Joint International Conference on Computing and Decision Making in Civil and Building Engineering
June 14-16, 2006 - Montréal, Canada

TransXML: ESTABLISHING STANDARDS FOR TRANSPORTATION DATA EXCHANGE

Paul Scarponcini, PE, PhD[1]

ABSTRACT

The objectives of the TransXML project are to develop standard, public domain XML schemas for the exchange of transportation data in four initial business areas. The Survey / Roadway Design schema areas address geometric roadway design, area features, and contract pay items. The Transportation Construction / Materials business area covers bid package, project construction status, construction progress, and materials sampling and testing. The Highway Bridge Structures business area focuses on bridge design and analysis. The Transportation Safety area addresses crash records and highway safety information analysis.

The Geography Mark-Up Language (GML) was selected for evaluation as a common framework. Its enterprise-wide feature-based approach proved advantageous in spite of the initial TransXML schema focus on engineering design and construction and GML's origin in the geospatial domain

UML models were created to garner consensus on data content and structure prior to XML schema development. Models were publicly posted for comment and the models were updated accordingly.

GML was generated for eight schema. Sample schemas and instance documents were created and posted for review to a public web site. Sample applications have been developed to demonstrate the applicability of the resulting schemas.

KEY WORDS

transportation, data exchange, XML, schema, GML

INTRODUCTION

In 2004, the Transportation Research Board initiated NCHRP (National Cooperative Highway Reaearch Proram) Project 20-64, *XML Standard for Exchange of Transportation Data (TransXML)* (NCHRP 2004). The objectives of the TransXML project are to develop standard, public domain XML schemas for the exchange of transportation data, as well as a framework for the development, validation, dissemination, and extension of current and future schemas. Though initially targeted at four business areas, the long term vision is for a comprehensive, cohesive set of schemas that span the entire transportation facility life cycle of planning, design, construction, operations, and maintenance.

The research team selected for the project consists of Cambridge Systematics (project lead), Bentley Systems, Inc.(technical lead), Info Tech, Michael Baker, and Charles Campbell. Their respective expertise in transporation safety, design, construction, bridge, and XML matched the focus business and technology areas for the project.

[1] Bentley Systems, Inc., 66 Willowleaf Drive, Littleton, CO 80127, paul.scarponcini@bentley.com

PROCESS

INITIAL ANALYSIS

The project is divided into two phases. Phase I focused on research and planning; Phase II on development and implementation.

Phase I began be examining existing transportation XML schemas, including LandXML, aecXML, JusticeXML, OpenGIS, and AASHTOWare, as well as emerging ITS data standards, such as location referencing were investigated. By leveraging this work, specific existing schemas that could be incorporated into TransXML were identified.

The four suggested initial business areas were then verified: Survey / Roadway Design, Transportation Construction / Materials, Highway Bridge Structures, and Transportation Safety. Within each of these areas, gaps and developmental needs for specific XML schemas were identified and prioritized. A plan was prepared for the development of these schemas by the project team in collaboration with industry, government agencies, and other stakeholders.

A six month interim report on Phase I tasks was submitted and approved by the NCHRP project panel. The proposed scope of work in each business area was published on the project web site (www.transxml.net) for public review and comment. Adjustments were made accordingly.

Ten schema areas were proposed in the four initial business areas. The Survey / Roadway Design schema areas address geometric roadway design, area features, and contract pay items. The Transportation Construction / Materials business area covers bid package, project construction status, construction progress, and materials sampling and testing. The Highway Bridge Structures business area focuses on bridge design and analysis. The Transportation Safety area addresses crash records and highway information safety analysis.

In two of these areas, existing XML schemas would be adopted. For geometric roadway design, a subset of LandXML would be recommended for inclusion. The National Highway Traffic and Safety Administration (NHTSA) schema for crash records would be recommended for crash records. XML schemas were to be developed for all other schema areas.

GML EXPERIMENT

In order to insure an interoperable set of schemas, a common framework was desired for prescribing how the overly flexible XML language would be employed. The Geography Mark-Up Language (GML) (Lake et al. 2004) was selected for evaluation. This appeared to be problematic since many of the initial schemas focused on engineering design and construction, while GML originated in the geospatial domain. Because of the enterprise-wide nature of TransXML, the feature-based approach of GML appeared promising if its apparent complexity could be overcome.

A sample problem was created involving design/construction pay items. A UML Class Diagram was developed to establish the precise requirements for data content and structure. Parallel XML and GML schema definitions and instance documents were then created and compared. GML proved to be advantageous in providing structure and consistency to the schema without undue overhead or geospatial bias. As a result, GML was selected as the framework for TransXML.

UML MODELING

The Unified Modeling Language (UML) (Rumbaugh et al. 1999) was selected as a method of specifying data content and structure for the targeted schemas. Conventions were established for using the language consistently on the project even though half of the teams were using Rational Rose and the other half Enterprise Architect. These followed the conventions established in ISO 19107 (ISO 2001) and are shown in Figure 1. Features represent abstractions of real world phenomenon. These are shown as UML Classes with attributes depicting the information about the features which will be supported by the XML schemas. Associations to other features or classes are shown with role names indicating the role which the associated class instance plays in the association.

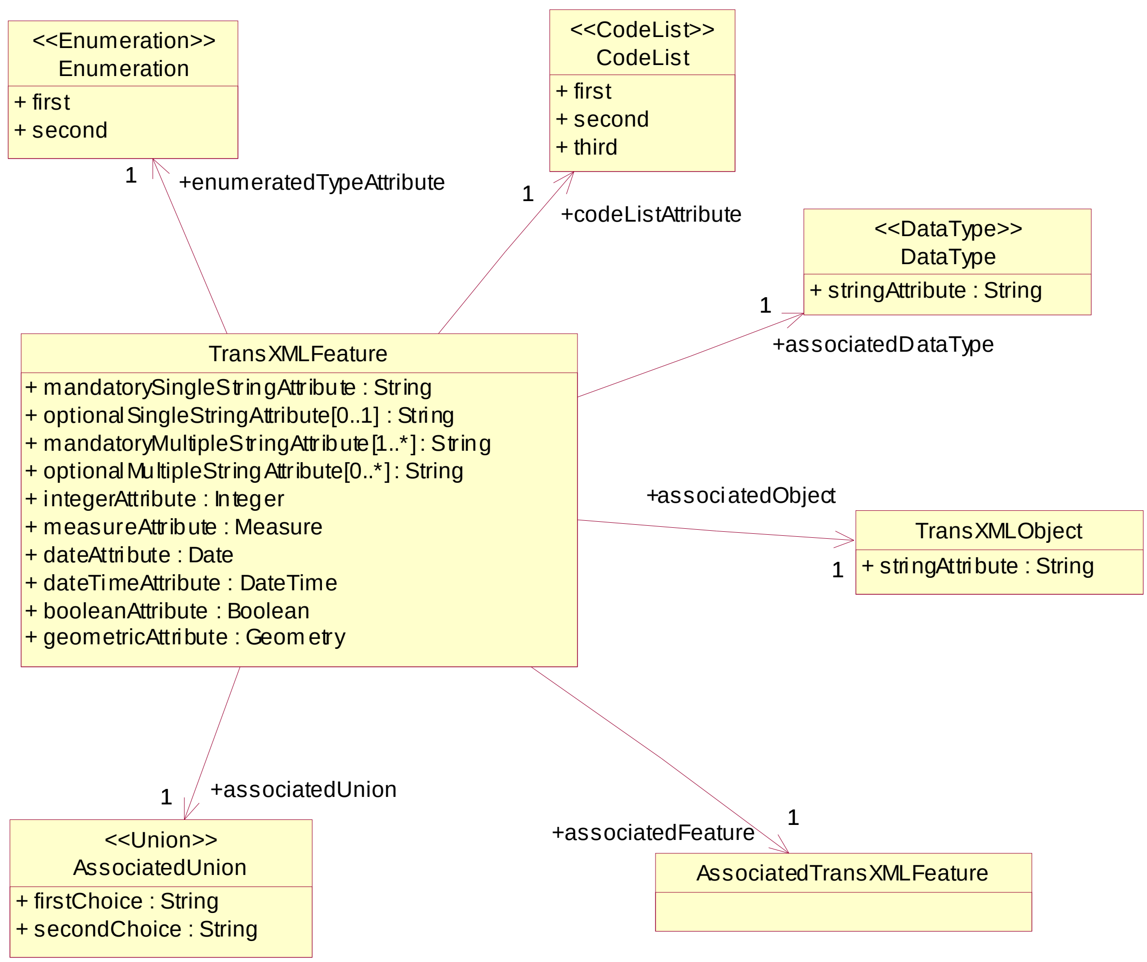

Figure 1: UML Modeling Conventions

UML models were created for each of the schema areas to garner consensus on the data content and structure prior to XML schema development. The models were publicly posted on the TransXML website for comment and were updated accordingly. The individual models were then consolidated into a single, consistent, comprehensive model in Rose, retaining their individual identities through the use of packages. Areas of overlap were harmonized, and commonalities were extracted out into a separate, shareable TransXML Base package. Location referencing was unified across all schemas by adopting the LR package from the recently approved ISO 19133 (ISO 2005) standard for linear referencing.

Dependencies between the resulting packages were identified and are shown in Figure 2.

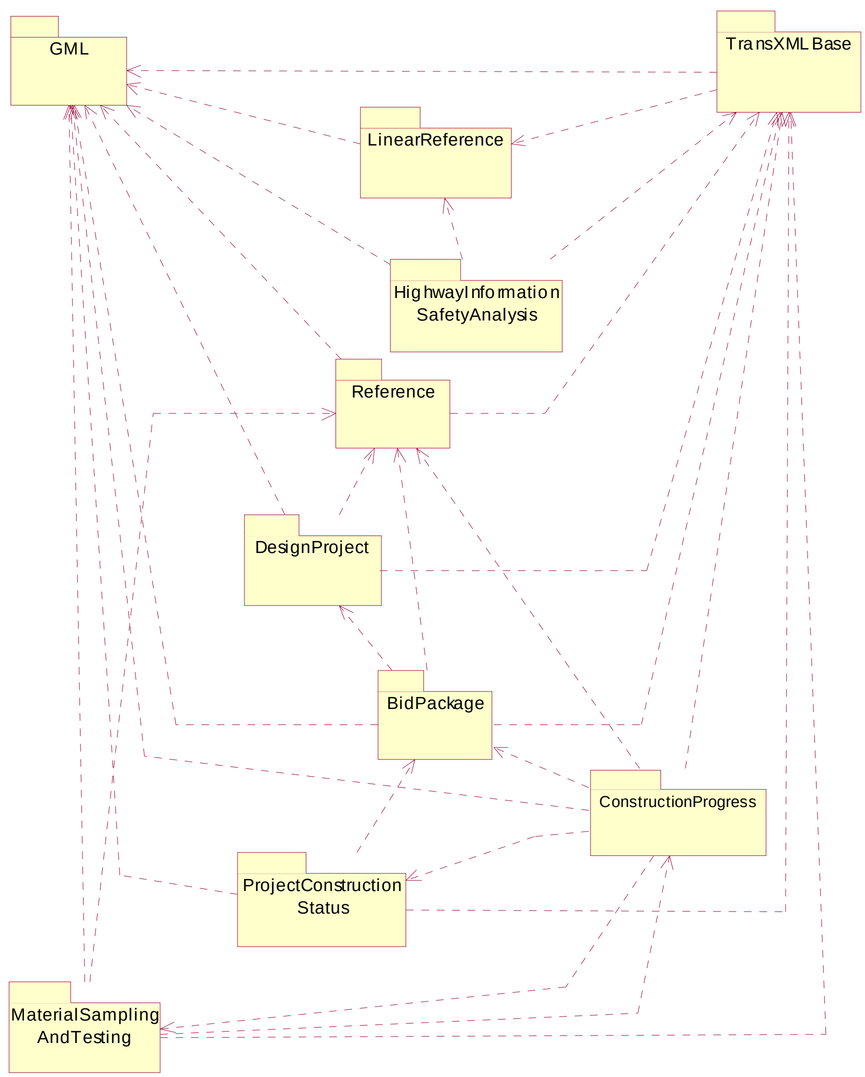

Figure 2: UML Package Dependencies

SCHEMA CREATION

XML schemas were then created from the UML model, one per package. They followed the rules for application schemas contained in the Open Geospatial Consortium, Inc. (OGC) Simple Feature GML (Open Geospatial Consortium, Inc. 2005) specification currently under development. The use of include statements accommodates the inter-package (schema) dependencies in Figure 2. The GML package shown is implemented as the OGC GML 3.1.1 schema.

Classes are specified as XML elements. If a class represents a feature, it is specified as an extension of the GML Feature from GML 3.1.1. GML Features have a mandatory id and an optional name, description and boundedBy envelope. Class attributes are specified as sub-elements under the class element and will contain the attribute value. They are easily differentiated from elements which represent classes by the leading lower case letter in the attribute name. A wide assortment of geometries have already been defined in GML and are available as GML Geometric Property types. Associated classes are represented first by a class element sub-element with the role name which then refers to an instance of the associated class. Part of the XML schema resulting from Figure 1 is shown below:

```
<xs:element name="TransXMLFeature" type="TransXMLFeatureType"
substitutionGroup="gml:_Feature">
</xs:element>
<xs:complexType name="TransXMLFeatureType">
    <xs:complexContent>
        <xs:extension base="gml:AbstractFeatureType">
            <xs:sequence>
                <xs:element name="mandatorySingleStringAttribute"
                 type="xs:string"/>
                …
                <xs:element name="geometricAttribute"
                 type="GMLGeometricPropertyType"/>
                …
                <xs:element name="associatedFeature"
                 type="AssociatedTransXMLFeaturePropertyType"/>
            </xs:sequence>
        </xs:extension>
    </xs:complexContent>
</xs:complexType>

<xs:complexType name="AssociatedTransXMLFeaturePropertyType">
    <xs:sequence>
        <xs:element name="AssociatedTransXMLFeature" minOccurs="0"/>
    </xs:sequence>
    <xs:attributeGroup ref="gml:AssociationAttributeGroup"/>
</xs:complexType>

<xs:element name="AssociatedTransXMLFeature"
 type="AssociatedTransXMLFeatureType" substitutionGroup="gml:_Feature"/>
```

```
<xs:complexType name="AssociatedTransXMLFeatureType">
    <xs:complexContent>
        <xs:extension base="gml:AbstractFeatureType">
            <xs:sequence/>
        </xs:extension>
    </xs:complexContent>
</xs:complexType>
```

XML schemas were posted on the TransXML website for public review, along with sample instance documents validated against the schemas by XMLSpy. Comments received have resulted in updates to the schemas.

SAMPLE APPLICATIONS

For each schema, a sample application has been created for demonstration purposes. These have been made available on the TransXML website.

SCHEMAS

SURVEY / ROADWAY DESIGN

Geometric Roadway Design

The most significant information exchanged during design and carried forward into subsequent life cycle phases is the geometric design of the roadway. Beginning with preliminary design, the roadway design evolves through the refinement of the horizontal and vertical alignments and the addition of pavement section, superelevation, cross sections, and geometrics. This is done in conjunction with other disciplines which impact or are impacted by the design, including but not limited to right of way, drainage, utility, hydraulic, traffic, environmental, and aesthetic concerns. XML can provide a method of sharing the roadway design as it evolves during the design process. It can also provide the basis for capturing as-built information as the roadway is actually constructed. This information can then be utilized during the maintenance and operation of the roadway.

The Geometric Roadway Design (GRD) schema defines the roadway design geometry, including the horizontal and vertical alignment, cross sections, super elevations, and geometric information. A subset of LandXML (Crews et al. 2002) already addresses the GRD area so a new TransXML schema was not developed in this area. Instead, project work includes development of a UML model and recommendations to LandXML.org on areas where LandXML might be improved.

Area Features

Instead of starting with a clean sheet of paper, it would be advantageous if designers could capitalize on the information collected during the planning phase. Much of this information is available with GIS software which is often incompatible with engineering design software.

The Area Features (AF) schema includes information about area features such as environmental areas, soils, wetlands, land use, flood plains, site improvement areas, right of way, and cadastral information. Of primary concern to the designer is the location of these areas with respect to the roadway project being designed. This information is typically held in a GIS, but it would be helpful to include this information as a backdrop to a CAD design drawing.

The AF schema relies heavily upon the existing GML geometry schemas. This enables the transfer of area features from GIS to CAD, as is demonstrated in the accompanying sample application.

Contract Pay Items

Contract pay items span the design and construction business areas. Pay items are the basis for estimating the cost of the project, comparing alternative design solutions, obtaining a contractor to construct the roadway, assessing the progress of the construction, providing the basis for partial (progress) payments during construction, billing the work completed, and potentially feeding maintenance and operation systems such as roadway inventory and asset management.

Though contract pay items predominate in the construction phase, they begin during design. The design engineer determines what pay items make up the project and determines how much of each pay item will be required to complete the job. This is based on a standard set of contract pay items with pre-defined units of measurement. The pay items and their quantities are then passed to the estimator to predict the cost of the project. Often the items and quantities are included in the contract documents for the project.

The Contract Pay Items (CPI) schema is the first step in addressing contract pay items. It includes available pay items (id, type, description, units of measure) and funding sources contained in the Reference Package as well as a list of contract pay items and their respective funding sources on a given project contained in the Design Project Package. Additional information added includes quantity and pricing for each pay item for a given estimate of a specific project or proposal.

It is intended that CPI will replace aecXML which currently has limited usage. CPI is integrated with the other pay item schemas in the Construction/Materials business area.

TRANSPORTATION CONSTRUCTION / MATERIALS

Bid Package

In preparation for letting a proposal, transportation agencies publish a proposal bid package. Contractors use this information to prepare their bids. Subcontractors and suppliers use it to identify potential business opportunities and submit quotes to primary contractors. In the event that project changes occur after publication but prior to the letting, the agency publishes these changes in proposal amendments.

The bid package includes general proposal and pay item information. Proposal amendments can include changes to any such element. The Bid Package (BP) schema builds on the CPI schema to support the letting process requirements for publishing bid packages. Additional proposal information includes the letting location and date, vendor qualification requirements, contract time information, and the additional attributes required for amendments.

Thousands of transportation proposal bid packages are published each year, often in paper form, to a community of tens of thousands of contractors, subcontractors, and suppliers. A standard transportation Bid Package XML schema will enable agencies to publish bid packages electronically in a standard form that can be directly loaded into bid preparation systems. As a result, information flows will be streamlined and redundant data entry and the associated opportunity for error will be substantially reduced.

Project Construction Status

External stakeholders such as the general public, elected officials, oversight or regulatory agencies, and other institutions such as utilities and railroads require or can benefit from access to timely information about the status of a transportation construction project. This information is managed within the agency in their construction management system and is provided to different stakeholders in different forms.

The information being exchanged includes project description, location, and fiscal, schedule and progress information, including milestone dates and those affecting traffic. The Project Construction Status (PCS) XML schema will enable the automated publication of transportation project status information in a standard format that can be presented in a variety of forms appropriate for the individual target audiences.

Construction Progress and Materials Sampling and Testing

A broad range of field devices are used to measure construction progress and track material use, sampling, and testing. Various elements of this information are communicated frequently among field, project office, test lab, and central office personnel throughout the construction project. A standard XML schema would enable integration of the diverse data collection and data management systems utilized to track this information, thereby streamlining information flows and reducing the opportunity for error. TransXML can provide a standard form for communication between systems that utilize the physical project view (items installed, materials sampled and tested) and the construction progress payment systems that utilize the business project view.

The information being exchanged includes pay item descriptive information, partial quantities placed, placement locations, materials samples collected, the field tests performed, and the outcome of those tests. During UML modeling, it was decided to split this area into two separate packages, Construction Progress (CP) and Materials Sampling and Testing (MST).

HIGHWAY BRIDGE STRUCTURES

Bridge Design and Analysis

The TransXML schema for transfer of Bridge Design and Analysis (BDA) should provide adequate information to allow the transfer of both bridge description information and analysis results. Uses for the transfer of bridge information include comparative analysis of the same bridge using multiple bridge analysis processes and the rating of bridges in the permit routing process. These uses span the design phase and maintenance phase of the bridge life cycle.

The AASHTOWare Virtis/Opis® (AASHTO 2005) bridge domain was considered to be a good starting point for the development of the BDA schema. It. provides a comprehensive description for the purpose of analyzing many bridge types. Previously, the Virtis/Opis software provided a reporting tool that produced an XML representation of the bridge domain. A TransXML schema definition was developed for validation purposes.

TRANSPORTATION SAFETY

Crash Records

The Crash Records (CR) schema describes the information recorded at the time of the crash (not the information which may be linked to the report after the crash). During the modeling phase, it was determined that a parallel effort by NHTSA had progressed significantly enough to obviate the need for continued work as part of this project. Instead, the NHTSA schema has been adopted for TransXML.

Highway Information Safety Analysis

The Highway Information Safety Analysis (HISA) schema describes safety-related highway inventory items that relate to a specific incident location. Though the FGDC Framework Data Content Standard (FGDC 2005) was considered as a candidate for the basis of the roadway representation, a simpler approach was adopted. The road location where the crash occurred (road segment, intersection, or ramp) is assumed to be homogeneous in its attributes, obviating the need for the more complex linear event representation.

FUTURE WORK

The XML schemas developed during this two-year project are intended to be only the starting point for TransXML. By establishing a procedure on how to develop additional schemas as well as rigid guidelines on their format and structure, it is anticipated that TransXML will grow to cover the entire transportation facility life cycle process.

CONCLUSIONS

XML schemas have been developed in eight schema areas for survey/roadway design, construction/materials, highway bridge structures, and transportation safety using a consistent methodology and framework. Adoption of the Unified Modeling Language to obtain consensus on data content and structure and the Geography Markup Language to provide a consistent framework and format for the schemas has so far proven to be advantageous. A side benefit was the development of a GML encoding for the ISO 19133 linear referencing standard. Sample applications have been developed in all schema areas.

Though a significant number of people have signed up on the website, minimal comments were received on the models and schemas which were posted there. Hopefully, seeing the schemas in the context of the sample applications will spark additional reactions.

ACKNOWLEDGMENTS

The author acknowledges the funding support by the NCHRP, the support of AASHTO, a major stakeholder in the project, the dedication of the review panel, public comment via the website, and the information contributed by other project team members.

REFERENCES

AASHTO 2005, *AASHTOWare Virtis/Opis*, The American Association of State Highway and Transportation Officials, 444 North Capitol Street, N.W. Suite 249 Washington, DC.

Crews, N., Hall, E., and Rebolj, D. 2002). *LandXML Schema Version 1.0 Reference*. http://www.landxml.org.

FGDC 2005. *Information Technology – Geographic Information Framework Data Content Standard*, Federal Geographic Data Committee, Reston VA.

ISO (2001). *ISO 19107*, *Geographic information — Spatial schema*, International Organization for Standardization, Geneva, Switzerland.

ISO (2005). *ISO 19107*, *Geographic information — Location based services – Tracking and Navigation*, International Organization for Standardization, Geneva, Switzerland.

Lake, R., Burggraf, D. S., Trninic, M. and Rae, L. (2004). *Geography Mark-Up Language*, John Wiley & Sons, Inc., San Francisco, CA.

NCHRP (2004). *Research Project Statement*, Project 20-64X, FY 2004, XML Standard for Exchange of Transportation Data (TransXML), http://www4.trb.org/trb/crp.nsf.

Open Geospatial Consortium, Inc. 2005). *GML simple features profile*, OGC 05-033.

Rumbaugh, J., Jacobson, I., and Booch, G. (1999). *The Unified Modeling Language Reference Manual*, Addison-Wesley, Reading, MA, 550 pp.