixed with ADM represent the same classes of the base metamodel presented in Figure 2. There are several constraints in the implementation of the metamodel. The relationship between features and feature types is restricted to a unique preferred feature type. The relationship between features and place names distinguishes between the preferred names from other possible roles of the place name. Footprints are not shared among features. The ADM_Feature class specializes the Feature meta-class, adding the field populated to store data about inhabitants for populated features. The ADM_PlaceName class specializes the PlaceName meta-class, adding the fields lowerCase and ascii to represent the lemma in lower case and transliterated to ASCII. Additionally, the field region represents the region subtag as defined in the IETF RFC 4646. The ADM_Footprint specializes the Footprint meta-class adding a second geometric field



(a) Model of the Geo-Net-PT 02 geo-administrative instance; in bold classes that define concepts.

(b) Model of the Geo-Net-PT 02 geo-physical instance; in bold classes that define concepts.

(c) Model of the Geo-Net-PT 02 network instance; in bold classes that define concepts; in italic the abstract class NET_FeatureAttribute.

Fig. 8: Geo-Net-PT 02: geo-administrative, geo-physical and network domains.

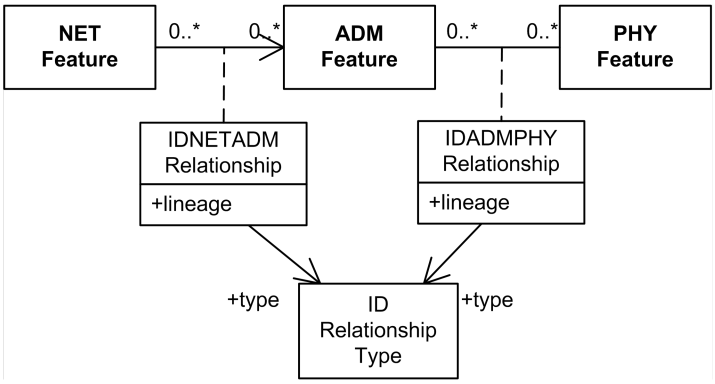


Fig. 9: Geo-Net-PT 02: interdomain relationships; in bold classes that define concepts.

named `extentGeo` that stores an alternative representation of the original geometry transformed to the geodetic coordinate reference system WGS 84. The support of fine grained lineage is restricted to features and place names. The support of attributes is restricted to feature, feature types and place names. The class `ADM_Populated` represents a feature attribute implemented in a separate table without lineage support.

Geo-Physical domain The abstract model of the implementation of this domain is represented in Figure 8b. Its implementation is similar to the administrative domain described above.

Network domain The abstract model of the implementation of the network domain is represented in Figure 8c. This instance implements a minimal subset of the features described in the metamodel. The class `Net_PlaceName` subclasses `PlaceName`, but contains only Web identifiers. The support of fine grained lineage is restricted to features. The support of attributes is restricted to features. Each attribute is persisted in its own table.

Relationships between domains One of the goals of GREASE-II is to assign geographic scopes to Web pages. For instance, the geographic scope of the Web site of the Lisbon municipality (`http://www.cm-lisboa.pt`) is the city of Lisbon. The current metamodel does not restrict the number and nature of interdomain relationships. This is an implementation decision. The Geo-Net-PT 02 implementation of the GKB metamodel explicitly supports unidirectional relationships between the network and the geo-administrative domains and bidirectional relationships between the geo-administrative and the geo-physical domains. These relationships are described in Figure 9.

5.2 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 namespace of this vocabulary is:

```
http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net-pt#
```

	Geo-Net-PT	Description	Specializes
Class	`gnpt:GeographicCode`	alphanumeric identifier	`gn:PlaceName`
	`gnpt:Escrow`	metadata placeholder	-
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`
	`gnpt:escrowFor`	surrogate of	-
Datatype property	`gnpt:population`	is inhabited by *literal value* people	-

Tab. 3: Classes and properties of the Geo-Net-PT vocabulary.

The prefix `gnpt:` is associated to this namespace in this report. The Geo-Net-PT vocabulary is available encoded in OWL 1 from `http://xldb.di.fc.ul.pt/wiki/Geo-Net-PT_02`. The terms of the Geo-Net-PT vocabulary are summarized in Table 3.

Subproperties based on the name relationship The Geo-Net-PT vocabulary defines thee subproperties: *preferred*, *alternative* and *identifier.* The later is presented below with the geographic codes. The properties that represent these roles are subproperties of the property `gn:name.`

Object property gnpt:preferred Identifies the resource of type `gn:PlaceName` used as 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.

Object property gnpt:alternative Is used to relate a feature with a resource of type `gn:PlaceName` not considered preferred.

With these properties, we can assert that the names *Lisboa, Lissabon* and *Lisbon* are preferred names for *Lisboa* in Portuguese, German and English, and the name *Cidade das Sete Colinas* is an alternative name Portuguese. one is the preferred and the others alternative versions.

```
:Lisboa gnpt:preferred :LisboaName , :LisbonName , :LissabonName  ;
        gnpt:alternative :CidadeDasSeteColinasName .
```

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.

Class gnpt:GeographicCode Identifies a feature name that is the identification of a feature by an alphanumeric code. The class `gnpt:GeographicCode` is a subclass of `gn:PlaceName`.

Functional object property gnpt:inSchema Denotes the schema, that is, a controlled vocabulary, where a geographic code is defined. The property `gnpt:inSchema` is a subproperty of the property `gn:lineage`.

Functional object property gnpt:identifies Asserts which resource of type `gn:Feature` is identified by a resource of type `gn:GeographicCode`.

Inverse functional object property gnpt:identifier Asserts that a resource of type `gn:Feature` has as geographic identifier the code described in a resource of type `gn:GeographicCode`. The property `gnpt:identifier` is a subproperty of the property `gn:name` and the inverse of `gnpt:identifies`.

The following example shows how to use the class `gnpt:GeographicCode` and its related properties to describe that the *Concelho of Lisboa* is identified by the administrative division code *1106* managed by Statistics Portugal (INE, `http://www.ine.pt/`). This fact can be described as follows:

```
:ConcelhoLisboa gnpt:identifier :DICO1106 .
:DICO1106 a gnpt:GeographicCode ;
          gn:lemma "1106"^^xsd:string ;
          gnpt:identifies :ConcelhoLisboa ;
          gnpt:inSchema :AdministrativeDivisionsINE .
```

Additional relations One of the goals of the Geo-Net-PT vocabulary is to describe conceptual, hierarchical and topological relations between features. The additional relations below are subproperties of the property `gn:relation`:

Object property gnpt:scope Asserts that the subject of the relation is a resource that identifies a real world object, such as a book, or digital resource, such as a Web page, whose content has a spatial extent that includes the object of the relation.

Object property gnpt:isLocatedOn Asserts that the subject is a feature located on the land surface of the target feature.

Symmetric object property gnpt:isAdjacentTo Asserts that its subject is a feature touching another feature in some topology.

Symmetric object property gnpt:isConnectedTo 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`.

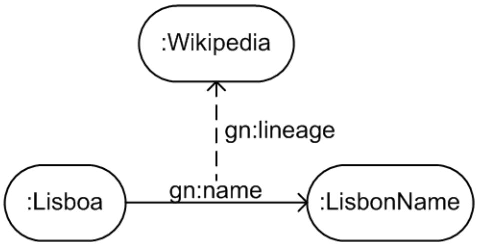


Fig. 10: *Wikipedia says Lisboa is Lisbon in English*; the straight arrow represents the relationship, the dotted arrow represents the provenance of the relationship.

Transitive object property gnpt:isPartOf Asserts that its subject is a feature that is a physical or logical component of the object feature.

Transitive object property gnpt:hasPart Asserts that its subject is a feature has as physical or logical component of the feature designated by the object. This property is the inverse of `gnpt:isPartOf`.

Datatype properties The Geo-Net-PT vocabulary has a single datatype property.

Property gnpt:population Asserts the inhabitants of the location.

Alternative to reified statements The metamodel allows a lineage attribute that attributes the source, for example `:Wikipedia`, to the name relation among the resource `:Lisboa` and the resource `:LisbonName` (a graphical representation of this is given in Figure 10).

We can create in RDF reifications to serialize this information:

```
:statement12345 a rdf:Statement ;
    rdf:subject :Lisboa ;
    rdf:predicate gn:name ;
    rdf:object :Lisbon ;
    gn:lineage :Wikipedia .
```

An alternative is creating a resource of type `gnpt:Escrow`, described below, representing the relation:

```
:Lisboa gn:name :escrow12345 .
:escrow12345 a gnpt:Escrow ;
    gnpt:escrowFor :Lisbon ;
    gn:lineage :Wikipedia .
```

Class gnpt:Escrow Models a resource that stands for other resource in the target of a relationship. These resources are instances of `gn:Feature`, `gn:PlaceName`, `gn:FeatureType` or `gn:Footprint` classes. An escrow is an alternative to reification when we need to add additional attributes to a relationship.

Functional object property gnpt:escrowFor Identifies the original object of the relationship.

This approach is based on the best practice, described by Noy and Rector (2006), for defining n-ary relations in the Semantic Web. It is an alternative to using reifications to represent the attributes of a relationship. We use the term *escrow* because this object acts as a digital escrow for the effective object of the relationship. It is also a surrogate of the relationship because is the placeholder of its attributes. Then, applications using the Geo-Net-PT vocabulary should interpret that triples like:

```
:s :p :escrow-uri .
:escrow-uri gn:escrowFor :o .
```

are surrogates for the statement:

```
:s :p: :o .
```

that is identified by the URI `:escrow-uri`. This is a case of property chain inclusion. OWL 1 does not support currently property chain inclusion, but that will be supported in OWL 2 (Golbreich and Wallace, 2009). As the current version of Geo-Net-PT is in OWL 1, these remarks only constitute advice for using the vocabulary and cannot be enforced.

The property chain inclusion that the *escrow* requires has implications in reasoning with symmetric and transitive properties. For example, if the symmetric property `gnpt:isAdjacentTo` is the predicate of a triple with an escrow object, applications using the Geo-Net vocabulary must interpret that triples like:

```
:s gnpt:isAdjacentTo :escrow-uri .
:escrow-uri gn:escrowFor :o .
```

entail both:

```
:s gnpt:isAdjacentTo :o .
:o gnpt:isAdjacentTo :s .
```

When the described situation occurs with transitive properties, such as `gnpt:isPartTo`, applications using the Geo-Net-PT vocabulary must interpret that triples like:

```
:s gnpt:isPartOf :escrow-uri .
:escrow-uri gn:escrowFor :o1 .
:o1 gnpt:isPartOf :o2
```

entail:

```
:s gnpt:isPartOf :o1 , o2 .
```

5.3 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. If there are human-readable descriptions available, the property `rdfs:comment` and the DCMI metadata terms are also used to provide a 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.

URI naming strategy The serialization of Geo-Net-PT 02 is designed to produce human readable URIs that encode unique identifiers of the relational storage whenever possible. This serialization process uses the namespace:

```
http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net-pt-02#
```

with the associated prefix `gnpt02:` as the base for URIs generation. Table 4 describes the mapping between relational identifiers of the implementation of the metamodel and the corresponding URIs.

If a concept has an available label, that label is the basis for the construction of the URI. The label is lowercased, transliterated to ASCII and whitespaces are replaced by underscores. Features, feature types and feature footprints add their respective relational identifiers. Place names add the language tag and a conditional hash code. This hash code is mandatory when the label is different from the lemma of the name.

`ADM_PlaceName` and `PHY_PlaceName` have an identical generation strategy. Identical names in different instances are treated as a single name in the serialization. In this case, these resources are related with two different `gnpt:InformationDomains`.

Information domains The serialization of Geo-Net-PT 02 identifies the three instances of the GKB metamodel in the Geo-Net-PT 02 repository. The resources `gnpt02:GeoAdministrative`, `gnpt02:GeoPhysical` and `gnpt02:Network` identifies the terms of the geo-administrative, the geo-physical and the network domain, respectively. These resources are instances of the class `gn:InformationDomain`. Each information domain is annotated with a human readable label and description.

Features Features are serialized as resources of type `gn:Feature` with the following properties:

- A human readable `rdfs:label`.
- Classified by a `gn:FeatureType.`
- Members of a `gn:InformationDomain.`
- Provenance captured with a `gn:lineage` property.

Model class	Strategy	Example
`ADM_Feature`	preferred name label + feature id	`gnpt02:lisboa-AF196518`
`PHY_Feature`	preferred name label + feature id	`gnpt02:rio_guizando-PF53459`
`NET_Feature`	feature id	`gnpt02:NF33571`
`ADM_PlaceName`	label + language + conditional hash	`gnpt02:lisboa-pt`
		`gnpt02:alcobaca-pt-1554844464`
`PHY_PlaceName`	label + language + conditional hash	`gnpt02:berlenga-pt`
		`gnpt02:corno_do_bico-pt-442568202`
`ADM_FeatureType`	label + feature type id	`gnpt02:freguesia-ATFRG`
`PHY_FeatureType`	label + feature type id	`gnpt02:rio-PTRIO`
`NET_Feature`	label + feature type id	`gnpt02:site-NTSTE`
`ADM_Footprint`	footprint id	`gnpt02:AFF241`
`PHY_Footprint`	footprint id	`gnpt02:PFP16400`

Tab. 4: Mappings between relational identifiers and URIs.

The details of the serialization of the features from the geo-administrative and the geo-physical domains are different from the details of the network domain.

Features from the geo-administrative and the geo-physical domain contain as many labels as place names. These labels are derived from the relations with place names asserted in the properties `gnpt:preferred` and `gpnt:alternative`. There is a label intended to be used as the main label of the resource that combines the preferred label and the preferred feature type. The features of the geo-administrative and geo-physical domains can be reused as spatial references. Thus, we use the class `geo:SpatialThing` from the Basic Geo vocabulary to describe these resources as having spatial extent. If a detailed description of the spatial extent is available as footprint, the property `gn:footprint` points to it.

Features from the network domain represent sites and domains. Sites and domains are places where data that can be annotated by the scope property resides, so each network feature is annotated as a resource of type `sioc:Space`. The label is based on the URL of the web page or in the registered name of the domain. If the feature describes a Web page, the property `rdfs:seeAlso` links the resource to the page.

The features in the network domain are not related with any place name, because the available names only describe their location on the Web: URL and domain names.

Place names and geographic codes Are serialized as resources of type `gn:PlaceName`, members of a `gn:InformationDomain` with a human readable label, its lemma and its language. Only place names from the geo-administrative and the geo-physical domains are serialized.

Some geographic codes are explicitly serialized as resources of type `gnpt:GeographicCode`. These names are related to features by the properties `gn:identifies` and `gn:identifier`. Also, the property `gnpt:inSchema` links the name with the schema where the geographic code is defined.

Feature types Are serialized as resources of type `gn:FeatureType`, members of a `gn:InformationDomain` with a human readable label. Feature types are also serialized independently with additional descriptions using the SKOS Core Vocabulary Specification (Miles and Bechhofer, 2009).

Footprints Are serialized as resources of type `gn:Footprint` in a `gn:InformationDomain` with a human readable label, based on the label of the Feature it refers to and a lineage description. The geometry description is encoded in GML (Portele, 2007). Footprints have spatial extent that can be pinpointed with a geodetic point. These characteristics are made explicit using the `geo:SpatialThing` class and the `geo:lat` and `geo:long` properties.

Reference systems Are resources of type `gn:ReferenceSystem`. The description of the reference system can be retrieved from a registry, such as the EPSG Geodetic Parameter Registry, 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 projection, are encoded in GML as the value of the property `gn:representation`.

Sources 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.

Relationship between features The serialization describes the following relationships:

Administrative hierarchy containment Only the direct hierarchical containment relationships between features in the geo-administrative domain are asserted. These are described by the transitive properties `gnpt:isPartOf` and `gnpt:hasPart`.

Spatial proper containment The serialization contains only proper part relationships between features in the geo-physical domain. These are described by the properties `gnpt:isPartOf` and `gnpt:hasPart`.

Administrative and spatial adjacency The adjacency relationships are only asserted between features of the same kind. This is described by the symmetric property `gnpt:isAdjacentTo`.

Spatial connectivity The connections of the hydrographic network in the geo-physical domain are described by the property `gnpt:isConnectedTo`.

Location Some features in the geo-physical domain have as georeference the administrative feature where they are located. This is described by the property `gnpt:isLocatedOn`. This relation is a direct inter-domain relationship as there exists a sequence of relations between resources of type `gn:Feature` belonging to different information domains.

Geographic scope of the site The identification of the geographic scope of some Web sites is part of the Geo-Net-PT data. The geographic scope of a Web site is derived from is content (see Silva et al., 2006). 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. This relation can be described as a direct inter-domain relationship.

Location of the owner of the site The network domain contains Web pages and sites. One of the pieces of information part of the Geo-Net-PT data is the owner location, which has been resolved to a geo-administrative `gn:Feature`. The ownership is described using the concept `sioc:User`, and the property `sioc:has_owner`. The location of the user is described using the property `foaf:based_near`. The location of the owner of the site can be described as an indirect inter-domain relationship.

The following example describes that Web page `http://www.cm-lisboa.pt/` has an owner, and its location is "near" the resource that identifies Lisboa:

```
gnpt02:LisboaWebPage a gn:Feature , sioc:Space ;
    rdfs:seeAlso <http://www.cm-lisboa.pt/> ;
    gn:inDomain gnpt02:Network ;
    sioc:has_owner [ a sioc:User ;
        foaf:based_near gnpt02:Lisboa . ] .
```

5.4 Descriptive statistics

Geo-Net-PT 02 defines 701,209 concepts, most of them administrative features and place names (see Table 5a). Some of these concepts have other types: 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. Figure 11 shows graphically the dimensions of the domains and the explicit relationships among them. Although the network domain is similar in size to the geo-physical domain, it is more linked than the geo-physical domain.

The administrative and physical features are classified by 81 feature types. Table 5b shows the most common feature types, some of them aggregated. 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.

Geo-Net-PT 02 has 21 different sources. Figure 12 shows the distribution of the sources asserted with the `gn:lineage` property in concepts and reified relationships. Table 5c summarizes these distribution, distinguishing between concepts and relations. The main source is CTT Correios (`http://www.ctt.pt/`), Portugal's mail services, which provides addresses. These outnumber all the other kinds of features. Excluding CTT Correios, about 5% of the data is derived using rules. Other relevant official sources are the FCCN (`http://www.fccn.pt/`), the APA (`http://www.apambiente.pt/`) and the IGP (`http://www.igeo.pt/`). FCCN provides data about Web sites. APA provides data about features in the geo-physical domain. IGP provides data about features in the geo-adminitrative domain. Wikipedia is a non-official source. This source provides ancillary data about the administrative structure.

The geographic descriptions are in 5 different coordinate reference systems. Table 5d shows that there are two different coordinate reference systems for footprints located in Portugal's mainland. Footprints provided by IGP are in ETRS 1989 TM06-Portugal, while footprints provided by APA are in the Lisboa Hayford Gauss IGeoE. The serialization of Geo-Net-PT 02 keeps the original values of the coordinates.

Concept	**Geo-Adm.**	**(%)**	**Geo-Phy.**	**(%)**	**Net.**	**(%)**	**Total**
gn:Feature	386,067	92.94	5,676	1.37	23,666	5.70	415, 409
gn:PlaceName	265,044	96.98	8,266	3.02	-	0.00	273, 310
gn:Footprint	4,597	100.00	-	0.00	-	0.00	4, 597
gn:FeatureType	62	69.66	25	28.09	2	2.25	89
gnpt:GeographicCode	4,597	58.91	3,207	41.09	-	0.00	7, 804
Total	660,637	94.18	17,174	2.45	23,668	3.38	701, 209

(a) Concepts; this table only considers direct assertions.

Geo-administrative	**Feature**	**(%)**	**Geo-physical**	**Feature**	**(%)**
Postal code	187,014	48.44	Stream	2,421	42.65
Street layout	146,422	37.93	Beach	588	9.83
Settlement	44,386	11.50	Museum	507	8.93
Civil parish	4,260	0.93	Archaeological site	414	7.29
Zone	3,594	0.08	Hotel	381	6.71
Municipality	308	0.01	Natural region	304	5.36
NUTs	40	0.00	Castle	256	4.51
Distrito	18	0.00	Spring	220	3.88
Province	11	0.00	Historic hamlet	217	3.82
Island	11	0.00	Reservoir	90	1.59
Region	2	0.00	Touristic resource	84	1.48
Country	1	0.00	*Other*	224	3.95
Total	386,067	100.00		5,676	100.00

(b) Features by feature type; in italics aggregations.

	Concepts	**(%)**	**Relations**	**(%)**	**Total**	**(%)**
CTT Correios	381,445	90.13	763,428	89.08	1,144,873	89.43
Production rules (topology)	-	-	63,838	7.45	64,524	5.04
FCCN	22,980	5.43	-	-	22,980	1.79
APA	6,420	1.52	8,338	0.97	14,758	1.15
Wikipedia	4,273	1.01	8,740	1.02	13,013	1.02
IGP	29	0.01	4,597	0.54	4,626	0.36
Production rules (footprints)	4,568	1.08	-	-	4,568	0.36
Others	3,498	0.83	8,090	1.91	11,588	0.91
Total	423,213		857,031		1,280,244	

(c) Lineage annotations grouped by target.

CRS Name	**Geometries**	**(%)**	**Extent**
ETRS 1989 TM06-Portugal (EPSG:3763)	4,346	56.69	Portugal mainland
Lisboa Hayford Gauss IGeoE (ESRI:102164)	3,297	41.09	Portugal mainland
ITRF93 UTM Zone 28N	67	0.86	Madeira
ITRF93 UTM Zone 26N	167	2.14	Azores Central and Oriental
ITRF93 UTM Zone 25N	17	0.22	Azores Occidental

(d) Coordinate Reference Systems; in parenthesis well-known CRS identifiers.

Tab. 5: Geo-Net-PT 02 statistics.

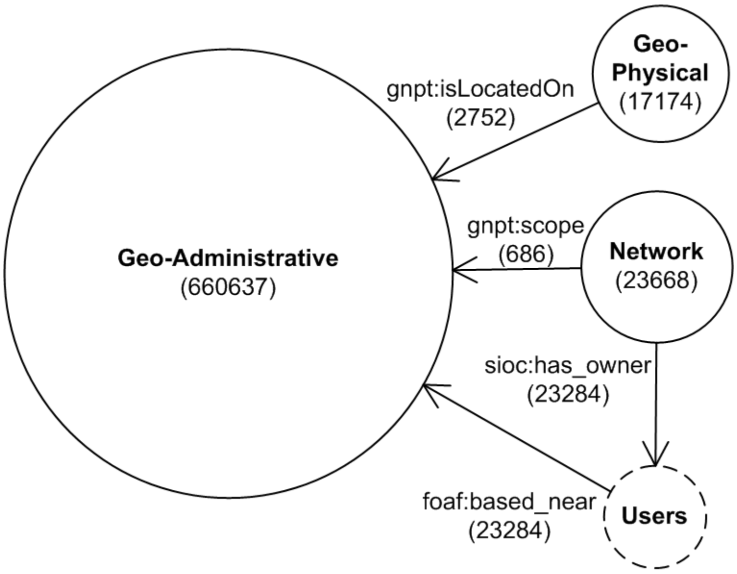


Fig. 11: Geo-Net-PT 02 statistics: Domains and inter-domain relationships; the size of the domains reflects the number of concepts.

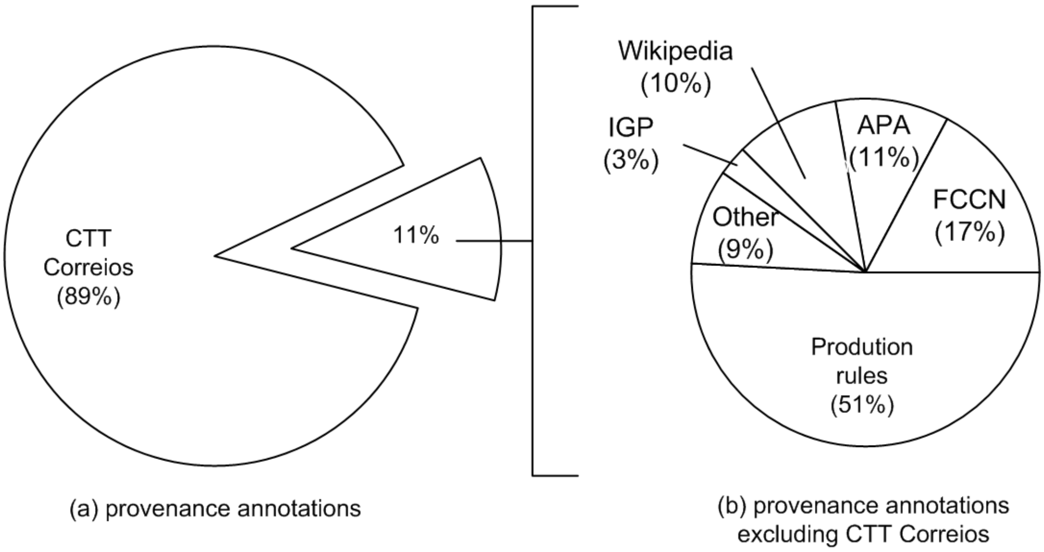


Fig. 12: Geo-Net-PT 02 statistics: Summary of lineage annotations; on the left including CTT Correios; on the right excluding CTT Correios.

5.5 Production environment

The RDBMS of choice for the GKB system (version 2.1) is PostgreSQL 8.3.6 (`http://www.postgresql.org/`). The support for geographic objects to the PostgreSQL is provided by PostGIS 1.3.6 (`http://postgis.refractions.net/`). 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 (`http://www.hibernate.org/`). The support for geographic objects is provided by the extension Hibernate Spatial (`http://www.hibernatespatial.org/`). The storage system of choice for the resource-oriented persistence system is TBD (`http://jena.sourceforge.net/TDB/`), a large scale non transactional file-based RDF storage.

6 Final Remarks

Geo-Net-PT 02 is available in the XLDB Node of Linguateca (`http://xldb.di.fc.ul.pt/wiki/Geo-Net-PT_02`) under a Creative Commons license (CC-BY) that allows its free use by researchers. The complete distribution consists of:

- The Geo-Net vocabulary.
- The Geo-Net-PT vocabulary.
- The Geo-Net-PT 02 feature types described using SKOS.
- The Geo-Net-PT 02 data with reified relations.
- The Geo-Net-PT 02 data with *digital escrows* instead of reified relations.
- The Geo-Net-PT 02 plain, that is, without reified relations.

Vocabularies are available in the formats RDF/XML and Turtle. The Geo-Net-PT 02 ontology is available in these formats and in Notation 3.

Acknowledgements

Contributors to this project over the years include (alphabetically) Daniel Amoedo, David Batista, Nuno Cardoso, Paula Carvalho, João Catalão, Francisco Couto, Bruno Martins and Patrícia Sousa, and other current or former members of the XLDB Group involved in GREASE and GREASE-II.

This work was partially supported by FCT (Portuguese research funding agency) for its LASIGE Multi-annual support, GREASE-II project (grant PTDC/EIA/73614/20 06).

The work of Francisco J. Lopez-Pellicer has been partially supported by Spanish Government (TIN2007-65341 and PET2008_0026) and the Aragon Government (PI075/08)

References

Coordinate Transformation Services. OpenGIS Consortium Implementation Specification OGC 01-009, OpenGIS Consortium Inc., January 2001. URL http://portal.opengeospatial.org/files/?artifact_id=999. Version 1.0.

Ben Adida and Mark Birbeck. RDFa primer. W3C working draft, W3C, June 2008. http://www.w3.org/TR/2008/WD-xhtml-rdfa-primer-20080620/.

David Batista. Prospecção de Conceitos Geográficos na Web. Master's thesis, University of Lisbon, Faculty of Sciences, September 2009.

Dave Beckett. RDF/xml syntax specification (revised). W3C recommendation, W3C, February 2004. http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/.

Dave Beckett and Tim Berners-Lee. Turtle – Terse RDF Triple Language. W3C team submission, W3C, January 2008. http://www.w3.org/TeamSubmission/2008/SUBM-turtle-20080114/.

Tim Berners-Lee. Notation 3. Personal notes, W3C, 1998. http://www.w3.org/DesignIssues/Notation3.

Mark Birbeck and Shane McCarron. CURIE syntax 1.0. a WD in last call, W3C, May 2008. http://www.w3.org/TR/2008/WD-curie-20080506.

Dan 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/ [Accessed 1 April 2009].

Marcirio Chaves, Mário J. Silva, and Bruno Martins. A Geographic Knowledge Base for Semantic Web applications. In Carlos A. Heuser, editor, *20th Brazilian Symposium on Databases - SBBD*, pages 40–54, Uberlândia, MG, Brazil, 3–7 Oct 2005 2005a. UFU. ISBN 85-7669-029-2. URL http://www.sbbd-sbes2005.ufu.br/arquivos/artigo-03-ChavesSilva.pdf.

Marcirio Chaves, Mário J. Silva, and Bruno Martins. GKB – Geographic Knowledge Base. Technical report, Departamento de Informática da Faculdade de Ciências da Universidade de Lisboa, July 2005b. URL http://www.di.fc.ul.pt/tech-reports/05-12.pdf. DI/FCUL TR 05–12.

Marcirio Chaves, Catarina Rodrigues, and Mário J. Silva. Data model for geographic ontologies generation. In *XATA 2007 - XML: Aplicações e Tecnologias Associadas*, February 2007.

Mike Dean and Guus Schreibe. OWL Web Ontology Language Reference. W3C recommendation, W3C, February 2004. URL http://www.w3.org/TR/2004/REC-owl-ref-20040210/.

Dublin Core Metadata Initiative. Dublin Core Metadata Element Set. Technical report, Dublin Core Metadata Initiative, 2003. Version 1.1. Available from: http://dublincore.org/documents/dces/.

Gaihua Fu, Alia Abdelmoty, and Christopher Jones. Design of a geographical ontology. Technical Report WP D5 3101, SPIRIT project, 2003.

Christine Golbreich and Evan K. Wallace. OWL 2 Web Ontology Language: new features and rationale. W3C propsed recommendation, W3C, Sep 2009. http://www.w3.org/TR/2009/PR-owl2-new-features-20090922/.

Tom Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993. ISSN 0001-2998. URL `http://tomgruber.org/writing/ontolingua-kaj-1993.htm`. See also *What is an Ontology?* http://www-ksl.stanford.edu/kst/what-is-an-ontology.html.

Tom Gruber. Ontology. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems*. Springer, 2009. ISBN 978-0-387-49616-0.

John 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.

Linda 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. ISBN 3-540-41023-6.

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

Gail Hodge. Systems for Knowledge Organization for Digital Libraries: Beyond traditional authority files. Technical report, Digital Library Federation, April 2000. URL `http://www.clir.org/pubs/reports/pub91/contents.html`.

Christopher B. Jones, Ross S. Purves, Paul D. Clough, and Hideo Joho. Modelling vague places with knowledge from the web. *Int. J. Geogr. Inf. Sci.*, 22(10):1045–1065, 2008. ISSN 1365-8816. doi: 10.1080/13658810701850547.

Wolfgang Kresse and Kian Fadaie. *ISO Standards for Geographic Information.* Springer, Berlin, 2004. URL `http://books.google.com/books?id=vvMBnAiYPIkC&printsec=frontcover&source=gbs_v2_summary_r&cad=0#v=onepage&q=&f=false`. ISBN 978-3-54020-130-4.

Francisco J. López-Pellicer, F. Javier Zarazaga-Soria, A. Mogollón-Diaz, Javier Nogueras-Iso, and Pedro R. Muro-Medrano. The gazetteer content model issue: Could Spatial Data Infrastructures provide it? In Sara Irina Fabrikant and Monica Wachowicz, editors, *AGILE Conf.*, Lecture Notes in Geoinformation and Cartography, pages 187–200. Springer, 2007. ISBN 978-3-540-72384-4.

Dimitar Manov, Atanas Kiryakov, Borislav Popov, Kalina Bontcheva, Diana Maynard, and Hamish 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. doi: 10.3115/1119394.1119395.

David M. Mark. *Human Factors in Geographical Information Systems*, chapter Human Spatial Cognition, pages 51–60. Halsted Press, New York, NY, USA, 1993. ISBN 1852932627. URL `http://www.geog.buffalo.edu/~dmark/DMScottchapter.html`.

Alistair Miles and Sean Bechhofer. SKOS simple knowledge organization system reference. W3C proposed recommendation, W3C, June 2009. http://www.w3.org/TR/2009/PR-skos-reference-20090615/.

Eric Miller and Frank Manola. RDF primer. W3C recommendation, W3C, February 2004. Available from: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

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. ISSN 0198-9715. doi: 10.1016/j.compenvurbsys.2003.12.004.

Natasha Noy and Alan Rector. Defining n-ary relations on the semantic web. W3C note, W3C, April 2006. http://www.w3.org/TR/2006/NOTE-swbp-n-aryRelations-20060412/.

A. Phillips and M. Davis. Tags for Identifying Languages. RFC 4646 (Best Current Practice), September 2006a. URL `http://www.ietf.org/rfc/rfc4646.txt`.

A. Phillips and M. Davis. Matching of Language Tags. RFC 4647 (Best Current Practice), September 2006b. URL `http://www.ietf.org/rfc/rfc4647.txt`.

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

Andy Powell, Mikael Nilsson, Ambjörn Naeve, Pete Johnston, and Thomas Baker. Dcmi abstract model [online]. Dcmi recommendation, Dublin Core Metadata Initiative, 2007. Available from: http://dublincore.org/documents/abstract-model/.

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

Andy Seaborne and Eric Prud'hommeaux. SPARQL query language for RDF. W3C recommendation, W3C, January 2008. Available from: http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/.

Mário J. Silva, Bruno Martins, Marcirio Chaves, Nuno Cardoso, and Ana Paula Afonso. Adding geographic scopes to web resources. *CEUS – Computers, Environment and Urban Systems*, 30(4):378–399, July 2006. doi: 10.1016/j.compenvurbsys.2005.08.003.

Technical Committee ISO/TC 211, Geographic Information/Geomatics. ISO 19111:2007: Geographic information – Spatial referencing by coordinates. Technical report, International Organization for Standardization, Geneva, Switzerland., 2007a.

Technical Committee ISO/TC 211, Geographic Information/Geomatics. Geographic information – Geography Markup Language (GML). Technical report, International Organization for Standardization, Geneva, Switzerland., 2007b. This is not the final standard.

Michael 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.