Expanding the Google Transit Feed Specification to Support Operations and Planning

November, 2011

Final Report

Contract Number: **BDK85 977-15**

DISCLAIMER

The opinions, findings, and conclusions expressed in this publication are those of the author, who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the views or policies of the Florida Department of Transportation or the Research and Special Programs Administration.   This report does not constitute a standard, specification, or regulation.

The report is prepared in cooperation with the State of Florida Department of Transportation and the U. S. Department of Transportation.

Expanding the Google Transit Feed Specification to Support Operations and Planning

Final Report

Prepared for
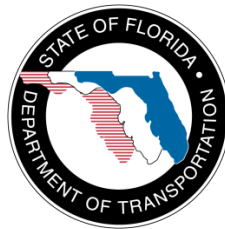
State of Florida Department of Transportation

Public Transit Office

605 Suwannee Street, MS 30

Tallahassee, Florida 32399-0450

Project Manager: Diane Quigley

Prepared by

Martin Catalá

National Center for Transit Research

Center for Urban Transportation Research (CUTR)
University of South Florida
4202 East Fowler Avenue, CUT100
Tampa, Florida 33620-5375
November, 2011
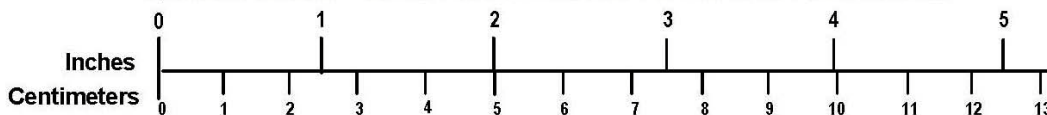
Contract Number BDK85 977-15

METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC
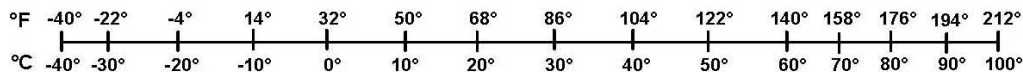
LENGTH (APPROXIMATE)

1 inch (in) = 2.5 centimeters (cm)
1 foot (ft) = 30 centimeters (cm)
1 yard (yd) = 0.9 meter (m)
1 mile (mi) = 1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2) = 6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2) = 0.09 square meter (m^2)
1 square yard (sq yd, yd^2) = 0.8 square meter (m^2)
1 square mile (sq mi, mi^2) = 2.6 square kilometers (km^2)
1 acre = 0.4 hectare (he) = 4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz) = 28 grams (gm)
1 pound (lb) = 0.45 kilogram (kg)
1 short ton = 2,000 pounds (lb) = 0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp) = 5 milliliters (ml)
1 tablespoon (tbsp) = 15 milliliters (ml)
1 fluid ounce (fl oz) = 30 milliliters (ml)
1 cup (c) = 0.24 liter (l)
1 pint (pt) = 0.47 liter (l)
1 quart (qt) = 0.96 liter (l)
1 gallon (gal) = 3.8 liters (l)
1 cubic foot (cu ft, ft^3) = 0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3) = 0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

$[(x-32)(5/9)]$ °F = y °C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm) = 0.04 inch (in)
1 centimeter (cm) = 0.4 inch (in)
1 meter (m) = 3.3 feet (ft)
1 meter (m) = 1.1 yards (yd)
1 kilometer (km) = 0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2) = 0.16 square inch (sq in, in^2)
1 square meter (m^2) = 1.2 square yards (sq yd, yd^2)
1 square kilometer (km^2) = 0.4 square mile (sq mi, mi^2)
10,000 square meters (m^2) = 1 hectare (ha) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gm) = 0.036 ounce (oz)
1 kilogram (kg) = 2.2 pounds (lb)
1 tonne (t) = 1,000 kilograms (kg)
= 1.1 short tons

VOLUME (APPROXIMATE)

1 milliliter (ml) = 0.03 fluid ounce (fl oz)
1 liter (l) = 2.1 pints (pt)
1 liter (l) = 1.06 quarts (qt)
1 liter (l) = 0.26 gallon (gal)
1 cubic meter (m^3) = 36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3) = 1.3 cubic yards (cu yd, yd^3)

TEMPERATURE (EXACT)

$[(9/5) y + 32]$ °C = x °F

QUICK INCH - CENTIMETER LENGTH CONVERSION

Inches	0	1	2	3	4	5
Centimeters	0 1 2 3	4 5	6 7	8 9 10	11 12	13

QUICK FAHRENHEIT - CELSIUS TEMPERATURE CONVERSION

°F	-40°	-22°	-4°	14°	32°	50°	68°	86°	104°	122°	140°	158°	176°	194°	212°
°C	-40°	-30°	-20°	-10°	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°

For more exact and or other conversion factors, see NIST Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50 SD Catalog No. C13 10286

Updated 6/17/98

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Expanding the Google Transit Feed Specification to Support Operations and Planning		5. Report Date 11/15/2011
		6. Performing Organization Code
7. Author(s) Martin Catalá, Samuel Downing, Donald Hayward		8. Performing Organization Report No.
9. Performing Organization Name and Address National Center for Transit Research University of South Florida 4202 E. Fowler Ave Tampa, FL33620		10. Work Unit No.
		11. Contract or Grant No. BDK85 977-15
12. Sponsoring Agency Name and Address Research and Innovative Technology Administration U.S. Department of Transportation, Washington, D.C. 20590 The Florida Department of Transportation Research Center 605 Suwannee Street, MS 30 Tallahassee, Florida 32399		13. Type of Report and Period Covered Final Report Project Period 9/22/2001 to 11/15/2011
		14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the Florida Department of Transportation and the U.S. Department of Transportation

16. Abstract

The development of Google's free online transit trip planner has been one of the most exciting developments in transit for many years. Transit agencies that store trip information into a specific file format (the General Transit Feed Specification) and forward the data to Google's transit team will have a robust, recognizable online trip planner for free. But perhaps equally exciting is the impact of the open data architecture, which is not hidden under the veil of proprietary software and has spurred many other developments, including extending the usefulness of the General Transit Feed Specification (GTFS) through the development of applications that leverage the GTFS data to further benefit transit agencies. This project identified opportunities to use GTFS data to support service planning and operational activity. The opportunities are limited to service-level evaluations, given the static nature of the GTFS data. Nonetheless, GTFS data can be used to support service planning efforts. Furthermore, the project identified opportunities to supplement the GTFS with performance-related information and developed a prototype application that integrated GTFS data with an automatic passenger counter (APC).

17. Key Words Google Transit, Service Planning, Geographic Information Systems, performance measures, information technology	18. Distribution Statement		
19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages 64	22. Price

Executive Summary

The development of Google's free online transit trip planner has been one of the most exciting developments in transit for many years. Transit agencies that store trip information into a specific file format (the General Transit Feed Specification) and forward the data to Google's transit team will have a robust, recognizable online trip planner for free. But perhaps equally exciting is the impact of the open data architecture, which is not hidden under the veil of proprietary software and has spurred many other developments, including extending the usefulness of the General Transit Feed Specification (GTFS). One clear example is the development of an open source application, TimeTable Publisher, developed by Tri-Met in Portland, Oregon. TimeTable Publisher allows anyone with GTFS data to create transit timetables in print and web format. The Federal Transit Administration (FTA) is promoting the use of TimeTable Publisher through the development of a webinar. Additional examples of applications built on the GTFS data include the development of transit planning mobile applications including iPhone and Android smartphone applications, and a trip planner that uses text messaging. The Brookings Institute used the GTFS data format as the source of a national study examining transit access to jobs. Also, more traditional GIS desktop applications are using the GTFS, with the development of an ESRI data model for the GTFS. Given the growing utilization of the GTFS data, this research project proposes an examination of opportunities and existing uses of the GTFS and the development of an application that will help leverage the GTFS data to further benefit transit agencies. Since the primary purpose of the Google Transit application is to create tools which assist transit agencies in better communicating with the transit customer, this project will focus on using the GTFS to support service planning and operation business activities.

The goal of the project is to identify how the General Transit Feed Specification data can be leveraged to further assist transit agencies with business activities, such as service planning and operations. This goal is achieved by identifying how data from transit Intelligent Transportation System (ITS) also known as Advanced Public Transit Systems and more specifically transit schedule data, are being used by the transit community and the development of a prototype application that incorporates the GTFS to support additional business functions within a transit agency.

As public transit agencies are asked to do more with less, agencies are looking towards technology to improve service and create data-driven decision making. Intelligent Transportation Systems technology in the transit industry is a significant source of data, and historically, many agencies use Advanced Public Transit Systems (APTS) for only operational activities, and the data are not used to support activities outside the primary purpose of the implemented APTS technology. However, this has changed in recent years, and agencies are starting to integrate transit APTS data with performance evaluation tasks and to coordinate service and asset management. The advent and popularity of Google Transit has contributed to this change. Google's free online transit trip planner has illustrated how data can be repackaged and used for other purposes. The data that drives the Google Transit application, the General Transit Feed Specification (GTFS), has been repackaged by developers to improve transit trip planning through the

development of mobile- based and web-driven applications. An example includes One-Bus-Away, which allows a user to identify services and businesses that are "one bus away" on their GPS-enabled mobile phones. Others have used the data to identify real estate near public transportation. These examples illustrate the rich resource that resides within GTFS data. To that end, this project identifies how the GTFS can be used to support transit agencies as they seek to improve their transit service.

As agencies seek to improve and streamline service, they look to data to support service evaluation, performance measurement and planning. The GTFS utilizes static schedule data; consequently, it can be used to support service evaluation but not system and route performance. The research identified examples of service evaluation metrics that can be calculated using GTFS data. These include:

1. Service Area Calculations
2. Service Coverage
3. Time and Distance Service Calculations
4. Route and Service Directness
5. Stop Location and Spacing Optimization
6. Service Frequency
7. Span of Service

GTFS is a snapshot of planned service; therefore, for planning purposes, it is capable of evaluating designed service levels only compared with real time data. Nonetheless, the GTFS data provides a clear illustration of an agency's service and can be very helpful in understanding new service proposals and cuts. To fully illustrate the opportunities of the GTFS, the project set out to develop a prototype application that did not duplicate existing applications and services.   Service level evaluations that illustrate the impact of service changes were among the prototype options. However, because desktop GIS software can use GTFS data to develop many service-level evaluation metrics, the project team chose not to duplicate that capability.

Mindful of the static nature of GTFS data and to avoid duplication, the project team looked to develop a tool that could expand the use of the GTFS beyond service level evaluations. The prototype would supplement the GTFS with performance-related data. Based on input from FDOT and the project team, a web-based application that integrated the GTFS data with ridership data from automatic passenger counters (APC) was developed.

The prototype application was designed to illustrate the power of supplementing the GTFS data to produce added value from the GTFS data. The prototype application used data from agencies that were willing and able to share their GTFS data and a corresponding APC dataset. The two datasets reflected a given time period and were integrated into an open source geographic information system (GIS)-enabled database, PostgreSQL. This database can represent spatial data in a web environment. The

tools imported and joined APC data and GTFS data for several agencies and developed some service and performance evaluation metrics. These metrics include:

1. Ridership by Hour
2. Ridership by Trip
3. Ridership by Stop
4. Trip Activity Ranking
5. Stop Activity Ranking
6. Activity by Period

The prototype displayed these metrics using charts, graphs and maps. The mapping function allows users to visualize ridership totals by trip. The application was deployed on the web site: http://ridership.transitgis.org

The web application is freely available for the transit industry to utilize, and the code for the web application can also be found at the above mentioned website.

The research effort identified opportunities to use the GTFS data for purposes other than itinerary planning. The project found that the transit agencies are not capitalizing on these opportunities through the creation of tools to support service planning. However, agencies are developing tools to improve data quality and integrity to help streamline the creation of GTFS data. In fact a byproduct of participating in the Google Transit program is improved data accuracy and quality. The GTFS data is not being used by agencies for purposes beyond delivering to Google Transit. The agencies recognize that route and system-wide service level evaluation measures can be calculated using GTFS data but are unable to capitalize due to resource limitations. The research team established that by integrating the GTFS data into a GIS, users are able to evaluate multiple measures by route, trip and network level. Further, the research identified that service level measures based on the temporal schedule data are easier to calculate with GTFS. Further opportunities to expand the GTFS by integrating it with performance data were identified, including the development of a prototype web application to further illustrate the opportunities. The application supplemented the GTFS data with automatic passenger counter (APC) data to produce information on ridership. The development of the prototype application illustrated the challenges of not having standardized data formats, as APC data from different agencies were not uniform and needed to be normalized to match each other, further illustrating the benefits of standardized data formats. The research effort successfully identified and demonstrated how agencies can use GTFS data for purposes beyond Google Transit to support planning activities.

This research project identifies opportunities to use the GTFS data to make more data-driven decisions, which in turn can provide improved and more efficient service. The examples of opportunities are illustrative of how the data can be used to further assist agencies with service planning. Further, the research illustrates the power of standardized data formats. As the GTFS is a clearly articulated and end

focused data format, it has made multiple additional services possible. The clear data documentation facilitated the development of the ridership prototype, and the challenges to development were due to the lack of standardized ridership data.

It is the hope of the research team that the benefits of acceptable and usable data formats are realized and contribute to a road map for the industry to develop better data specifications that support ongoing service evaluations and performance.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Background

Advances in information technologies have created enormous opportunities for transit agencies in the United States. Early technology systems were primarily used for internal business activities and designed for operational and planning purposes. These systems often operated in isolation resulting in duplication of services and activities. With the advent of enterprise information systems, some transit agencies have overcome the "silo effect", leveraging the technology and data, thus producing information for multiple operational and planning activities. However, this is the exception within the transit industry, and in many cases, the implementation of transit technologies continues to operate in isolation. Successful implementations of transit technologies within the industry are sporadic and the advantages of transit technologies have not been universally realized. Wide-spread integration of transit technologies to support multiple purposes, such as planning and operational activities, has not occurred. Nonetheless, the use of transit technology continues to grow within the industry, particularly for the high-value and high-profile technologies such as online transit trip planners, real-time information systems and smartphone applications.

Due to the positive impact of online transit planners, the availability of transit itinerary planning services within the industry has grown rapidly. Contributing to this growth is the development of the Google Transit trip planner and the rapid development environment of the Internet and mobile web applications. Agencies looking to gain an edge to capture more customers and increase the satisfaction of existing customers have embraced Google transit and other Web 2.0 technologies as a marketing effort and customer service application.

The foundation of Google Transit's success is its data format, validation tools and free service. The data format called the General (formerly Google) Transit Feed Specification (GTFS) is a highly documented and clearly defined data schema or format. The open format and clearly defined data allows software vendors and developers to easily produce applications that can read and write to the data format. Validation tools and properly formatted, error-free data assist agencies to maintain the integrity of such applications, and help agencies share accurate information on their transit stops, routes and schedules. All of these services are currently provided to agencies for free. As long as the agency shares its data with Google, it enjoys the benefit of a free, online trip planner.

The open data has contributed greatly to the widespread utilization of Google Transit trip planner, but the data has produced a secondary effect: the growth of an application developer community leads to the development of innovative solutions using the GTFS data. Public transit agencies, software vendors, third party application developers, and regional and statewide planning organizations are utilizing GTFS data to support trip planning, service evaluation and regional planning activities. The applications vary from off-the-shelf applications and mobile transit trip planning applications to real estate web sites showing the location of public transit. Many of the major schedule and run-cutting vendors directly export to the GTFS format. Additionally, several startup companies have created web-based tools to

create GTFS data. The Transportation Research Board's (TRB) IDEAS program sponsored a GTFS generator tool, for example. Perhaps the most prolific area of growth is the mobile application arena, where developers are capitalizing on the rapid growth of the location-based smartphone market. A small industry of downloadable applications onto multiple phone platforms has blossomed relying on data provided in the GTFS format. Innovative solutions such as dynamic transit time tables, measuring transit availability and accessibility, and open-source multi-modal trip planners have emerged leveraging the GTFS data.

Of equal interest is the growth of applications and innovation within the transit industry. The GTFS has spurred innovation within agencies. The transit industry, hungry to capitalize on the services of Google Transit has created multiple applications which help improve processing its data into the GTFS format. This has led to improved data integrity and closer integration between technology systems. Agencies using data from scheduling systems have created scripts and applications which communicate and interact with historically closed systems to generate GTFS data. Software vendors have created add-on tools which export and read the GTFS format. Frequently these applications are not "public facing" applications, but do improve the efficiency of how agencies control and use its information.

The ability to develop these innovative solutions is due to the uniform and open data driving each application. The GTFS is a robust data format that has changed to accommodate the increasing needs of the user community, and is clearly documented. Further, Google and the user community have created validation tools to ensure some level of data quality to integrate into Google's trip planner. The data format, the user tools and the user community results in a pool of expertise that allows the users to expand the utility of the data.

Capitalizing on transit technologies is critical for the transit industry. Agencies that are successfully implementing technology are expanding the breadth of how and who utilizes the data. These examples serve as a backdrop for this research. Just as other data and information technologies have been leveraged to improve operational efficiency and service planning, as well as improve customer service and service efficiency, the GTFS offers a wealth of information that can be used or combined with other data to improve business operations and increase transit demand.

This research investigates how transportation data systems, including GTFS data, are being leveraged to support multiple activities. The successes, lessons learned and recommendations will serve as a foundation for the goal of expanding the uses of schedule and route information that resides within the GTFS. Further, this research will illustrate the power of leveraging GTFS data to support operational and planning activities through the creation of a prototype application.

Chapter 2 - Literature Review

Countless research efforts have been conducted to articulate the best practices for implementing technology in the transportation industry. The research is from a broad spectrum of transportation technologies and can be generally described as APTS, which includes automatic vehicle location (AVL), automatic passenger counters (APC), fare-card technologies, and information systems.

The role and challenge of APTS data is well documented. Transportation organizations have used data by-products of their APTS systems to support ongoing operational and planning activities. Databases from APTS applications serve as repositories to be used for analysis and ultimately improved decision making. Moreover, the authors of the Federal Highway Administration's archived data management study conducted a survey of multiple organizations that are creating an archive of the APTS roadway data to support planning and operational studies.   These organizations are realizing the opportunities from the APTS archives:

> Transportation professionals, especially those in operations and maintenance, are increasingly coming to understand that the data generated by Intelligent Transportation Systems and subsequently archived are useful well after they are no longer 'real-time' data. (USDOT, 2005)

The use of archived APTS data is not limited to roadway APTS systems. Multiple studies have identified instances of transit APTS applications being used to support multiple activities. Multiple syntheses reports put out by the Transit Cooperative Research Program (TCRP) identify opportunities that transit APTS technologies present.

Many of the studies report on the impact transit APTS technologies have on improving route and system performance, as well as reliability and transit demand. Most commonly, real-time information systems are used by operations and dispatch to identify bus issues such as schedule adherence and vehicle performance. And while many of these systems are designed to perform operation and dispatch functions, several agencies capitalize on the data to support additional activities. Kimpel et al. (2004), reported on the use of AVL/APC data to support route performance monitoring and schedule improvements. The AVL system, used by the dispatch center to provide real-time operations management, was also used to capture arrival and depart times, dwell times and speed between stops, as well as boardings and alightings. Using these data schedule improvements optimized route performance.   In addition, Horning et al.'s study (2009) on estimating running time and demand for a bus rapid transit corridor utilized AVL and APC data to support its analysis. Information on passenger boardings and alightings, distance and travel time between stops, and number of bus stops all derived from APTS were used for the demand estimations. Bertini examines how to leverage an agency's bus dispatch system (BDS), a combination of automatic vehicle location (AVL) technologies and automatic passenger counting (APC) systems to create transit performance measures. Using data from the BDS multiple performance measures were identified.

Table 1: Performance Measures Using BDS Data

• Scheduled hours of service, • Actual hours of service, • Number of scheduled trips, • Number of actual trips, • Number of scheduled miles, • Number of actual miles operated, • Number of passengers carried,	• Total boardings and alightings, • Average passenger load during each trip, • Number of passengers per mile, • Average scheduled speed (mi/h), • Average speed (mi/h), and • Number of operators.

Automated fare collection systems such as smart cards are traditionally implemented for transit to support administrative functions of controlling access to service and revenue management. However, a study by Trépanier et al utilized smartcard data to develop performance measures. Information on passengers such as passenger miles, hours, and trip length were used to develop performance measures.

Multiple TCRP reports identified opportunities for implementing and maximizing how such systems and data can be used to support multiple activities.   TCRP Reports 24 and 73 provide a synthesis of the state of practice with AVLs and an update to the synthesis, respectively. Furthermore, planning efforts to track revenue, ridership, and monitor schedule adherence can be performed using AVL data. These data can be used to make minor schedule adjustments based on observed route running times and dwell times from the AVL system. Report 73 identified that the combination of dwell time information with traffic signal information can be used to identify intersections and corridors for traffic signal priority efforts. Other operational activities that can benefit from the use of AVL data include the ability of dispatchers to control bus operations, increased accuracy in schedule adherence monitoring and monitoring driver behavior and performance.

The TCRP Report 77 on the state of practice in APC found that agencies used the data for many of the same purposes: support planning and operational management functions.

> *"The primary benefits of APCs included data disaggregated at the stop, segment, and trip levels; better quality of ridership data; availability of running time data to adjust schedules; and a better basis for decision making." (2008)*

The robust and fine detailed data provided by the APC systems offered countless efforts to better understand the performance of the transit system.

The TCRP Report 126 on leveraging APTS data to support marketing efforts found that transit APTS data can be used to help with transit planning activities. In particular, it found that passenger counters, fare

box and smartcard data can assist with marketing efforts. The development and creation of sampling plans and intercept surveys using smartcard data were also reported in TCRP 126 (2008).

Leveraging transit data produced by APTS technologies can help agencies produce actionable information, leading to improved decision making, efficiencies and service quality for the agency. Improving access to data can also result in innovative uses of data that are beyond the originally intended purpose. The synthesis reports and studies all rely on having access to APTS data to perform the analysis and research. APTS systems with more open data and architecture create opportunities for agencies to capitalize their APTS investments. This is true for the online transit trip planners as well. The development of open route, stop and schedule data results in the rapid development and utilization of online transit trip planners. Moreover, the availability of the data to other potential users of schedule data results in innovative solutions that addresses trip planning and other activities.   Understanding the history of the online trip planners provides a glimpse into how the data was developed and ultimately used to support activities beyond trip planning.

Online Transit Trip Planners

Transit authorities used transit trip planners internally for years, but it was not until the 1990s that web-based transit trip planners were made available to the public (Watkins & Ferris, 2010). As of June 2002, there were "30 transit web-based trip planners in the United States, with 22 serving single agencies and eight serving multiple agencies…recently, about one new trip planner per month has been made available to the public" (Radin et al, 2002). Radin et al. noted that based on the number of transit authorities interested in developing - or in the process of developing - web-based transit trip planners, there was no need to sell the development of transit trip planning to transit authorities (Radin et al. 2002). However, in the following years, development costs and inadequate "knowledge about GIS, APTS, trip planning vendor terminology and maintenance of web sites" hindered the ability of many agencies to successfully develop and implement transit trip planners for use by the public (Watkins & Ferris, 2010).

In December 2005, Google revolutionized web-based transit trip planning with the introduction of the Google Transit trip planner. As part of their company philosophy to cultivate innovation and creativity, Google instituted the "20 percent time" program, which allows engineers to spend 20 percent of their time on any project that interests them personally ("Engineer's Life-US Jobs-Google," 2010). Google Transit trip planner was started as a 20 percent program project by several Google engineers located in San Francisco, New York, and Zurich, Switzerland. They wanted a quick and easy method using one web site to plan a trip using public transportation. The Google Transit trip planner provided access to public transit schedules and routes so any user can plan a trip using public transportation ("Public Transit via Google," 2009). The city of Portland, Oregon was chosen as the launch site for the program due to its technologically savvy transit authority. In June 2006, Google Transit trip planner was integrated into

Google Maps and Earth, and in June 2008, Google Maps for mobile was launched giving users the ability to plan transit trips using applications on mobile devices ("Public Transit via Google," 2009).

A driving force behind the technology of the Google Transit trip planner is the data. Google, working with a group of transit experts, developed a data schema that would serve as the fuel for the trip planning service. This data schema, the General (formerly Google) Transit Feed Specification (GTFS) contained schedule data to feed into the Google Trip planning engine. Since its inception, the GTFS has undergone changes to the format of the data and this is due to the open design of the GTFS. With feedback and input from the user community, Google considers all suggestions and developments to the GTFS. Further, among the changes includes the adoption of a neutral name- the General Transit Feed Specification. The intent of this change is to encourage the user community to more readily embrace the format, and realize that the format is open to change based on the user community and not Google alone.

General Transit Feed Specification (GTFS)

In an effort to provide access to accurate, timely transit information, the Google engineers developed an open format data product known as Google Transit Feed Specification (GTFS). The GTFS format established a "de facto standard for exchanging transit schedule data" (*IEEE Pervasive Computing*, Ferris et al., 2010). And while the GTFS is not a standard, it is a widely recognized data schema that effectively identifies transit itineraries for the transit customer. GTFS, which now stands for General Transit Feed Specification, "defines a common format for public transportation schedules and associated geographic information" (Google code general transit feed specification, 2010). Transit authorities provide data such as schedule times and stop locations in the GTFS format to Google, which is then consumed or uploaded by Google into its trip planner engine for use in public websites such as Google Maps (Google code, 2010; *IEEE Pervasive Computing*, Ferris et al., 2010). This allows both the general public and other government agencies quick and easy access to the information and provides opportunities for third-party developers (*IEEE Pervasive Computing*, Ferris et al., 2010) to use the data in unique and innovative ways.

Changing GTFS

The GTFS is popular and widely used because of the transparent documentation and the underlying logic and service the data provides. However, the strength of the format is the result of an evolving design schema. Through the use of a web- based user group, the GTFS has been changed several times and currently has over 30 proposals under consideration. Moreover, it is important to note that the proposals are being considered by the user community and Google. If the user community can demonstrate and illustrate the benefit of modifying or changing the GTFS, then the user community and Google will implement those changes. To date, there have been six formal and accepted changes to the GTFS. Below is a list of the changes, followed by a brief description of each.

- Agency_phone (added August 8, 2008)
- CSV format clarifications (added March 30, 2009)
- Preferred transfer points (added August 8, 2008)
- Station groupings for stops (added August 8, 2008)
- Stop codes (added February 29, 2008)
- Trip_short_name (continued) (added March 30, 2009)

Agency Phone

In August 2008, the issue of how to include a transit agency/operator telephone number for specific stops in the GTFS feed arose. As riders sometimes needed to contact their transit agency/operator for pick up or drop off at certain stops, there was a need to include a telephone number for the rider within the feed. In response to this issue, the GTFS staff updated the feed to include the field- "agency_phone"- to agency.txt. The field is a string value that allows for a single telephone number of a transit agency/operator to be included. The field needs punctuation to group the numbers correctly, but descriptive text should not be included within the field.

CSV Format Clarifications

In March 2009, two issues regarding glitches with unusual syntax were addressed. According to the GTFS staff, not all syntax was interpreted in the same way by all parsers. The first issue occurred with a header such as the following: stop_id, stop_name, stop_lat, stop_lon. The spaces before the names were fine, but the space after "stop_lat" was considered by many parsers, including Excel, to be a part of the value. The second syntax issue was that each line did not consistently end in CLRF or LF. To correct the first issue, a patch was developed to ignore initial whitespace before quotes. In addition, the patch also provided a warning if any field in the first row of a file started or ended with whitespace after the initial whitespace was skipped. The second issue was corrected by adding "EndofLineChecker" to ensure that each line ended in CLRF or LF, that lines did not contain other random characters that look like new line separators, and that every line in the file ended with the same end of line sequence. The GTFS staff concluded that RFC4180 (http://tools.ietf.org/html/rfc4180) should be the standard format for GTFS CSV files. The two deviations from the format were as follows: there were often one or more space characters after the comma that was between fields, and tools that parse GTFS should skip these spaces; moreover, some GTFS files started with a utf-8 byte order marker that parsers should skip.

Preferred Transfer Points

In August 2008, GTFS users suggested adding an optional file to the feed specifications to define preferred transfer points. The GTFS group proposed to simplify transfer time at a station, a possibility would refer to stations listed in the stops.txt file format proposal. The different types of transfers could help the algorithm decide which stop is best to provide a transfer when two routes share several stops. The route direction fields could also be added to further specify transfers.

Station Groupings for Stops

In August 2008, the GTFS group addressed the issue of how to capture underground transit station entrances, which sometimes had multiple entrances a significant distance apart. According to the GTFS group, adding station and entrance information would make it possible for GTFS-using applications to display station entrances on map views, and incorporate entrance information into travel instructions. The stops.txt file would be expanded to refer to more location types.

Stop Codes

In February 2008, the GTFS group proposed adding the ability to specify a short code for each stop that transit riders can use to identify that stop. Transit systems use short codes on signs, schedules, and phone-based transit information systems that provide schedules or real-time information. The "stop_code" field would be added to the stops.txt file.   A short code (usually a number) that uniquely identifies the stop would be included only if the code is displayed on stop signage, timetables, and/or other rider-oriented materials. This can be left blank for stops without a code.

Trip Short Name

In March 2009, the GTFS working group proposed adding a new field, "trip_short_name," to trips.txt to better identify services that presented a meaningful trip name to riders. For individual trips that have a short rider-facing name in schedules and station signage (for instance, commuter rail train numbers), this field provides that name. If riders did not commonly identify trips by name, the field would be left blank. A "trip_short_name" value, if provided, should uniquely identify a trip within a service day. It should not be used for destination names or limited/express designations. (trip_short_name)

Each of the changes were reviewed by the group of users and tested for their usefulness. Once the new format has been tested and its usefulness articulated through concrete examples then the changes are submitted to the specification.

Additionally, there are over 30 open proposals to the GTFS that are being considered by the Google Transit Changes Group. This is an active group that includes many of the leaders of the transit community that are working on ensuring that the Google Transit trip planner works effectively and for the customer.

Leveraging GTFS Data

One of the leading uses of the GTFS transit data (*IEEE Pervasive Computing*, Ferris et al., 2010) has been online public transit applications that can be accessed via transit riders' mobile telephones. As of April 2010, on the website City-Go-Round, there were over 158 public transit applications that could be downloaded, either for free or a nominal fee, to a user's mobile phone for instant access to public transit information in cities across the United States and internationally (City-Go Round Transit App Gallery, n.d.). The increased availability and use of transit applications could potentially increase transit ridership and help riders with disabilities navigate public transit (*IEEE Pervasive Computing*, Ferris et al.,

2010). The development of transit applications based on GTFS data shows that transit data, in a straightforward industry-standard format, has the potential to transform the public transit industry. In fact, the open format of the GTFS simplifies the development effort, which in turn encourages innovation and development of new and novel implementations of schedule data.

Much of the innovative solutions using GTFS are external, customer facing applications; however, several developers and users have put the GTFS data to other uses. The developers at FrontSeat, the makers of WalkScore.org, have created an application called TransitScore which is a neighborhood-based application to help residents get a better understanding of the available transit services. The developers at FrontSeat are envisioning other transit applications including the development of TransitShed, which would provide a commuting time shed for a given point.   TransitShed (See figure 1), provides the user with a series of bands that illustrate multiple travel sheds from a given location.

Figure 1: Transit Travelshed Map
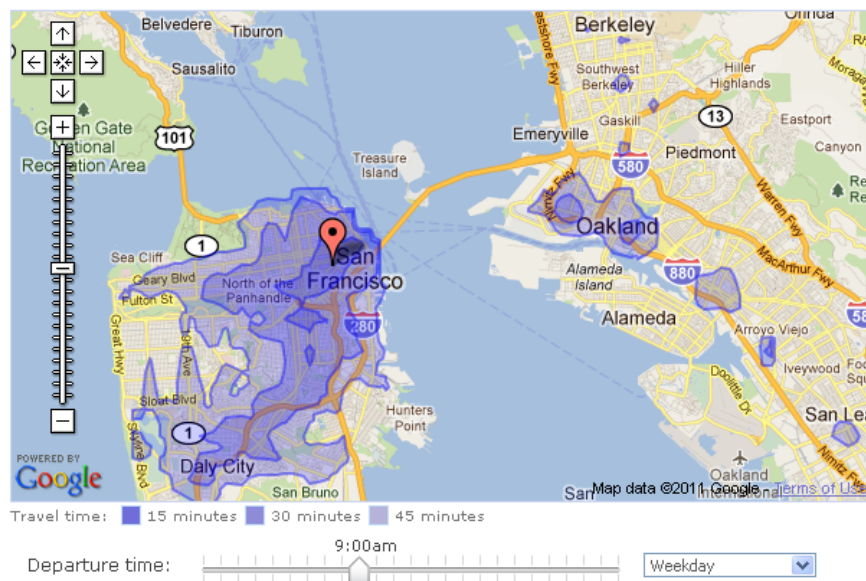
Another example is the Mapnificent web site which uses data from the GTFS to identify travel sheds around a given point. Figure 2 illustrates the locations that can be accessed by transit within 15 minutes. Further, the Mapnificent web site can be used to identify overlapping travel sheds for two locations to identify areas where both locations are able to access within a given time (See figure 3).

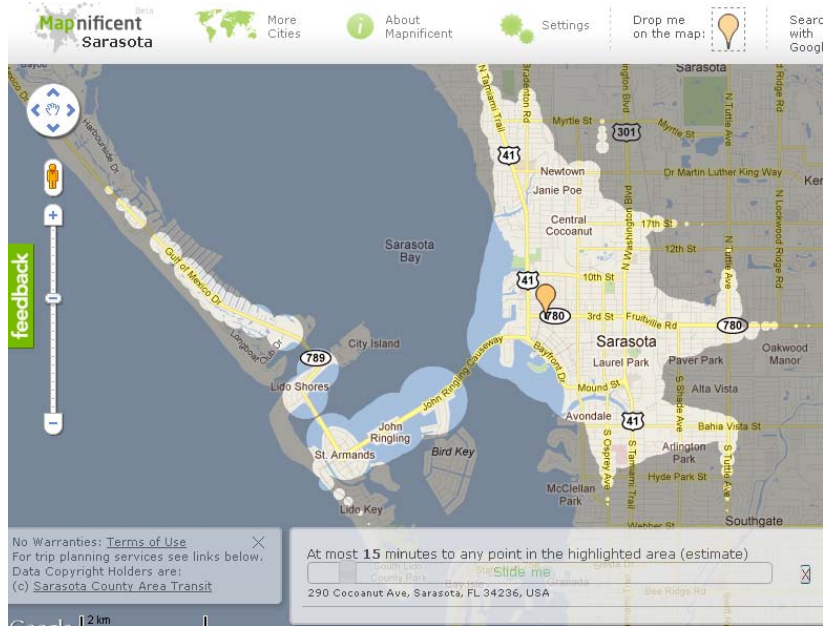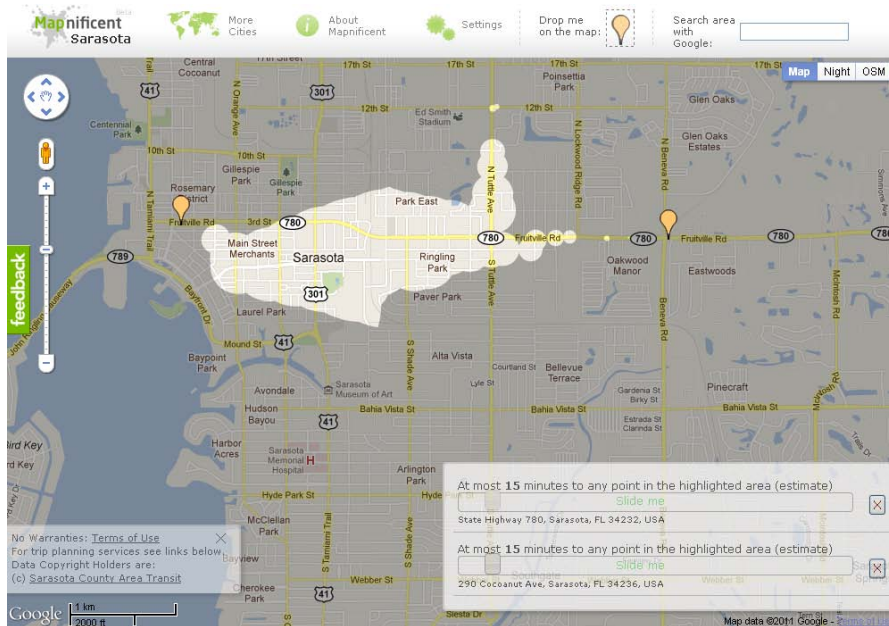**Figure 2: Mapnificent 15 Minute Travelshed**

Figure 3: Mapnificent 15 Minute Meet Up Map

The unique element of both of these applications is that they rely on shared GTFS data and they are not for trip planning, but rather to identify and visualize the transit service in a given area.

The use of GTFS data by organizations and software developers has resulted in a great deal of innovative applications and solutions. Moreover, public transit agencies have also contributed to the collection of such. Tri-Met, the transit agency in Portland, OR, popular for its open data and innovative use of technology, has contributed to the pool of applications through the creation of TimeTable Publisher. TimeTable Publisher creates web-based and print timetables using data within the GTFS. Figure 4 shows an output developed by a consultant using the TimeTable Publisher application. The advantage of the TimeTable Publisher application is that it is an open source application, and as such, others are able to use the application and develop around it.

Figure 4: Sample TimeTable Publisher Output

El Cerrito Plaza BART Station	6:45am	—	—	—	—	—	—	—	—	—	—
Richmond Field Station	6:57am	8:30am	10:35am	11:35am	**12:35pm**	1:35pm	2:35pm	3:35pm	4:35pm	5:00pm	—
Regatta Facility	7:00am	8:35am	9:40am	10:40am	11:40am	**12:40pm**	1:40pm	2:40pm	3:40pm	5:05pm	6:05pm
Buchanon Street @ Jackson Street	7:05am	8:40am	9:45am	10:45am	11:45am	**12:45pm**	1:45pm	2:45pm	3:45pm	5:10pm	6:10pm
Hopkins Street @ MLK, Jr. Way	7:10am	8:45am	9:50am	10:50am	11:50am	**12:50pm**	1:50pm	2:50pm	3:50pm	5:15pm	6:15pm
Downtown Berkeley BART Station: Shattuck Avenue @ Addison Street	—	—	—	—	—	—	—	—	—	5:25pm	6:25pm
Evans Hall: Hearst Mining Circle Side	7:35am	9:10am	10:10am	11:10am	**12:10pm**	1:10pm	2:10pm	3:10pm	4:10pm	5:30pm	—
Oxford Street @ University Avenue	7:40am	9:15am	10:15am	11:15am	**12:15pm**	1:15pm	2:15pm	3:15pm	4:15pm	5:40pm	—
Allston Way	7:42am	9:17am	10:17am	11:17am	**12:17pm**	1:17pm	2:17pm	3:17pm	4:17pm	5:42pm	—
Buchanon Street @ Jackson Street	7:53am	9:28am	10:28am	11:28am	**12:28pm**	1:28pm	2:28pm	3:28pm	4:28pm	5:53pm	—
Regatta Facility	7:00am	8:35am	9:40am	10:40am	11:40am	**12:40pm**	1:40pm	2:40pm	3:40pm	5:05pm	6:05pm

It is not expected that transit agencies will have the skill pool similar to Tri-Met to develop such tools. However, some agencies may be leveraging the GTFS data to support other activities. Identifying the examples of innovation and how agencies are capitalizing on the data resource within the GTFS can serve as a deep resource for the transit industry. As the private sector is developing applications for customers, transit agencies may be developing applications or add-ons to the GTFS that might support planning and operational activities.

To help the transit industry leverage the data within the GTFS to support internal business activities, this research project examined how the GTFS was currently created at transit agencies to uncover how it can be used to support operational and planning activities.

Chapter 3 - Transportation Agency Interviews

The open data format of the GTFS has led to innovative solutions in the private sector and third party software developers. Developers and programmers using the GTFS are finding innovative ways of integrating the data format into a variety of applications. Mobile phone transit trip planners are the most notable among the applications. Equally notable are non-transit organizations that are using the GTFS data to measure access to jobs, to support marketing for real estate, and to create open source software applications. Understanding how transportation agencies are using GTFS data will provide insights of how the data is being – or can be – leveraged to provide better information to support planning and operational activities.

At the outset of this project effort, the research team assessed two major users of the public transportation schedule, route and stop data, transit agencies and third party software developers. However, during the research effort a third user group emerged: regional and statewide planning organizations. These agencies were looking to utilize standardized transit route, stop and schedule data to support regional planning activities and coordinate construction and maintenance services. As such, interviews of two state DOTs were added, the District Seven office of the Florida Department of Transportation (FDOT) and the Policy and Planning Division from New York State Department of Transportation.

Regional Transportation Planners

Many transit agencies serve a portion of a wider region's mobility, and understanding how a transit system interacts with the broader transportation system and region is important to statewide and regional transportation planners. How agencies acquire and integrate this data with their overall planning efforts provides an insight to how useful transit schedule, stop and route information can serve a regional and statewide planning group. Two organizations, the District Seven office of the Florida Department of Transportation and New York State's Office of Planning, serve as examples of how transit data like the GTFS can be used to support purposes other than trip planning

Florida Department of Transportation District Seven Office

The FDOT District Seven office has created an enterprise GIS to support many activities in the planning and operations departments. Furthermore, the project team met with the FDOT team that regularly utilizes transit route and stop location data to coordinate safety and planning activities. The primary data needed for the ongoing activities included:

Table 2: FDOT Transit Data

Transit Data in District Seven Office Enterprise System	
Bus Stop Locations	Park and Ride Data
Ridership Data	Fare Data
Route Location and Patterns	Schedule and Span of Service Data

Integrating transit data in the enterprise system presented challenges for the GIS team. Perhaps the most daunting was the need to manually collect and import the data into the enterprise system. Moreover, the effort required requests for data on a regular basis which were made through a series of contacts and personal relationships. Compounding the effort was the need to acquire the data from multiple transit agencies, and the agencies did not conform to a uniform stop, route and schedule format. Consequently, it was necessary to normalize the data to fit into the enterprise system.

In spite of the required staff hours, the effort was worthwhile because the data provided service to several efforts within the district office. The primary use of the data was to identify needs for pedestrian safety countermeasures. Identification of high traffic volumes, pedestrian incidents, multiple lane roadways and bus stop locations supported research efforts to help focus the safety team's efforts. The data was also used by traffic operations to coordinate activities in the operations department of the District. Further, the district wanted to use the data to coordinate FDOT improvements with the transit agencies. Identifying scheduled sidewalk improvements that impacted transit stops and routes would help FDOT and the transit agency coordinate the improvements to include shelters and American with Disabilities Act (ADA) improvements. Through this effort, leveraging the construction funds of both agencies would result in improved services and facilities, as well as a reduction of costs for both agencies.

To ensure that this program would be able to continue to have up-to-date transit data, the District Seven office has partnered with the University of South Florida's Center for Urban Transportation Research (CUTR) to develop procedures and scripts to regularly update the transit stop, route and schedule data using the GTFS feeds from the area agencies. There are three current agencies within the District 7 boundaries that participate in the Google Transit program, and the ability to quickly integrate GTFS data into the enterprise GIS system would provide savings for both agencies: for the District Seven office by automating the integration of the transit data into the enterprise system, and for the transit agencies that will not need to acquire and transmit the transit data to the district office. Furthermore, this coordination will simplify the process for coordinating service changes due to construction and improving facility improvements.

With the completion of the project, the District Seven office of FDOT envisions using this data to integrate with modeling efforts of the district and to create visualization tools to illustrate the regional transit service. The availability of a uniform transit stop, route and schedule format created an opportunity for FDOT to develop tools to import and leverage the GTFS data to support additional service opportunities.

New York State Department of Transportation Office of Policy and Planning

The New York State Department of Transportation (NYSDOT) has a 511 system to inform the public of commuting conditions and alternatives for the entire state. The 511 system (NY511) contains data about roadway conditions and some transit information. In an effort to provide more thorough transit data, the planning department set out to provide a transit trip planning tool for commuters on the NY511 system. In order to migrate the state's transit systems into the 511 system, a uniform data model of transit route and schedule data was required. Further, it was realized that to migrate all of the state's transit agencies in the system NYSDOT needed to create tools and incentives for the agencies.

A transit stop, route and schedule data schema for all of the State's transit agencies to store its transit data in was developed by NYSDOT. The data schema known as the Schedule Data Profile (SDP) was designed to store transit data to be fed into the state's 511 system. NYSDOT also created a portal for the data to be stored, accessed and edited. Thus, the Next Generation Transit Information Portal (NGTIP) was created. Access to the data within the NGTIP and the ability manage and edit the data was an additional hurdle to ensuring that the transit data within the 511 system was up-to-date. To address this obstacle, NYSDOT developed a web-enabled database interface to manage all the transit data. The Web Data Maintenance System (WDMS) was created to give agencies the tools to manage and edit the route, schedule and stop information residing in the NGTIP and the 511 system. Below is a screen shot of the managing transit facility page. The WDMS is designed to manage all aspects of the stops, routes and schedules of a transit agency.

Figure 5: WDMS Facility Manager

The tools and data infrastructure developed by NYSDOT provide transit agencies with tools to edit, manipulate and share its transit data. The robust data structure, the SDP, developed by NYSDOT allows the agencies to have its data on the NY511 system. Furthermore, the SDP data format allows the agencies to export its SDP data into the GTFS format.   As a result of the WDMS and NGTIP, all NY State transit agencies will have multiple transit trip planners available for its customers. The NYSDOT gets robust transit data to use in its planning and operations activities. Much like the District Seven office of the FDOT, NGTIP will allow agencies and the NYSDOT to coordinate asset improvements, provide a platform for regional analysis and provide the agencies with data that will support additional planning activities. The NYSDOT is planning on developing further data schemas that support additional activities including coordinating fare card, real-time information, and a planning module. NYSDOT's effort is illustrative of how leveraging transit data into a robust and clear format can provide service to many organizations and activities.
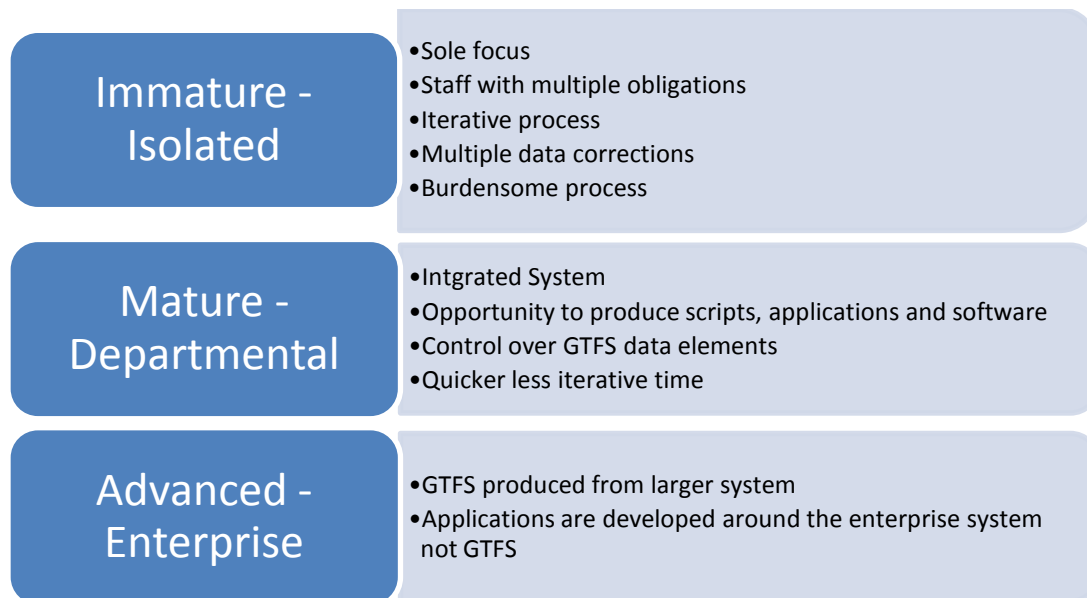
Public Transit Agencies

Transit agencies, as discussed in the literature review, recognize the power of leveraging data from technology to support multiple purposes. Although the success and practice is not widespread, the literature discusses the opportunities that exist when data is leveraged to support a variety of departments and functions. The interviews of transit agencies that deploy GTFS were designed to identify how and if agencies were discovering the same multiple purposes for the GTFS data. In certain ways, the interviews mirrored the findings in the literature review. Innovation was not widespread, however examples of innovation provide compelling evidence of how the GTFS can be used to a greater extent.

The experiences of public transit agencies using GTFS offer unique insights as to how GTFS is created, managed, and used. Agencies exhibit a broad spectrum of innovation. Some are developing new tools and procedures to improve the process for generating and using the GTFS and to ensure greater data integrity, while others are conducting the bare minimum of activities to ensure that the GTFS is generated and shared with Google Transit in a timely manner. The latter of these agencies did not report innovative tools or applications with GTFS data, while those agencies looking to improve and streamline the process of developing the GTFS experienced higher levels of innovation, creating tools to improve the GTFS generation process. The tools also contributed to improved data integrity and quality control over the route, stop and schedule data. The processes and systems used by transit agencies to generate the GTFS are illustrative of innovative solutions to improve the management of the schedule, route, and stop data. It is important to note that the interviews were designed to identify applications and innovative solutions as well as opportunities to use the GTFS data. They are not intended to serve as agency case studies, but rather a source of innovative solution and ideas of how GTFS data can be better developed and utilized to support activities beyond trip planning.

The agencies are grouped into three generalized categories to provide a clearer understanding of the innovation and needs of the agencies. The categories are:

1. Immature or non-integrated IT (direct export to GTFS from Software)
2. Mature staff departmental (Export and with minor GTFS manipulation)
3. GTFS is part of a larger enterprise system

Figure 6: Agency Categories

Immature - Isolated	•Sole focus •Staff with multiple obligations •Iterative process •Multiple data corrections •Burdensome process
Mature - Departmental	•Intgrated System •Opportunity to produce scripts, applications and software •Control over GTFS data elements •Quicker less iterative time
Advanced - Enterprise	•GTFS produced from larger system •Applications are developed around the enterprise system not GTFS

The interviews provided examples of innovative solutions. The solutions developed all served the general purpose of improving data through efficient systems, and improved quality control over the data. The interviews also produced many wishes. Agencies lacking the time, skills or both indicated a wealth of opportunities to use the GTFS. Most of the applications and ideas proposed were focused on managing GTFS data and few agencies identified opportunities to use the GTFS to support other activities, such as planning and operations. Consequently, the pool of solutions and uses to leverage the GTFS will be expanded by analysis and input from the research staff and the FDOT.

The agency's approach towards Google transit and its organizational structure impacted the level of innovation within the agency. Staff who operated in isolation and with multiple obligations, yet worked for agencies with an isolated goal of implementing Google Transit, experienced lowers levels of innovation. Organizational structures that supported greater data and information integration, allocation of training and development time, and focused on tasks exhibited higher levels of innovation.

Immature Agencies

For the agencies categorized as immature, the staff had a great deal of technical understanding of the data that made up the GTFS information - in short, the stops, routes, and schedules. However, for the agencies within this group, the staff was charged with getting the GTFS up and running to implement Google transit, but this was but a small part of the staff's overall responsibility. The responsibilities ranged from driver schedules, run-cutting, general IT support, APTS support and bus stop inventory management. As a result, time to develop database and technical skills to best manage and create the GTFS was very limited. Agencies operating in this type of environment exclusively utilized commercial scheduling applications. As mentioned earlier, scheduling and run-cutting applications offer modules and tools that allow for the direct export to the GTFS format. All the stops, routes, trips and schedules were managed, created and examined using the commercial off- the- shelf applications. The scheduling software typically would generate the agencies' GTFS data to be validated using Google's validation tools. The validation tools would frequently report errors requiring the data to be repaired in the scheduling software, and then exported again to be evaluated by Google's tools. The errors varied from formatting issues of the data, problems with route colors (GTFS requires that each route be assigned a color), stop locations being too close together, to issues regarding transfers. These agencies frequently reported having an iterative process for developing the data. Any errors identified were addressed using the scheduling software and the entire process would be repeated. For these agencies, this process was time consuming, arduous and was described as the most challenging for agencies. Given the time constraints and the need to balance multiple other responsibilities, the staff members at these agencies were not in a position to develop innovative solutions, and consequently offered no examples of innovative applications, scripts or processes to develop and use the data. However, this group did have strong understanding of how the data could be used in other areas. Not surprisingly, these requests were to help with the process of developing and managing the GTFS data. These agencies had several common wishes, most of which would improve efficiencies. Below is a list of the desired tools:

1. GTFS Merging tools – several agencies indicated difficulty with the merging tool provided by Google Transit, and on several occasions, when an updated schedule of stops and routes were provided to Google, the data resulted in a lapse of service on Google Transit. Most of the time this is due to one of two issues: a lag in the data because it takes Google Transit approximately two weeks to integrate the changes into its system, or the data had bad calendar dates resulting in an active GTFS with service dates that were past or not current. Several of the agencies wished there were better tools to help with the merge process.

2. GTFS visualization tools – after a service change was proposed or completed, agencies expressed the desire to visualize the GTFS changes on a map or chart. This was particularly important when significant changes and cuts occurred. A visualization tool may help staff identify and understand the changes, and these visualizations can be used by planners to recognize and understand the proposed cuts.

3. Tracking Driver assignments – many of the staff that worked for agencies within this category were responsible for development of the runs and the "pieces of work" for drivers. Consequently, the ability to track driver responsibilities and assignments was a highly sought feature. The staff recognized the benefit of understanding bus assignments help when drivers were absent, and understood that the data in the GTFS contained information that could support this effort.

4. Customer Service Integration – the users of this data saw the power of Google Maps to help with customer service requests and tracking of complaints and issues.

Agencies within the immature category exhibited a vision for the agency to embrace technology and use its services for multiple purposes. However, given the limited resources, the opportunities to create innovative applications and use of the GTFS were simply out of reach for the staff.

Mature IT Agencies

Agencies categorized within the Mature IT group had a broad range of experiences. Within this category, the innovation ranged from tools for automation, to data integration with APTS systems and planning tools. Several agencies within this group used technology and software scripts to automate and simplify the process of developing the GTFS. The staff at these agencies displayed a higher level of programming and software skills. These agencies also provided institutional support such as time opportunities to develop such scripts and an overall recognition from management of the benefits of programming and script development.

Agencies that exhibited more innovative approaches operated in an environment where technology solutions to streamline the process were a recognized responsibility of the department. This provides the staff the ability to spend resources to improve data generation and development. These agencies still relied on the commercial scheduling applications, but working and managing the data was a part of the obligations within the department. The scheduling software served as the producer of the parts that make up the GTFS. Any schedule changes that occur with an agency's system occurred within the scheduling software. For the agencies relying on this software, there were differences in how and who developed the data, which further impacted its position along the innovation spectrum. For example, one agency developed all the routes, trips and runs for the system within the scheduling software. However, the location and maintenance of the bus stop inventory was maintained elsewhere. Consequently, the stop inventory within the scheduling software was noticeably not up- to- date. Staff members responsible for maintenance of the schedule, routes trips and runs had little control over the quality and timeliness of the stop inventory. Organizationally the two departments did not develop protocols that supported each department, resulting in a GTFS that did not correctly reflect the stop locations for each route. This structure was common among agencies that exhibited lower levels of innovation within this group.

Agencies with slightly more innovation typically developed tools, scripts, small applications, and customized software tools to better manage the stops, routes, schedules and runs. These tools focused on streamlining the process to produce and manage the GTFS data. Nearly all the innovative solutions were used to streamline the process of developing the GTFS data. As stated earlier, many agencies, particularly at the beginning of the process, found the process very iterative and arduous.

Continuing along the spectrum of experiences, other agencies would develop intermediary tools and scripts that facilitated data integrity checks along the way. The staffing opportunities to develop tools and skills to create efficiencies existed within these agencies. Although there were still multiple responsibilities for these staff members, there was a clearer opportunity to develop and create systems and processes that improved the process of getting from the scheduling data into the GTFS. These agencies identified opportunities to streamline the process for generating GTFS data. Route, stop and schedule data continued to be managed using the commercial scheduling applications. However, staff members developed processes that integrated the development and creation of the GTFS data using additional software applications such as Microsoft Access and Excel. Other agencies used programming technology and scripting language to directly communicate with the scheduling software and generate the GTFS. These applications and scripts would simplify the process of generating the GTFS data. Agencies within this group would develop applications or scripts to help with the exporting and generation of the GTFS data. This streamlining would improve the process for generating and checking the schedule and route information. However, this did not remove the iterative efforts of getting data from the scheduling software into GTFS format, it only made it easier and more automated. In some instances, the agencies used MS Access to develop that data in the best format- to include improved head sign data, in particular. The headsign data is used by the Google application to report the location and direction where the bus is going for the trp. This helps customers using the tool to plan trips. Other agencies would develop scripts that are part of the scheduling software. These scripts would communicate with the scheduling software database. Using the scripts, the agencies are able to customize how the data is extracted from the software. Validation errors are less frequent for the agencies that employed scripting language.

Agencies with the skill and the time to develop stand-alone mini applications in programming environments like php, Java, Python, and .NET exhibited the highest innovation. These applications directly communicate with the scheduling database and extract, format and compile the information into the GTFS. Typically, these applications are part of a data integrity process put in place to support transit APTS applications such as APC and AVL systems. These systems improve their processing and development time significantly. Nonetheless, these agencies experienced the same problems with the data being incorrect and requiring additional data improvements. However, the frequency of the errors and the total time to produce the GTFS is reduced.

There were several examples of innovative applications and tools created by this group of agencies which include:

1. Exporting tools such as scripts and programs built on Java and the .Net environment that communicated with scheduling software to generate the GTFS.
2. Additional data tables were developed to coordinate the GTFS data with automatic vehicle location (AVL) systems and real-time systems
3. GTFS Calendar Tool – this tool developed by San Diego Transit was used to identify the impact of service changes through a calendar. The tool allowed planners to see how service has changed through a summary of services for each day. This tool is also used to understand changes to service over holidays. Figure 6 below illustrates the application with description of the service changes visualized on the calendar.

Figure 7: Calendar Checker from San Diego

Feature	Calendar Checker Interpretation Key
1	On each line is the service_id and then in parentheses the number of trips and the number of stop_times for that service_id.
2	A line with + and green text signifies an exception from calendar_dates.txt where service is being added for that day.
3	A line with a – and red text signifies an exception from calendar_dates.txt where service is being removed for that day.
4	A line without + or – signifies regular service specified in calendar.txt.

This tool allows the planners and schedulers to get a snap shot of the service provided on a given day and to identify the service changes due to special events, holidays and other events that might require a change in service.

The agencies in the Mature IT category displayed higher technical skills and understanding of data integration. Several of the agencies had enterprise systems which ran parallel to their efforts, and sometime the scripts and applications communicated with the enterprise system. The agency staff within this category requested several tools to further their efforts to leverage transit data to support planning and operational activities, they include:

1. Use GTFS and AVL system to develop schedule adherence tool
2. Integrating GTFS with Automatic Passenger Counter (APC) data
3. Visualization for system-wide performance evaluation
4. Tools to manage and organize driver runs and shifts.

Enterprise Agencies

The agencies interviewed from the final group provided different feedback from the other two groups. Agencies from this group did not see GTFS as a repository from which to leverage, but rather an output from a larger repository of data.   All the agencies within this group indicated that the enterprise data system would be used to leverage their data to support other activities. In fact, these agencies identified providing schedule data for online trip planners as a way they could leverage their enterprise data. Agencies in this category developed tools that interact with the enterprise system. The development of the TimeTablePublisher (TTPUB) was an outgrowth of an existing system that generated time tables from the enterprise system. Consequently, the tools created from these systems were far beyond the reach of this project. However, agencies within this category recognized the power that the GTFS data format offered for developing web enabled tools. Recognizing this had been a great deal of movement within these agencies to share their GTFS data with outside developers, which in turn expanded the utility of the data at no cost to the agency. There were several agencies that mentioned the power of the web enabled applications, and identified the Mapnificent application mentioned earlier to identify service characteristics and travel sheds. The interviewee suggested that tools similar to TransitScore and Mapnificent can be used by policy makers to communicate the impact service cuts had on communities and lost opportunities. The suggestions for GTFS data from this group were:

1. Create a visualization tool for policy makers and planners to see the impact of the service.
2. Integrate performance data with the GTFS. Again, the power of the web-based visualization was the core of this recommendation. This performance data included integration with passenger loads, boardings and alightings, and real-time and schedule adherence.

All the transit agencies provided insights into how the GTFS data could be used to support other activities. At all levels of technical sophistication, there was an overall recognition of the need to expand the use of technology to improve data integrity and as a tool to distribute data to other members of the organization. The transit agencies provided many suggestions and insights as to how expanding the GTFS can achieve some efficiency. However, the agencies did not provide a robust list of applications for leveraging and using the GTFS to support planning and operations. The research team

developed a list of opportunities that added to the wish list and examples of applications provided by the agencies.

Chapter 4 - Opportunities to Expand GTFS

The transit agency interviews revealed various opportunities to leverage the GTFS to support planning and operational activities. It is important to note the GTFS contains schedule data but no situational data, such as current location, passenger loads or arrival time. It exclusively contains information about planned service. Consequently, to expand the use of the GTFS to evaluate a route's performance would require supplemental transit APTS data. However, opportunities to evaluate service levels based on GTFS exist. The transit industry has identified multiple measures to evaluate levels of service. Transit service evaluation and performance measures provide opportunities to use the GTFS for purposes beyond trip planning. The opportunities are categorized into two groups:

1. Leveraging existing GTFS data
2. Supplementing the GTFS data with other organizational data

These categories highlight what agencies with GTFS can do immediately with its existing GTFS dataset versus what can be done with the GTFS when it is combined with other data.

Leveraging Existing GTFS Data

Perhaps the single most significant opportunity to utilize the GTFS is to support service planning and evaluation. Service planning ensures a transit system is able to effectively and efficiently meet the demand of its customers. Transit service levels can be evaluated using GTFS data and service standards established in The National Center for Transit Research (NCTR's) Best Practices in Transit Service Planning and TCRP's Synthesis 10, Bus Route Evaluation Standards. The measures can be categorized into to two groups:

1. service evaluation and
2. performance measures

Performance measures, for the purposes of this report, reflect the activity of the transit service. The data within the GTFS contains scheduled services but no performance data. Consequently, the GTFS cannot report on the performance of a route but it is able to provide data to evaluate the service of the route. The following represents measures that can be based on data residing within the GTFS.

Service Evaluation

Service evaluations are measures that examine the design, layout and schedule of a transit system. These are static and planned measures that help agencies ensure its service is designed to run efficiently and effectively.

Classification Systems

Service evaluations examine the design of a transit route. Typically these evaluations are stratified by service types. Classification systems facilitate the evaluation of routes based on its function. This

ensures that route and system evaluations properly compare routes that provide similar service types. For example, an express route would provide limited service coverage compared to a downtown circulator. Performance indicators for those routes should not be compared due to the different service it provides to the customers. Currently, route type delineations are expressed in the GTFS table routes.txt. However, these types are not functional classifications but mode classifications.   Below are the varying modes within the GTFS routes.txt table:

Table 3: Route Type in GTFS

Data Value	Type Description
0	**Tram, Streetcar, Light rail**. Any light rail or street level system within a metropolitan area..
1	**Subway, Metro**. Any underground rail system within a metropolitan area
2	**Rail**. Used for intercity or long-distance travel
3	**Bus**. Used for short- and long-distance bus routes.
4	**Ferry**. Used for short- and long-distance boat service
5	**Cable car**. Used for street-level cable cars where the cable runs beneath the car.
6	**Gondola, Suspended cable car**. Typically used for aerial cable cars where the car is suspended from the cable
7	**Funicular**. Any rail system designed for steep inclines

Evaluations based on the route type within GTFS would be insufficient due to the lack of detail needed to compare services within each route type. However, the GTFS can be used to evaluate a time based classification system. Generally, there are five commonly used time based classifications:

1. Peak
2. Non-Peak
3. Night
4. Saturday
5. Sunday

Routes can be categorized using data within the stops_times.txt table and the arrival_time and departure_time fields. Using the data within the GTFS can help evaluate routes based on a time based classification system. To utilize other classification systems would require the creation of additional data tables to be linked with the GTFS.

Service Availability
The extent and level a passenger has access to a transit route is a measure of service availability. Typically these measures are based on:

1. Service Area Characteristics

2. Service Coverage
3. Route Layout and Design and
4. Stop Location and Spacing

Internally the GTFS does not contain any data that directly addresses service availability. Nonetheless, the data can be used with a desktop or web enabled geographic information system (GIS) to evaluate service availability.

Service Area Characteristics

Identifying demographic characteristics of a route and system ensures the service is aligned with where the potential customers are located. The calculation of the total, percent and density of population and employment characteristics within the service area is possible with the data residing with the GTFS. This type of evaluation calculates the total population, population group, (i.e. population living in poverty) or employment within a given service area buffer, typically one-quarter and or three-quarter miles on either side of a route. This evaluation would require the use of additional technologies, such as GIS or a spatially enabled database.

Service Coverage

The service coverage is a measure of the extent a transit system or route is serving its customers. Again, the GTFS does not directly contain service population data, but it does contain data that can be used to calculate and evaluate the extent of the service coverage. A typical service coverage measure is percent of population within ¼ mile of a transit route or stop. This evaluation would require the use of a GIS to interpret the text tables which contain spatial data. Using the data from the shapes.txt, routes.txt and trips.txt the route and its alignment can be produced and a service area calculation using GIS.

Time and Distance Calculations

The traditional service availability measures of coverage and service characteristics were limited in that it did not account for a time component. Temporal information would provide additional information that would make the service availability measure more meaningful. For example, how much of a population group (i.e. persons living in poverty) can access employment opportunities within sixty minutes using public transportation.

Route Layout and Design

This design standard considers spacing and location between routes. Distance between routes should be considered to avoid creating duplicated service. Much like the other evaluation measures, the GTFS data can be used to conduct this evaluation, but it is not inherently located within the GTFS dataset. Using GIS or other spatially enabled database tools, the distance or the average distance between routes can be evaluated. The specifics of this measurement would have to be more clearly defined but the data within the GTFS table, shapes.txt provides the points that represent the polyline of the route.

Using a combination of tables, shapes.txt, routes.txt and trips.txt the route and its alignment can be produced and evaluated in a GIS.

Stop Location and Spacing

Bus stop spacing is the distance between adjoining service stops on a route. Spacing standards are expressed in one of two ways, stop per mile or distance between stops. Agencies trying to strike a balance between accessibility and vehicle speed can use the GTFS data to calculate and evaluate the distance between stops. Poorly spaced stops can increase the overall travel time and consequently reduce transit demand. Agencies considering stop consolidation and realignment are able to identify specific stops that violate a spacing standard. Table 1 is an example of stop spacing standards.

Table 4: Stop Distance Standards

Type of Service	Distance between stops (feet)		Stops per mile	
	Minimum	Maximum	Minimum	Maximum
Local routes in residential areas	600	1250	4	8
Commercial Sites	500	1250	4	10
Limited Stop zone	2500	8000	0.75	2

Calculating the distance between stops using the GTFS data is clearly possible. However, it is not a trivial undertaking if an agency does not have the staff to dedicate time and effort to develop a process for evaluating this data from the GTFS. However, given that many spacing standards are reliant on service type classifications the need to develop a route classification table would be necessary. Nonetheless, with growing use of geographic information systems (GIS) and the development of web based tools to perform spatial analysis it is very achievable.

Route or Service Directness

Service or route directness is mathematical efficiency measure that examines the "degree to which a route deviates from the shortest path between the start and end points of the route." (Gregg et al., p. 21) Agencies using this measure is must balance time efficiency (shortest and quickest path) with service area (i.e. ¼ mile buffer of a route). Frequently route directness is measured as factor of increased time or distance. In other words, a route deviation must not increase the distance and or travel time for a customer by a particular factor or value. In other instances, the factor is based on travel time compared to an automobile, for example:

> "Route deviation is also controlled by limiting transit travel times to a maximum of twice the automobile travel time for the same trip regardless of the number of such route deviations that exist." (Bus Route Evaluation Standards: TCRP Synthesis 10, p. 12)

GTFS data provides the location of a beginning and end point of a transit route. Web based applications (graphserver and open trip planner) as well as desktop GIS applications, can calculate the transit travel time between the two end points of a route or trip. These results would represent the "ideal" transit travel time, based on the scheduled travel time. The results can be compared to shortest path tools that reside within both desktop and web based applications.

Travel Time and Capacity

Travel time and capacity measures relate to the "service frequency, service directness, … and span of service." (NCTR Service Standards, p.19) Given that the GTFS contains schedule data several of these measures can be evaluated using the GTFS.

Service Frequency

Service frequency measures the number of times a bus stop is visited in a period of time. This measure is often expressed as headway, referring to the time interval between two successive departures. Often agencies will employ a headway policy which describes the agencies' headway standard. Frequently these standards are divided based on the service type (local, commuter, trunk, BRT) and time of day (i.e. peak, off-peak and late night). Given that the GTFS contains data on the time and trip information it is possible to identify the scheduled service standards of a given GTFS dataset.

Number of Transfers

Number of transfers is another measure of service directness: the most direct transit service has the fewest number of transfers. Transfers add to the total travel time for a customer and in turn impacts transit demand. Total transfers between two points via public transit can be calculated through the GTFS data.

Using the GTFS data to provide service level evaluations provides transit agencies the opportunity to leverage the GTFS data to better understand and evaluate its community's service.   Performance evaluations require the integration of supplemental data, such as automatic passenger counters, and automatic vehicle locators. Using data from transit APTS applications would provide additional measures. Evaluations of performance such as on-time performance, passenger trips per hour, revenue miles and passenger miles are indicators that the GTFS cannot currently support. The requirement of additional transit data presents a challenge for some agencies that do not employ transit APTS applications. However, for agencies that are able to relate GTFS data and transit APTS data together are able to leverage the GTFS data to provide and calculate performance measures.

Span of Service

Span of service relates to system and route capacity. It reflects the number of hours, beginning and end of service provided by route, day of the week, and system. The evaluation of the span of service can be done using the GTFS data. Using database tools the information can be extracted by calculating the

minimum and maximum stop times from the stops_times.txt table for a given route. This would also require the joining of the trips.txt and routes.txt tables.

Supplementing GTFS Data

Given the robust nature of the GTFS it can be an ideal data platform to link additional data for the purposes of service evaluations and performance measures.

Service Evaluations

Understanding service evaluation measures can provide transit planners an opportunity to assess the location and schedule of service. As stated earlier, these evaluations can be based on data residing within the GTFS. However, if the GTFS data is integrated with other data, then the capacity to leverage the GTFS to support other activities is greatly increased. For the purposes of this report service evaluations address the scheduled service. Service delivery standards address such measures as, on-time performance, passenger amenities, customer service and safety issues. (Gregg et al., p.28)

Service Amenities

The presence of transit stop amenities provides planners with a measure to evaluate the quality of the service provided. The GTFS does not currently contain stop amenities. To measure the extent of amenities provided by an agency, system or route, additional data would be required. The GTFS currently contains a unique record for each stop within the stops.txt table. Using this data, agencies can link and join the GTFS stops.txt data with a stop inventory table that contains the amenities and information to better evaluate this information. It is noteworthy that this added table is already considered by the GTFS Changes Group.

Customer Service Evaluations

Passenger complaints and comments about routes and services are a critical measure for agencies to consider. The ability to track and analyze complaints by trip, stop and route are important for an agency that attempts to limit the number of complaints. By utilizing the unique stops and trip identifiers that are within the GTFS, agencies can capture this data and link it with the stops, route, and trips within GTFS. Tracking and analyzing customer complaints provides the agency with tools to better address customer needs and concerns.

Safety Evaluations

Tracking and evaluating passenger safety is important to transit providers. The ability to track the location of accidents and integrate the accident records with the route, trip and stop stored within the GTFS would provide an extraordinary tool for agencies. By analyzing the accident information, by time of day and location (proximity to stops) agencies may better develop safety counter measures.

Performance Measures

The performance of a transit system is critical for agencies to develop and design effective transit service. As stated earlier, the GTFS data cannot evaluate the actual route and system performance. To accomplish this GTFS data would have to be joined with additional data. In most cases this data could come from transit APTS activities.

Passengers Per Hour

Passengers per hour is a performance measure that addresses route and system productivity. Using data from an agency's automatic passenger counters (APC) total passengers by hour can be calculated. Many APC systems have data reporting items; however, the ability to map this information would be very helpful. Mapping and visualizations of passengers per hour by trip, between timepoints, and route could be visualized using data residing within the GTFS.   Multiple variations of this measure can be derived from this metric.

Passengers Per Trip

The TCRP study identified the passenger per trip measure based on survey of transit agencies. Passengers per trip is the total number of riders on a one-way trip. (TCRP 10, p.19) Using the GTFS data and ridership data from an agency's APC, the calculation of passengers per trip can be evaluated.

Chapter 5 - Candidate GTFS Applications

The GTFS contains a rich source of information about a transit agencies service. Agencies are mostly using the GTFS to participate in the Google Transit program. For a variety of reasons the agencies have not recognized the opportunities to utilize the data for other purposes. The data can be used to get a better understanding of the service levels and can be used in combination with other data to create new information.

Developing a prototype software application will illustrate the opportunities to leverage the GTFS data. The agencies identified multiple opportunities to leverage the GTFS. These opportunities were reviewed by the project team and FDOT staff. The following section describes these opportunities.

GTFS Merging Tools

There are many tools that assist with merging GTFS datasets. Many of the concerns with merge functions were often issues with user errors. To develop a tool to help with this function would duplicate efforts that are addressed.

GTFS Mapping and Visualization Tools

Many of the agencies requested the ability to visualize and map GTFS data. Through the use of charts, maps, and tables service characteristics of a GTFS can be easily understood. Given the robust nature of GTFS community, there are several tools to help with visualizing and mapping GTFS data.

Mapping GTFS

Mapping environments include desktop GIS applications and web enabled geospatial applications. All applications must create an interpreter to translate the GTFS text files into a GIS ready format.   One such mapping application is Open Street Map, an open source web mapping tool. The Open Street Map data relies on crowd sourcing and community input. A tool developed by the University of South Florida's Center for Urban Transportation Research (CUTR) created a tool to import GTFS data into open street map.   The GTFS-OSM-Sync application imports the GTFS into the open street map web application. http://code.google.com/p/gtfs-osm-sync/

Figure 8: Open Street Map Representation of Bus Stops
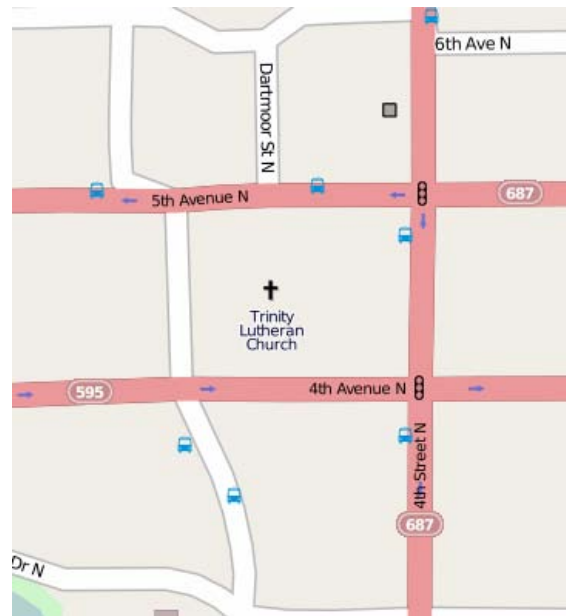
As stated earlier, there are many open source and web enabled tools to visualize GTFS data. An open source software project called Google TransitDataFeed contains multiple tools for reading and writing GTFS data. The project offers multiple tools to support the management of an agency's GTFS. One of the tools, The Schedule Viewer, uses Python to support Google Transit users to improve and trouble shoot the GTFS data quality. Through the use of maps and graphs the tool provides insight to errors in an agency's GTFS feed. Below is a screen capture of the mapping tool. The code with user manual can be found in Google Code repository:

http://code.google.com/p/googletransitdatafeed/wiki/ScheduleViewer .

Figure 9: Schedule Viewer Tools
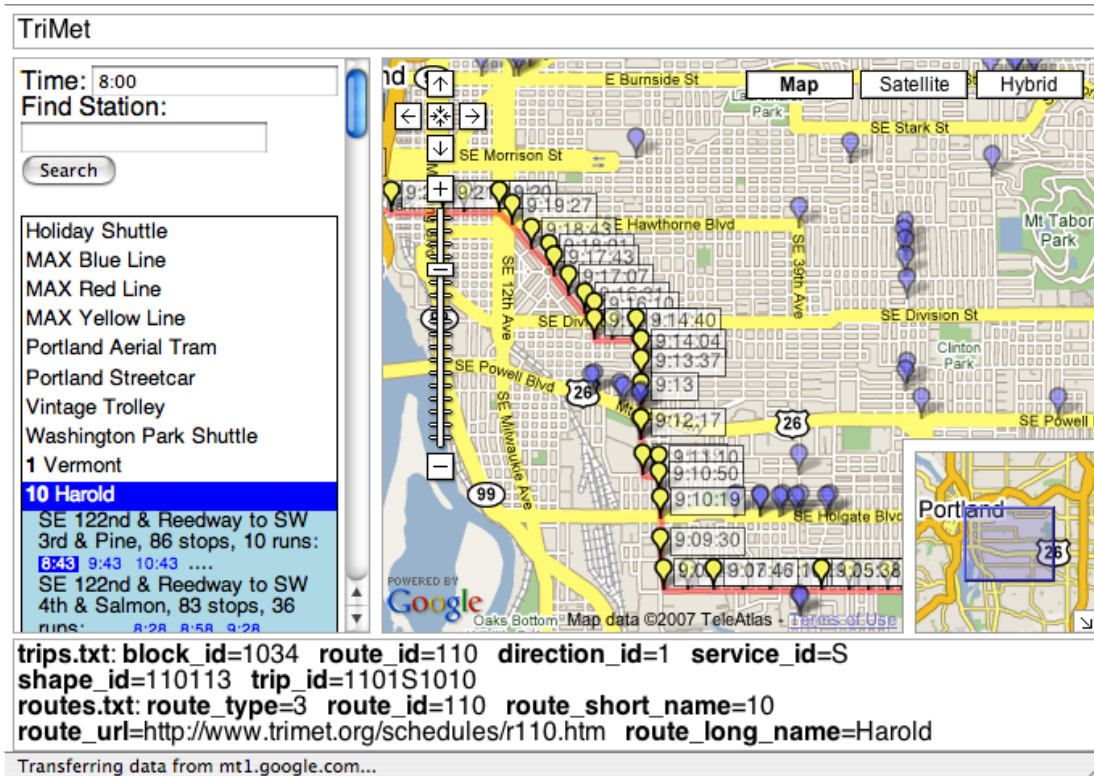
The GoogleTransitDataFeed project also provides tools to generate the GTFS in a KML format. The KML format can be viewed in Google Earth and Google Maps. The KML format is a standard format that can also be viewed using desktop GIS applications such as ESRI's ArcGIS. Further, the ability to map the GTFS data using desktop GIS is possible using data residing within the GTFS tables, shapes.txt and stops.txt. The stops.txt table contains the x, y coordinate for each bus stop. With little effort the ability to create point GIS data from x, y coordinate data is minimal. Creating the route is slightly more challenging given that the shapes.txt file contains the x, y coordinates for the vertices of the transit route. However, ESRI's ArcGIS and a free plugin for ESRI called ET GeoWizards for GIS (http://www.ian-ko.com/downloads/downloads.htm) can be used to develop the route alignments.   Using the fields shape_pt_lat, shape_pt_lon, and shape_point_sequence a user could create the string of points into the a polyline the accurately represents the route alignment.

GTFS Visualizations

Charts and graphs of GTFS data provide a snapshot of service that can be easily understood. The GoogleTransitDataFeed project also provides charting visualization of the GTFS within the schedule viewer application. This chart is designed to understand the distribution of trips and is visual check for

data continuity. This chart (Marey Graph) is very helpful in identifying patterns, and generally, consistent patterns suggest good data.

Figure 10: Marey Graph
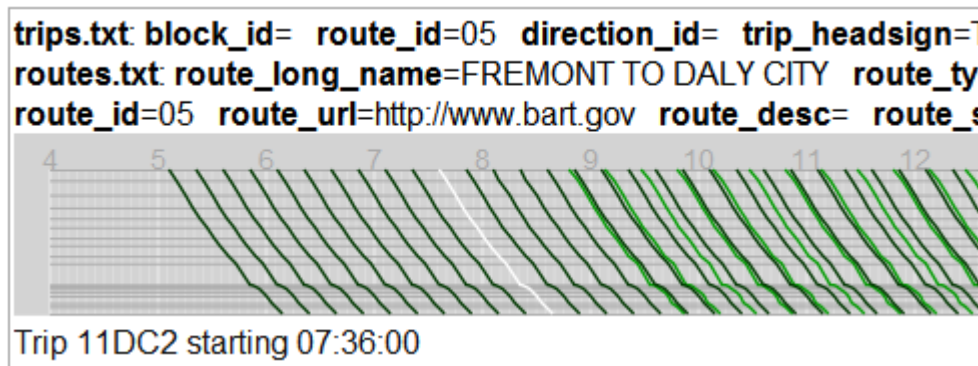
Another illustration of service levels comes from the technologist and public transportation blogger Kurt Raschke who used the GTFS data to represent vehicles per hour chart. This provides a clear representation of activity at a given transit stop. The tools used to develop these graphics are available for others to download and use at: http://transport.kurtraschke.com/2011/05/gtfs-visualize-service .

Figure 6: Vehicle Per Hour Chart
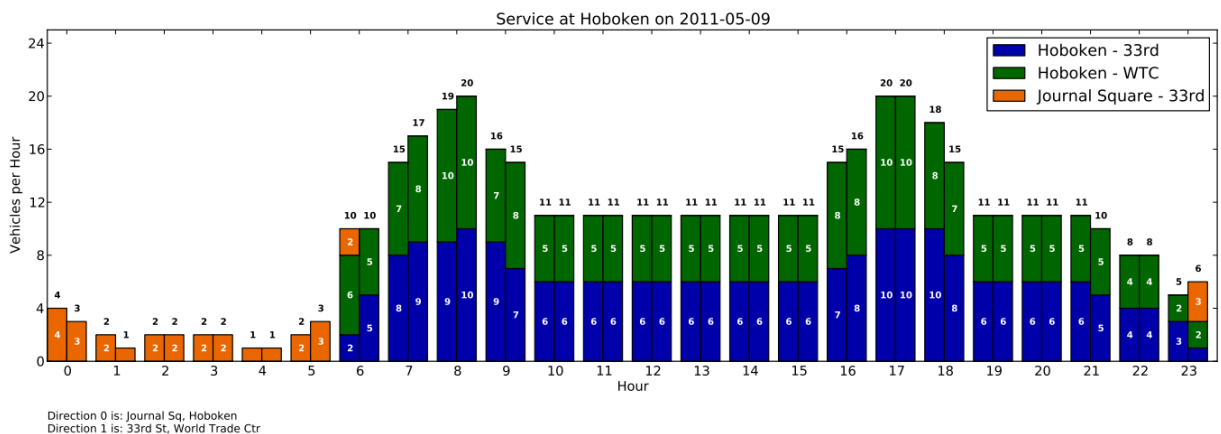
 Mapping and data visualizations are effective tools for communicating the robust data and information within the GTFS by providing clarity to service levels which the data doesn't naturally produce. This can help articulate the impact of an agency's service changes and service cuts.

Driver and Shift Management

Scheduling and run cutting applications serve as the source for many agencies' GTFS data. Given that these applications support driver and shift management issues, it seems natural for agencies to request

expanding the usefulness of these applications. However, it seems that many of these issues were grounded in complications with managing staff hours and shift responsibilities. Producing a tool to support this important function would serve the transit community well.   The complexity of the managing shifts and work hours would be too complex to develop under this project.

Customer Service Integration

Customer service departments often use Google Maps to help customer service agents. The ability to attached customer requests, complaints and comments to stops and routes in the system was requested by several agencies. Integrating GTFS data to allow for comments and issue tracking is a very powerful tool. This type of integration would require additional data tables to work. Currently the GTFS does not contain service complaints and comments. However, the power of the Google Maps to deliver transit customer comments would provide a web enabled tool which can be used by multiple employees to input, track and analyze.

Regional and System Performance Measures

Visualizing and evaluating transit performance measures using GTFS data offers great promise to expanding the utility of GTFS data. **Given the power of web enabled applications using GTFS data and the robust information within the GTFS made this very enticing. Further, there are several instances of web applications that can be built upon to fully develop such a tool. This recommended application came from nearly every category of transit agency.

GTFS Data to Support Service Planning

The ability to leverage the GTFS data to support service planning is an important goal of this research effort. Agencies that employ systematic service evaluations are able to use the data within the GTFS to support those efforts.   Typically performance measures are based on defined standards and principals. The following section identifies service measures and standards that can be evaluated using an agency's GTFS.

Evaluating system performance through service standards is an important tool for transit planners. When services and system modifications are being considered the use of services standards are invaluable.   Using data within an agencies GTFS multiple service standards can be evaluated. The following standards were identified in the "Best Practices in Transit Service Planning" report produced by the National Center for Transit Research.

Integration of GTFS with Desktop GIS Applications

Several agencies indicated the desire import and analyze GTFS in a traditional desktop GIS environment. The main goals were to gain greater control over the visualization of the data and perform geo-spatial analysis on the schedule data. Additionally, several other opportunities would arise from this ability.

This includes the ability to calculate the distance between stops, which can be used for stop and amenity management and consolidation.

Bus Stop Amenity Management

Many agencies realize the power of managing bus stop amenities. The GTFS currently does not support stop amenities, however there have been several proposals for storing the data. Managing amenities would integrate the GTFS with customer service, planning and maintenance. Several interviewees suggested the development of a web based invoicing system that reflected customer complaints, driver comments and planning inputs. This effort would require the development of additional data tables that would join with the GTFS. The Google Transit community on changes to the GTFS includes a proposal to include a stop amenity table. The proposal and the issues surrounding the debate can be found at: https://sites.google.com/site/gtfschanges/proposals/stop-amenity. Below is a snapshot of the proposed data structure.

Figure 7: Sample of Proposed Amenity Table

Value	Name
1000	Shelter
2000	Furniture
2100	Bench
2200	Trash Can
2300	Bike Storage
2310	Bike Rack
2320	Bike Locker
2400	Vending Machine
2410	Ticket Vending Machine
2420	Food Vending Machine
2500	Phone

Integration with Real-time Information Systems

Agencies with real-time information systems realize the power of integrating its real-time arrival information with Google Transit. There are several issues with integrating these systems with Google Transit. However, during the course of this project, Google Transit released an update that allowed for the integration of real-time information with Google Transit. Details and description of the new

specification can be found at: http://code.google.com/transit/realtime/. Developers have already created tools to import the real-time feed into a database. The database may serve as a platform for evaluating performance measures such as schedule adherence and on-time performance. This development may show a great deal of promise for furthering the tools available to agencies.

Integration with Automatic Passenger Counter Data

The power of web visualization was the foundation of this request. Many agencies wanted to be able to combine the power of the web interface with visualizing ridership data. Given that much of the source data that feed into the GTFS is also used with AVL/APC applications several agencies identified this as a requested item.

Each of these application ideas has varying merit. Several criteria were considered when identifying which application will be developed. Among the criteria were time/cost to develop the application, benefit to developing application and input from FDOT.

Through consultation with FDOT and the project team, it was decided to develop a tool that would combine related datasets to expand the use of the GTFS data. The tool would create a web-based tool that would integrate automatic passenger counter (APC) data with an agencies GTFS data.

Developing Prototype Application

Much of the data within the GTFS is about scheduled service. Many service level measures can be obtained from the GTFS. However, understanding actual performance is a limitation of the GTFS data. Given that limitation, the project team intended to identify an opportunity that adds to the GTFS data by linking ridership data from automatic passenger counters to an agencies GTFS data.

By integrating GTFS with APC data, agencies are able to evaluate additional performance measures such as ridership activity by system, route, trip and stop. Other measures such as passengers per hour and passengers per trip can be evaluated by combining APC data with GTFS.

The project team acquired APC data from three separate transit agencies. Each of the agencies agreed to share its APC and its corresponding GTFS data for a given time frame. The use of APC data presented several initial challenges. The first and foremost was the alignment of the data. In order for the data to be matched up it was necessary to identify the fields which the two datasets matched. Furthermore, the APC data from the agencies have different degrees of "granularity". That is to say, some datasets are summaries to reflect the average boarding and alighting over a given time. The differences occur because some agencies have a full fleet of busses equipped with APC data and in other cases the data is a sample. For agencies that have a sampled data, there are missing trips of data, and the information is weighted based on a sampling technique and proprietary algorithms to determine the average daily boarding for trips and stops that were not surveyed. The other datasets have raw trip and stop totals but not for the entire system. They were based on single or unknown sampling rates.

Application Framework

The project team developed a web based application that is intended to utilize both APC data sets.   The first challenge was to develop a database schema that would accommodate both data types while still reporting on ridership information.

The second hurdle was identifying relevant ridership data that would be useful to transit agencies and FDOT. Based on finding from the literature review the following variables were identified and were the basis of the features of the application. Below are the features to be included in the application:

### *Mapping of Boarding and Route Activity*

The ability to visualize transit route activity is significant. Visualizing transit activity in a geographic context provides more insight in the performance and activity of a given route, trip or stop. Figure 12 illustrates the ridership dashboard. The dashboard is intended to quickly identify ridership activity.

Figure 8: Transit Ridership Dashboard
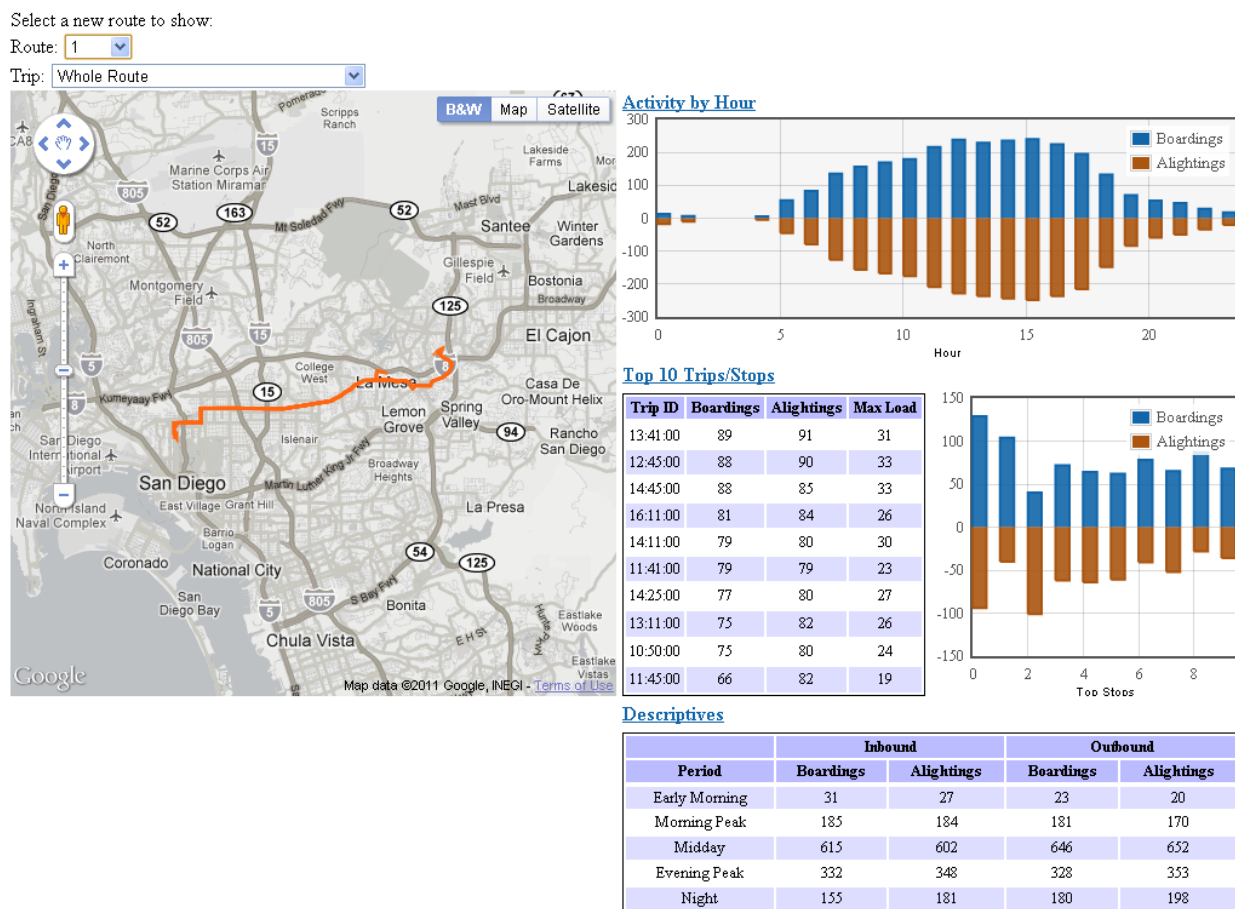
The dashboard is intended to be interactive to allow the user to access different levels of information. Through the use of pull down menus a user is able to evaluate ridership data for each route and trip. Figure 13, illustrate the pull down menus to identify data for separate routes and trips.

Figure 9: Dashboard Pull down Menus
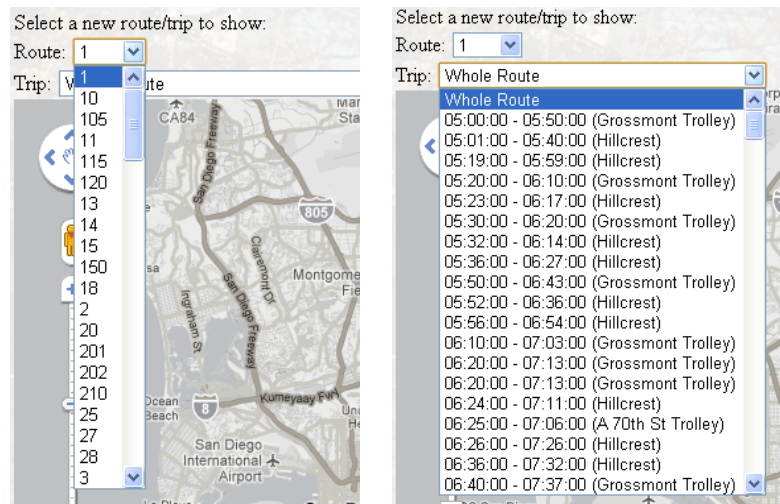
Ridership by Route

The ability to get a quick snapshot of the activity of a route is important. The ridership will be average daily across all trips of boardings and alightings. Route ridership is made up of the map of activity and a chart of boardings and alightings. The representation of the route in the mapping portion of the application, utilizes the GTFS table shapes.txt and route.txt to ensure the APC data accurately reflect the data in the GTFS.
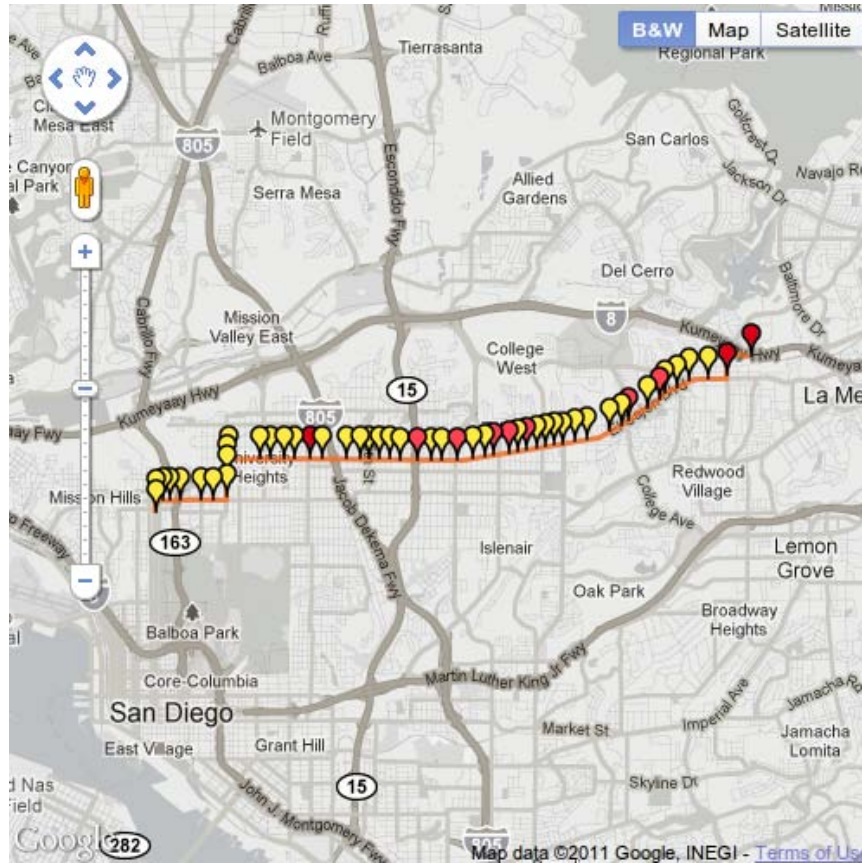
Figure 10: Trip Level Boarding Activity

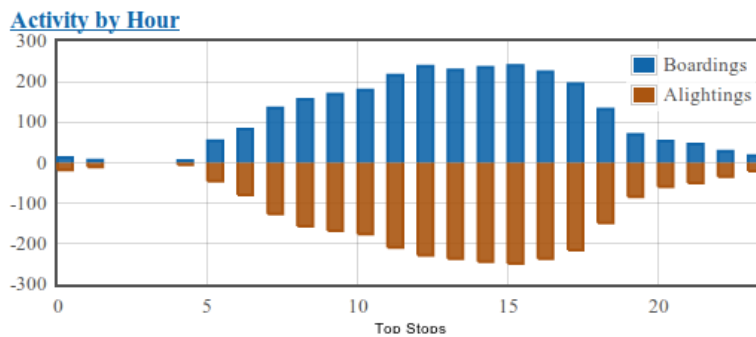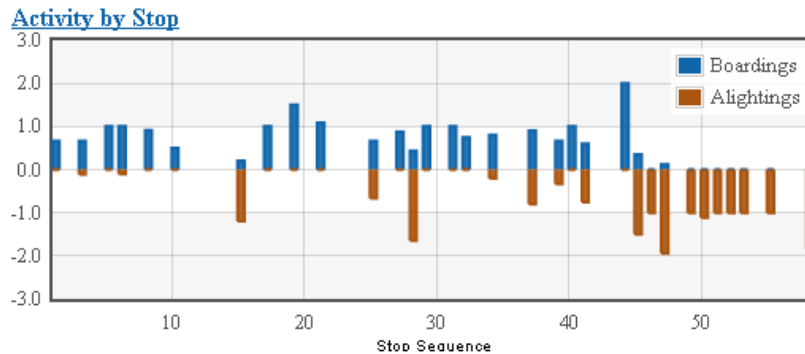**Figure 11: Route Level Activity by Hour**

**Figure 12: Trip Level Activity by Stop Sequence**

Ridership by Trip

Ridership by trip allows the user to identify the multiple trips which make up the transit service. This information is critical for users to be able to identify well performing trips and can help with efforts to consolidate and cut service. Additionally, with load information included in the APC data it also allows for users to identify trips with high loads so perhaps service changes can be made to accommodate high load factors.

Ridership by Stops Per Route and Trip

The identification of ridership at the stop is important. The data shows stop activity based on the route and trip. This again is helpful with planning efforts.

Stops with Highest Ridership

Identifying stops with high ridership provides agency insights that help with asset management and resource allocation. By identifying the stops with high activity agencies are able to ensure that improvements are targeted at stops that service the customer the most. The web application provides the user the ability to identify the stop ID with a mouse over. Further, by clicking on the bar, the google map zooms to the stop and highlights it allowing the user to match the highest performing stops with geospatial alignment.
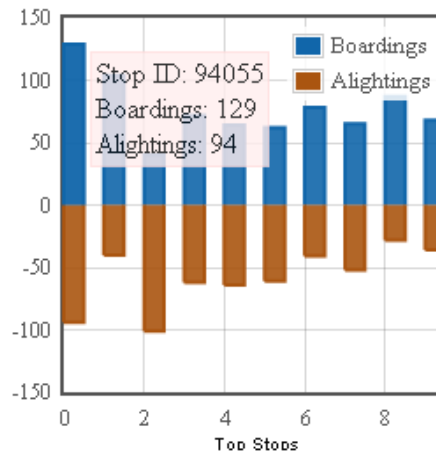
Trips with Highest Ridership

Identifying the trips with the highest activities allow planners to identify trips that require more capacity or schedule opportunities. Currently the trips only identify trips within a route.

Trip ID	Boardings	Alightings	Max Load
13:41:00	89	91	31
12:45:00	88	90	33
14:45:00	88	85	33
16:11:00	81	84	26
14:11:00	79	80	30
11:41:00	79	79	23
14:25:00	77	80	27
13:11:00	75	82	26
10:50:00	75	80	24
11:45:00	66	82	19

Further the web application allows for a drilling down of the data to see the ridership activity by stop. By using this graph the user is able to quickly assess the activity of a particular trip.

Route Ridership by Hour

For each route with APC data this graph illustrates hourly activity for the route. This gives the user the ability to visualize the trend and activity for a given route.

Activity by Hour
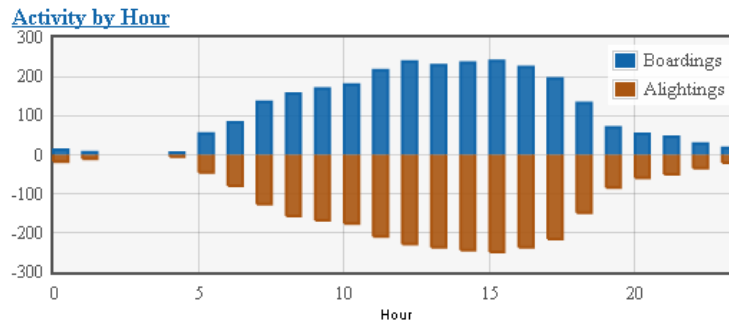
Prototype Application Findings

The prototype application was designed to illustrate the power of expanding the GTFS. By integrating the service level data within the GTFS data with performance related data the power of the two datasets are illustrative of how the GTFS can be leveraged.

It is important to note that this type of application may not be of service to agencies that have robust enterprise systems which may have the capacity to produce this data. However, for most agencies that do not have the large information technology support to produce enterprise wide data systems this tool may provide a mechanism to regularly evaluate service levels and activity.

Further, the application is limited by the availability of APC data in a format that can be joined with the agency's GTFS data. There are no accepted data models for APC output, and although the datasets are very similar, an APC validation tool would be required to allow this prototype to be fully operational. Nonetheless, the application provides a foundation from which to begin this process. It is important to note that the application was developed in the open source, so all the code and data formats are available for download and use by all in the user community.

Chapter 6 - Findings and Conclusions

As public transportation agencies continue to manage with fewer resources it is important for agencies to take advantage of technologies to create a transit service that attracts more and retains existing customers. The creation of free web based transit itinerary planning from Google, Bing and others have created an opportunity for agencies to simplify transit trip planning. Agencies that have embraced these customer facing resources have needed to improve data quality to provide the web based trip planners with robust, timely and accurate data. The GTFS format helps agencies do this in a straightforward and highly documented manner. Participating in Google Transit has created a secondary benefit for agencies, which is an up-to-date inventory of its transit service. The inventory and the resulting GTFS data can be used to support additional activities within the organization.   In particular, the GTFS can be used to evaluate typical service measures such as, service availability, service directness, service frequency, span of service and service coverage. Furthermore, the GTFS data incorporates temporal data elements that add a new dimension of evaluation that was previously challenging to perform. By combining location and time elements, the GTFS can be used to measure accessibility based on time and location. For example, agencies using GTFS are able to calculate the population that can commute using public transit to a downtown within 30 minutes. Measures such as these were not readily available due to the lack of data. The GTFS has created a new opportunities to evaluate transit service.

The ridership prototype application is designed to illustrate how the GTFS can be leveraged with other transit APTS data to produce data in a greater context. The power of data to support decision making is evident for agencies. As cuts and service changes are proposed, agencies that combine the ridership data with the GTFS data are able to identify the routes and trips that perform the poorest. Consequently, the changes can be designed to impact the fewest number of riders and to maintain the most.

As stated in earlier sections, the GTFS data allows agencies to evaluate service levels through service planning activities such as calculating the number of transfers to a destination, calculating population and employment service area, evaluate stop spacing, evaluating route and system span of service, route directness and travel time isochrones. However, a GTFS is limited to evaluating scheduled service and not service performance. The ability to address route and system performance requires transit APTS technologies such as APC, AVL and real-time arrivals. Transit APTS technologies and GTFS require the same data integrity and rigor, and in many cases the same stop, routes and schedule data repository are used to produce and interact with both systems. Consequently, GTFS data and transit APTS data can be joined to reflect and analyze some performance measures.

The research illustrates the opportunities to use GTFS data to support operational and planning activities.   The GTFS is a data specification that provides service to transit agencies. The data drives the service that Google Transit provides but it also provides an incentive for the agencies to utilize the data structure.   The public transit industry should learn from this model by designing tools that support a

direct activity will result in data specifications that support the utility of the tool. For example, a bus stop inventory tool designed to help agencies evaluate Title VI and Americans with Disabilities Act (ADA) compliance, would require a clearly prescribed data structure. If such a tool provided service to the industry then they would follow the data structure. The result would be a normalized data format that can be used to evaluate and compare services and progress. Further research should be conducted to identify opportunities to evaluate and create tools that are built around the GTFS.

References

Bertini, R., & El-Geneidy, A. (2003). Generating transit performance measures with archived data. *Transportation Research Record, 1841*(1), 109.

City-Go-Round. Retrieved April 1, 2010, from http://www.citygoround.org/

Engineer's life - US Jobs- Google. (2010). Retrieved February 23, 2010, from http://www.google.com/jobs/lifeatgoogle/englife/

Ferris, B., Watkins, K., & Borning, A. (2010). Location-aware tools for improving public transit usability. *IEEE Pervasive Computing, (9)*1, 13-19. Retrieved February 17, 2010, from http://onebusaway.org/research/

Google transit partner program. Retrieved February 12, 2010, from http://maps.google.com/help/maps/transit/partners/index.html

Kimpel, T., Strathman, J., & Callas, S. (2004, April). Improving scheduling through performance monitoring using avl/apc data. Retrieved August 29, 2011, from http://www.its.pdx.edu/upload_docs/1248894227awBVBIdrTc.pdf

Public transit via Google. Retrieved February 23, 2010, from http://googleblog.blogspot.com/2005/12/public-transit-via-google.html

Radin, S., Jackson, D., Rosner, D., & Peirce, S. (2002). *Trip planning state of the practice*. FTA-TRI-11-02.6. Volpe Transportation Systems Center; Federal Transit Administration; Federal Highway Administration.

Transit Cooperative Research Program. *AVL systems for bus transit* (TCRP Synthesis 24). National Academy Press, Washington, D.C., 1997. Retrieved August 31, 2011, from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_syn_24.pdf

Transit Cooperative Research Program. *AVL systems for bus transit: Update* (TCRP Synthesis 73). Transportation Research Board of the National Academies, Washington, D.C., 2008. Retrieved August 31, 2011, from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_syn_73.pdf

Transit Cooperative Research Program. *Leveraging ITS data for transit market research: A practitioner's guidebook* (TCRP No. 126). Transportation Research Board of the National Academies, Washington, D.C., 2008. Retrieved February 12, 2010, from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_126.pdf

Transit Cooperative Research Program. *Passenger counting systems* (TCRP Synthesis 77). Transportation Research Board of the National Academies, Washington, D.C., 2008.

Retrieved August 31, 2011, from
http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_syn_77.pdf

Trépanier, M., Morency, C., & Agard, B. (2009). Calculation of transit performance measures using
smartcard data. *Journal of Public Transportation, 12*(1), 79-96.

U.S. Department of Transportation. Intelligent Transportation Systems. *Archived data management
systems: A cross cutting study linking operations and planning data*. Washington:
Government Printing Office, 2005. Retrieved August 29, 2011, from
http://ntl.bts.gov/lib/jpodocs/repts_te/14128/14128.pdf

Watkins, K., & Ferris, B. (2010). Explore: An attraction search tool for transit trip planning.
Proceedings of the 2010 Transportation Research Board Annual Meeting. Retrieved
February 17, 2010, from http://onebusaway.org/research/

Bibliography

1. APTA Research and Technology Committee. *Research and technology strategic plan*. American Public Transportation Association, Washington, D.C., 2009.
2. Burt, M.W., Cluett, C., Schweiger, C.L., Coogan, M.A., Easley, R.B., & Easley, S. *TCRP Report 84: e-transit: Electronic business strategies for public transportation, Vol. 8: Improving public transportation technology implementations and anticipating emerging technologies.* TRB, National Research Council, Washington, D.C., 2008. Retrieved February 17, 2010, from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_84v8.pdf
3. Bus Stop Inventory Task Force (2000). Bus stop inventory best practices and recommended procedures. Transit Standards Consortium, Inc.
4. Cambridge Systematics, Inc. (2009). Enhancing the american community survey data as a source for home-to-work flows. Retrieved August 29, 2011, from http://www.statewideplanning.org/_resources/254_NCHRP-08-36-81.pdf
5. Cervero, R. (1990). Transit pricing research: A review and synthesis. *Transportation, (17)*2, 117-139. Retrieved February 19, 2010, from ProQuest database.
6. Corporate Information. (2010). Retrieved February 23, 2010, from http://www.google.com/corporate/history.html
7. *Evans, J. Chapter 9 - Transit Scheduling and Frequency. In *Traveler Response to*
    a. *Transportation System Changes*. (TCRP 95). Transportation Research Board, Washington D.C.: 2004.
8. Ferris, B., Watkins, K., & Borning, A. (2010). OneBusAway: Results from providing real-time arrival information for public transit. *Proceedings of CHI 2010*. Atlanta, GA, USA, April 10-15, 2010. Retrieved February 16, 2010, from http://onebusaway.org/research/
9. Ferris, B., Watkins, K., & Borning, A. (2009). OneBusAway: A transit traveller information system. *Proceedings of Mobicase 2009*. San Diego, CA, USA, October 26-29, 2009. Retrieved February 16, 2010, from http://onebusaway.org/research/
10. Furth, P., Hemily, B., Muller, T., & Strathman, J. (2003, June). *TCRP Web Document 23: Uses of archived AVL-APC data to improve transit performance and management: Review and potential.* Transit Cooperative Research Program, Transportation Research Board of the National Academies. Retrieved August 31, 2011, from http://gulliver.trb.org/publications/tcrp/tcrp_webdoc_23.pdf
11. Golani, H. (2007). Use of archived bus location, dispatch, and ridership data for transit analysis. *Transportation Research Record: Journal of the Transportation Research Board, 1992*(1), 101-112.
12. Harder, B.T., & Tucker, S.L. *Scoping study for a national strategic plan for transportation information management.* National Cooperative Highway Research Program, TRB, National Research Council, Washington, D.C., 2003. Retrieved February 16, 2010, from National Transportation Library Web site: http://ntl.bts.gov/lib/23000/23900/23972/scoping_study_final_report.pdf
13. Horning, J., El-Geneidy, A., & Hourdos, J. (2009, December). Estimated running time and demand for a bus rapid transit corridor. Retrieved August 29, 2011, from http://www.its.umn.edu/Publications/ResearchReports/pdfdownload.pl?id=1266

14. Kizoom, N., & Miller, P. (2008). *A transmodel based XML schema for the Google Transit Feed Specification - with a GTFS/Transmodel comparison*. Retrieved February 12, 2010, from http://www.dft.gov.uk/transmodel/schema/doc/GoogleTransit/TransmodelForGoogle-09.pdf

15. National Cooperative Highway Research Program. *Performance measures of operational effectiveness for highway segments and systems* (NCHRP Synthesis No. 311). Transportation Research Board of the National Academies, Washington, D.C., 2003. Retrieved August 29, 2011, from http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_syn_311.pdf

16. Okunieff, Paula (2006, March). *Dynamic timetable generator* (IDEA Project 39). Transportation Research Board of the National Academies, Washington, D.C., 2006. Retrieved August 29, 2011, from http://onlinepubs.trb.org/onlinepubs/archive/studies/idea/finalreports/transit/Transit39_Final_Report.pdf

17. Southworth, F., Vogt, D., Curlee, R., Chatterjee, A., & Wegmann, F. (2002, March). An assessment of future demands for and benefits of public transit services in tennessee. Retrieved August 29, 2011, from http://cta.ornl.gov/cta/Publications/Reports/R02-114617.pdf

18. Taylor, B., Iseki, H., Miller, M., & Smart, M. (2009). Thinking outside the bus: Understanding user perceptions of waiting and transferring in order to increase transit use. California PATH Research Report UCB-ITS-PRR-2009-8. Retrieved February 19, 2010, from http://www.path.berkeley.edu/PATH/Publications/PDF/PRR/2009/PRR-2009-08.pdf

19. Transit Cooperative Research Program. *Bus route evaluation standards* (TCRP Synthesis 10). National Academy Press, Washington, D.C., 1995. Retrieved August 31, 2011, from http://onlinepubs.trb.org/onlinepubs/tcrp/tsyn10.pdf

20. Transit Cooperative Research Program. *Geographic information systems applications in transit* (TCRP Synthesis 55). Transportation Research Board of the National Academies, Washington, D.C., 2004. Retrieved August 31, 2011, from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_syn_55.pdf

21. Transit Cooperative Research Program. *Real-time bus arrival information systems* (TCRP Synthesis 48). Transportation Research Board of the National Academies, Washington, D.C., 2003. Retrieved August 31, 2011, from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_syn_48.pdf

22. Transit Cooperative Research Program. *Research Results Digest 56: A summary of TCRP report 88: A guidebook for developing a transit performance-measurement system*. Transportation Research Board of the National Academies, Washington, D.C., 2003. Retrieved February 12, 2010, from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rrd_56.pdf

23. Transit Cooperative Research Program. *Management Information Systems* (TCRP Synthesis 5). National Academy Press, Washington, D.C., 1994. Retrieved August 29, 2011, from http://onlinepubs.trb.org/onlinepubs/tcrp/tsyn05.pdf

24. Transit Cooperative Research Program. *A handbook of proven marketing strategies for public transit.* (TCRP No. 50). TRB, National Research Council, Washington, D.C., 1999. Retrieved February 12, 2010, from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_50-a.pdf

25. Transit Standards Consortium Inc. (2000). *Bus stop inventory: Best practices and recommended procedures*. New Jersey. Retrieved February 12, 2010, from http://tsconsortium.org/TSC/TSC2003/BSIOrders/BSIOrderForm.htm

26. U.S. Department of Transportation. Federal Transit Administration. *Multimodal trip planner system, final evaluation report*. Washington: Government Printing Office, 2010. Retrieved August 29, 2011, from http://www.fta.dot.gov/documents/MMTPS_Final_Evaluation_05-24-2011%283%29.pdf

27. U.S. General Accounting Office. *Intercity passenger and freight rail: Better data and communication of uncertainties can help decision makers understand benefits and trade-offs of programs and policies*, GAO-11-290. Washington, DC: General Accounting Office, 2011. Retrieved August 29, 2011, from http://www.gao.gov/new.items/d11290.pdf

28. U.S. General Accounting Office. *Public transportation: Transit agencies' actions to address increased ridership demand and options to help meet future demand*, GAO-11-04. Washington, DC: General Accounting Office, 2010. Retrieved August 29, 2011, from http://www.gao.gov/new.items/d1194.pdf

GoogleTransitDataFeed Errors

Feed Validator errors

http://code.google.com/p/googletransitdatafeed/wiki/FeedValidatorErrorsAndWarnings