Standardising a core data set for Crown emergency services and administration in NZ

Dr R. T. Pascoe  
Dept. of Computer Science, University of Canterbury, New Zealand  
richard@cosc.canterbury.ac.nz

R. Murcott  
National Topographic Hydrographic Authority, LINZ, New Zealand  
rmurcott@linz.govt.nz

24 October, 2001

Abstract

Several governmental agencies in New Zealand have recognised a substantial overlap in their use of geospatial data sets for the purposes of emergency services and government administration. To exploit the benefits available by collaborating on the acquisition and maintenance of a single data set, the ISO 19100 series of (draft) standards were employed in the development of an application schema for this common data set.

Here, we describe our insights into defining an application schema that supports the function of locate and verify within the context of emergency services (eg where exactly is this incident) and administration (eg determine the valid electoral district from the voter’s address) by New Zealand government agencies.
New Zealand

Somewhat geographically ‘distant’ from others around the globe . . .

Similar in area to . . .

- California: 157,776 km$^2$
- UK: 243,137 km$^2$
- New Zealand: 266,820 km$^2$
- Japan: 373,049 km$^2$
1. ESA Project

The National Topographic Hydrographic Authority (NTHA) is a regulatory business unit of Land Information New Zealand (LINZ). NTHA leads an inter-agency programme (sponsored by the Officials Committee for Geospatial Information) to improve particular core Crown geospatial data which is used for emergency services responses and other core Crown administrative functions; known as ESA data.

Although this ‘core’ data is used by everyone in many application environments, the data is especially important for emergency services call-centre locate and verify functions which provided one of the more compelling reasons for improving the definition of core data.

1.1. Project goals

The ESA project contributes to this data improvement programme by designing a complete specification of the ESA data set. Our objective was to define the minimum characteristics of the adequate set of spatial data needed for authoritative locate and verify functions. In doing so, the project contributes to defining the minimum investment in core location data to be made by the Crown.

Analysis was not constrained by existing data models or systems, neither was portrayal a consideration. The intent was to produce a data specification that was conceptual and implementation neutral.
1.2. ESA Data

The nation has both digital cadastral and topographical coverage. But contemporary needs of government for location data indicate improvements are required. Less emphasis on portrayal and more on completeness, consistency, and quality of the ‘core’ or ‘fundamental’ data at a national level.

This data includes dwelling numbers or identifiers, road names and network and all types of named geographical features (natural and built) and is to be formed by integrating the core elements that are distilled from a variety of existing sources, such as:

- NZ Police
- NZ File Service
- NZ electoral
- NZ Geographic Board
- Local Territorial Authorities
- NZ Topography
- NZ Cadastre
1.3. Project Members

The immediate project members comprised two groups of people:

**the working group** within which there were representatives of the Crown agencies concerned with emergency services and administration. This group contained a variety of people having quite different roles within the agency they represented. Some were systems analysts, some were system administrators, while others were users of the geospatial data being considered.

**the modelling team** within which there were two experienced geographic data modellers, one project coordinator, and one technical advisor. One modeller came from the commercial sector while the other came from the local government. Neither had much experience with object modelling. The technical advisor had taught UML for 3 years at University level and had been researching geographic data transfers for 12 years.

<table>
<thead>
<tr>
<th>Team Member</th>
<th>Affiliation</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richard Murcott</td>
<td>LINZ, NTHA</td>
<td>Project Coordinator</td>
</tr>
<tr>
<td>Janet Bateson</td>
<td>Critchlow Associates</td>
<td>Principal Data Modeller</td>
</tr>
<tr>
<td>Bryan Clarke</td>
<td>Christchurch City Council</td>
<td>Data Modeller</td>
</tr>
<tr>
<td>Dr Richard Pascoe</td>
<td>University of Canterbury</td>
<td>Advisor, UML modelling &amp; ISO/TC211 standards</td>
</tr>
<tr>
<td>Neil Pullar</td>
<td>Cadastre Limited</td>
<td>Land information &amp; project services</td>
</tr>
</tbody>
</table>
2. ESA Application Schema

The general structure or ‘nuts and bolts’ of the ESA application schema has been distilled to create the following diagram. Please note this diagram is incomplete in many respects.
The class diagram below illustrates the diagrams within the ESA spatial schema.

There are over 40 diagrams in the complete application schema...which is a substantial UML model to review, maintain, and understand!
3. Definition of Features

During the project two levels of abstraction emerged. In one, features having the same attributes and properties were represented by a single class that was associated with an enumeration, each enumerated value representing one type of feature.

In the other, there would be a superclass containing all common properties and associations, and a subclass for each type of feature. The subclasses had no attributes or associations of their own.

In the former abstraction, there was ambiguity as to whether the features Siding, PrivateRailway, and MainLine should be documented in the feature catalogue because each feature lacked a corresponding class definition. In time, use of the first abstraction was replaced by the second in part to allow these features to be described in the feature catalogue and in part because the individual subclasses had distinguishing associations with other classes.

Another abstraction would be to have a generic feature class with attributes in which to store the type of feature, the relevant list of attributes for this feature, and the relevant associations. Given appropriate entries in a feature catalogue, integrity constraints could be expressed using the Object Constraint Language (OCL).
4. Schema Accessibility

During the project there were two concerns related to accessibility of the application schema being designed. The first was that of using UML to model the ESA while the second was distributing the application schema in a form that could be imported into a variety of modelling tools.

4.1. Use of UML

Although modelling geographic data sets using some object modelling paradigm is becoming increasingly more frequent in New Zealand, the majority of modelling is accomplished in terms of Entity Relationship Diagrams (ERDs) and relational databases. Established modellers with the experience necessary for designing national data specifications are more likely to be comfortable with ERDs and relational databases than with UML and object databases or XML.

Adopting the ISO 19100 series of standards for defining the ESA Data Specification extended the skill set of the participating modellers and required members of the working group to become familiar with class diagrams as defined by UML and ISO 19103. In general, this investment slowed project participants accepting that aligning to the ISO 19100 series of standards was advantageous to the overall goals of the project.

Part way through the project, the modelling team began to use object diagrams, another element of the UML, to make their modelling more concrete by relating the class diagrams to real data. Use of object diagrams for communication became an important tool.
4.2. Use by modelling tools

Enabling easy importation of the ISO 19100 series of standards into a modelling tool would have reduced the initial modelling effort and enhanced alignment to the standards. Enabling easy importation of the ESA application schema into a variety of modelling tools would enable an agency to augment this specification to encompass all of their data requirements.

Storing the application schema as an XMI document appears to be the most effective mechanism by which the ESA application schema could be distributed. There are other advantages in terms of developing tools for designing and implementing software to display, manage, and manipulate data conforming to particular application schemas.
5. Other insights in brief

5.1. Testing Application Schemas

During development and review of the ESA application schema use case scenarios and object diagrams were used to verify that the application schema supported the requirements of the Crown agencies. Although no formal use case analysis occurred before or during this project, over 100 use case scenarios were reverse engineered from historical examples of agencies fulfilling the locate and verify function. These scenarios were expressed as one or more queries and the appropriate data to be retrieved for these queries. For each use case scenario, one or more object diagrams were created to show instances of classes defined by the ESA application schema. These instances represented the real data from the scenario. Situations where the schema lacked sufficient structure to represent these data, or where the data could not be retrieved in the manner needed to satisfy the queries associated with the scenario indicated a need to further refine the ESA application schema.

5.2. Business Rules

Although business rules could, to a degree, be documented through considered use of the UML for class diagrams, this subtle use of UML often required explanation to the point where an alternative approach was adopted that project participants could more easily relate. For a similar reason, neither the Object Constraint Language (OCL) nor a functional language were used. Instead, we:

1. expressed business rules as statements in a natural language,
2. labelled these business rules, and
3. displayed the labels on the class diagrams.
6. Summary

In summary:

- There are a variety of approaches to defining features, and we have as yet to determine all circumstances under which one approach is better than another.

- Accessibility of the application schema is of concern in terms of general awareness of UML and of importing schemas (class and package diagrams) into a variety of modelling tools.

- There is a steep learning curve during the initial use of the ISO 19100 series of standards in a ‘real world’ project.

Overall, aligning the ESA data specification to the ISO 19100 series of standards has been of benefit. For example, an external reviewer familiar with only the ISO 19100 series of standards could relatively easily appreciate our application schema. Time will tell as to the ease with which agencies, unfamiliar with the ISO 19100 standards, use the ESA Data Specification.