Integration of Coordinate Information in CIDOC CRM

Contributers:

Gerald Hiebel, Surveying and Geoinformation Unit, University of Innsbruck

Øyvind Eide, Department of Digital Humanities, King's College London / Unit for Digital Documentation, University of Oslo.

Mark Fichtner, Zoologisches Forschungsmuseum Alexander Koenig, Bonn

Klaus Hanke, Surveying and Geoinformation Unit, University of Innsbruck

Georg Hohmann, Germanisches Nationalmuseum, Department of Cultural Informatics, Nürnberg

Siegfried Krause, Germanisches Nationalmuseum, Department of Cultural Informatics, Nürnberg

October 2011

Content

1 Introduction

Integrating spatial information in the CRM can be divided into two tasks:

- The conceptual modeling to define subclasses for coordinates (E47) or other spatial information (E48,…) and to map from other conceptualizations for spatial information to the CRM
- The encoding of coordinate information

On both subjects options are discussed in this document. The OGC (Open Geospatial Consortium, www.opengeospatial.org) has published a series of standards that can be divided in abstract and implementation standards and applied to the two tasks of conceptual modeling and encoding. A great number of OGC standards have been adopted as ISO standards through the work of the ISO TC 211 (http://www.isotc211.org/).

During the research into methods for the integration of coordinate information in the CIDOC CRM three main options were found:

- Represent spatial data in RDF without using OGC standards
- Represent spatial data in XML using OGC/ISO standards
- Represent spatial data in RDF/OWL based on the OGC candidate standard of "GeoSPARQL"

The first option is applied in an ongoing research project (CLAROS) and will be presented with references and characteristics. Before the second and third option is presented an introduction to OGC/ISO standards will provide information on the relevant standards to represent coordinates. The third option is an approach based on the specification of the OGC (Open Geospatial Consortium, www.opengeospatial.org) candidate standard of "GeoSPARQL".

2 Represent spatial data in RDF without using OGC standards

2.1 RDF Representation „Basic Geo (WGS84 lat/long) Vocabulary"

This is a basic RDF vocabulary that provides the Semantic Web community with a namespace for representing lat(itude), long(itude) and other information about spatially-located things, using WGS84 as a reference datum.

The vocabulary is getting significant usage, both (as intended) within RDF documents, but also as a namespace used within non-RDF XML documents, such as RSS 2.0

2.1.1 Status of the „Basic Geo (WGS84 lat/long) Vocabulary"

It was created as an informal collaboration within W3C's Semantic Web Interest Group. This work is not currently on the W3C recommendation track for standardization, and has not been subject to the associated review process, quality assurance, etc. If there is interest amongst the W3C membership in standards work on a location/mapping RDF vocabulary, this current work may inform any more formal efforts to follow.

Overview

This vocabulary begins an exploration of the possibilities of representing mapping/location data in RDF, and does not attempt to address many of the issues covered in the professional GIS world, notably by the Open Geospatial Consortium (OGC). Instead, it provides just a few basic terms that can be used in RDF (eg. RSS 1.0 or FOAF documents) when there is a need to describe latitudes and longitudes. The motivation for using RDF as a carrier for lat/long info is RDF's capability for cross-domain data mixing. It can describe not only maps, but the entities that are positioned on the map. And it can use any relevant RDF vocabularies to do so, without the need for expensive pre-coordination, or for changes to a centrally maintained schema.

Examples

A basic, standalone example:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

        xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">

  <geo:Point>

    <geo:lat>55.701</geo:lat>

    <geo:long>12.552</geo:long>

  </geo:Point>

</rdf:RDF>
```

http://www.w3.org/2003/01/geo/#vocabulary

2.1.2 pro and cons

+ namespace available

+ easy to use for geographic point data

+ popular in the linked data community

+ integration examples for google maps available

+ examples of implementation with CIDOC CRM available

- not currently on the W3C recommendation track

- no intention of connecting it to OGC and GIS world

- coordinates have to be in WGS84

- currently only point data represented

Summing up:

It is an ready to use option to integrate point data in WGS84 (Standard Ellipsoid for the world, used in GPS) into the CRM. Not usefull if you don't have WGS84 Coordinates representing points on the surface of the earth or any other coordinates maybe not related to the earth.

Examples: Claros Project

(http://explore.clarosnet.org/XDB/ASP/clarosExplorer.asp?help=true)

2.2 RDF Representation through a "GeoAnnotation class"

A similar approach as in the „Basic Geo (WGS84 lat/long) Vocabulary" is used in the PELAGIOS project where spatial information is integrated in an ontology through an GeoAnnotation class. It is based on the "Open Annotation" approach (http://www.openannotation.org/) and was integrated with data from the CLAROS project.

2.2.1 About Open Annotation

Annotating is a pervasive element of scholarly practice for both the humanist and the scientist. It is a method by which scholars organize existing knowledge and facilitate the creation and sharing of new knowledge. It is used by individual scholars when reading as an aid to memory, to add commentary, and to classify. It can facilitate shared editing, scholarly collaboration, and pedagogy. Over time annotations can have scholarly value in their own right. Yet scholars remain dissatisfied with the options available for annotating digital resources. Scholars wanting to annotate have to learn different annotation clients for different content repositories, have no easy way to integrate annotations made on different systems or created by colleagues using other tools, and are often limited to simplistic and constrained models of annotation. The importance of annotating as a scholarly practice coupled with the real-world limitations of existing practices and tools supporting annotation of digital content has had a retarding effect on the growth of digital scholarship and the level of digital resource use by scholars.

The overarching goals of this project (consisting of multiple phases) are:

- To facilitate the emergence of a Web and Resource-centric interoperable annotation environment that allows leveraging annotations across the boundaries of annotation clients, annotation servers, and content collections. To this end, interoperability specifications will be devised.

- To demonstrate through implementations an interoperable annotation environment enabled by the interoperability specifications in settings characterized by a variety of annotation client/server environments, content collections, and scholarly use cases.
- To seed widespread adoption by deploying robust, production-quality applications conformant with the interoperable annotation environment in ubiquitous and specialized services, tools, and content used by scholars -- e.g.: Zotero, AXE, LORE, Co-Annotea, Pliny; JSTOR, AustLit, MONK

2.2.2 GeoAnnotation in PELAGIOS

A new class was created called GeoAnnotation. Intuitively, this class represent the oac:Annnotation(s) that point to (oac:hasBody) geographical objects (wgs84_pos:SpatialThing).
It is therefore defined in abstract OWL syntax as

```
class (GeoAnnotation partial oac:Annotation)

class (GeoAnnotation complete restriction(oac:hasBody
someValuesFrom(wgs84_pos:SpatialThing))
```

In other words, GeoAnnotation is the class of annotations that have a body which is a SpatialThing. In triple form:

```
<http://pelagios-project.org/ontology/oac-geo/GeoAnnotation>  rdf:type
 owl:Class
<http://pelagios-project.org/ontology/oac-
geo/GeoAnnotation>  rdfs:subClassOf oac:Annotation
<http://pelagios-project.org/ontology/oac-
geo/GeoAnnotation>  owl:equivalentClass _:node417
_:node417           rdf:type               owl:Restriction
_:node417           owl:onProperty         oac:hasBody
_:node417           owl:someValuesFrom     wgs84_pos:SpatialThing
```

http://pelagios-project.blogspot.com/2011/03/pelagios-workshop-sparql-demo-rdfa.html

2.2.3 pro and cons

+ examples of implementation available

+ examples of integration of different coordinate representations

- not currently on the W3C recommendation track

- no intention of connecting it to OGC and GIS world

- coordinates have to be in WGS84

- currently only point data represented

There was no closer examination of this approach, more research or knowledge of the Open Annotation Approach would be necessary to evaluate it properly.

Examples: PELAGIOS Project

http://pelagios-project.blogspot.com/

3 Open Geospatial Consortium (OGC) and ISO 19100 series Standards on Geographic information

The Open Geospatial Consortium (www.opengeospatial.org) (OGC) is an international voluntary consensus standards organization, founded in 1994. In the OGC, more than 400 commercial, governmental, nonprofit and research organizations worldwide collaborate in a consensus process encouraging development and implementation of open standards for geospatial content and services, GIS data processing and data sharing.

Most of the OGC standards depend on a generalized architecture captured in a set of documents collectively called the Abstract Specification, which describes a basic data model for representing geographic features.

Atop the Abstract Specification members have developed and continue to develop a growing number of specifications, or standards to serve specific needs for interoperable location and geospatial technology, including GIS. The OGC has a close relationship with ISO/TC 211 (Geographic Information/Geomatics). Volumes from the ISO 19100 series under development by this committee progressively replace the OGC abstract specification.

There is an extensive volume of ISO Standards (37 +) dealing with Geographic Information. They can be grouped in Abstract Specifications defining conceptual models to represent geographic information and implementation specifications defining how to implement these conceptual models. A good overview is found in a PDF created for the "Qualification Improvement of the Lithuanian Geographical Information Managers"

http://www.geoportal.lt/download/gii_mokymai/GII_03_mokomoji_medziaga/En/Paskaitu_konspektai/GII-03_training_material.pdf

A short description of the most important standards relevant to represent geometries and geographic coordinates is given. These are also the standards GeoSPARQL is based on.

3.1 Abstract Specifications – Conceptual models

In this section we want to give a brief description of the most important Abstract Specifications that are used to represent coordinates. The specifications use UML (http://en.wikipedia.org/wiki/Unified_Modeling_Language) to describe the classes with their attributes and relations. Figures in Appendix A show UML diagrams of the mentioned standards to give an idea of the scope and the elements included in the specifications.

3.1.1 ISO 19107 Geographic information - Spatial Schema

ISO 19107:2003 specifies conceptual schemas for describing the spatial characteristics of geographic features, and a set of spatial operations consistent with these schemas. It treats vector geometry and topology up to three dimensions. It defines standard spatial operations for use in access, query, management, processing, and data exchange of geographic information for spatial (geometric and topological) objects of up to three topological dimensions embedded in coordinate spaces of up to three axes. ISO 19137

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26012

3.1.2 ISO 19137 Geographic information - Core profile of the spatial schema

This standard is used to define a subset of ISO 19107, in order to facilitate implementation.

ISO 19137:2007 defines a core profile of the spatial schema specified in ISO 19107 that specifies, in accordance with ISO 19106, a minimal set of geometric elements necessary for the efficient creation of application schemata. It supports many of the spatial data formats and description languages already developed and in broad use within several nations or liaison organizations.

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32555

3.1.3 ISO 19111:2007 Geographic information -- Spatial referencing by coordinates

ISO 19111:2007 defines the conceptual schema for the description of spatial referencing by coordinates, optionally extended to spatio-temporal referencing. It describes the minimum data required to define one-, two- and three-dimensional spatial coordinate reference systems with an extension to merged spatial-temporal reference systems. It allows additional descriptive information to be provided. It also describes the information required to change coordinates from one coordinate reference system to another.

In ISO 19111:2007, a coordinate reference system does not change with time. For coordinate reference systems defined on moving platforms such as cars, ships, aircraft and spacecraft, the transformation to an Earth-fixed coordinate reference system can include a time element.

ISO 19111:2007 is applicable to producers and users of geographic information. Although it is applicable to digital geographic data, its principles can be extended to many other forms of geographic data such as maps, charts and text documents.

The schema described can be applied to the combination of horizontal position with a third non-spatial parameter which varies monotonically with height or depth. This extension to non-spatial data is beyond the scope of ISO 19111:2007 but can be implemented through profiles.

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=41126

3.1.4 ISO 19109:2005 Geographic information -- Rules for application schema

This standard introduces “features” as abstractions of real world objects with a relation to a place. As the OGC candidate standard of “GeoSPARQL” uses features as its core class, it is important to understand the semantic meaning of features within the ISO 19100 standards to make appropriate mappings to CRM classes.

ISO 19109:2005(E) defines rules for creating and documenting application schemas, including principles for the definition of features. Its scope includes the following:

- conceptual modeling of features and their properties from a universe of discourse;
- definition of application schemas;
- use of the conceptual schema language for application schemas;

- transition from the concepts in the conceptual model to the data types in the application schema;
- integration of standardized schemas from other ISO geographic information standards with the application schema.

The following are outside the scope:

- choice of one particular conceptual schema language for application schemas;
- definition of any particular application schema;
- representation of feature types and their properties in a feature catalogue;
- representation of metadata;
- rules for mapping one application schema to another;
- implementation of the application schema in a computer environment;
- computer system and application software design;
- programming.

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39891

3.2 Implementation specifications

Based on the conceptualizations and classes defined in the Abstract Specifications, standards have been developed to implement them for the exchange and use of Geographic information. GML is the most extensive standard covering most of the Abstract Specifications (not only the ones described above) in its 437 pages in GML 3.2.1.

The ISO 19125 - Simple Feature Profile and ISO 19142 -Web feature services use subsets of the Abstract Specifications.

KML is currently in the process of being aligned with OGC Specifications.

3.2.1 ISO 19136:2007 Geographic information -- Geography Markup Language (GML)

The Geography Markup Language (GML) is an XML encoding in compliance with ISO 19118 for the transport and storage of geographic information modeled in accordance with the conceptual modeling framework used in the ISO 19100 series of International Standards and including both the spatial and non-spatial properties of geographic features.

ISO 19136:2007 defines the XML Schema syntax, mechanisms and conventions that:

- provide an open, vendor-neutral framework for the description of geospatial application schemas for the transport and storage of geographic information in XML;
- allow profiles that support proper subsets of GML framework descriptive capabilities;
- support the description of geospatial application schemas for specialized domains and information communities;
- enable the creation and maintenance of linked geographic application schemas and datasets;
- support the storage and transport of application schemas and data sets;

- increase the ability of organizations to share geographic application schemas and the information they describe.

Implementers may decide to store geographic application schemas and information in GML, or they may decide to convert from some other storage format on demand and use GML only for schema and data transport.

NOTE If an ISO 19109 conformant application schema described in UML is used as the basis for the storage and transportation of geographic information, ISO 19136 provides normative rules for the mapping of such an application schema to a GML application schema in XML Schema and, as such, to an XML encoding for data with a logical structure in accordance with the ISO 19109 conformant application schema.

ISO:
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32554

OGC:

http://portal.opengeospatial.org/files/?artifact_id=20509

http://schemas.opengis.net/gml/3.2.1/

http://schemas.opengis.net/gml/3.2.1/gml.xsd

3.2.2 ISO 19125-1:2004 Geographic information -- Simple feature access -- Part 1: Common architecture

ISO 19125-1:2004 establishes a common architecture for geographic information and defines terms to use within the architecture. It also standardizes names and geometric definitions for Types for Geometry.

ISO 19125-1:2004 does not place any requirements on how to define the Geometry Types in the internal schema nor does it place any requirements on when or how or who defines the Geometry Types. ISO 19125-1:2004 does not attempt to standardize and does not depend upon any part of the mechanism by which Types are added and maintained.

ISO:
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40114

OGC: http://portal.opengeospatial.org/files/?artifact_id=25355

3.2.3 OGC Standard: Geography Markup Language (GML) simple features profile

This profile of GML implements the simple feature specification. It is frequently used in the Web Feature Service specification to serve geographic data as a service.

OGC: http://portal.opengeospatial.org/files/?artifact_id=42729

OGC: http://schemas.opengis.net/gml/3.1.1/profiles/gmlsfProfile/1.0.0/gmlsf.xsd

3.2.4 ISO 19142:2010 Geographic information -- Web Feature Service

ISO 19142:2010 specifies the behaviour of a web feature service that provides transactions on and access to geographic features in a manner independent of the underlying data store. It specifies discovery operations, query operations, locking operations, transaction operations and operations to manage stored parameterized query expressions.

ISO:
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=42136

OGC : http://portal.opengeospatial.org/files/?artifact_id=39967

3.2.5 OGC Standard: KML (formerly Keyhole Markup Language)

KML Preamble

Google submitted KML (formerly Keyhole Markup Language) to the Open Geospatial Consortium (OGC) to be evolved within the OGC consensus process with the following goal: KML Version 2.2 will be an adopted OGC implementation standard. Future versions may be harmonized with relevant OGC standards that comprise the OGC standards baseline. There are four objectives for this standards work:

- That there be one international standard language for expressing geographic annotation and visualization on existing or future web-based online and mobile maps (2d) and earth browsers (3d).
- That KML be aligned with international best practices and standards, thereby enabling greater uptake and interoperability of earth browser implementations.
- That the OGC and Google will work collaboratively to insure that the KML implementer community is properly engaged in the process and that the KML community is kept informed of progress and issues.
- That the OGC process will be used to insure proper life-cycle management of the KML Standard, including such issues as backwards compatibility.
-

The OGC has developed a broad Standards Baseline. Google and the OGC believe that having KML fit within that family will encourage broader implementation and greater interoperability and sharing of earth browser content and context.
KML is an XML language focused on geographic visualization, including annotation of maps and images. Geographic visualization includes not only the presentation of graphical data on the globe, but also the control of the user's navigation in the sense of where to go and where to look.
From this perspective, KML is complementary to most of the key existing OGC standards including GML (Geography Markup Language), WFS (Web Feature Service) and WMS (Web Map Service). Currently, KML 2.2 utilizes certain geometry elements derived from GML 2.1.2. These elements include point, line string, linear ring, and polygon.

OGC : http://portal.opengeospatial.org/files/?artifact_id=27810
OGC : http://schemas.opengis.net/kml/2.2.0/ogckml22.xsd

4 Represent spatial data in XML using OGC/ISO standards

4.1 Use GML representation of spatial objects as "E73 Information Object" attached to an "E53 Place"

This approach is used in the AnnoMAD System, a modular framework created by PIN (PIN, Universitá degli Studi di Firenze, Italy) and The Cyprus Institute (STARC, The Cyprus Institute, Cyprus) for the management and the integration of free-text archaeological data and geographic information related to excavations.

pro and cons:

+ full GML integration (all spatial and temporal objects defined in GML available)

+ representation and usage in GIS possible

+ examples of implementation available

- understanding of complex GML syntax / structures

- GML object as black box, no semantic differentiation. A GML object can contain anything from points, lines, polygons, solids or a combination of those in any coordinate system. There are no semantic means to determine what's the content of the GML object like through a SPARQL query. Therefore these GML object can't be compared or related to other GML objects or queried for their spatial content. Of course you can define implicit rules which GML elements are used and how they are applied to certain features like rivers or buildings.

Example: AnnoMAD System, Project http://www.vast-conference.eu/node/121

PPT Presentation: www.athenaeurope.org/getFile.php?id=627

```
<crm:E53.Place rdf:about="F231">
  <crm:P67B.is_referred_to_by>
    <crm:E73.Information_Object>
      <gml:Polygon>
        <gml:outerBoundaryIs>
          <gml:LinearRing>
            <gml:coordinates>
              <(coordinates here)/>
            </gml:coordinates>
          </gml:LinearRing>
        </gml:outerBoundaryIs>
      </gml:Polygon>
    </crm:E73.Information_Object>
  </crm:P67B.is_referred_to_by>
</crm:E53.Place>
```

A. Felicetti, M. Samaes, K. Nys & F. Niccolucci / AnnoMAD: Integrating Full-text Excavation Data and Geographic Information in The 11th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2010)

4.2 Use KML representation of spatial objects as "E73 Information Object" attached to an "E53 Place"

The same methodology as in the former section can be used with a KML representation instead of GML. KML is widely used on the net for geographic information, promoted by google, can represent textured 3D models and was adopted as OGC standard in 2008. Styling of information is available in KML.

pro and cons:

+ global user group and promotion through google

+ direct integration in Google Earth, google maps and a growing number of other Geobrowsers, Digital Globes and GIS Systems

+ examples of KML implementations widely available

- KML object as black box, no semantic refinements, same issue as with GML object. Moreover KML can have styling information which can represent and transport implicit semantic information.

Example for KML Syntax from OGC KML 2.2 Specification

```
<kml xmlns="http://www.opengis.net/kml/2.2">
  <Document>
    <Placemark>
      <Polygon>
        <altitudeMode>clampToGround</altitudeMode>
        <outerBoundaryIs>
          <LinearRing>
            <coordinates>-135,78.5,300000 -135,12.5,300000 -45,12.5,300000 -
45,78.5,300000 -135,78.5,300000</coordinates>
          </LinearRing>
        </outerBoundaryIs>
      </Polygon>
    </Placemark>
  </Document>
</kml>
```

5 Represent spatial data in RDF/OWL based on the OGC candidate standard of "GeoSPARQL"

The GeoSPARQL candidate Standard provides a framework how to implement the OGC Standards (Abstract Specifications and Implementation Specifications) with semantic technologies through RDF/OWL encoding with definitions what SPARQL queries should be available.

The candidate OGC "GeoSPARQL: A Geographic Query Language for RDF Data" Standard defines spatial extensions to the W3C's SPARQL protocol and RDF query language. http://www.opengeospatial.org/standards/requests/80

SPARQL is a protocol and query language for the Semantic Web. SPARQL is defined in terms of the W3C's RDF data model and will work for any data source that can be mapped into RDF, which potentially includes sources of geospatial data. The OGC GeoSPARQL standard supports representing and querying geospatial data on the Semantic Web. GeoSPARQL provides the foundational geospatial vocabulary for linked data involving location and defines extensions to SPARQL for processing geospatial data. This standard serves as a common target for vendors to implement and provides rich functionality for building geospatial applications.

5.1 Overview

GeoSPARQL follows a modular design. A *core* component defines top-level RDFS/OWL classes for spatial objects. A *geometry* component defines RDFS data types for serializing geometry data, RDFS/OWL classes for geometry object types, geometry-related RDF properties, and non-topological spatial query functions for geometry objects. A *geometry topology* component defines topological query functions. A *topological vocabulary* component defines RDF properties for asserting topological relations between spatial objects, and a *query rewrite* component defines rules for transforming a simple triple pattern that tests a topological relation between two features into an equivalent query involving concrete geometries and topological query functions.

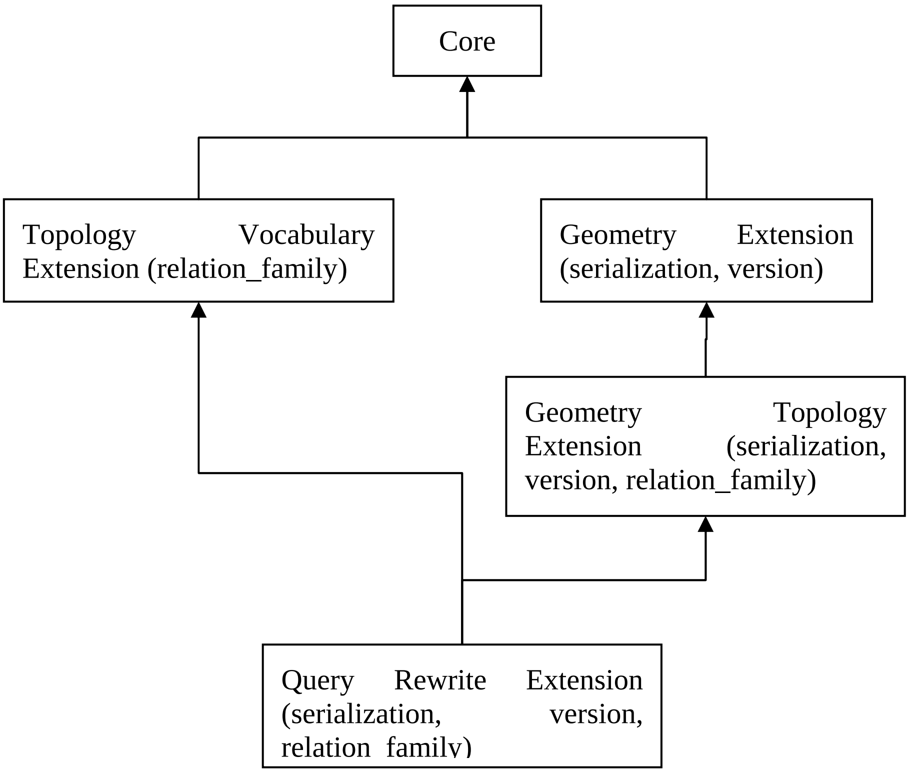


Figure 1: Requirements Class Dependency Graph (GeoSPARQL 2011)

5.2 Core Component

Two main classes are defined: `geo:SpatialObject` and `geo:Feature`. `SpatialObject` is the superclass of everything feature or geometry that can have a spatial representation and the root class within the hierarchy of the GeoSPARQL ontology.

`Feature` is defined as superclass of everything feature in GeoSPARQL.

The definitions within the ISO 19100 series give a more detailed view what is meant:

"A feature is an abstraction of a real world phenomenon" [ISO 19101];

A feature is a geographic feature if it is associated with a location relative to the Earth. Vector data consists of geometric and topological primitives used, separately or in combination, to construct objects that express the spatial characteristics of geographic features.

Attributes of (either contained in or associated to) a feature describe measurable or describable properties about this entity. Unlike a data structure description, feature instances derive their semantics and valid use or analysis from the corresponding real world entities' meaning.

Documenting feature instances, types, semantics and their properties is often detailed in an information model. An information model details how to take real world ideas or objects and make them useful to a computer system. In the geospatial world the focus is on depicting things in the real world as points, lines, or polygons (the geometry ”primitives” we use to assemble location data about those real world objects) and their attributes (information about those objects). When linked together, a pair (geometry and attributes) representing one or more real world objects, is called a feature.

In ISO 19101, there is the concept of the Domain reference model. The DRM provides a high-level representation and description of the structure and content of geographic information.

Examples from the CRM domain:

streets, buildings (E26)

administrative boundaries like countries or communities (E28)

zones of activities, travelling routes (E7)

5.3 Geometry Extension

This requirements class defines a vocabulary for asserting information about geometry data, and it defines query functions for operating on geometry data.

As part of the vocabulary, an RDFS dataype is defined for encoding detailed geometry information as a literal value. A literal representation of a geometry is needed so that geometric values may be treated as a single unit. Such a representation allows geometries to be passed to external functions for computations and to be returned from a query.

A single root geometry class is defined: `geo:Geometry` as a subclass of the `geo:SpatialObject` class defined in the core component. In addition, properties are defined for describing geometry data and for associating geometries with features.

To represent the actual coordinates of a geometry, a so called Serialization is used. That means that the coordinates are stored in a format which defines the sequence of the characters. Two formats are specified. One is Well Known Text (WKT) Serialization as defined by Simple Features or ISO 19125 and the other is a GML Serialization as defined in ISO 19136. These specifications (ISO 19125, ISO 19136) are also the base for subclasses of the geometry class. An RDF/OWL class hierarchy can be generated from the WKT or GML schema that implements GM_Object from ISO 19107 or its core profile ISO 19137.

Coordinate Reference System information is encoded in the WKT or GML Serialization. In future there is the possibility for further extensions covering KML or GeoJSON Serialization.

Figure 2 is an illustration of the Core component classes (SpatialObject, Feature) and the Geometry class with its properties.

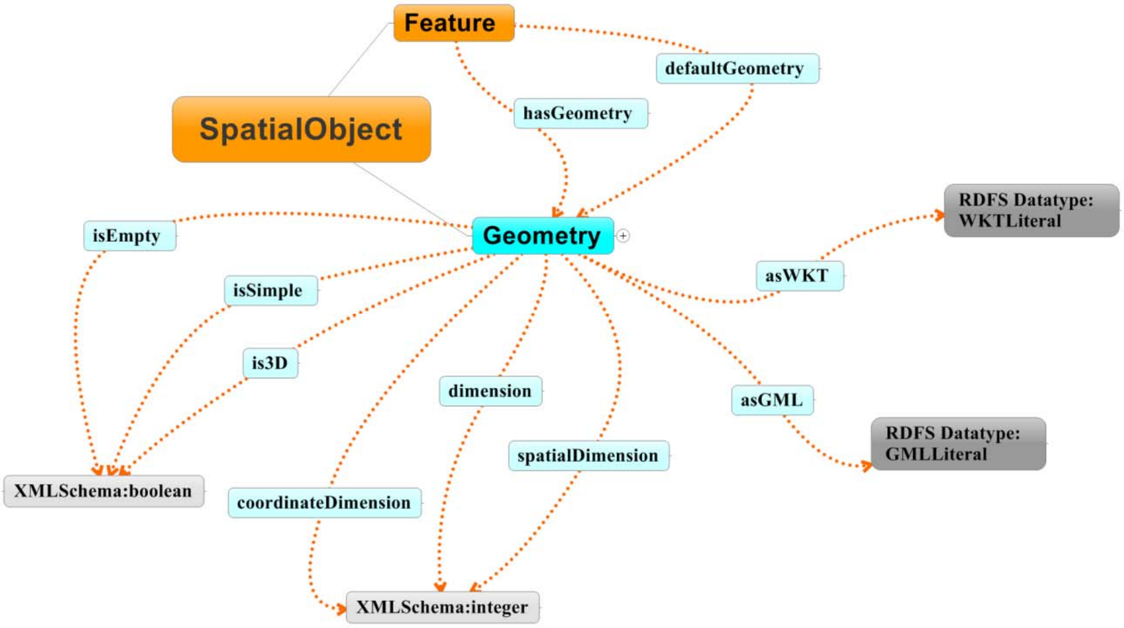


Figure 2: Core component classes (SpatialObject, Feature) and Geometry Class with properties

pro and cons:

+ combination of semantic world (RDF/OWL) and OGC/ISO standards

+ full GML integration (all spatial and temporal objects defined in GML available)

+ representation and usage through WebFeatureServices in GIS possible (transformations necessary)

+ promising in the point of view to become a standard - if it becomes OGC standard, it may become an ISO standard

- only candidate standard yet

- no examples of implementation available

- understanding of semantic GeoSPARQL structures (based on OGC/ISO standards) and GML syntax necessary

- implementation cost very high for the first implermentations

- no namespace yet

6 Concept to integrate the GeoSPARQL Core Component and Geometry Extension in CIDOC CRM

In this chapter we try to develop a concept how GeoSPARQL could be integrated into the CRM. Most of all, this concept is an invitation to discuss the following ideas and challenges within the CRM community.

There are two tasks to be realized:

- Mapping of GeoSPARQL classes to CRM classes
- Encoding of the actual Coordinate Information

6.1 Mapping

In order to have semantic functionalities and be aligned with the GeoSPARQL standard a mapping to CRM is necessary. There are three issues to be discussed how to map the GeoSPARQL classes Features, Geometries and Spatial Reference Systems to CRM classes.

6.1.1 Features

There is the question how to integrate the feature class in the CRM. As the definition in 5.2 shows a feature is most of the times an object representing an E53 Place in combination with other E1 objects. In ISO 19109 features do have "feature types" specifying and classifying the meaning of features. For the domain of the CRM these feature types correspond to the CRM classes. But a feature can also just be instances of E53 Place without any other additional semantic meaning. The Getty Thesaurus of Geographic Names (TGN) with placenames and coordinates would be one example of features that could be represented only through E53 Place.

A possible approach would be to subclass feature to E53 with the rule, that for any other semantic meaning of a feature (specified through the feature type) a corresponding CRM object has to be created in addition to the E53 instance. A less semantically rich alternative would be to create an E55 Type from the feature type.

6.1.2 Geometries

The GeoSPARQL candidate standard offers two alternatives to represent geometries. Either WKT or GML. In this concept we would opt for GML as the GML specification is richer, more widely used (especially in WFS services) and uses XML encoding.

The challenge is now to build a RDF/OWL class hierarchy from the GML schema that implements GM_Object from ISO 19137 Spatial Schema core profile (as the scope is actually sufficient). The Geography Markup Language (GML) simple features profile (see chapter 3.2.3) implements the relevant parts of ISO 19137. The suggestion is to implement for a test scenario only classes that are necessary to hold the geometric information as subclasses to the E47 Coordinate Information class.

6.1.3 Spatial Reference Systems

Although in GeoSPARQL Spatial Reference Systems are defined in the Serialisation of WKT or GML we would suggest to explicitly store the information in a subclass of E75 Conceptual Object Appellation.

6.2 Encoding of the actual Coordinate Information

In GeoSPARQL there are currently two options to encode the actual Coordinate Information. The richer alternative is GML and the suggestion is to build a test scenario using GML.

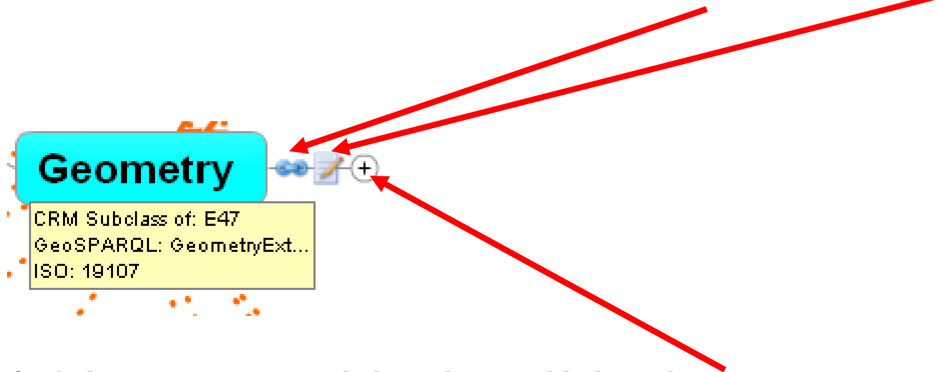
Maybe alternatives as KML and „Basic Geo (WGS84 lat/long) Vocabulary" should be tested as well although they are not in the GeoSPARQL recommendation (yet?).

6.3 Concept of GeoSPARQL integration into the CRM

In Figure 3 we make a suggestion how to integrate the GeoSPARQL Core Component (orange) and the GeoSPARQL Geometry Extension (blue) with CIDOC CRM. For each class and property of GeoSPARQL the suggested CRM superclass or superproperty is given. For the classes the ISO standards providing the conceptual background are listed with the GeoSPARQL component name.

Figure 3 is available as pdf called "GeoSPARQL_4_CRMv2.pdf" containing comments and hyperlinks for the classes and properties.

It is an interactive pdf, to klick or mouseover on Hyperlinks and comments.



If subclasses are not extended you have to klick on the "+"

For test purposes an OWL file (based on the Erlangen CRM) was created with Protegé 3.3.1. containing these classes. The files are called:

GeoSPARQL_4_CRMv2.owl

GeoSPARQL_4_CRMv2.repository

GeoSPARQL_4_CRMv2.pprj

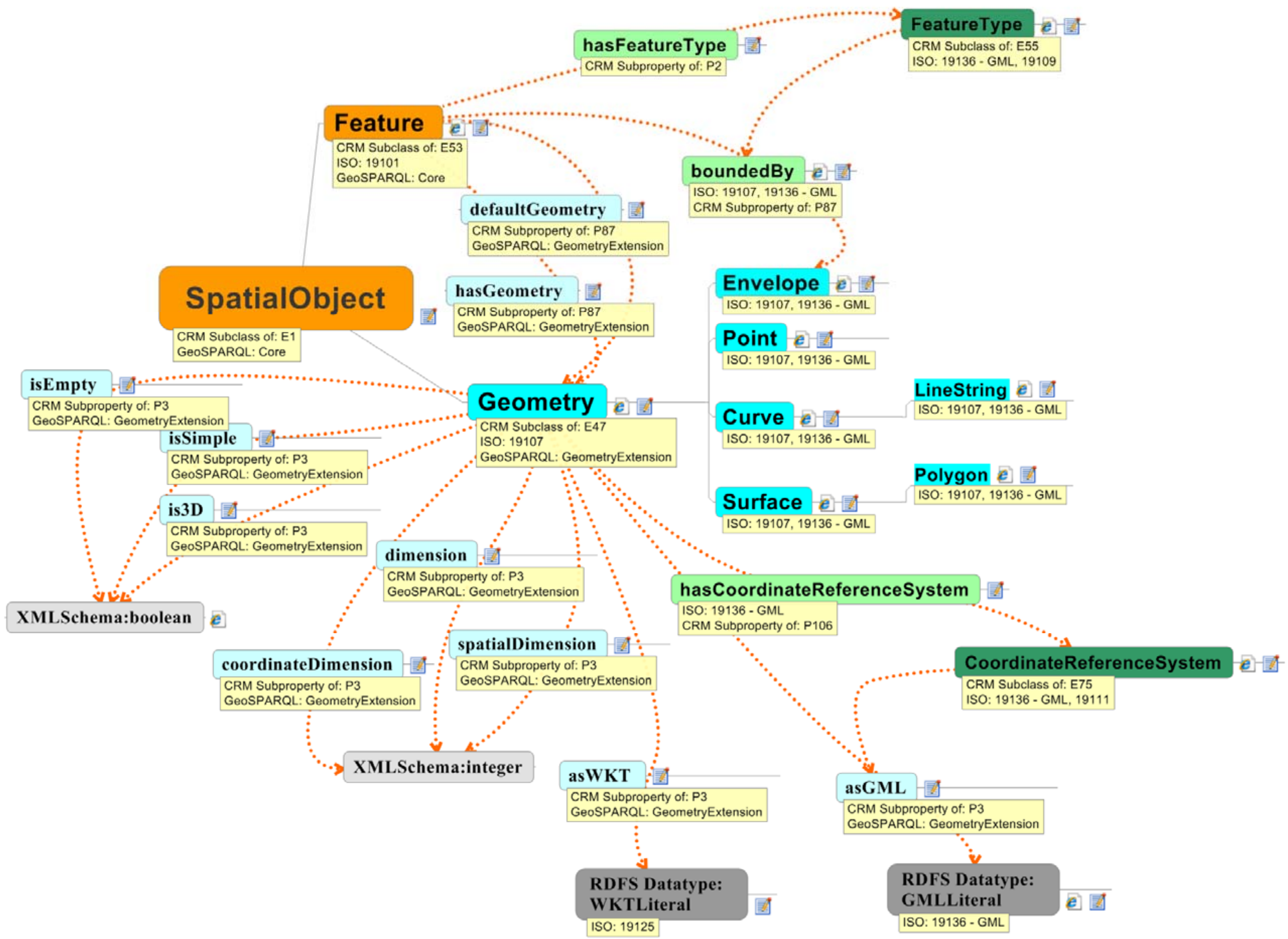


Figure 3: GeoSPARQL Core Component and Geometry Extension with CIDOC CRM and ISO relations

7 Conclusions

8 Appendix A – UML Diagramms for Abstract Specifications

8.1 ISO 19107 - Spatial Schema and ISO 19137 - Core profile of the spatial schema

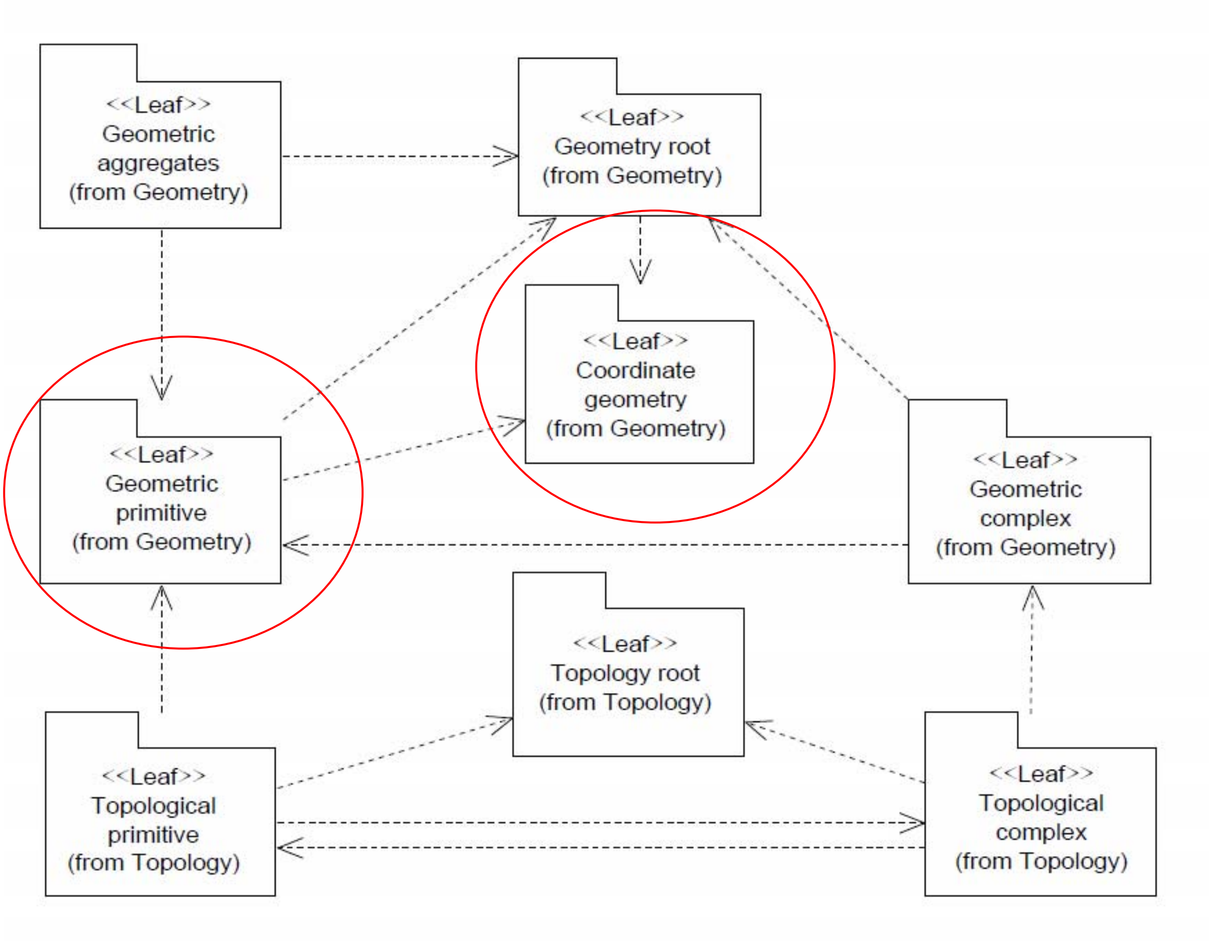


Figure 4: ISO 19107 Geometry packages (Red circles for the subset of the ISO 19137 core profile)

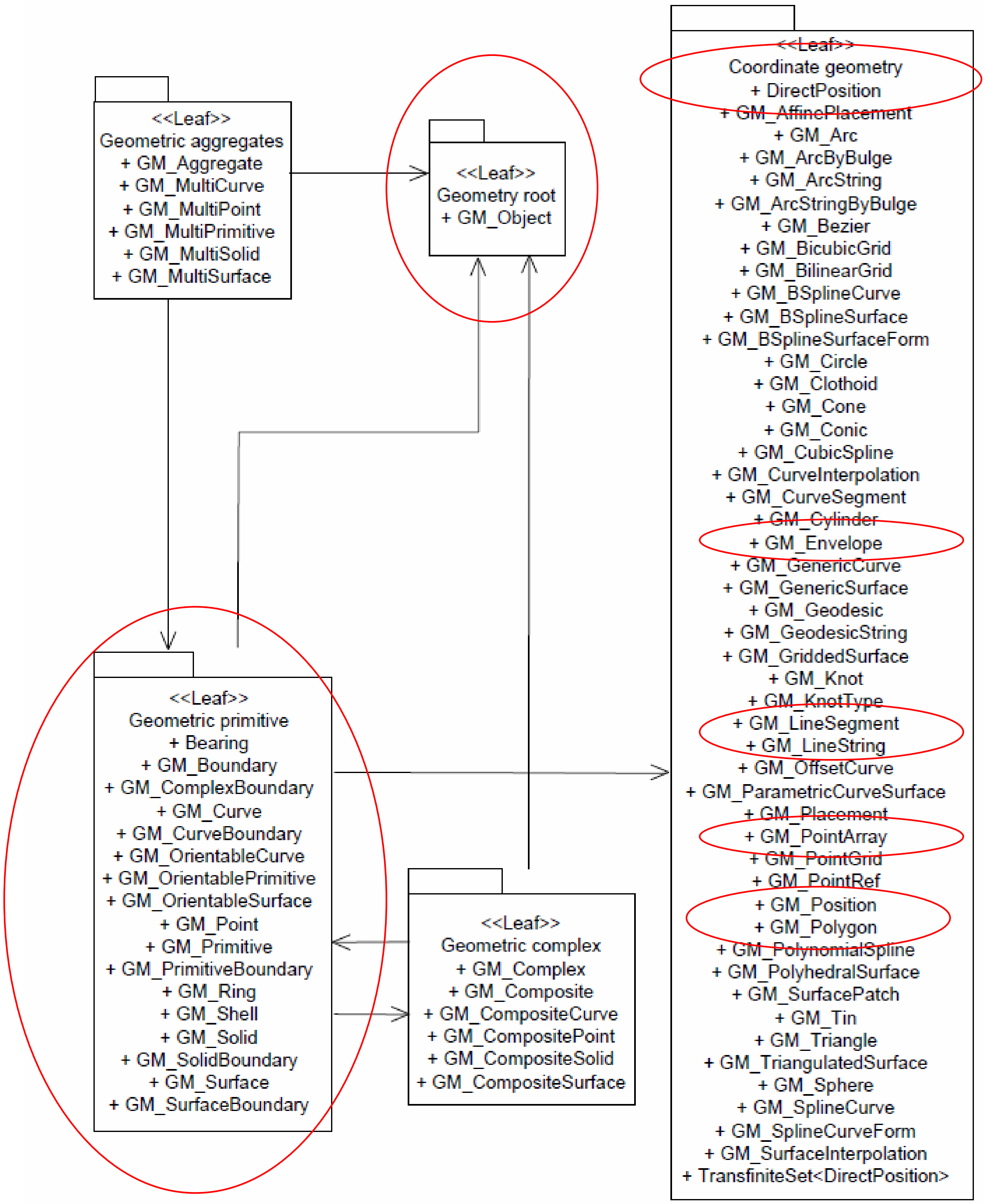


Figure 5: Geometry package: Class content and internal dependencies (Red circles for the subset of the ISO 19137 core profile)

8.2 ISO 19111:2007 Geographic information -- Spatial referencing by coordinates

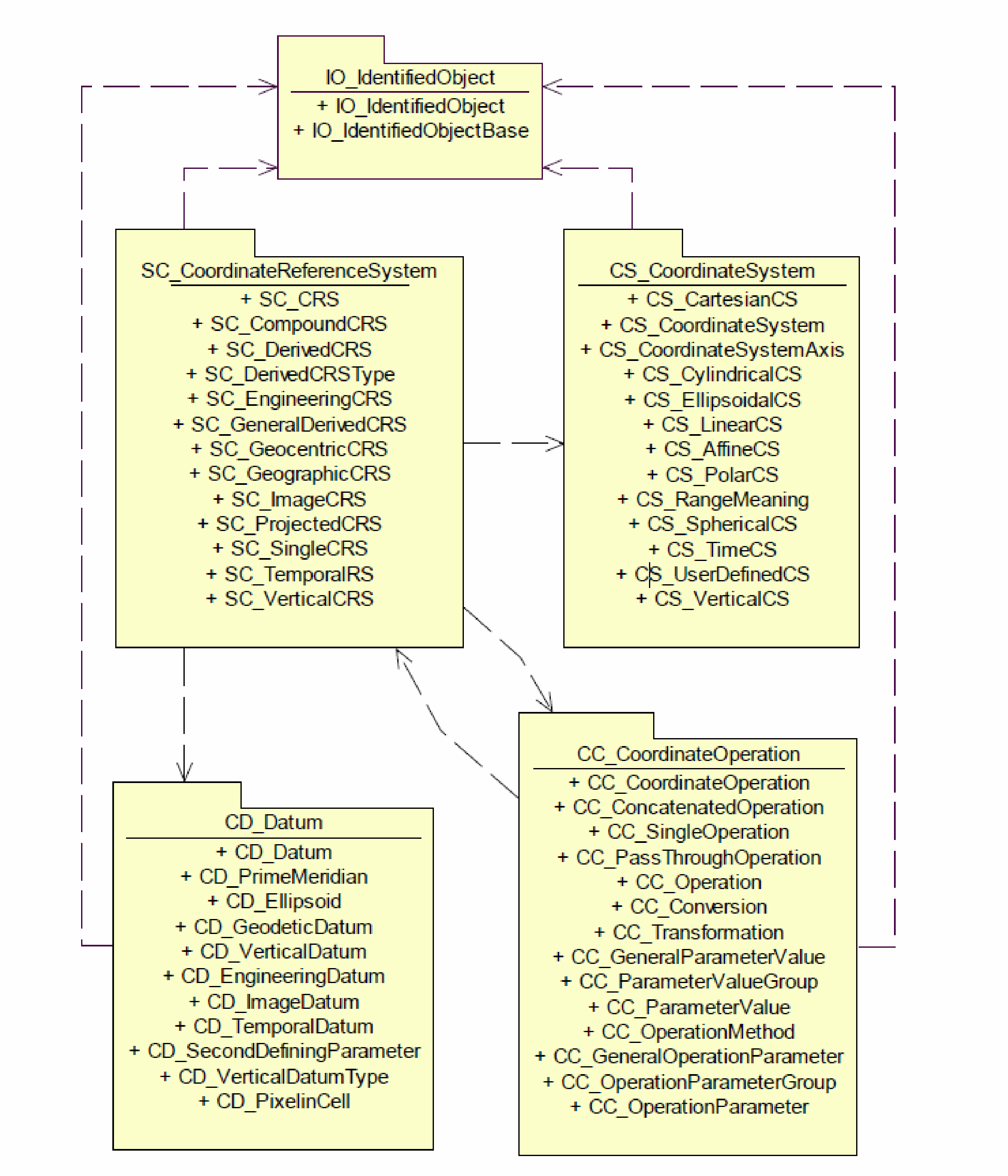


Figure 6: Five primary UML packages for Spatial referencing by coordinates