A Paradigm Shift in Travel Forecasting: Let Web 2.0 Feed the Network Model

Wolfgang Scherr (corresponding author)
Delaware Valley Regional Planning Commission (DVRPC)
190 N Independence Mall West, 8th Floor
Philadelphia PA 19106
Phone: 215-238-2885; Fax: 215-592-9125
Email: wscherr@dvrpc.org

Greg Burton
University of Illinois
601 East John Street
Champaign, IL 61820
Phone: 408-476-0630; Fax: n/a
Email: burton12@illinois.edu

Christopher Puchalsky
Delaware Valley Regional Planning Commission (DVRPC)
190 N Independence Mall West, 8th Floor
Philadelphia PA 19106
Phone: 215-238-2949; Fax: 215-592-9125
Email: cpuchalsky@dvrpc.org

Submitted on August 1, 2010 for the TRB Annual Meeting
Revised on November 12, 2010

Committee on Transportation Planning Applications (ADB50)

Paper # 11-3348

Word Count: 5,742 plus 5 figures and 2 tables

Keywords:
travel forecasting, travel demand model, network model, transportation supply data, planning data, web 2.0, GoogleTransit, GTFS, OpenStreetMap, VISUM, volunteered geographic information, VGI

Abstract
Travel forecasters need to process a wealth of information to properly inform their models on the demand side as well as on the supply side. In particular for the supply side, typically referred to as the network model, the information depth is significant, and it is necessary to import external data sources to avoid the unreasonable cost of data updates.

The Delaware Valley Regional Planning Commission (DVRPC) is in the process of a multi-year model upgrade project. An innovative approach has been used by DVRPC to develop detailed and accurate highway and transit networks from internet-published data sources. The street data is derived from OpenStreetMap (OSM), a crowd-sourced routable network. Transit data is derived from Google Transit Feed (GTFS), a data supply used by many internet applications related to public transportation. Both OSM and GTFS are considered a part of the "web 2.0". Web 2.0 data is published on the internet and generated with active participation of internet users.

The authors believe that DVRPC's new network model is the first one in North-America built mainly on "web 2.0" data. The result is a forecasting network of high geographic detail that includes all street classes down to local neighborhood streets. Transit service with full operational detail is integrated in the street network. This paper describes DVRPC's experience with the integration of web 2.0 data into a consistent network model and shares promising results of the ongoing calibration process. The cost and benefits of the use of web 2.0 data in travel forecasting networks are discussed.

1) INTRODUCTION

The field of travel forecasting, once static for approximately 30 years, is currently undergoing a revolution through the adoption of activity-based models and dynamic assignment algorithms. The basic highway and transit network representations that these models rely on, however, have not advanced as aggressively. While some demand models are already producing detailed travel diary-type simulations, some supply side models continue to rely on stick-based models with insufficient spatial detail and accuracy. One reason for this neglect of the supply side is the difficulty and cost of building large network models.

It is the state of the art to represent transportation supply in a travel forecasting model with a high level of spatial detail from the Interstate highways down to every neighborhood street and cul-de-sac. To obtain this level of detail, the traditional method of manual network coding is no longer an option. Therefore, network data need to be imported from existing data sources. The most important requirement for street data is "routability". Routability entails segmentation of street objects from intersection to intersection, distinct coding of one-way versus two-way streets, and attributes for travel speed classes. Not every street layer used in geographic information systems (GIS) fulfills the routability requirement. Transit data need to include geographically accurate stop locations, the stop sequences for routes and service patterns, run-times between stops, and headways or schedules.

Data sources and data detail have evolved with the history of travel demand forecasting. In the 1960s and 1970s, models were "stick-based". In other words, the network representation was an abstract node-link graph as required by the route-search algorithms. The network model was obtained by manual coding. Since the late 1980s, models were interfaced with geographic information systems (GIS) to improve the assumptions of link length. During the 1990s, some network models became truly GIS-based as the old stick-based networks were replaced with exact street-centerline representations, and transit was represented with detailed stops and routes. On the transit side, an additional improvement to data quality comes from a direct integration of transit operations data and transit schedules in regional travel forecasting models, which has been practiced in some metropolitan areas since the late 1990s (*1*).

Over the past five years, an entirely new source of transportation supply data has emerged: open-source data generated on internet sites. These new data are part of a larger phenomenon known as "web 2.0", which refers to internet content created with high levels of user participation (*2*, *3*). The web 2.0 data used in this project include the OpenStreetMap (OSM), a crowd-sourced routable street network (*4*), and the Google Transit Feed in GTFS format, which describes transit service as routes and schedules (*5*).

The Delaware Valley Regional Planning Commission (DVRPC) is the metropolitan planning organization for the region of Greater Philadelphia. DVRPC has maintained a travel forecasting model since the 1970s. In 2009, a multi-year model upgrade project was started. As part of the upgrade project, the software platform VISUM (*6*) has been introduced and the legacy stick-network is currently being replaced by a geographically detailed network model.

This paper presents an innovative approach by DVRPC to create geographically accurate and detailed highway and transit networks from OSM and GTFS as the main two data sources. OSM provides street data for most of the region. GTFS provides operational transit data, including GPS-based stop locations by direction and a complete one day schedule for every service pattern. To the knowledge of the authors, it is the first time that a large metropolitan area in North-America has used web 2.0 network data as the basis for the regional forecasting model.

The development of DVRPC's network model from web 2.0 data sources is discussed in the remaining sections. Section 2 presents the various types of data sources available for network models with a focus on the open-source data that DVRPC has chosen. Then, section 3 discusses the steps of processing and integrating the open-source data. A summary of DVRPC's experience is given in section 4. Section 5 discusses future data updates, followed by conclusions and recommendation in section 6.

2) WEB 2.0 DATA FOR TRAVEL FORECASTING

When DVRPC decided to replace the legacy stick-network with a completely new network model of high geographic detail, an important requirement was to identify existing data sources and to avoid extensive data editing. Otherwise, the time and budget constraints of the DVRPC's model upgrade project could not have been met.

2.1) The Market of Network Data Sources

DVRPC analyzed various potential data sources for the new network model, which can be roughly divided into three categories:

1. Government-owned data, typically maintained by the GIS staff of agencies.
2. Private/proprietary street data from companies like NAVTEQ and TeleAtlas.

3. Web 2.0 data, which have emerged only very recently.

After extensive testing of data from all of these sources and a comparison of the expected cost and benefits, DVRPC decided to use web 2.0 data as the main input for the new network model.

2.2) Web 2.0 Transportation Network Data

DVRPC's two main data sources for the network model, OSM and GTFS, are both considered "web 2.0" internet resources (*2*). "Web 2.0" is a term used for a new generation of internet sites and online applications that provide user-driven content and emphasize user collaboration. Most web 2.0 applications are communication tools (social media such as Facebook), while others generate information (like "wikis"). In addition to social media, web 2.0 also includes several efforts to generate and share geographic data. Such geographic data are also referred to as "voluntary geographic information" or VGI (*3*), as opposed to government-maintained GIS. There are many examples of transportation agencies using the resources of web 2.0, mostly focused on communication tools and social media (*2, 7*). Recently, web 2.0 applications have been introduced to compute travel times and shortest paths for transportation planning or traffic engineering projects – mostly in academic research, for example (*8*).

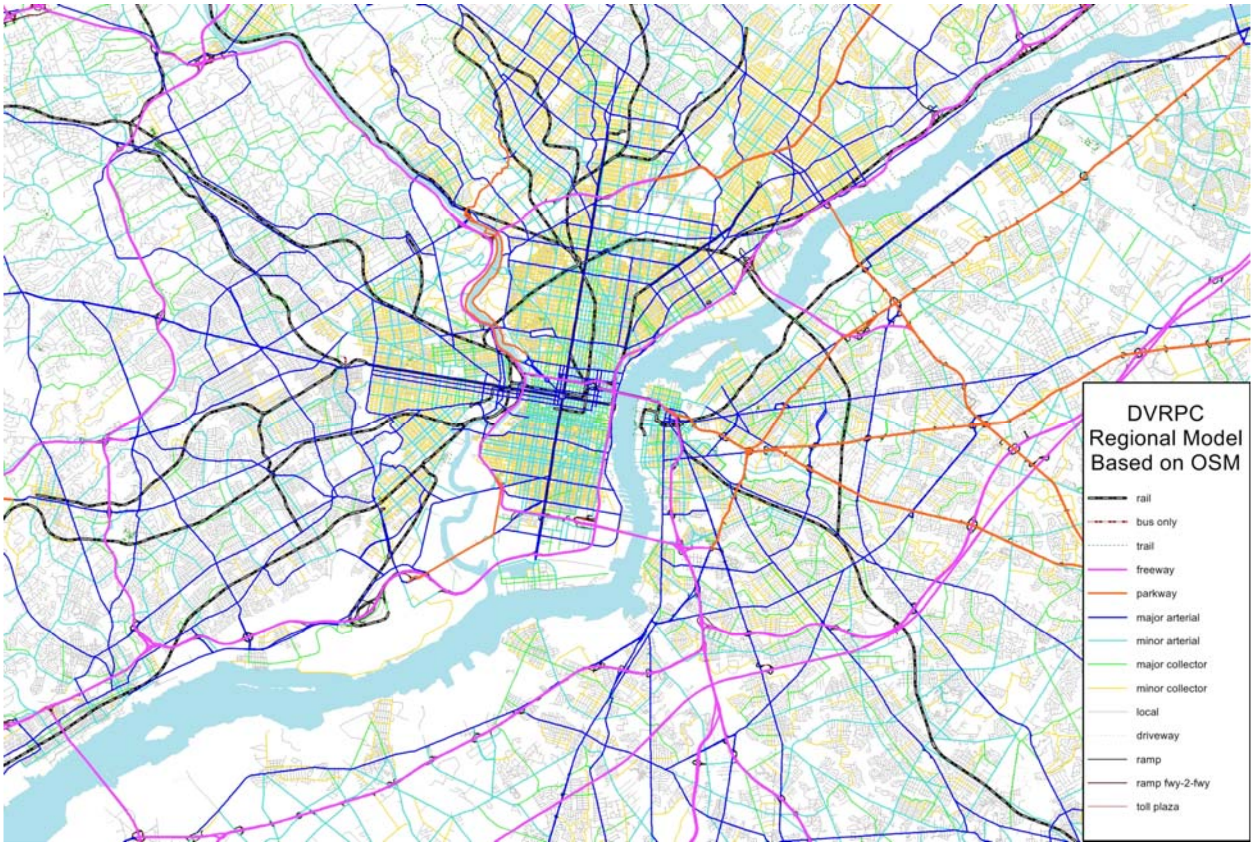
OSM and GFTS are both transportation supply data published on the internet as open-source, i.e. everyone can access and download the data. There are differences between the two cases, but they are united in four aspects that are important for their use in forecasting models:

1. They leverage the work of third parties and consolidate the "work of many hands" in a common platform. In the case of GTFS, the third parties are transit agencies; in the case of OSM, they are private internet users.
2. They have established a documented data format which is a uniform standard for all users.
3. The data are published on internet sites with an established procedure for constant data updates.
4. Free software tools are provided for data management, validation and visualization.

Later in this paper, we will discuss the benefits that arise from these four aspects. In the following section we provide a brief description of OSM and GTFS.

Open Street Map (OSM)

OpenStreetMap can be accessed at www.osm.org. It was founded with the objective of providing free routable street data and free GIS data to any individual or institution. The use of the data is subject to a Creative Commons agreement. OSM is considered truly "crowd-sourced", as every user is encouraged to contribute to the data set, e.g. by uploading GPS-generated data. At the time of this paper, there are over 250,000 registered users according to OSM (*4*). OSM was established by a non-profit foundation in the U.K. that provides the organizational framework to OSM. The recent announcement of AOL/MapQuest to use OSM data as the basis for their online travel directions service in the U.K. (*11*) indicates that the data quality of OSM has achieved high standards in some European countries (*9*, *10*). The situation in the U.S. is different, as most of OSM's street network in the U.S. originates from the federal government's 2005 TIGER data which was imported into the OSM in 2007 and has since been enhanced by volunteers who refine the alignments and add routability attributes to the data (Figure 1).

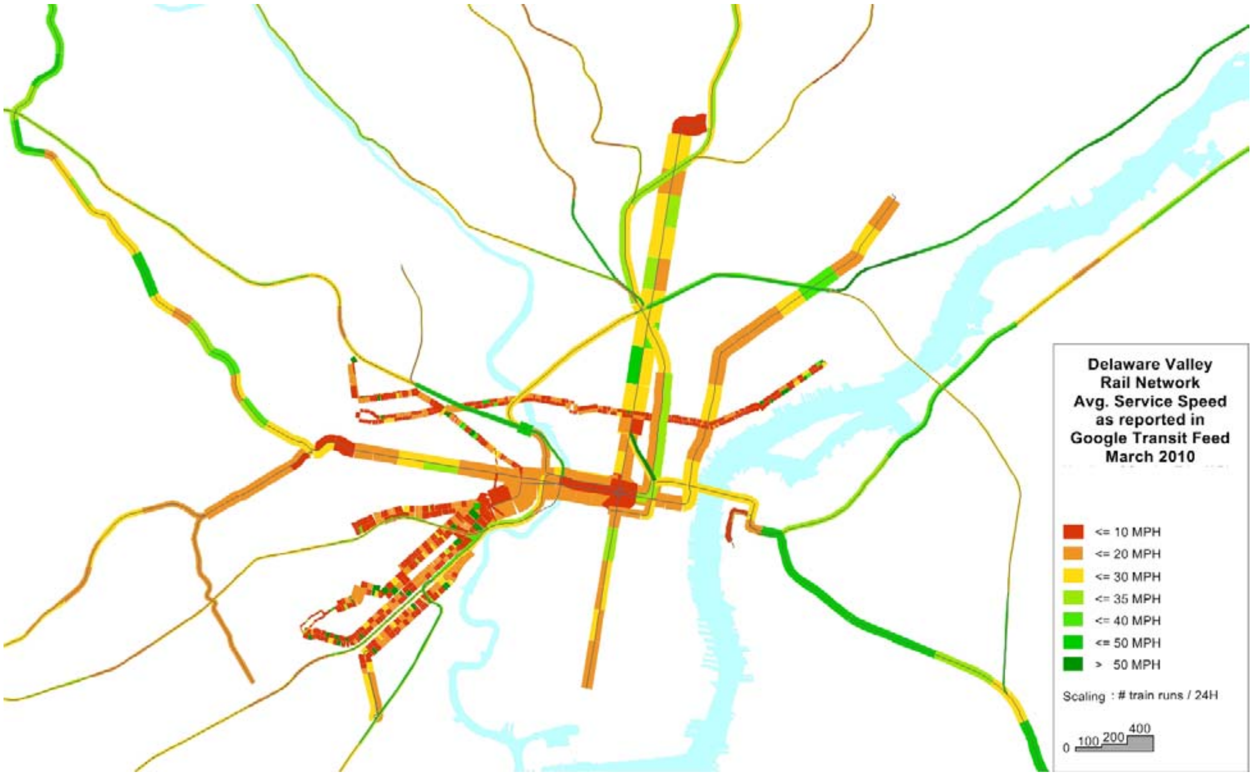


Source: DVRPC, © in parts OpenStreetMap, CC-BY-SA
FIGURE 1 Detailed Street Network Model Based on OSM.

Google Transit Feed (GTFS)
Google Transit™ is a set of online applications that is accessible today in Google Maps and Google Earth. The first beta release in December 2005 featured the region of Portland, OR and the schedules of TriMet (*12*). An important outcome of this application was the definition of the General Transit Feed Specification (GTFS), which was first released with a Creative Commons license in 2006 (*12*). Today, over 170 transit agencies in the United States and Canada generate GTFS data and many use the online forum provided by Google to publish the data (*13*). In only five years since 2006, the GTFS format has been adopted by the transit industry as a standard for sharing schedule data, and most transit software packages are now offering interfaces for GTFS import and export. Many third party developers of online applications have started to use the data published in GTFS format. The two largest transit agencies in the DVRPC region, SEPTA and NJ Transit have published every schedule update in GTFS format since 2009 (Figure 2). Other transit operators in the DVRPC region are in the process of preparing their first GTFS launch. As a result, GTFS allows DVRPC to obtain data on most transit services in the region in one common format.

2.3) Complementary Data
While OSM and GTFS provide the majority of data for DVRPC's network model, complementary data were used in the following cases. Transit service data of smaller operators who do not yet publish GTFS were modeled from printed or electronic schedules. For two counties in the DVRPC region, local GIS departments were able to provide routable street data

sets that were preferred over OSM input. Approximately 80% of all links in the model have been derived from OSM.



Source: DVRPC, SEPTA, NJ Transit, DRPA, © in parts OpenStreetMap, CC-BY-SA
FIGURE 2 Rail Network based on GTFS Data from SEPTA and NJ Transit.

3) DATA INTEGRATION IN A FORECASTING NETWORK FRAMEWORK

The process of obtaining OSM and GTFS data is fairly simple. GTFS can be downloaded typically as one data package per transit agency. OSM data is obtained by running a free downloader tool called Osmosis to extract the road network data from a desired geographic area, which is delineated by a polygon feature that is provided to the downloader tool. The entire process is free and well-documented. While the acquisition of OSM and GTFS data requires minimal effort, the process of integrating them is more complex. The remainder of this chapter describes this process in detail.

3.1) Translation of OSM Network Data into a Node-Link Graph in VISUM

Importing the OSM network data into travel-demand forecasting software like VISUM requires the reconciliation of the two data models, which have substantive differences. For example, VISUM models a link as a simple street segment between exactly two nodes, while OSM allows for complex “way” objects that represent street segments that span multiple nodes. A Python script was designed by DVRPC to handle these differences when converting the OSM data to a VISUM-compatible format. Care was taken to preserve the lineage between the new “links” in VISUM and the OSM “ways” from which they were derived.

3.2) Determination of Link Attributes for All Streets

Then followed a major effort of the project, which consisted of determining the relevant link attributes. DVRPC's travel forecasting model estimates speed and capacity of all links based on the proxy attributes of link class, number of lanes, and area type. OSM does not provide these attributes in the detailed format needed for the model, but it provides a reasonable basis that can be supplemented with manual network coding. For example, the OSM network has street class data for most links, including *freeway, primary, secondary, residential*, etc. DVRPC requires a different, more detailed set of street classes (*freeway, parkway, major/minor arterial, major/minor collector, local, ramp*), but there is high correlation between OSM and DVRPC classes, which was used to obtain a first setting of the link-class attribute on all streets. Then, the major road network (from freeways down to major collectors) was manually revisited to determine link class and number of lanes. The information used includes: aerial photography, DVRPC's legacy stick-network, which contains reliable link attributes validated over 40 years of use in forecasting projects, plus the in-depth knowledge of the regional street network by DVRPC's planning staff. While the large number of links (over 500,000 in the DVRPC region) seems overwhelming at first, it is important to know that 70% of all links in the DVRPC network are "local", which means that there is no need to care much about street attributes. Only the other 30% needed to be revisited regarding class and number of lanes. Throughout this process, the OSM-derived network was scrutinized for the existence of one-ways as well as for connectivity errors and corrected accordingly. The final result is a comprehensive, geographically-based road network that incorporates all relevant information for regional forecasting.

3.3) Translation of GTFS to VISUM

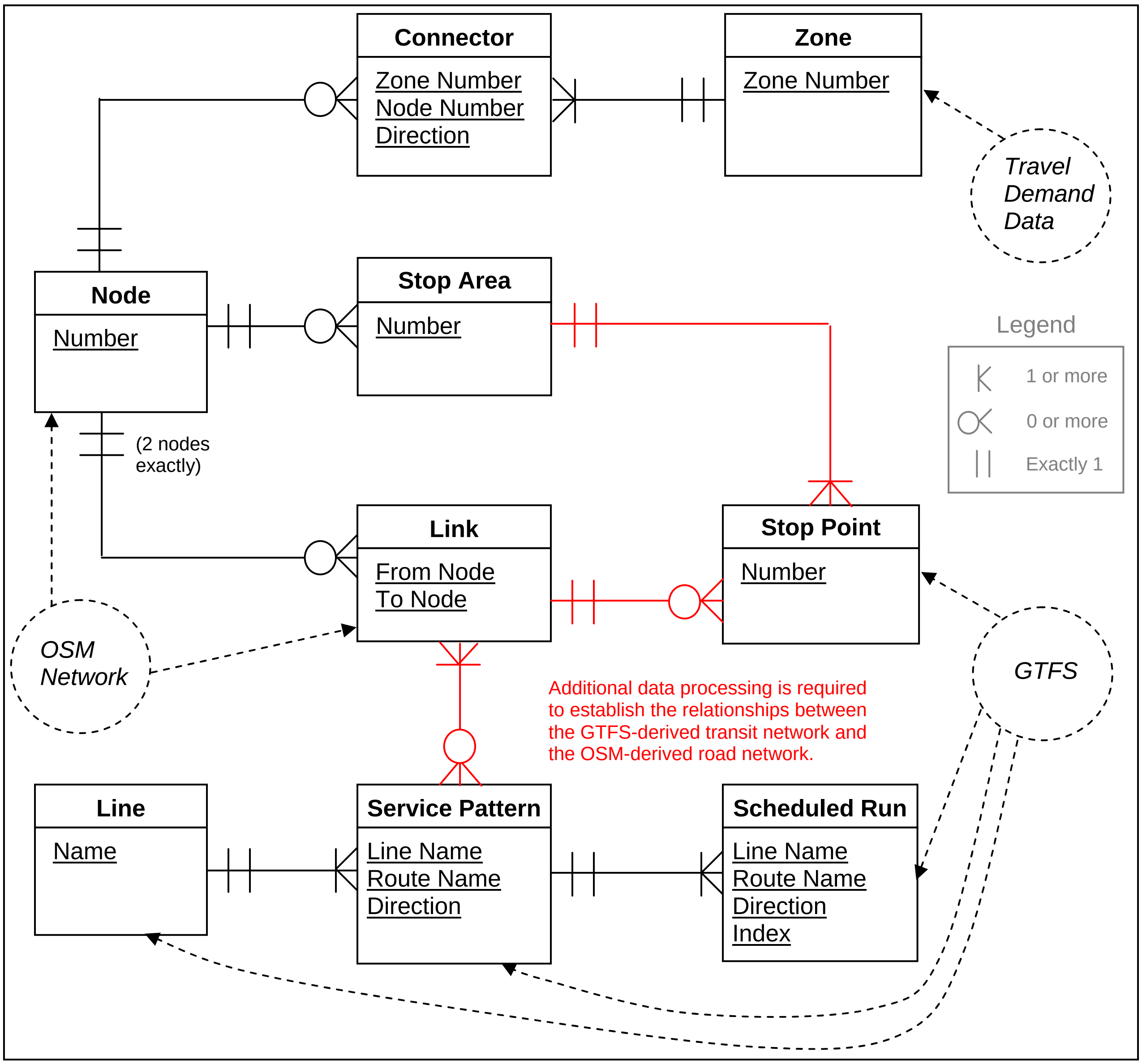
As with OSM, the GTFS data model differs from VISUM's data model, which means that they must be reconciled. For example, GTFS specifies the arrival and departure times of every bus stop independently for every trip that the bus line makes between the beginning and end of the route. In VISUM, the time *between* stops is specified once as part of the service pattern, while the actual journey of the bus contains only the departure time at the beginning of the journey. The conversion is non-trivial and requires significant effort. Fortunately, VISUM provides an importer that translates GTFS transit data into the VISUM transit format without any loss of data. In addition, DVRPC developed a Python program that refines the raw import by renaming routes and patterns, and by renumbering stops so that overlaps of stop IDs from different operators are resolved.

3.4) Integration of OSM-derived Street Network with GTFS-derived Transit Data

After the OSM and GTFS data have been translated to VISUM, there remains the task of integrating them with each other. The OSM-derived data has no knowledge of the transit network, and the GTFS-derived data has no knowledge of the road network. The stop points of the transit network must be merged onto the appropriate links in the road network, and the service patterns must be made to follow the correct series of road links between stop points (Figure 3).

With approximately 18,500 stop points in the region (and many service patterns connecting them with multiple time profiles), the task of integration onto the road network had to be done mainly by automated means. DVRPC developed Python scripts that join stop points onto the road network based on spatial proximity to links. A filtering mechanism was developed based on the names of stop points (e.g. Market St and 5th St) and nearby roads (e.g. Market St). This prevents stop points from being merged onto incorrect links at intersections where the initial placement of the stop point may be ambiguous.

To ensure that service patterns move between stop points on the same roads that actual transit vehicles use, a GIS layer of transit routes is overlaid on the network, and variations from the correct route can be identified by visual inspection. The distance between most stop points is short enough that there is only one reasonable path between them, creating little opportunity for error. But where bus lines have long distances between stop points, the resulting gaps create opportunities for the service patterns to traverse the road network incorrectly. These cases normally required manual effort to adjust the course of the service pattern.

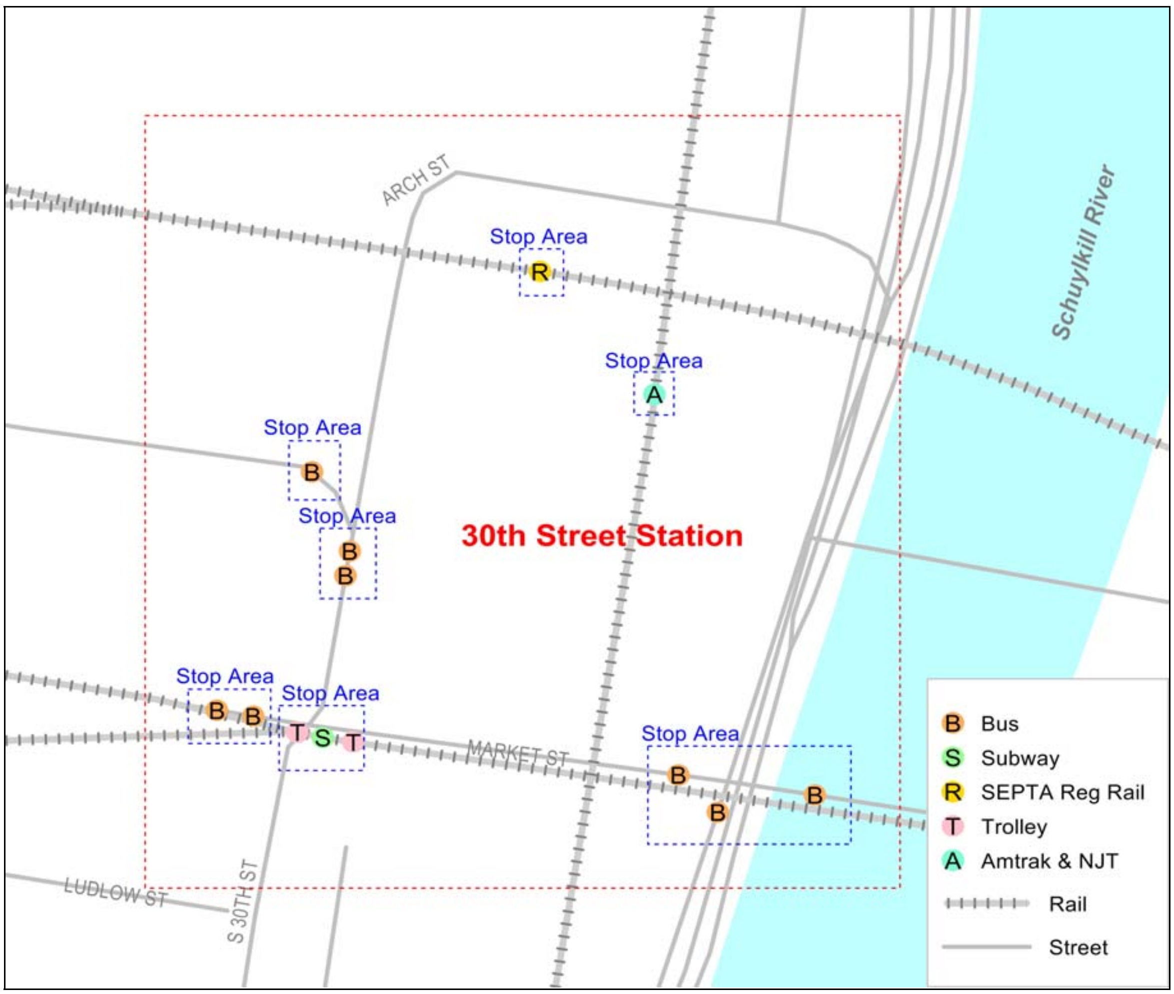


Source: DVRPC, based on VISUM data model by PTV

FIGURE 3 Entity-Relationship Diagram for the DVRPC Network Objects Model.

3.5) Aggregation of Stop Points to Stop Areas

GTFS only specifies one type of stop, which is the equivalent of a “stop point” in VISUM, which is the GPS location of a bus stop or the platform of a subway stop. For the purposes of travel forecasting, VISUM needs to account for passenger transfer times between “stop areas”. The stop area in VISUM is a network object used to group stop points into an area that requires no time to move within. An example would be an intersection where several bus routes have stop points—by alighting from one bus, a passenger is instantly in position to transfer to another bus. An actual example of stop areas and stop points in Philadelphia is shown in Figure 4. The 30th Street Station is a hub for several modes of public transportation, including buses, trolleys, subway, regional trains, and AMTRAK. The subway and trolley stop points are so close together that they can be modeled as a single stop area, since the travel time between them is close to zero. There is sufficient distance between the regional rail and Amtrak stop points to warrant independent stop areas, so that the travel time between them can be captured in the model.



Source: DVRPC, SEPTA, NJ Transit, © in parts Open StreetMap, CC-BY-SA

FIGURE 4 Integration of Street and Transit Topologies (30th St Station, Philadelphia).

3.6) Calibration of Path Building and Path Flows

It needs to be said for completeness, that the next step in the process is to calibrate the path building and path flow algorithms to allow for assignment and computation of skim matrices.

This step is not significantly different from a traditional network development with the exception that the amount of data adds to complexity and attention needs to be given to the resulting computation times (see Table 1).

3.7) Quality Assurance and Control

For the most important highways and interchanges in the region, the quality of the network connectivity is critical to the highway assignment results. There were several ways in which quality issues were identified and corrected on these important highways and ramps. First, individual shortest path searches were performed around all interchanges to verify that the ramps were functional. In some cases, the route takes a circuitous route around the interchange, revealing an issue such as a missing connection, a one-way attribute specifying the wrong direction, or an incorrect type assignment that closed the road to vehicular traffic. These issues were corrected with the help of Google Maps, aerial photos, and the legacy model. Another indirect method of identifying network errors is the process of importing bus lines. A bus line that takes a circuitous route that diverges from its normal course often reveals a connectivity error that prevented it from taking the correct path. Finally, crude highway assignment tests can quickly reveal network errors by highlighting bottlenecks in traffic volumes. Quality control of the GTFS import included comparisons with printed route schedules and systematic identification of incomplete route data.

4) COST AND BENEFIT OF WEB 2.0 BASED NETWORK DEVELOPMENT

4.1) Benefits from Using Web 2.0 as Main Network Data Sources

The authors found several advantages to using GTFS and OSM as main sources for the new network model. As a result, DVRPC decided to use OSM and GTFS data for the new network model. The main benefits are the following:

Open-Source Nature of the Data

OSM and GTFS are freely available and have no licensing restrictions. While DVRPC was also highly impressed by the data quality of private sources such as NAVTEQ and TeleAtlas, the restrictions in sharing the data would not have been compatible with DVRPC's general policies of sharing data with other agencies and the public.

Data Quality and Level of Detail

In the case of the Greater Philadelphia region, DVRPC found that the OSM data quality is adequate for the purpose of a forecasting model. Only a few routability issues were found on the level of freeways and arterials (e.g. a ramp link that was not connected to a node). Not surprisingly, the data quality of street attributes was found to be better in the dense urban centers of the region. In outlying areas, for example, the OSM data set did not always have all of the one-way attributes set correctly.

GTFS is updated periodically by the transit agencies or operators. DVRPC is optimistic that the smaller operators in the region will also join the GTFS community.

The level of detail of the data is very high for both OSM and GTFS, as can be seen in Figures 1 and 2. Figure 1 shows that OSM streets include local streets and even driveways. Figure 2 shows a graphical evaluation of operating speed and service frequency based on the GTFS data that have been imported into the travel model. The data includes details at the operations level, including individual scheduled runs of buses and trains and GPS detail of stops. In both cases, this level of detail exceeds the minimal data requirements of a travel forecasting model. However, DVRPC decided to keep this high level of detail in the travel model to take advantage of the following opportunities:

1. The highway assignment can be performed on a fine zonal structure.
2. The network allows measurement of pedestrian walking distances.
3. The highway network is compatible with traffic analysis tools.
4. The transit data allow for timetable-based assignment techniques.
5. The schedule-based model allows for transit operations studies such as timed-transfer systems or transit capacity analysis.

Standardization

Both data sources provide the advantage of one standard format for the entire DVRPC region. A single translation process can be applied for all sub-divisions of the extended model area which includes over 20 counties in four states. Standard formats are typically well supported by software tools. The travel demand modeling software VISUM, which is used by DVRPC, has provided an interface to GTFS since 2007 and will soon release an interface to OSM. Data acquisition is highly simplified by the standard formats and by the release procedures that are already in place for both GTFS and OSM, minimizing bureaucratic hurdles and communication problems.

Long-term Perspective

The future of online data sources is very promising: both data quality and coverage are continuously increasing.

4.2) Challenges with Web 2.0 Data

There are certainly also challenges related to the open source nature of the data used. Four major types of challenges are:

Data Quality

A concern with OSM is that the network coding guidelines are not followed by each contributor in the same way, creating a possible challenge to consistency in network coding. While the data quality of OSM has been systematically researched in other countries (*9, 10*), no such study has been carried out so far in the U.S. The situation in the U.S. is however different, as most of the street data set originates from an import of federal TIGER data. As a result, the quality concerns in the U.S. are not related to coverage but mainly to connectitivy and routability. DVRPC found that the routability of the OSM data set was almost perfect and needed little repair in the urban core of the DVRPC region. As described before, several link attributes had to be added to the OSM data to fulfill the specific requirements of forecasting models, but this would have had to be done regardless of the data source used. Over time, it can be expected that the increasing number of registered users and contributors to OSM, as well as the feedback from OSM-based applications, will improve the data.

Data Integration
OSM and GTFS are two independent and complex data sources, which need to be reconciled into one consistent data model. The previous section of the paper discusses DVRPC's approaches to data integration in more detail.

Update Capability
Constant updates are one of the primary benefits of OSM as well as of GTFS, but maintaining the link from the forecasting model to the web-published data is non-trivial. DVRPC's perspective on dealing with data evolution and updates is described in a later section of the paper.

Copyright Issues
A copyright note needs to be added to each publication of maps with OSM content, as can be seen on some of the figures in this paper. This is a minor challenge compared to the copyright restrictions of privately-sourced network data.

4.3) Quantitative Effort and Performance
As Table 1 shows, the number of nodes and links in the network model is approximately 10 times the number in the stick-based legacy model. The number of transit objects (stop points, service patterns) is approximately 3.5 times what was represented in the legacy model. These numbers will significantly affect complexity and computation time of the regional model. Table 1 shows a comparison of a highway assignment between the old and the new network model. Both have been performed with the identical peak period trip table and the same system of 2000 TAZs. The assignment with the new network model is still somewhat "improvised" as the TAZs have been connected generically to the network without rigorous calibration. A more qualified comparison between the networks will be available in a few months, when a systematic assignment calibration will have taken place. For the final model, DVRPC will introduce a new system of 3400 TAZs, which will allow for better loading of local streets.

TABLE 1 Quantitative Comparison of the Legacy and the Web 2.0-Based Models

Network Statistics	Legacy Model	Web 2.0-based Model *
Number of nodes	20,000	196,000
Number of links **	45,000	509,000
Total link length (miles) **	32,700	57,200
Total link length (km) **	52,600	92,100
Number of transit stop points	5,200	16,200
Number of transit stop areas	5,200	7,500
Number of transit service patterns	1,700	10,700
Highway assignment run time (PB-UE) ***	42 min	208 min
Highway assignment run time (OB-UE) ***	9 min	188 min

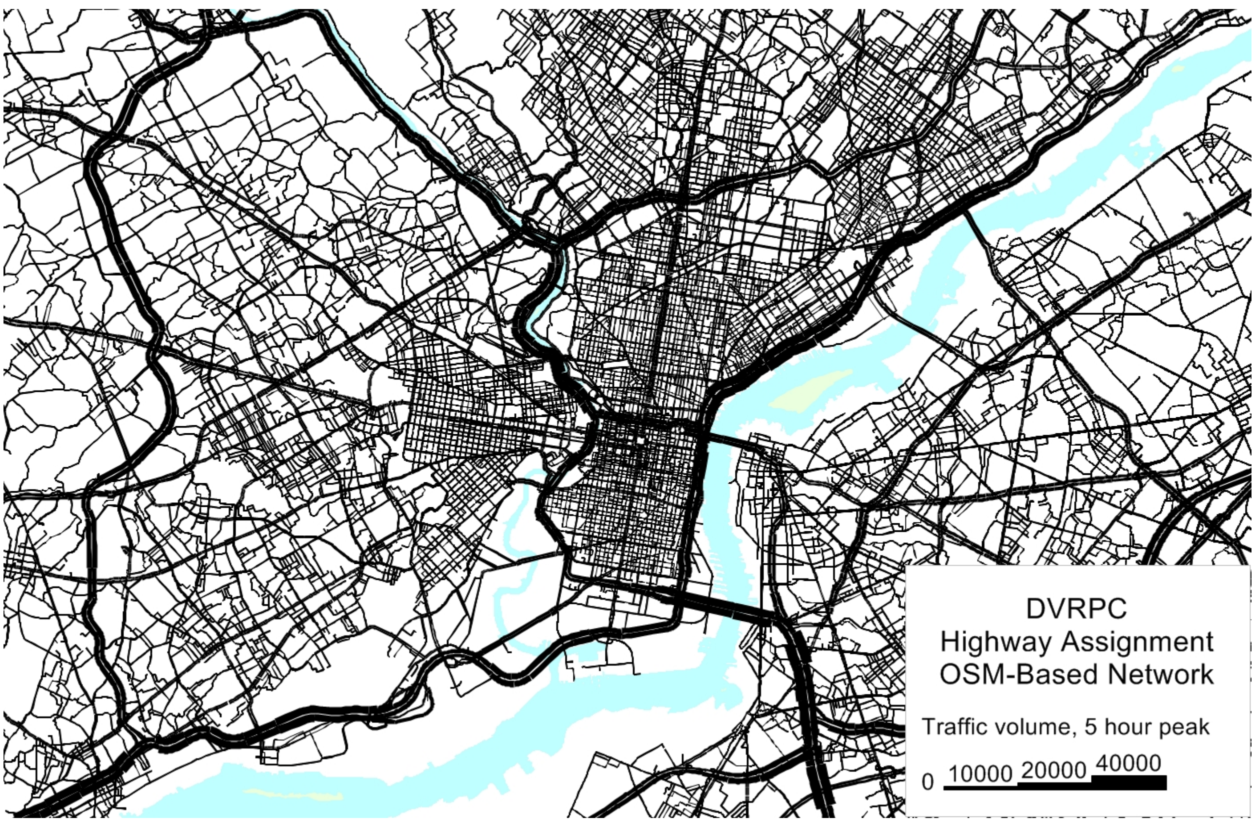
*: To properly compare with the legacy model, only the network objects inside of the DVRPC region are counted.
**: Counting all link directions that are open to at least one highway or transit mode
***: Preliminary results, user equilibrium for a relative gap of 1.0E-3, based on a fixed trip table of peak-period, all-purpose demand, on 2,000 TAZ. The computer hardware: 64-bit computer, 8 cores, 2.4 GHz for each core. PB-UE = path-based user-equilibirum, OB-UE = origin-based UE.

The effort of developing the web 2.0 based network model has been 14 person-months by the time of submission of this paper. It is expected to end up being close to 30 person-months when the calibration will be finished (table 2). About 40% of the total labor was done by interns. Overall, DVRPC feels that the cost of the project is reasonable and that the use of web 2.0 data has helped to constrain the cost.

TABLE 2 Effort of Model Data Integration and Calibration in Person Months

Project Tasks	Full-time employee	Intern
Import of raw data (OSM, GTFS, GIS)	1.8	0.5
Link attribute coding	3.4	1.6
Integration of hwy & transit topology	2.3	2.5
Modeling of transit transfers, fare *	3.3	0.6
Creation & integration of 3500 new TAZ	0.5	3.2
Calibration, highway assignment *	2.9	1.6
Calibration, transit assignment *	3.2	1.3
Total	**17.4**	**11.2**

*: Estimated results, as task is not yet completed



Source: DVRPC, © in parts Open StreetMap, CC-BY-SA

FIGURE 5 Highway Assignment Volumes Computed on OSM-Based Street Network.

5) DATA UPDATES AND FEEDBACK

Web 2.0 sources like OSM and GTFS are continually updated and improved by the user community. It is DVRPC's goal to keep the travel forecasting model up-to-date with improvements of the internet-based data sources. Similarly, DVRPC is also interested in improving the internet-based data by feeding back observations and improvements that were made while refining the forecasting model. However, data updates in either direction can be challenging. This section discusses DVRPC's approach to data exchange in both directions.

5.1) Update from GTFS to the Model

Schedule updates in GTFS format are posted by the operating agencies on specific web sites, where the current schedule can be accessed at anytime. Updating the model with the most recent GTFS schedules is very well supported by the data model and has already been successfully tested. Each transit stop ID in DVRPC's network model has been derived from the stop ID used in the respective GTFS source. An important step at the beginning of a schedule update is to identify new stops and add them to the network model. Then, new routes, service patterns or vehicle trips can easily be updated because they all refer to stops that are already integrated into the road network. In addition, the process of updating schedules is supported by the VISUM software, which can complete routes by performing a shortest path search between stops in the case of new or changed alignments of bus routes.

5.2) Feedback from the Model to GTFS

DVRPC can provide feedback to the transit community to improve their GTFS data. On a small scale, this has already happened in direct communication between DVRPC's modelers and technical staff at the transit agencies. The information that was exchanged included small data errors or coding suggestions. In the future, it is possible that DVRPC could feed back more systematic data, for example exact street-based alignments of bus routes, or transfer-connections between stops and the respective transfer times.

DVRPC has also started to offer technical assistance to smaller transit agencies to facilitate the generation of GTFS data. This is another way in which the MPO can contribute to the improvement of internet-shared transportation data in the region.

5.3) Update from OSM to the Model

In the forecasting model, DVRPC preserves the original ID of nodes and links to maintain a connection with OSM data. However, the two data sets diverge over time, and data objects cannot be easily matched by ID. Conflation of street network data is challenging, but DVRPC expects that the available tools will improve.

5.4) Feedback from the Model to OSM

A similar challenge is certainly the data exchange back to the OSM. Once an effective conflation process is in place, there are several street attributes that DVRPC can feed back to the OSM, such as road class, speed, one-way restrictions, number of lanes or information on the bus service on the streets. It is more difficult to feed back corrections of connectivity problems that DVRPC found in the OSM data set after it was imported into the travel model. This is because it involves the modification or creation of network objects, rather than just a simple attribute change. While

DVRPC is willing to make the data available to the OSM, neither an agreement nor a procedure is in place how the OSM community or a third party would perform the data update.

5.5) Sharing Model Data for Other Purposes

When the network model project is finished, DVRPC will investigate in more detail how the web 2.0 based model can be used by other government applications in the region. One particular proposal is to use the network model as the basis for a region-wide routable street-centerline file that the regional GIS community can use for various applications.

6) CONCLUSIONS AND RECOMMENDATIONS

DVRPC has successfully processed OSM and GTFS data into one consistent network model using the VISUM platform. While the data integration work has been finished by the time of the revision of this paper, the overall model development project is still ongoing. Currently, highway and transit assignment and path-choice are being calibrated. Next steps will include the integration with a four-stage travel demand model and the enlargement of the model area beyond the DVRPC region.

An open question at this time is if DVRPC will keep the full depth of the highway network down to neighborhood cul-de-sacs and back-alleys, or if some pruning will be undertaken for the purpose of reducing computation times and memory requirements of the regional forecasting model. Over the coming weeks, additional run-time tests with the four-stage forecasting model will inform this decision.

At this stage of the project, it is already obvious that it has been highly beneficial for DVRPC to use OSM and GTFS as major data sources for the new model. Without OSM, it is likely that DVRPC would have opted to manually upgrade the legacy stick-network, as all other data sets have either too many copyright restrictions attached or would have required too much manual network coding. The integration of GTFS has not only streamlined the process of updating transit data; it has also helped to strengthen the working relationships between DVRPC and the transit operators in the region. Both OSM and GTFS have allowed DVRPC to increase the spatial accuracy of the forecasting model and to achieve operational detail for both highway and transit.

Most likely, OSM and GTFS can also be a useful resource for government agencies and transportation consultants in other regions. In the current situation, the benefits depend on the advancement of both sources in the particular region of interest. In the long run, the authors are convinced that coverage and data quality of internet-based transportation data will increase and as a result there will be more and more opportunities for government applications.

ACKNOWLEDGEMENTS

We thank Martin Fix and Charles Dougherty for advice with new data sources and the approach to the project. Also, we thank Klaus Noekel and Nicholas Canzoneri for technical support with GTFS import. We acknowledge the contributions of Matthew Gates, Jonathan Ferullo, Drashti Joshi, Christopher Muhs and James Manuel to the network model development.

REFERENCES

1. Fisher, I., K. Lew, W. Scherr and C. Xiong. Planning of Vancouver's Transit Network with an Operations-Based Model. Paper presented at 55th North American Meeting of the Regional Science Association International. New York, NY, Nov. 2008.
2. Volpe Center. *Current Uses of Web 2.0 Applications in Transportation. Case Studies of Select State Departments of Transportation.* Report prepared for FHWA, U.S. Department of Transportation, 2010. http://www.gis.fhwa.dot.gov/documents/web20report. Accessed July 24, 2010.
3. Goodchild, M. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, Vol. 2, 2007, pp. 24-32.
4. Open Street Map. Beginner's Guide, Press Room. http://wiki.openstreetmap.org. Accessed July 24, 2010.
5. General Transit Feed Specification. Information for Riders, Developers and Transit Agencies. http://code.google.com/transit/spec/transit_feed_specification.html. Accessed July 24, 2010.
6. *VISUM 11.0 Manual*. PTV AG, Karlsruhe, Germany, 2009.
7. Routes to New Networks. National Center for Transit Research. http://www.gosocialtransit.com/routes_to_new_networks.pdf. Accessed July 25, 2010.
8. Figliozzi, A. and R. Conrad. Algorithms to Quantify the Impacts of Congestion on Time Dependent Real World Urban Freight Distribution Networks. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
9. Haklay, M. How good is Voluntary Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B, Planning and Design*, Vol. 37, 4, 2010, pp. 682-703.
10. Zielstra, D. and A. Zipf. OpenStreetMap Data Quality Research in Germany. Sixth international conference on Geographic Information Science, 2010, Zuerich, Switzerland, 2010.
11. AOL Press Room. July 9, 2010. http://corp.aol.com/2010/07/09/mapquest-opens-up-in-europe-with-open-source-mapping-with-uk-l/. Accessed July 30, 2010.
12. TriMet. Transit Investment Plan. http://trimet.org/pdfs/tip/tip.pdf. Accessed on July, 25, 2010.
13. GTFS Exchange Site. http://www.gtfs-data-exchange.com. Accessed on July, 25, 2010.