Open GIS Consortium

35 Main Street, Suite 5
Wayland, MA 01778
Telephone: +1-508-655-5858
Facsimile: +1-508-655-2237

Editor:
Telephone: +1-703-830-6516
Facsimile: +1-703-830-7096
ckottman@opengis.org

The OpenGIS™ Abstract Specification
<u>Topic 14: Semantics and Information Communities</u>

Version 4

The OpenGIS™ Abstract Specification

Revision History

Date	Description
4 April 1999	Carry forward 98-114r1 and rename to 99-114; update copyright for 1999; use revised document template; update cross-references to references, sections, figures and tables; add content from Section 6.1 of the OpenGIS Guide to Section 1.2.

This page is intentionally left blank.

Table of Contents

1. Introduction

1.1. The Abstract Specification

The purpose of the Abstract Specification is to create and document a conceptual model sufficient enough to allow for the creation of Implementation Specifications. The Abstract Specification consists of two models derived from the Syntropy object analysis and design methodology [1].

The first and simpler model is called the Essential Model and its purpose is to establish the conceptual linkage of the software or system design to the real world. The Essential Model is a description of how the world works (or should work).

The second model, the meat of the Abstract Specification, is the Abstract Model that defines the eventual software system in an implementation neutral manner. The Abstract Model is a description of how software should work. The Abstract Model represents a compromise between the paradigms of the intended target implementation environments.

The Abstract Specification is organized into separate topic volumes in order to manage the complexity of the subject matter and to assist parallel development of work items by different Working Groups of the OGC Technical Committee. The topics are, in reality, dependent upon one another— each one begging to be written first. *Each topic must be read in the context of the entire Abstract Specification.*

The topic volumes are not all written at the same level of detail. Some are mature, and are the basis for Requests For Proposal (RFP). Others are immature, and require additional specification before RFPs can be issued. The level of maturity of a topic reflects the level of understanding and discussion occurring within the Technical Committee. Refer to the OGC Technical Committee Policies and Procedures [2] and Technology Development Process [3] documents for more information on the OGC OpenGIS™ standards development process.

Refer to Topic Volume 0: Abstract Specification Overview [4] for an introduction to all of the topic volumes comprising the Abstract Specification and for editorial guidance, rules and etiquette for authors (and readers) of OGC specifications.

1.2. Introduction to Semantics and Information Communities

The OpenGIS™ notion of Information Communities was devised to enable groups such as ecologists and civil engineers to efficiently manage the semantics (or feature schema mismatches) of their own geodata collections and get maximum benefit from each other's geodata collections, despite semantic differences.

An Information Community is a collection of people (a government agency or group of agencies, a profession, a group of researchers in the same discipline, corporate partners cooperating on a project, etc.) who, at least part of the time, share a common digital geographic information language and share common spatial feature definitions. This implies a common world view as well as common abstractions, feature representations, and metadata. The feature collections that conform to the Information Community's standard language, definitions and representations belong to that Information Community.

1.3. References for Section 1

[1] Cook, Steve, and John Daniels, Designing Objects Systems: Object-Oriented Modeling with Syntropy, Prentice Hall, New York, 1994, xx + 389 pp.

[2] Open GIS Consortium, 1997. OGC Technical Committee Policies and Procedures, Wayland, Massachusetts. Available via the WWW as <http://www.opengis.org/techno/development.htm>.

[3] Open GIS Consortium, 1997. The OGC Technical Committee Technology Development Process, Wayland, Massachusetts. Available via the WWW as <http://www.opengis.org/techno/development.htm>.

[4] Open GIS Consortium, 1999. Topic 0, Abstract Specification Overview, Wayland, Massachusetts. Available via the WWW as <http://www.opengis.org/techno/specs.htm>.

2. The Essential Model for Semantics and Information Communities

2.1. Geospatial Information Communities

We will define a Geospatial Information Community (GIC) in several steps. First we list the requirements we impose upon an information community. Then we introduce the notion of isomorphisms between Project Worlds using the notion of subworlds. Using this notion, we define *community metadata*. Finally, GICs are defined in terms of the community metadata with which they are compatible.

2.1.1. The requirements upon an information community

Each Project World has a fixed vocabulary, a fixed collection of schema (namely, the Project Schema and Feature Schema and their subschema), and an unambiguous set of feature instances. If we change a word in our Project World vocabulary, say from "house" to "dwelling," the result is a different Feature Schema, and therefore, a different Project World. We need a tool that, when we have two implementations with such trivial differences, enables each to be able to share information with the other as if it were native.

The situation should be the same if we make a minor change to the Geometry Schema, say by modeling a curve with a spline instead of an arc, the result is a different Project World. Two implementations should also be able to share information if they use different Spatial Reference Systems, or use slightly different methods to determine the width of a road.

It is toward these ends that we make the formal definitions in Section 2.1.2.

2.1.2. Subworlds Defined

To make the concept of Information Community sharp, begin by considering the set, A, of all possible Project Worlds. In Topic 5 of the Abstract Specification [2], we have defined the notion of Project World very tightly, so that each Project World has exactly one implementation as an OpenGIS™ Feature Collection. Each Project World has, therefore, among other things, a spatial reference system, a set of feature instances, and a list of properties and property values.

It should be possible to move information easily and without semantic loss from Project Worlds having naive schema into Project Worlds with more sophisticated and inclusive schema. Moving information the other way requires the truncation and loss of information. A Project World that is more naive than another is called a subworld of the other.

Note that a sophisticated schema should not be denied potential subworlds just because they fall outside the physical extent of its Project as specified in its Project Schema. We ignore the physical extent of projects when comparing them to check if one is a subworld of the other.

Here is a more detailed definition of "

Definition: Let S_1 and S_2 be two Project Worlds in **A**. Let S_2* be the Project World obtained by extending the physical extent of Project S_2 (if necessary) until it covers the extent of Project S_1. We say S_1 is a *subworld* of S_2 if there are three functions, R_1, R_2, and R_3, that behave as follows:

    i. R_1 is a one-to-one change-of-spatial-reference from the reference system of S_1 to that of S_2*

    ii. R_2 is a mapping from the feature instances of S_1 into those of S_2* such that F is a feature of S_1 occupying a point P if and only if $R_2(F)$ is a feature of S_2* occupying $R_1(P)$,

    iii. R_3 is a mapping from the set of property/value pairs of all features in S_1 into the set of property/value pairs of all features of S_2* that preserves semantics, and is canonical with R_2.

2.1.3. Subworlds Explained

The notion of a subworld needs a little explanation.

The extension from S_2 to S_2* is necessary because we are interested in whether the Feature Schema and Project Schema of S_2 is robust, sophisticated and inclusive enough to model all the features found in S_1 <u>without regard</u> for the specific extent of Project S_2. Therefore, we suspend the portion of the Project Schema of S_2 that defines its physical extent for the definition of Subworld. The reason is clear: we are defining the ability to import information, and if one can import a certain

schema in one region of the earth, then one can import information in the same schema at any other location.

The rest of the definition involves three functions, R_1, R_2, and R_3, and three conditions.

The first condition allows the two Project Worlds, S1 and S2*, to have different spatial reference systems. However the two spatial reference systems must be resolved by a one-to-one change-of-spatial-reference function, R1. Notice that this change-of-spatial-reference maps from the subworld to the superworld, and that it could possibly be an into (not onto) function. This allows a 2-dimensional Project World to be a subworld of a three-dimensional Project World.

The second condition guarantees that, after the change-of-spatial-reference, every feature in S1 geometrically "neatly fits" a feature of S2* with which it is paired. R2 is the pairing function. This allows differences in geometric constructions, such as the use of polylines in one Project World, and the use of splines in another, up to the tolerance of the condition "occupies." Just how snug the image of a feature should fit its pair is a delicate issue related to the use of different geometries in different implementations. We leave it to the reader to use good judgment, and defer further discussion.

In the case that S1 is two dimensional, and S2* is 3-dimensional, the "neatly fits" testing is done in a projection of S2* upon S1.

The third condition states two things. First there is a (perhaps one-to-many) mapping R3 from the property/value pairs of S1 to those of S2* with the following condition:

The Semantic Condition: if P/V is a property/value pair in S_1, then the semantics of P/V are a subset of the semantics of R_3(P/V).

This is just a formal way to state that the Project World, S2, is more sophisticated than S1. It is, of course, a human function to affirm whether the Semantic Condition is met or not.

The second part of the third condition states that if a feature F1 in S1 is paired by R2 with a feature F2 in S2*, then the semantics of any property/value pair of F1 are a subset of the semantics of property/values of F2. This means that within every category of information captured for feature F1, the information held within that category is a subset of (or can be derived from) the information captured about F2.

The canonical mapping is shown symbolically in Figure 2-1.
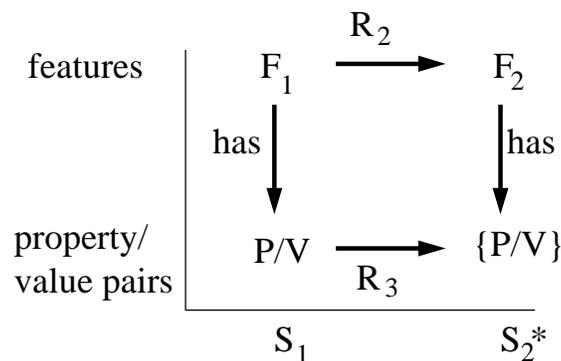
Figure 2-1. The canonical map.

The third condition in the definition of subworld is designed to keep language (e.g. the use of Latin names, or the use of the jargon of a particular discipline) from being a barrier that divides information communities. It is intended that R3 preserve semantics, but not necessarily the actual names of properties nor the units or domains of their values.

R2 is actually a special case of R3, in as much as the geometry property (taking an OpenGIS WKS for a value) is a special case of property. (Here, the "semantics of geometry" is the space that the geometry occupies.) In other contexts, the mappings R3, would be an instance of a Thesaurus. This makes R2 a kind of spatial Thesaurus.

If S1 is a subworld of S2, then geospatial information can move from S1 to S2 without information loss. Moreover, the transfer can be automated with the help of formalization and crosswalks between the schema involved.

2.1.4. Isomorphic Project Worlds

Definition: Two Project Worlds, S_1 and S_2, in *A* are *isomorphic* if each is a subworld of the other.

The isomorphic relation clearly satisfies the symmetry and reflexivity conditions. For now we will assume it satisfies the transitivity condition as well. (We will revisit this point in a future section on tolerances.) Therefore, the isomorphic relation is an equivalence relation. The notion of an equivalence class of Project Worlds is closely related to that of a GIC.

It is important to note that isomorphisms between Project Worlds depend on the compatibility of their feature schema, and the lexical semantics of their project schema. In particular, the feature collection identification, the project semantics, and the use semantics do not play a role. (They play a role, perhaps, in a user's decision to desire a feature collection, but no role in the user's technical ability to ingest it into the user's system, or into an existing Feature Collection there, <u>with understanding</u>.) Let us, therefore, denote the union of the feature schema and lexical semantics of a Project World the *community metadata*.

Since each Project World specifies a set of community metadata, the isomorphism relation on Project Worlds also creates equivalence classes of community metadata. Each equivalence class defines an information community. To be formal, let Q denote an equivalence class of community metadata.

Definition: A system, S, is a member of the *Geospatial Information Community* associated with Q, denoted $S \in GIC(Q)$, if there is a Project World, P, such that:

 i. The community metadata of P is a member of the equivalence class Q, and

 ii. S models the OpenGIS Feature Collection associated with P.

That is, if a system accommodates one member of a set of equivalent community metadata objects, then it belongs to the GIC associated with that equivalence class.

Figure 2-2 helps explain the situation. The grid squares represent equivalence classes of community metadata. System 1 recognizes four of these equivalence classes. (Perhaps System 1 supports maps of a common theme at four different scales.) System 2 recognizes five of the equivalence classes. (Perhaps System 2 supports maps at one scale of five different themes.) In Figure 2-2, Systems 1 and 2 share membership in one equivalence class called Q. Therefore Systems 1 and 2 share membership in the GIC associated with Q, GIC(Q).
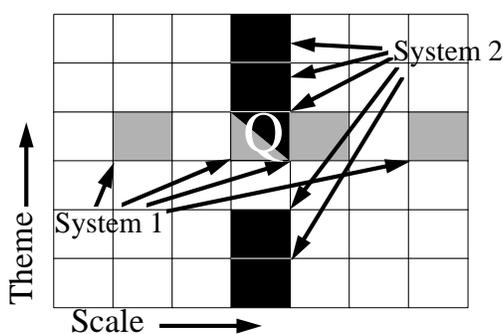

Figure 2-2. The Space of Information Communities =Equivalence Classes of Community Metadata

Information communities are partially ordered by the inclusion relation on their community metadata equivalence classes. That is, we can say GIC(R) Í GIC(Q) if R Í Q, or equivalently, if there is a Project World, P1, whose community metadata is a member of R, and a Project World, P2, whose community metadata is a member of Q, and P1 is a subworld of P2. Two information communities that are incomparable under this relation can be bridged by the use of the intersection or union of appropriate Project Worlds. The intersection is accomplished by deleting all community metadata in each except that common to both. The union is accomplished by

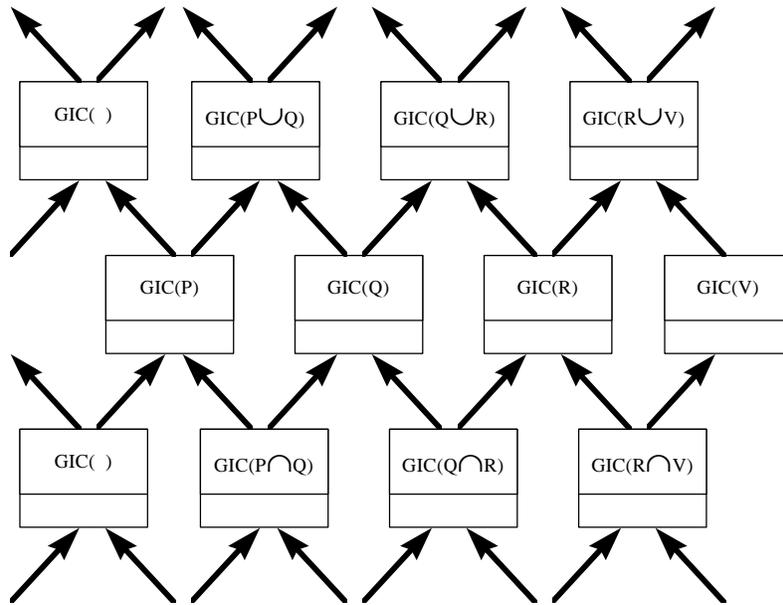expanding their Project Worlds until each encompasses all information found in both information communities.

Figure 2-3. Relationships between Information Communities.

Figure 2-3 shows a portion of the partially ordered network of Geospatial Information Communities.

We close this section with a brief discussion on the emergence of standard Geospatial Information Communities in the United States. There are several powerful movements that are standardizing community metadata. For example, the U.S. Federal Geospatial Data Committee, empowered with an Executive Order (number 12906) and a mandate from the Office of Management and Budget (Circular A-16), is forwarding standard community metadata in twelve areas.

Base Cartography	Geodetic	Soils
Bathymetric	Geologic	Water
Cadastral	Ground Transportation	Vegetation
Cultural and Demographics	Wetlands	Nat. Res. Inv.

The amount of research into standards in these areas has been considerable. The Tri-Service CADD/GIS Technology Center, for example has been developing a standard for Community Metadata for Facilities information called the Tri-Service Spatial Data Standard. See [1]. It is available for ftp at the Tri-Service web site, but occupies about 30 megabytes compressed.

Note: Section 2.2, following this note, seems more properly to belong in Topic 13. The editor recommends this section be deleted from Topic 14 in the next revision cycle, and, if appropriate, picked up by the Catalog Working Group and placed in Topic 13.

2.2. A Discussion of Discovery Technology

GICs need a technology that empowers them to announce the existence of themselves and their information, so that other individuals outside that GIC may discover them and assess whether there may be interest in sharing their information. Since these services are traditionally provided by libraries, we postulate a Library World specific for digital geospatial information.

2.2.1. Catalogs

The Catalog is the fundamental tool for digital geospatial information discovery. Catalogs are populated by entries that are queryable. The entries are usually derived directly from Feature Collection Signature object types. It is possible, however, that additional entries may be generated by augmenting the Feature Collection Signature with additional information, such as that which might be obtained by passing a special filter over the Feature Collection object type itself.

Figure 2-4 shows some of the important object types associated with the Catalog object type.
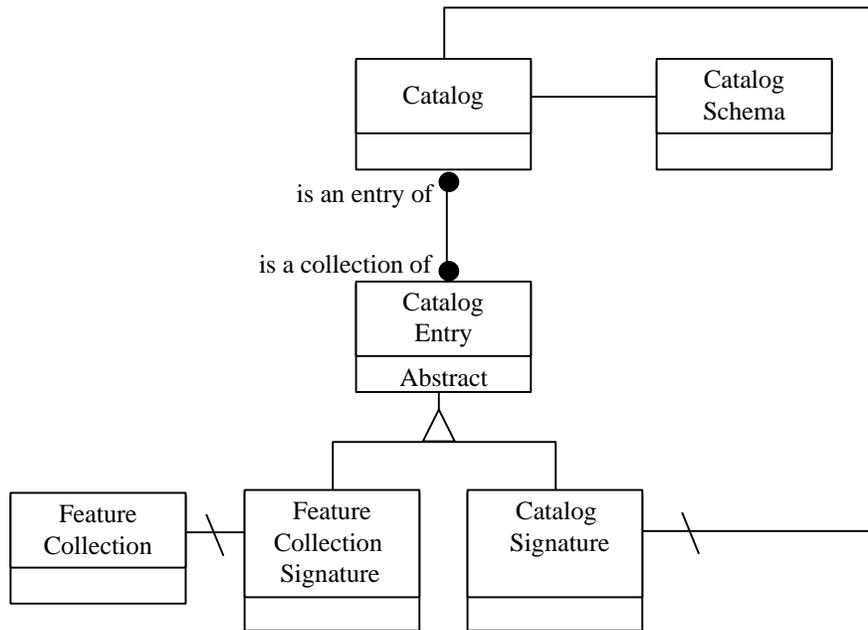

Figure 2-4. Object Types near the Catalog

The Catalog Entry object type should embody the key information answering the questions: where, what, who, when, how, and why.

The Catalog Signature functions to describe the properties of the services offered by the Catalog . A Catalog Schema imposes the required structure on the Catalog Signature.

Catalog Entries are, in fact, either Feature Collection Signatures or Catalog Signatures, depending on whether the catalog contains information about Feature Collections or other Catalogs.

2.3. The Status of the Abstract Specification for Semantics and Information Communities

There are many important research issues in the area of Semantics and Information Communities. OGC plans to develop close ties with selected university programs dedicated to fundamental research in these areas.

First implementations of interfaces in the areas of Semantics and Information Communities is expected from the Domain Special Interest Groups and Working Groups: Transportation, Telecommunications, and Defense and Intelligence. Each of these domains has a "world view" that engenders a discipline dictionary, and a supporting semantic library or thesaurus.

2.4. References for Section 2

[1] Tri-Services CADD/GIS Technology Center, 1995. Tri-Services Spatial Data Standards, Vicksburg, Available via the WWW as http://mr2.wes.army.mil/docs/sds.htm

[2] OpenGIS™ Abstract Specification, OpenGIS™ Project Documents 99-100 through 99-116, available through www as <http://www.opengis.org/techno/specs.htm>.

3. Abstract Model for Semantics and Information Communities

> **Note:** It is the editor's intention that this section be a domain-independent abstraction of the work that is beginning in the Domain Task Force. That is, this Topic should provide Core foundations for structures needed in specific domains. Abstract feature class hierarchies, abstract entity-to-entity relationships, and similar Core technologies are envisioned.

3.1. References for Section 3

[1] OpenGIS™ Abstract Specification, OpenGIS™ Project Documents 99-100 through 99-116, available through www as <http://www.opengis.org/techno/specs.htm>.

4. Future Work

Note: As sound abstractions of specific domains are instituted, work will begin to "bridge" the domains. At that time, a Topic on Semantic Translation may be contemplated.

5. Appendix A. Acronyms and Glossary

This section is TBD.

6. Appendix B. Well Known Structures

It is intended that WKS, as they are developed for this Topic, be placed in this Section.