Topic Research: Public Transport Public Transport Information Services; ITS Data Management; Data Fusion and Data Mining; and Architecture and Standards

Building Public Transport Information Services for Downstate New York Travelers and Operators

Submitted to:

14th World Congress on Intelligent Transport Systems

Authors:

Paula Okunieff
Consensus Systems Technologies
43 Jamaica Street
Boston, MA 02130
paula.okunieff@consystec.com
Tel: +1 617-983-3364

James Davis
New York State Department of Transportation
50 Wolf Road
Albany, New York 12232
jdavis@dot.state.ny.us
Tel: +1 518-457-8355

ABSTRACT

Two nearly universal obstacles to developing and sustaining public transport service information are the complexity and lack of standard definitions, as well as the organization and exchange of data between applications that produce and those that use schedules as their core functionality. Regional dataset integration remains a costly and resource-intensive activity regardless of data model and dictionaries promulgated by regional and standards organizations.

This paper describes a project initiated by the New York State Department of Transportation to develop a prototype portal framework that manages and enables exchange of schedule data among agencies and effectively uses this multi-provider information as the basis for downstream applications. This paper describes the approach used to collaboratively define a cost-effective framework, tools and guidelines for integrating public transport schedule data.

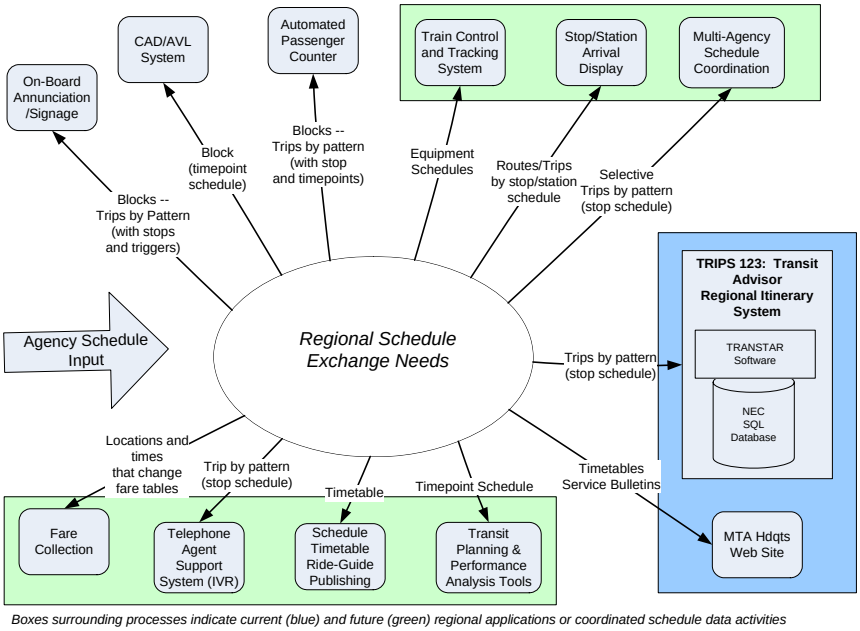
INTRODUCTION

PROJECT VISION

Travelers have many options to choose from in deciding how to reach their destinations. With the advent of the Internet and related technologies, instant, comprehensive, readily available information about products and services has increasingly become a customer expectation in consumer decisions. Transit has ventured into this new world with sophisticated websites, custom ride guides, and next-bus/train arrival systems. There have been many successful, and favorably received, examples of this new breed of customer information. However, two nearly universal obstacles to developing and sustaining these initiatives are a) the complexity and lack of standardization in the definition, and b) the difficulty in organization and exchange of schedule data between applications that produce schedules and those that use schedules as a core element of their function.

In initiating the project to develop a "proof-of-concept" *Transit Schedule Data Exchange Architecture* (TSDEA) portal and *Schedule Data Profile* (SDP), the New York State Department of Transportation (NYSDOT) sought to provide an efficient framework for managing and exchanging schedule data among agencies and effectively using this multi-provider information as the basis for downstream applications.

The project's objectives include to "provide a fully integrated regional transit schedule information exchange framework that enables high quality, reliable service information to travelers, transit agency staff and other regional/ statewide stakeholders who can benefit from transit schedule information." The project outcomes are the specifications for a well understood, unambiguously defined schedule data reference model, and a framework or data portal to validate, organize and distribute schedule and related datasets. Figure 1 summarizes the complexity and requirements for schedule data exchange for internal and regional distribution.



Boxes surrounding processes indicate current (blue) and future (green) regional applications or coordinated schedule data activities

Initiated in April 2005, the project included two technical phases. In the first phase, a concept of operations and requirements analysis was developed. Phase 2, initiated in July 2006, consisted of demonstrating the validity of the SDP requirements through translating a diverse set of schedule data to SDP documents, implementing integrity and validity checks,

automating a registration process, and using these SDP documents in downstream applications. The project demonstration phase was completed in August 2007.

DESCRIPTION OF ENVIRONMENT

The "Downstate" region of New York State, including parts of New Jersey and Connecticut is the focus of this initial phase of the TSDEA prototype because of the significant density and diversity of transit services. This remarkable range of service options has the greatest opportunity and need for integration and coordination. The challenge of multi-agency schedule information sharing was demonstrated in the development of the TRIPS123 Traveler Information system, in its objective to provide one stop shopping for itinerary information across three states, 29 counties and, at last count, 52 transit carriers, ranging from municipal bus operators to commuter rail to ferries to the New York City Transit Subway. The ongoing support and challenge of the TRIPS123 project was how to improve data validity, increase the amount of data from new transit providers and how to support multi-versions and temporary schedules in the application. The effort to achieve these goals helped drive the TSDEA project. To this end, the TSDEA project focused on the transit carriers active in the 7 county and five New York City boroughs of Metropolitan Transportation Commuter District (the five boroughs of New York City, Orange, Rockland, Dutchess, Putnam, Westchester, Nassau and Suffolk Counties). In addition, the prototype implementation includes participation by New Jersey Transit, Connecticut Transit, and several private carriers.

There were 2.4 billion passenger trips within this region in 2003. Travelers using more than one carrier for a given journey are a significant and growing segment of this total. Suburban bus carriers routinely provide feeder service to either the Long Island Rail Road or Metro North Railroad. Westchester County Bee Line Bus has estimated that approximately 30 percent of its daily riders transfer to either Metro North or New York City Transit service. MTA Long Island Bus estimated an even greater transfer rate, in excess of 60 percent to Long Island Rail Road or New York City Transit services. Rockland and Dutchess Counties both operate bus services that feed Metro North Railroad that carry between 13-15 percent of the County's total ridership. When placed in the context that a substantial percentage of regional travelers destined for Manhattan transfer at the Penn Station, Grand Central Terminal or the PANYNJ Bus terminal to the subway for the final leg of their journey, the value of multi-agency schedule coordination and integrated schedule and facilities information becomes apparent.

PROJECT RESULTS

Safety, security, economic competitiveness and environmental conditions are all enhanced by the efficiencies a common information base can provide. Although still in its early phase, efficiencies were achieved through reduction of duplicative effort, economies of scale in schedule data distribution and better information supporting downstream applications such as TRIPS123. Agencies now have a test bed to develop tools to coordinate schedules and standardize interfaces for Carriers that want to be integrated into the area's traveler information service. Independent, traveler information service providers are able to request and acquire integrated transit data for the region, consequently allowing them to provide multi-agency trip plans to their customers.

Perhaps the greatest value from the project is the provision of validated, integrated data. The availability of consistent and quality-checked rail, bus and ferry schedule information in electronic format enables the existing carriers and Information Service Provider entrepreneurs to develop useful new services to present in various ways to operators and travelers.

APPROACH TO A DATA EXCHANGE ARCHITECTURE

SYSTEMS ENGINEERING APPROACH

A collaborative, consensus-based approach following a system engineering methodology was used in developing the TSDEA and SDP. Through a series of Technical Working Group (TWG) meetings and interviews, stakeholders defined their procedures and processes for generating and using schedule information. The TWG contributed to defining a Concept of Operations that included stakeholder roles, operational scenarios and high level requirements for the TSDEA. More detailed data requirements were defined to meet several high priority existing and proposed downstream applications. These applications are encapsulated in three Use Cases including:

- Trip planning,
- Dynamic generation and presentation of public timetables, and
- "Ad hoc" and pre-planned special event scheduling.

The results of the Concept of Operations and Use Cases drove the requirements for the SDP reference data model and XML schema, and Transit Schedule Data Exchange Architecture.

SCHEDULE DATA PROFILE (SDP)

The SDP is a conceptual data model that is mapped to entity-relationship and hierarchical models (implemented as an XML schema). In addition, key metadata structures are defined in the SDP to capture nuances, extensions and constraints of each agency dataset. This agency metadata will enable the TSDEA to implement a comprehensive web service architecture complete with schedule data discovery, catalog and brokering functions.

Some key innovations found in the SDP include:

- An approach to describing *location*, location references and locations related to other locations using a "location" table.
- A description of *Transit Facility* that enables a representation of a complex multi-agency, multi-modal facility while allowing data "ownership" of parts of the transit facility such as NYCT subway stations, Amtrak, Long Island Railroad, and New Jersey Transit stations, and countless public and private carrier stops along the facility, in addition to their shared walkways, stairs, entrances, and amenities.
- A description of "recommended" transfer points and connection instructions. The transfers may be impacted by dynamic conditions such as walking through a rail car to make a connection between two trains.
- A structure for updating only one route within an existing schedule.
- A provision for defining an "ad hoc" schedule for
    - Unplanned rerouting due to service control/recovery

    - Planned rerouting and staging services due to known events (e.g., special events and construction)
- Specifications for validating semantic, format and referential integrity requirements.

In addition to the SDP Requirements document that describes guidance on how to implement key provisions of the SDP, during the demonstration phase, guidance on how to map existing transit schedule and related data was developed. The recommendations were derived from mapping existing schedule data from a range of transit providers within the region to the SDP. For example, nine examples of transit facilities/stops of varying complexity were documented, in addition to several transit network and service descriptions, location referencing methods, route, schedule versioning and revisions definitions.

Several of the lessons learned were deployed as a wizard to support the translation of native data to the SDP as well as guidebooks discussing SDP requirements, translations, procurement, and programming issues. Transit agencies worked with NYSDOT, and its consultant (the author) to develop scripts that translated native data sets to the SDP. This process exercised the SDP to ensure that it meets the key upstream data requirements as well as feeds the TSDEA schedule data registration, catalog and storage processes.

TRANSIT DATA EXCHANGE ARCHITECTURE (TSDEA)

The vision of the TSDEA is to provide an online data repository for regional schedule and related data. The role of the repository is to register, catalog, store, integrate, archive, broker requests for, and distribute schedule data from authenticated data providers and consumer applications. A portal environment that supports registration, discovery, cataloging and brokering through web services provides the infrastructure needed to meet these goals.

A prototype infrastructure was implemented using two open source transit applications to test varied schedule data sets implemented in the consistent web services framework. The Dynamic Timetable Generator using a standards based TCIP interface [1] and a tool developed by TriMet, the Timetable Publisher [2], were used to show how web services may be implemented using a common data management and semantic framework. The results were instructive. Rail and bus data generated by different scheduling software, stored in different formats provided similar results. The XML interface provided the first level of validation, the process tools to translate native data to SDP verified key referential integrity constraints. The SDP model supported simple and complex descriptions that could be leveraged by small and large operators.

BENEFITS

Several key benefits were achieved from this demonstration. They include:

- One reference data model enables multiple operators to leverage a single application
- Framework is scaleable to multiple public transport service providers varying over a wide range of characteristics (e.g., mode, size, complexity of transit facilities and location resources)
- Provides a test bed to develop tools for integrated services and other planning needs
- Open source software and web services enable modular applications that may evolve over time based on data collected and functionality desired

- Scalability of infrastructure and applications for larger regions potentially including statewide/multi-state deployment

In particular, the conceptual data model, business rules for applying them, and metadata descriptions are key to ensuring consistency of varying agency datasets and the ability to integrate them across the region.

ACKNOWLEDGEMENT

The authors would like to thank the participants of the Regional Stakeholder Technical Working Group who contributed to the success of the project.

The project was funded by the NYSDOT under the "Technical Support and Strategic Plan Development Services for NYSDOT Statewide ITS Program" contract.

REFERENCES

[1] TRB. *Transit IDEA Report*. January 2005. Available at www.trb.org/publications/sp/transit-idea_report_jan2005.pdf

[2] Purcell, Frank, "Timetable Publisher". GOSCON, 12 October 2006. Available at http://oregonstate.edu/conferences/presentationsGOSCON/GOSCON_Purcell.pdf