Moving Towards Regional Public Transportation Systems through Integration and Data Quality Standards: Going Beyond the Transit Data Dictionary

By
Paula Okunieff
Consensus Systems Technologies Corp.
Context for Standards

- Relationship between architecture & standards
- Standards work together
- Interoperable and stackable
  - No one standard can meet all requirements for a system
  - Want standards to be specified in layers (mix and match)
Mapping Systems to Architecture

Centers

- WMATA Core Data
- Local TA #n Core Data
- WMATA RTBI Data Stream
- Local TA #n RTBI Data Stream
- RITIS
- DDOT
- VDOT
- MDSHA
- CapCOM / CapWIN
- Claris
- WMATA Incident and other RT
- Local TA #n Incident and other RT

Data Tier (note 4)

- Core Data (schedules, patterns, bus stop inventory, transfers, fare, transportation network & landmarks, etc.)
- Real time bus information caching (level 1 or 2)
- Other Real Time Information
- Transit incidents and other real time status or detour information

WMATA RTBI Service

- RTBI Services
- RTBI Channel services
- RTBI DMS Application

Network Communications Performance
Note 1: As needed
Note 2: Real time performance and guaranteed Quality of Service (QOS)
Note 3: Near real time performance and high QOS

Data Storage
Note 4: Data Tier may reside at each local transit agency or be centralized at WMATA

Travelers

ConSysTec
Web Services Standards Stack

- Adopt a flexible standards stack based on Industry Standards
- Define a Web Services Standard “Profile”

<table>
<thead>
<tr>
<th>Category</th>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry and Business Semantics</td>
<td>TCIP*, Spatial Content Framework Standards Part 7d</td>
</tr>
<tr>
<td>Service Presentation / Event Delivery</td>
<td>Portal Standards</td>
</tr>
<tr>
<td>Service Orchestration</td>
<td>WS Choreography</td>
</tr>
<tr>
<td>Security</td>
<td>WS Security, SSL</td>
</tr>
<tr>
<td>Service Discovery</td>
<td>UDDI</td>
</tr>
<tr>
<td>Service Description</td>
<td>WSDL, WS-Policy</td>
</tr>
<tr>
<td>Message Processing</td>
<td>SOAP, NTCIP 2306 (profile of SOAP 1.1)</td>
</tr>
<tr>
<td>Data Extraction</td>
<td>XQuery, SQL/XML</td>
</tr>
<tr>
<td>Data Semantics</td>
<td>XML Schema</td>
</tr>
<tr>
<td>Data Syntax</td>
<td>XML, XPATH</td>
</tr>
<tr>
<td>Transformations, Encoding &amp; Rendering</td>
<td>WAP, WML, Voice XML, XSLT</td>
</tr>
<tr>
<td>Transport</td>
<td>HTTP</td>
</tr>
</tbody>
</table>
## Partitioning the Standards “Space”

### Viewpoint

<table>
<thead>
<tr>
<th>Levels of Abstraction</th>
<th>Service Invocation</th>
<th>Information Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Implementation specifications:</strong> Recipes, Practice</td>
<td>Interface:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Web Service specification (SOAP/WSDL)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Datex</td>
<td></td>
</tr>
<tr>
<td><strong>Abstract model:</strong> Guidance for Design</td>
<td>Behavior:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exchange Patterns</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Query</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Publish/Subscribe</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Broadcast</td>
<td></td>
</tr>
<tr>
<td><strong>How</strong></td>
<td><strong>What</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Encoding:</strong></td>
<td><strong>Content:</strong></td>
<td></td>
</tr>
<tr>
<td>• Data Structures</td>
<td>• Abstract Reference Model</td>
<td></td>
</tr>
<tr>
<td>• XML Schema</td>
<td>• Message Content, Data Dictionary</td>
<td></td>
</tr>
<tr>
<td>• ASN.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## US ITS Standards in the Standards Space

<table>
<thead>
<tr>
<th></th>
<th>Service Invocation</th>
<th>Information Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Implementation specifications</strong></td>
<td><strong>Interface:</strong> C2C NTCIP 2306 (SOAP/WSDL) TCIP Volume 4</td>
<td><strong>Encoding:</strong> TCIP Volume 3 (XML Schema) All other ITS standards GML</td>
</tr>
<tr>
<td><strong>Abstract model</strong></td>
<td><strong>Behavior:</strong> TCIP Volume 1, Chpt 7 C2C (exchange patterns)</td>
<td><strong>Content:</strong> TCIP Volumes 1* &amp; 2 All other ITS standards IEEE Data Dictionary/Msg Template</td>
</tr>
</tbody>
</table>

*ConSysTec*
TCIP Dialog Functional Areas
(Business to Business and some Business to Consumer)

Dialog Patterns: Broadcast, Request/Response, Publish/Subscribe, etc.
Conformance: Requirement (PRL)/Implementation (PICS) Specifications Formats

Encoding Alternatives
Where are the European Standards?

<table>
<thead>
<tr>
<th>Implementation specifications</th>
<th><strong>Service Invocation</strong></th>
<th>Information Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interface:</strong> SIRI (SOAP/WSDL)</td>
<td><strong>Encoding:</strong> SIRI (XML Schema)</td>
<td></td>
</tr>
<tr>
<td><strong>Behavior:</strong> SIRI (exchange patterns)</td>
<td><strong>Content:</strong> TransModel (TPEG, Datex2)</td>
<td></td>
</tr>
</tbody>
</table>
HOW DO STANDARDS APPLY TO MULTIPLE AGENCY DEPLOYMENTS?
Features and Characteristics

For a Single Agency

- Naming/Indexing Policies
- Data Development Env.
- Valid?
- Data Production Env.
- Trip Planner
- Other data sets
  - Transport. Network
  - Landmarks, Addresses, Walking/obstacles

ConSysTec
Regional Integration Features

Agency 1..n Service Data Production

Naming/Indexing Policies

Valid?

Regional Service Data Development

Valid?

Regional Service Data Production

Regional Transit Features
stops, stations, transfers

Common representation and relationships

Other data sets
Transport. Network
Landmarks, Addresses, Walking/obstacles

Trip Planner
New York Penn Station (only for Long Island Rail Road)

- Amtrak and NJ TRANSIT Trains
- Private Buses to New Jersey, DC & Boston
- Private Carriers to Airports
Approach to Transit Data Integration in the US
Seattle
Integrated Corridor Management System Model

Acquire & Store Information
- Data Acquisition and Persistent Data Store
  - Log
  - Archived
  - Central Information Repository

Process Information
- Traveler Information System Engine
- Corridor Management Decision Support System
- Incident Reporting
- Construction Management and Event Reporting System

Publish Information
- Traveler Information -- Plan, RT, Alerts
  - Internet
    - Firewall
    - Permissions
      - Web & Location Services
      - Channel Services (IVR)
      - Other Services/Processes

Client Control (for NTCIP Devices)

Web Services to DMS/VMS, CCTV, Signal Control

Agency Access

- CCTV (TrafficBuster)
- Hazard Advisory Radio
- WSDOT Freeway Traffic Management Center
  - Ramp Metering
  - Highway Detectors
  - DMS/VMS
  - License Plate Readers
- Seattle Arterial Traffic Management Center
  - Arterial Detectors
  - Traffic Signal Controllers
  - DMSA/VMS
- TNET Regional GIS and Asset Data
- Event & Construction Reporting Tools
- KCM Transit Operations Center [LRT, Bus, RapidRide]
Camden County Transportation Management Coordination Center

- Maximize funding sources for Community Mobility
- Optimize schedules across providers
- Enhance customer satisfaction
Data Sharing Concept for NY

Boxes surrounding processes indicate current (blue) and future (green) regional applications or coordinated schedule data activities.
What is the architecture concept?

Schedule Data

- LIB SDP file
- NYCT SDP file
- MNR SDP file

TSDEA

- Registry
- Catalog/Metadata
- Message Editor

Native to SDP Data Scripts / Validation and Integrity Checker

Dynamic TT Gen.

- API_DTG (TCIP)
- API_TTPub

Timetable Publisher

- SDP
- GTFS

TRIPS 123

Google Transit Schedules
Case Study on Data Quality

- Example from NYSDOT Transit Schedule Data Exchange Architecture (TSDEA) Project
  - Objective: Collect and store multi-agency schedule data that supports multiple downstream applications
  - Developed **Schedule Data Profile** (SDP) to ensure consistent transit data quality for use in regional transportation systems.
What is Data Quality?

- “Fitness-for-use of a particular piece of data for your application”
  - From Transit Location Referencing Guidebook

- What if the data will be used for multiple downstream applications and services?
  - Data requirements “fit for use” for all applications
How is Quality Achieved?

- Quality Checking
- Role of Custodian and Users
- Requirement description of downstream uses including defining data business rules, identity, persistence, regional authorities, etc.
Data Quality “Owner”

Option 1

Owner Transit Agency

Native Data

Check and format

Regional Application

Regional Trip Planner

Option 2

Owner - Transit Agency

Native Data

Check and format

Regional App.

Regional Trip Planner
Quality Elements

- **Semantics**
  - Unambiguous data definitions
  - Relationship among concepts
  - Business Rules
- **Syntax/format**
- **Accuracy, Lineage, Currentness, Logical Consistency, Complete**
Concept to Design
Going beyond the Data Dictionary

- **Semantics:** “data concept meaning”
  - Model relationship of concepts,
  - Describe business rules for using data, and
  - Interpret equivalent (“Isomorphic”) relationships.

- **Format Requirements**
  - Relational database
  - Exchange while minimizing quality checking
  - Facilitate exchange of large files
Design to Implementation

- Various Implementation Methods
  - Physical Database Query capability (SQL)
  - XML Schema (e.g., SOAP & XML Document)
  - Comma Delimited (CSV files)

- Quality checks
  - Levels of Quality Checks
  - Syntax
  - Referential Integrity (uniqueness & logical consistency)
  - Dates
  - Other business rules related to data concepts (e.g., day type, pattern, facility, location)
TSDEA Levels of Quality Checking

- **Level 1: Registration** –
  - Ensures that the file contains a well formed and complete SDP XML document.

- **Level 2: Authorization** –
  - The file content has passed quality checks that are based on business rules and requirements. The file content is deemed logically consistent (semantically and logically accurate).

- **Level 3: Regionally Consistent** –
  - File content has passed tests to ensure consistency with regional naming conventions and representations.
Paula Okunieff
Consensus Systems Technologies Corp.
43 Jamaica Street
Boston, MA 02130 USA

Email: paula.okunieff@consystec.com
Telephone: +1 617-983-3364