Development of Common Geospatial Data Content Standards

*Concepts and techniques for standardization*

Douglas Nebert
FGDC

Data Standardization

- **Application Requirements**: Business rules, usage
- **Conceptual Model**: Abstract model, UML
- **Implementation Model**: ER Diagrams, XML Schema, ISO Application Schema
- **Encoding/Packaging Format**: XML, DIGEST VPF, SDTS
- **Transfer Protocol**: ISO 9660, http, ftp
- **Transport**: CD-ROM, tcp-ip
Data Development

There are two broad categories of geospatial data to consider building content standards for:

- Base data, also known as Framework, Fundamental, Foundation data
- Thematic data: usually additional themes with more specific utility to a discipline

It may not be easy to distinguish base from thematic data – the design process can be the same.

Framework Themes

Themes providing the core, most commonly used set of base data are known as Framework Data:

- Geodetic Control,
- Orthoimagery,
- Elevation and Bathymetry,
- Transportation,
- Hydrography,
- Cadastral, and
- Governmental Units.
Additional Data

- Geographic names (toponymy) layer
- Land cover/vegetation/wetlands
- Cultural and Demographic Statistics
- Buildings and Facilities
- Natural hazards
- Soils and Geology
- Utility distribution networks

Where to begin?

- Cartographic base map content tends to be driven by provider requirements to generate cartographic products
- Today’s geospatial information is useful for analytic purposes in databases, GIS, and in supporting models – and to make maps
- Requirements for representative data design must recognize multiple requirements for use including but not limited to the generation of traditional maps
Application Requirements

- Application requirements are used to design data and interfaces that will support generalized access to geospatial data for multiple participants.
- Contribute features, attributes, relationships and constraints as input to the conceptual modeling.

- Business rules, usage
- Abstract model, UML

Application Requirements

- Can the content of this information be used by multiple organizations?
- Who are the producers and consumers of such information? Who are the domain experts?
- What applications could use this information if shared?
- What attributes or structural characteristics are needed to support multiple uses in GIS and in mapping?
Feature Catalog

► One first step toward developing a conceptual model of geographic information is to construct a Feature Catalog

► Feature Catalog includes:
  ■ Feature types, definitions
  ■ Attributes, definitions, data types
  ■ Domains, expected values and types

Catalog as Abstract Model

► Feature Catalog describes what information is included in a given data theme and what properties and values are stored there

► A feature catalog is not an implementation model but can, with rules, be used to create one or more implementation models

► Implementation guidance supplements abstract or conceptual models
Business requirements

- Treat the development of data as an element in the design of a community information system.
- Must support generic functional needs of users and providers in solving problems.
- Such "use-case" development identifies possible functions of a distributed or federated system and the data needs behind it.

Design Process

1. Requirements
2. Design
3. Review
4. Comments
5. Refine
6. Forward Draft for Review and Approval

Modeling Advisory Team (Team of Experts)

Application-Neutral Content Model

Encoding (XML)
Conceptual Model

- A conceptual or logical design of the information that preserves the native groupings of the data
- Is implementation- and software-independent to provide a stable base for current and future implementations
- Describes graphically and with narrative the design assumptions and conditions
- Currently expressed using the Unified Modeling Language (UML)

What is UML?

- Unified Modeling Language
- UML is an industry standard language for visualizing, specifying, constructing, and documenting artifacts of a software-intensive system
- Platform-neutral environment for abstract modeling of data and processes
- Adopted as the Conceptual Schema Language for ISO TC 211
UML Diagrams

- Use Case Diagrams
- Class Diagrams
- Object Diagrams
- Component Diagrams
- Deployment Diagrams
- Collaboration Diagrams
- Sequence Diagrams
- Statechart Diagrams
- Activity Diagrams

UML Capabilities

- User Interface
- Integrated Development Environment

Conceptual Models:
- Oracle Table Schema
- Java Program Code
- XML Schema Document

Implementations:
- UML
- Procedures and Structures
- Content and Structures
Start modeling each theme

- Review existing requirements, models, and systems from stakeholders to define scope of applicability
- Identify a team of stakeholder theme experts to include both producers and users of digital geographic data
- Hire modeling and facilitation expertise to work with the experts interactively to build model

**Possible Process Flow**

Requirements, Workflows, Existing data models → Task Groups

- Call for Participation: Experts, Collaborators, Producers, Users
- Modeling Team

Draft Model → Stakeholder Experts

New Model → Reviewers

Final Model → "Standard"

Standard Document Preparation
Use Cases

- Intended to capture processes in a workflow to solve specific problems
- Generalized use cases (archetypes) can be developed to satisfy multiple application requirements
- Define processes and actions (services) required by stakeholders
- Help frame and validate requirements for common data content behind the function or service

Conceptual Data Modeling

- Based on provider and consumer requirements for GIS and mapping, focus on a specific theme of information
- Convene a group of experts with modeling support and have them bring any relevant systems designs or requirements documents
- Strive to build model that supports a common, not universal, set of needs
- Publish model and narrative in a standard
Content Modeling Baseline

- Feature types (classes) included
- Unique feature identifier system
- Basic attributes
- Controlled vocabulary, codes, authorities
- Valid at a range of scales and resolutions
- Multiple representations of same features possible
Going from the Abstract to Implementation

- Conceptual modeling yields the natural organization of the data but not a specific implementation.
- For interoperability in the exchange of data, an agreement on encoding and format is required.
- CASE tools and scripts can convert UML designs into specific implementation schemas.

Implementation Model

- Defines the specific content organization that could be carried in a preferred transfer format.
- Need to express how the information will be structured in a given encoding/packaging format.

<table>
<thead>
<tr>
<th>Implementation Model</th>
<th>Encoding/Packaging Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER Diagrams, XML Schema, ISO Application Schema</td>
<td></td>
</tr>
<tr>
<td>XML, DIGEST VPF, SDTS</td>
<td></td>
</tr>
</tbody>
</table>
Application Schema

- Name for the rules that define the content, relationships, attributes, domain values and constraints in a specific implementation environment
- UML may be converted into XMI to load the model design into a different modeling software
- UML may be converted to an XML/GML Schema Document

Content + Format

- Conceptual Data Model
- Implementation model/schema for Format A
- Data encoding
- Validation
Creating a Standard

A standard facilitates interoperability if it includes both the conceptual data model and one or more implementation annexes with specific guidance for content validation.

Interoperability

System 1  <--- export  --- API  <--- import  --- System 2
          Core
          Framework
          Encoding

System 1  <--- export  --- API  <--- import  --- System 3
          Core
          Framework
          Encoding

System 1  <--- export  --- API  <--- import  --- System 2
          Core
          Framework
          Encoding

System 1  <--- export  --- API  <--- import  --- System 3
          Core
          Framework
          Encoding
Transfer and Transport

- These procedures apply whether you are going to make data available on your network, over the Internet, on CD-ROM, or as printed maps.
- Declaring the means of providing the data within the community is helpful.

| Transfer Protocol | ISO 9660, http, ftp, print |
| Transport         | CD-ROM, tcp-ip, mail        |

Geospatial Services

- Increasingly geospatial data can be accessed in real-time over local area networks and the Internet as if it were local data.
- Multiple organizations can benefit from the data being staged and maintained once and used many times.
- Desktop software and portals can use these services over the Web.
Establish WFS on agreed content nationwide

- Mission System A
  - Web Feature Browser/Client Application
  - GML (XML) translation utilities
  - WFS
  - Mission System B
    - Native Format
    - Public schema
    - Private schema
    - Transformation rules