

Regional Strategic Data Assessment: Application of Emergent Technologies and Timely Information Sharing for the New York Metropolitan Transportation Council.

Word Count 7,012
Submission Date: July 20, 2007

Kuo-Ann Chiao

Director of Technical Group
New York Metropolitan Transportation Council
199 Water Street, 22nd Floor
New York, NY 10038
Telephone 212-383-7212
Fax 212-383-2418
kchiao@dot.state.ny.us

Munnesh Patel

Manger Data Unit, Technical Unit
New York Metropolitan Transportation Council
199 Water Street, 22nd Floor
New York, NY 10038
Telephone 212-383-2528
Fax 212-383-2418
mpatel@dot.state.ny.us

Richard Barone

911 Academic Program Student Assistant
New York Metropolitan Transportation Council
199 Water Street, 22nd Floor
New York, NY 10038

Abstract:

This paper provides a background of current data collection, storage, and utilization practices at NYMTC. It critically examines these practices, finding them to be overall deficient and inefficient. To address these shortcomings we lay out a five-part strategy, based on data compatibility, definition, collection, storage, and accessibility, to move the organization forward. Next, we examine current initiatives underway at NYMTC that will further address our current data management issues and proposed projects in the pipeline. Finally, we conclude by examining the benefits of these new processes and technologies to the overall region as a whole. From this we find that it is paramount that data is made available in a timely and organized format in order to properly inform decision makers.

Introduction:

The New York Metropolitan Transportation Council (NYMTC) is responsible for providing a platform for discussion to address regional transportation related issues in the New York metropolitan area. The NYMTC also aids in the process to distribute federal funding for transportation related capital projects. The NYMTC region consists of New York City, Long Island, and the lower Hudson Valley. This area encompasses 2,440 square miles and has an approximate population of 11.3 million, which is about 65% of New York State's population. To best serve local needs of our communities, NYMTC maintains three Transportation Coordinating Committees (TCC): New York City TCC, Mid-Hudson South TCC and Nassau/Suffolk TCC. These committees recommend sub-regional transportation priorities and provide opportunities for the private sector, general public, local government and interested stakeholders to become involved in the planning process on a more local level. The actual voting members of the NYMTC are the counties of Nassau, Rockland, Suffolk, Westchester, the Metropolitan Transportation Authority (MTA), New York City Department of Planning (NYCDOP), New York City Department of Transportation (NYCDOT), and New York State Department of Transportation (NYSDOT).

As the Metropolitan Planning Organization (MPO) for the New York Metropolitan area, one of the main objectives of the NYMTC is to use current, accurate data and state of the art analytical tools to articulate a shared vision in the Regional Transportation Plan (RTP) and to implement that vision collectively through the Unified Planning Work Program (UPWP) and the Transportation Improvement Program (TIP). To this end, the collecting and analyzing of transportation (and related) statistics is essential for NYMTC to effectively measure the performance of transportation services and provide an important reference for planners and operators. The data is used in the development of many of NYMTC's products (RTP, TIP, UPWP, Travel Demand Models, Conformity Analysis, Congestion Management Process, and Annual Reports, etc.).

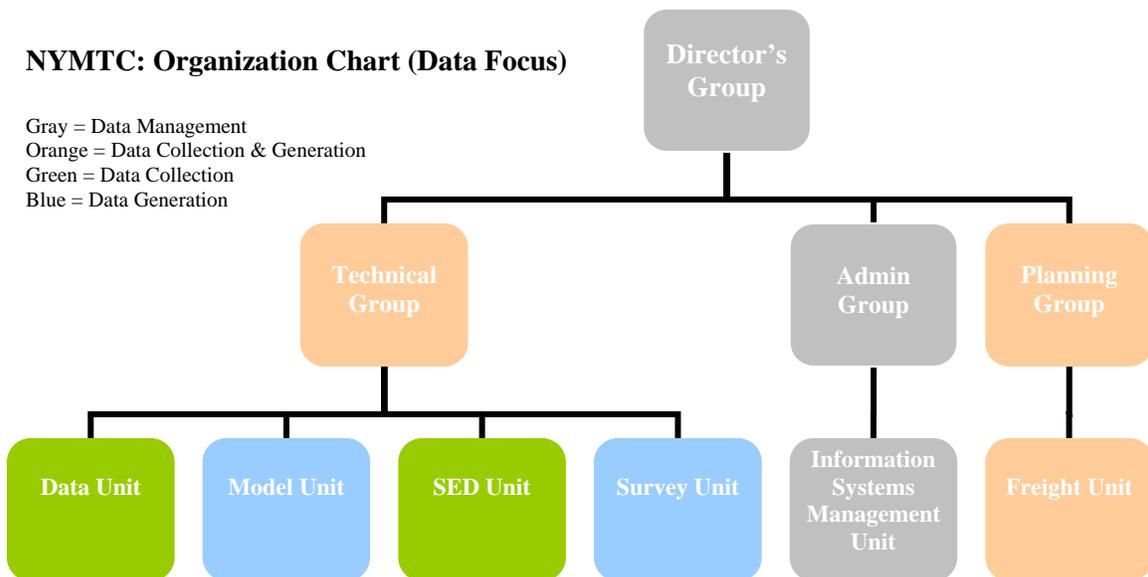
NYMTC's staff has been working with its member agencies to assemble and collect various transportation data over the past thirty years. A series of data reports are produced based on the databases NYMTC's staff compiles on a regular basis. As the popularity of these reports has increased, so has the amount of time required to compile the vast amounts of information required. Additionally, NYMTC staff receives countless requests for information on a daily basis, which leads to a reduction in the time available for staff to focus on improving our data products.

The goal of this paper is to address these concerns by reviewing and critically evaluating NYMTC's current data practices, taking into consideration the additional data collection and visualization requirements set forth in the Safe, Accountable, Flexible, Efficient, Transportation Equity Act: A Legacy for Users (SAFETEA-LU) legislation. With the goal of developing a long-term strategy to mitigate these concerns and, at the same time, reviewing the current steps NYMTC is taking to utilize "state-of-the-art tools and emerging technology" to try to improve its offerings to the public and its membership.

Critical Analysis of Current Data Handling, Storage & Collection Practices

NYMTC, like many other organizations that handle vast amounts of transportation data generated by various operating agencies, lacks centralized data storage and standardized data collection/management practices. Currently, the organization operates as multiple independent data silos that are independent of each other, with little or no consistency in how the data is collected, stored or formatted. The three groups responsible for producing, collecting, or managing data within NYMTC are the Administrative, Technical and Planning departments (The TCCs also sometimes interact directly with the counties and local municipalities to collect data and complete studies).

The Administrative group, specifically the Information Management Systems Unit, is tasked with providing and managing storage for the data that NYMTC generates and collects. This includes management of the existing NYMTC website and available online public reports. Technical and Planning groups are the major generators and aggregators of data within NYMTC. The Technical group collects and generates a “lion’s share” of the data for NYMTC. The Data and SED units, with Technical, utilize much of this data for reporting and to feed the regional transportation model, called the Best Practice Model or BPM. The Modeling unit generates data, utilizing the BPM and Post-Process (PP) Suite, which is used for regional transportation forecasting and determining conformity with the Clean Air Act. The SED unit also generates data through consultants, currently Urbanomics, to provide additional input forecast data for the BPM. Where the Technical group tends to collect raw data for further processing or in some cases generates data, the Planning group mostly conducts studies, contracting out the data aggregation, generation and sometimes analysis to private consultants or the Technical group. There is one major exception that exists within the Freight Unit of the Planning group. NYMTC purchases a proprietary database for freight planning purposes. This database is produced by Global Insight and provides NYMTC with intermodal freight traffic and commodity flows for the region.



In order to more clearly examine how NYMTC currently handles data and what data the organization collects and/or generates, we have defined three subsections. These are; data formats/storage, data definition/procedures, and data categories/types. The first two sections delve into the current processes we employ at NYMTC, while the last section expounds on the types of data we collect, in order to provide the reader with a more comprehensive picture of the amount and variety of the data collected by the organization.

Data Formats/Storage -

The problem, as mentioned earlier, is not the lack of data but how the data is stored, organized and managed. Much of the data is currently in formats that do not allow for easy manipulation or analysis. These include both printed and digital Adobe PDF files. Many of the reports completed or contracted by the Planning department are in these formats, preventing an easily accessible historical archive of data from being produced. The Technical groups' Data unit, in some cases, manually enters information into spreadsheets that they collected from various sources (facsimile, PDFs or phone conversations), creating tabular data, which is later printed and bound to create the "final" product. None of this data is made available in tabular format and stored in a centralized database for the public or other staff to easily access. This prevents the organization from easily and quickly producing historical data sets or cross referencing different data sets produced by the various units. The below table shows the different formats and file types that a portion of the NYMTC technical staff utilizes to complete their reports. The large number of hardcopy files (a file that is printed out and/or that the staff does not manipulate or copy data from electronically) is surprising, as is the variety in file types.

Data Survey - Collected									
Group	Unit	Report/Survey Name or File Name	Type	Format	Collection Method	Collection Freq	Size (KB)	GeoRef	Source
Technical	Survey	Household Travel Survey Data	OTHER	Digital			up to 80,000	Yes	Consultant
Technical	Data	Report: Travel Patterns in the New York Metropolitan Area, File: Long Island Ferry	WORD	Digital	Email (Attachment)	quarterly		No	Consultant
Technical	Data	Report: Travel Patterns in the New York Metropolitan Area; File: Suburb Buses Ridership - Suffolk County Buses	Excel	Hardcopy	Fax	quarterly		No	Consultant
Technical	Data	Report Name: 2005 Truck Toll Volumes, File Name: Comparison of Quarterly PANY&NJ Truck Toll Volume by Facility	Excel	Digital	Email (Attachment)	annually		No	Consultant
Technical	Data	Report Name: 2005 Truck Toll Volumes, File Name: Comparison of Quarterly MTA B&T Truck Toll Volume by Facility	Excel	Digital	Email (Attachment)	annually		No	Consultant
Technical	Data	Section A - Bus Transit by Sector	Excel	Digital	Email (Attachment)	annually	hundreds	No	Member
Technical	Data	Section B - Subway by Sector	Excel	Digital	Email (Attachment)	annually	hundreds	No	Member
Technical	Data	Section C - Suburban and Intercity Rail by Sector	Excel	Digital	Email (Attachment)	annually	hundreds	No	Member
Technical	Data	Section E - Vehicles by facility	Excel	Digital	Email (Attachment)	annually	hundreds	No	Other
Technical	Data	Section F - Ferry & Tramway Passengers by Facility	Excel	Digital	Email (Attachment)	annually	hundreds	No	Other
Technical	Data	Section G - Bicycle Volumes by Sector	Excel	Hardcopy	Email (Attachment)	annually	hundreds	No	Other
Technical	Data	NYC Transit Fleet size	PDF	Hardcopy		annually		No	Member
Technical	Data	Metro North Vehicle miles traveled	PDF	Hardcopy		annually		No	Member
Technical	Data	LIRR annual Passengers	PDF	Hardcopy		annually		No	Member
Technical	Data	JKF annual passengers	PDF	Hardcopy		annually		No	Other
Technical	Data	Housing Permits	PDF	Hardcopy		annually		No	Other
Technical	Data	Passenger Car Registrations	OTHER	Hardcopy		annually		No	Other
Technical	Data	Population	OTHER	Hardcopy		annually		No	Consultant
Technical	Data	Table 1 Total Number of Motor Vehicle Crashes	Excel	Hardcopy	Email (Attachment)	annually	multi meg	No	Consultant
Technical	Data	Table 46 - Transit Accidents	PDF	Hardcopy	Email (Attachment)	annually	multi meg	No	Member
Technical	Data	Accidents By Gender and Age	Excel	Hardcopy	Email (Attachment)	annually	multi meg	No	Other
Technical	Data	Accidents by Time of day and day of week	Excel	Hardcopy	Email (Attachment)	annually	multi meg	No	Other
Technical	Data	Accidents by Human Factors	Excel	Hardcopy	Email (Attachment)	annually	multi meg	No	Other
Technical	Data	Accidents by Vehicle factors	Excel	Hardcopy	Email (Attachment)	annually	multi meg	No	Other
Technical	Data	Accidents by Environmental factors	Excel	Hardcopy	Email (Attachment)	annually	multi meg	No	Other
Technical	Data	Table 3 Fatality rate per 100,000 Population	Excel	Digital	Email (Attachment)	annually	multi KB	No	Consultant
Technical	Data	Table 16A Male/Female Drivers Involved in Crashes Pe	Excel	Digital	Email (Attachment)	annually	multi KB	No	Other
Planning	Freight	Port/Freight Database	Databases	Digital	Purchase	annually	multi meg	Yes	Consultant
Planning	Freight	BikePATH.Table_M03-006.	PDF	Digital	CD-ROM	annually	18	Yes	Consultant
Planning	Freight	BIKEPATH.Table_M04-042	PDF	Digital	CD-ROM	annually	50	Yes	Consultant
Planning	Freight	BikeStations.Table_train_stations	PDF	Digital	CD-ROM	annually	370	Yes	Consultant
Technical	SED	Tm 3.1	Excel	Digital	Email (Attachment)	annually	104	Yes	Consultant
Technical	SED	TM 8-41-2 - Table 5	Excel	Digital	Email (Attachment)	annually	185	Yes	Consultant
Technical	SED	NYMTC SED Adjusted Draft 091603	Access I	Digital	CD-ROM	annually	9,292	Yes	Consultant
Technical	Model	NonState.xls	Excel	Digital	Email (Attachment)	annually	121		Member
Technical	Model	OffState.xls	Excel	Digital	Email (Attachment)	annually	247		Member

A survey of our data collection practices revealed that the majority of the data is provided by member agencies to the staff. A smaller proportion of data is collected by the staff directly through outside organizations or by consultants on the staff's behalf.

Data Definition/Procedures -

NYMTC currently does not have a uniform procedure for collecting data from its membership or outside agencies. This process often varies on a report-by-report basis and/or the particular approach preferred by the assigned staff member. A number of staff members rely heavily on publicly available data that is accessible via the internet, while others prefer to initiate phone conversations to record the specific statistics they require. Additionally, some staff members utilize email to electronically request digital files (only to then print them out) while others are faxed reports from agencies. This process is clearly inconsistent, making for an overall inefficient and likely duplicative process.

Including descriptive information about the attributes of the collected data is another area that requires a closer examination. The fact that most of our data is collected in either static formats or within Microsoft Excel spreadsheets only makes acquiring this information more difficult. There is no metadata attached to the files defining the source, author, describing the data, or any other pertinent information. The reports produced by NYMTC source tables directly, not specifically sourcing each column/field of data, but do provide credit in a more general manner in the form of a bibliography at the end of the report. The spreadsheets themselves do not include this information, with the exception of a database created by one of the staff members to source the data collected by his particular unit within the technical group. The information residing within this database could serve as a foundation for a data dictionary or organization wide metadata repository.

Data Categories/Types -

NYMTC currently collects over two dozen different data sets. This data can be organized under one of four major categories, with additional sub-categories assigned for a further level of classification. The four major categories are Socioeconomic, Highway, Transit, and Supporting Data. This survey of NYMTC's data was originally completed in 2005 and further amended in spring 2007.

Socioeconomic Data

A. Demographics

1. Population – Decennial censuses and inter-censal estimates prepared by the Census Bureau. This is augmented by estimates submitted by member agencies and County level long range forecasts, which are developed by consultants and must be reviewed by Forecasting Working Group (FWG) to reach a consensus. Following the FWG comments, revised forecasts are presented to the NYMTC

Program, Finance, and Administrative Committee (PFAC), which in turn adopts or incorporates them into the Regional Transportation Plan.

2. Household – Decennial censuses from the United States Census Bureau. Agencies should submit their own estimates with methodologies.
3. School Enrollment – New York and New Jersey Departments of Education
4. Vehicle Registrations – Vehicle data collected from New York State: Department of Motor Vehicles (NYSDMV) by the county includes registered passenger cars, commercial registrations and motorcycles. Rental cars and taxis are included in passenger vehicles.
5. Driver Licenses – Data is compiled by county and gender from NYSDMV.

B. Economic

1. Employment – County level data on non-wage and salary (payroll) by industry categories based on the North American Industry Classification System (NAICS) are obtained from the three State Departments of Labor and, on self-employed proprietors, comes from U.S. Bureau of Economic Analysis (BEA).
2. Labor Force – County level data on civilian labor force, unemployed labor force, and unemployment rates obtained from the three State Departments of Labor and Bureau of Labor Statistics (BLS).
3. Consumer Price Index – BLS annual data and percent changes.
([//data.bls.gov/servlet/SurveyOutputServlet?jrnsessionid=1063635452091347445](http://data.bls.gov/servlet/SurveyOutputServlet?jrnsessionid=1063635452091347445))

C. Regional Development

1. Land Use / Floor Space
2. Permit Data – Member agencies report residential projects with 50 or more residential units and commercial projects (retail and/or office buildings) with 50,000 sq. ft or more. Data includes the name of the development, street address, town, number residential units or square feet of commercial-office space, and proposed year of construction.
3. Major Projects – Agencies provide NYMTC with copies of Environmental Assessments (EAS) or Environmental Impact Studies (EIS) for major projects that will have a significant impact on the transportation system.

Highway Data

A. Traffic Counts

1. Volume Counts - Coverage counts by agencies must contain a minimum of 48 to 72 hours and be performed on a 3 to 5 year cycle on all State and County Highways and Highway Performance Monitoring Systems (HPMS) segments. These counts should be performed in accordance with the New York State Standards for traffic data collection outlined in EI 01-001.
2. Classification Data - Agencies conduct classification counts on approximately 10 to 15 percent of their highway systems. Federal Highway Administration (FHWA) 13 classes should be used. Vehicle Classification counts should be conducted for a minimum 48 hour period with at least 2 complete counts for each interval by direction and by lane during the work week. Free flow conditions are required for effective machine classification. Where conditions are not free flow, or speeds are lower than 25 mph, manual vehicle classification counts should be obtained instead.
3. Vehicle Occupancy – Data is used to evaluate Transportation Demand Management (TDM) strategies, High Occupancy Vehicle (HOV) lanes, hub-bound travel, and air quality. Data categories should be collected as Single Occupancy Vehicle (SOV), 2, 3+, buses, and motorcycles at a minimum. This data must also be geo-coded to the survey location before being submitted to NYMTC.
4. Tolls – Revenue and vehicle usage is reported from operators of toll facilities around the metropolitan area. The data is obtained monthly by vehicle axle classification.

B. Congestion Monitoring

1. Spot Speed Data - A spot speed study is the observation of traffic speed at one point or spot on the roadway. Agencies monitor approximately 10 to 15 percent of their highway systems for speed data. This data can be collected concurrently with classification data with the proper Automatic Traffic Recorder (ATR) equipment.
2. Delay – Measured in vehicle or passenger hours, delay is the excess time required to make a trip in comparison to the time the trip would have taken under free flow conditions. Delays can be recurring, attributed to congestions and bottleneck or nonrecurring, resulting from incidents, crashes or weather related.
3. Freight Delays – Measured in tons or cargo value, is the excess time above the projected time required to make deliveries of goods or products. For highways this data is estimated from percentage of commercial vehicles in the traffic

stream.

4. Corridor Travel Times - Travel time studies by time of day for corridors are important for determining congestion levels and estimating system travel time reliability. Agencies collect this data by highway segment, reporting the data by average speed for the each segment and by time of day.

C. Accident Data

1. Number of Accidents – This data is collected by NYSDMV by county and includes: number of vehicle crashes, crashes resulting in fatality, injury and property damage, crashes by human, vehicular and environmental factors, total number of vehicles in crashes with fatality, injury and property damage, crashes by gender and age, vehicular/pedestrian crashes at intersection and mid-block, pedestrian fatalities in vehicle/pedestrian crashes, bicyclists killed/injured in vehicle/bicycle crashes, crashes by road jurisdiction, crashes by day of week, crashes by time of day, crash data by manner of collision, centerline miles, auto registration, and registration by vehicle type. Transit data collected by other member agencies include: transit accidents, injuries, and fatalities.
2. Locations of Accidents – Again, the data are collected by NYSDMV by county and by reference marker along state highways.
4. Rates of Accidents – Rates are determined from the number of accidents and from traffic counts provided by each agency. Rates are calculated in accidents per million vehicle miles of travel for highways, or for intersection by million vehicles entering.

D. Infrastructure / Assets

1. Signal Systems – Provided by agencies and includes the number and percentages of traffic signals that are in coordinated systems for air quality analysis.
2. Highway Miles/Lanes – Annual updates of roadway inventories completed by agencies that includes the number of local highway miles and their functional classification. This data is required for determination of federal fund allocations and for evaluating whether a transportation project is eligible for federal funds.
4. Roadway Condition – Agencies rate their highway's structural condition. State highways are evaluated annually, while counties rate their highways on one to three year cycles. This information assists in programming pavement improvement projects.
5. Useful Life – Agencies determine service life of various pavement treatment options and provide justification for the application of each option.

6. Maintenance/Replacement Cost – Agencies develop estimates of maintenance and replacement cost for their infrastructure assets. This information is essential for developing long term financial needs, which is required for the Regional Transportation Plan (RTP).

Transit Data

A. Subways, Buses, and Commuter Rail

1. Ridership - Passenger ridership data by month for NYC Subway, Long Island Rail Road, Long Island Bus, Metro-North Railroad (separately for Harlem Line, Hudson Line, New Haven Line, and West of Hudson lines) and Staten Island Railway are provided in the Metropolitan Transportation Authority monthly report prepared by MTA Budget Division. (Hardcopy)
2. Financial - This data is collected for the: Subway, Long Island Rail Road, Long Island Bus, Metro-North Railroad and Staten Island Railway: Revenue Passengers data for each month for Subway, Long Island Rail Road, Long Island Bus, and Metro-North Railroad (separately for Harlem Line, Hudson Line, New Haven Line, and West of Hudson lines) and Staten Island Railway, are available in Metropolitan Transportation Authority monthly report prepared by MTA Budget Division (Hardcopy).
3. Schedules/Headways – Is provided by the MTA and other operators for each mode of public transportation and for each line.
4. Vehicles/Capacities - Is provided by the MTA and other operators for each mode of public transportation and for each line, as required by SAFETEA-LU.

B. Ferries

1. NYC DOT – Public - Ferries operated by New York City: Department Of Transportation (NYC DOT). There are no fares collected for this service between NYC & Staten Island. Ridership monthly data and schedule are available from NYC DOT (Hardcopy).
2. Ferries – Private - There are several operating companies in the region (NYC-NJ and Long Island ferries). Ridership monthly data is available (except Long Island ferries) in electronic format from PANY&NJ. Financial data, such as passenger revenue and fares for each operating company and for each line, may be requested in electronic format from PANY&NJ.

C. Equipment Needs

1. Condition Data – Average age of transit fleets.

2. Useful Life – Expected number of years of service from transit vehicles.
3. Maintenance/Replacement Cost – Transit agencies develop estimates of maintenance and replacement cost for their vehicles and equipment. This information is essential for developing long term financial needs required for the Regional Transportation Plan (RTP).

Additional Supporting Data

A. Non-Motorized

1. Bicycle Data - Studies conducted by agencies that include counts or observations of bicycle travel in the Region by location and/or any demographic characteristics of bicycling.
2. Pedestrian Data - Studies conducted by agencies of counts or observations of pedestrian movement in the region.

B. TIP

1. Project Data – Scope and Financial Data. Agencies should identify the complete scope of work with limits for each project to be displayed in the TIP, whether for federal funding or informational purposes, and include enough detail such that anyone can determine the intent of the project and its location. As there are many users from implementing agencies, air quality agencies (EPA, DEC), federal oversight agencies, and the general public, there needs to be enough information to determine "what the projects accomplishes."
2. Air Quality Data – If a project is deemed to have a potential environmental impact and valid for the BPM process (typically, but not limited to Non-Exempt projects), the data/characteristics of the project will need to be collected from the sponsoring agency. The agency must complete the Air Quality online entry forms (through the PIMS) necessary to update the transportation network in order for the model to perform the required conformity analysis.
3. GIS Spatial Info - The spatial data (Geo-Coding) is required for all non area-wide projects for the GIS interactive TIP feature on the NYMTC web-site. This is automatically performed through the online PIMS software.

C. Inventories & Surveys – These are normally conducted region-wide over a period of years in order to better inform decision makers and update the BPM.

D. Transportation Demand Management (TDM) - Agencies report any current efforts underway pertaining to these strategies as well as any air quality or congestion reduction benefit. TDM are strategies that result in the more efficient use of pre-existing transportation infrastructure and facilities. Some of these strategies include:

ridesharing/carpooling, public education about transportation issues, various driving disincentives and ridesharing incentives, commute trip reduction programs, transit use, zoning and ordinances that promote transit use and mixed-use development, park-and-ride facilities, and other measures to reduce pollution and congestion.

E. Airports - Data pertaining to revenue passengers, aircraft flights and cargo tonnage, is collected for the following airports: Kennedy, LaGuardia, Newark, Long Island MacArthur, Stewart, and Westchester County. This data is supplied by the Port Authority of NY&NJ and local airport operators.

F. Freight Data/Tonnage/Value

1. Trucks/Vehicle Miles of Travel (VMT) – Volume of trucks by class, route or location, and truck VMT
2. Rail – Carloads, containers, and twenty foot equivalent units (TEU's)
3. Waterborne – Containers ports and seaports.
4. Air Freight – Volume and value.

G. Customer Perceptions – These include the results and summaries of surveys data that query customers about the transportation service provided by agencies in the NYMTC region.

H. Intelligent Transportation Systems (ITS) - Collect, store, process and share information related to the movement of people and goods. Examples of ITS include systems for traffic management, public transportation management, emergency management, traveler information, advanced vehicle control and safety, commercial vehicle operations, electronic payment and railroad grade crossing safety. NYMTC's ITS Integration Strategy Project is a roadmap for transportation systems integration in the ten county area over the next twenty years. The ITS integration strategy is being developed through a cooperative effort by the region's transportation agencies, covering all modes and all roads in the region.

I. Financial Data

1. Programmed – The financial data for the TIP that is programmed on the current TIP or as an update includes phases, amounts, obligation date, and funding source. The financial data must be constrained to include fiscal constraint by year and fund source. Occasionally, the constraint can be across the three years of the program, providing fiscal constraint is coordinated through the respective TCC and NYSDOT-MO.
2. Expenditures – The historical information regarding the previous TIP projects is known as the "Disposition of the previous TIP" and is a federal requirement. The "Disposition", as it is known, typically shows the expenditures, slippage, deletion, or any other explanation informing the performance or status of an earlier TIP project entry.

It is evident, based on our above survey of data handling/storage/collection practices and the amount/type of data we collect, that NYMTC is in desperate need of a comprehensive strategy to more effectively and efficiently manage its data. Specifically, it requires one that would layout how the organization should approach each portion of the data management puzzle and also be flexible enough to deal with the large variety of information, collected from multiple sources, that the agency currently is tasked with aggregating.

Strategy to Leverage Information Technology & Define Processes to Overhaul Data Practices

The prior section provided an overview of the kinds of information NYMTC collects and discussed the challenges the organization is facing when it comes to managing this information. Overall, the organization lacks a comprehensive policy for dealing with its data, from initial data collection to providing accessibility to this data, in a centralized manner, to the public and even its own internal staff. It is critical that NYMTC address these shortcomings by clearly articulating a policy that defines its data management practices. To this end we have defined a five-part strategy that is designed to address each piece of the data management puzzle. NYMTC should first focus on how it manages its data from the bottom up; relying on the assumption that once the foundation is shored up the remaining parts of the structure should more easily be made sound. This is our reasoning behind focusing on data Compatibility and Definition as the first two strategies. Once we defined our data and ensured that it is in a format we can utilize, then we proceed to outline a process to Collect the information and Store it, with the final step to make it Accessible to all stakeholders in the NYMTC region. These steps are covered in further detail in the section below:

1) *Compatibility: Standardize Data Format – NYMTC should acquire data in digital formats with linking fields*

Ideally, all the data that NYMTC collects should be in tabular or tagged (markup) format with defined field names/columns for the rows of data. Acceptable formats would include XML, XLS, CSV, DBF, and other tabular or tagged formats supported by our storage server. Equally important the tabular data should contain linking fields, or an attribute that would allow it to be combined with other datasets. In many cases an address, mile marker, or some other Geo-codeable reference would be sufficient. The ability to geo-code the data is critical to providing NYMTC with the ability to layer the information it collects, creating new associations that could better inform decision makers.

2) *Definition: Require Metadata – All data should include titles, keywords, and/or field/attribute descriptions*

In order for stakeholders to properly interpret the meaning of our reports it is important that the data used to create them be properly defined. The possibility of data being misconstrued is real and can have very serious consequences. To prevent this from occurring NYMTC has instituted guidelines for Metadata. The organization must enforce this practice to ensure that data and files are being marked with the required identification elements.

This is an example of what Meta Data is already stored in Microsoft Excel

Microsoft already has metadata embedded within the XLS file format. The data collected is minimal and currently includes:

- Your name
- Your initials
- Your company or organization name
- The name of your computer
- The name of the network server or hard disk on which you saved the document
- Other file properties and summary information
- Non-visible portions of Object Linking and Embedding (OLE) objects
- The names of previous document authors
- Document revisions
- Document versions
- Template information
- Hidden text
- Comments
- Macros
- Hyperlinks
- Routing information

There are also opportunities to develop Meta Data within Excel, Word, and Access natively, utilizing the pre-canned functionality of the applications. NYMTC could then programmatically mine the metadata, creating a report or repository to store the information for indexing.

Additionally, a data dictionary would be required for cataloging the tabular data that NYMTC collects from agencies and local governments. This document would act as a reference to the tables, including some pieces of metadata, but going a step further by providing detailed descriptions and examples of the data contained within each field of the table.

3) Collection: Create a Uniformed Process for Data Collection

A data analyst should be assigned the task of organizing the agency's data collection efforts. Guidelines must be developed based on the above strategies 1 & 2 and be submitted to all local governments and agencies that NYMTC receives data from. The analyst would be responsible for organizing the delivery process, ensuring it is accomplished in a timely manner, and when at all possible completely automating it (through either a server dump, direct data base import, or queried as a web-service).

4) *Organization: Install an Enterprise Level Data Repository/Warehouse*

The NYMTC Enterprise level data warehouse/repository should allow for the aggregation and integration of data from different sources and specified formats. The data warehouse should be able to collect data over the internet utilizing different communication layer protocols. These include, but are not limited to, web services, XML, EDI, FTP, SFTP, HTTP, HTTPS, and email. The data warehouse must also perform checks for data integrity, to ensure that the data is not corrupt or contain errors.

The NYMTC data warehouse architecture should meet, but not be limited to, the following:

- The architecture must be capable of adapting to future changes without major redesign. Individual components may change, but the architecture needs to maintain a stable structure.
- Implementation must be a scalable approach allowing building components as time, resources and funds permit. It must allow for measurable progress towards goals.
- It must be affordable and rely on components that are available, stable and obtainable by government purchasing processes or already in the possession of NYMTC.
- It must support transactional, Decision Support System (DSS), exploratory, mining and future types of processing. It must support data sharing while maintaining privacy and confidentiality under a secure environment.

NYMTC would require the database to import the following file types:

- .DBF (for SPSS and ESRI)
- .CSV (generic file type – comma separated values – to provide maximum compatibility with other databases)
- .SHP (shapfiles to be used in ARCGIS or other GIS applications)
- TransCAD Outputs - .NET, .MAP, ...etc (to be used by members to analyze BPM output)
- Microsoft Word (.doc) and Excel (.xls) Formats

The data warehouse would be the central data repository for the entire organization, allowing staff to quickly query the most current information in order to generate their reports. In the long-term, if made accessible, this repository would become a regional transportation data clearinghouse, providing NYMTC's

membership and the public access to a vast amount of data that could be used in aiding the decision making process.

5) *Accessibility: Develop GIS and Additional Web Applications to Allow for Maximum Access to Data.*

In order for the council and public to access the information NYMTC has now collected, formatted, defined, and centrally stored, it is necessary to create an application(s) that can easily query the data warehouse. The world wide adoption of the internet and the default installation of a web browser on every computer, makes utilizing web-based applications the most logical approach for accessing the data. SAFETEA-LU requires MPOs to visualize their information whenever possible; therefore, creating a Geographical Information Systems (GIS) based application for display of spatial data is required. Additionally, an application to generate data sets for further analysis, a table builder like “American Fact Finder” utilized by the U.S. Census Bureau, would provide an alternative to viewing data spatially or for data that does not contain any geo-reference information. The next two sections cover in further detail the GIS application and a web-based table builder application.

GIS Application Development

NYMTC must develop a browser based GIS user interface for the visualization of the BPM and other geospatial data as interactive maps, per the requirements set forth in SAFETEA-LU. The application should allow users to view, select, analyze, print, and download the data and information. This GIS application should source the NYMTC data warehouse and additional data sources.

Figures 1 & 2 provide examples of our vision, referencing the well designed New York City MAP and the non-profit NYC Oasis. A major component of the interface design would be the ability to aggregate different groupings of layers into selectable (pre-defined) themes, as shown in the NYC Oasis example. This interface must also allow for future customizations/changes.

Figure 1: NYC: Department of Information Technology & Telecommunications – City Map

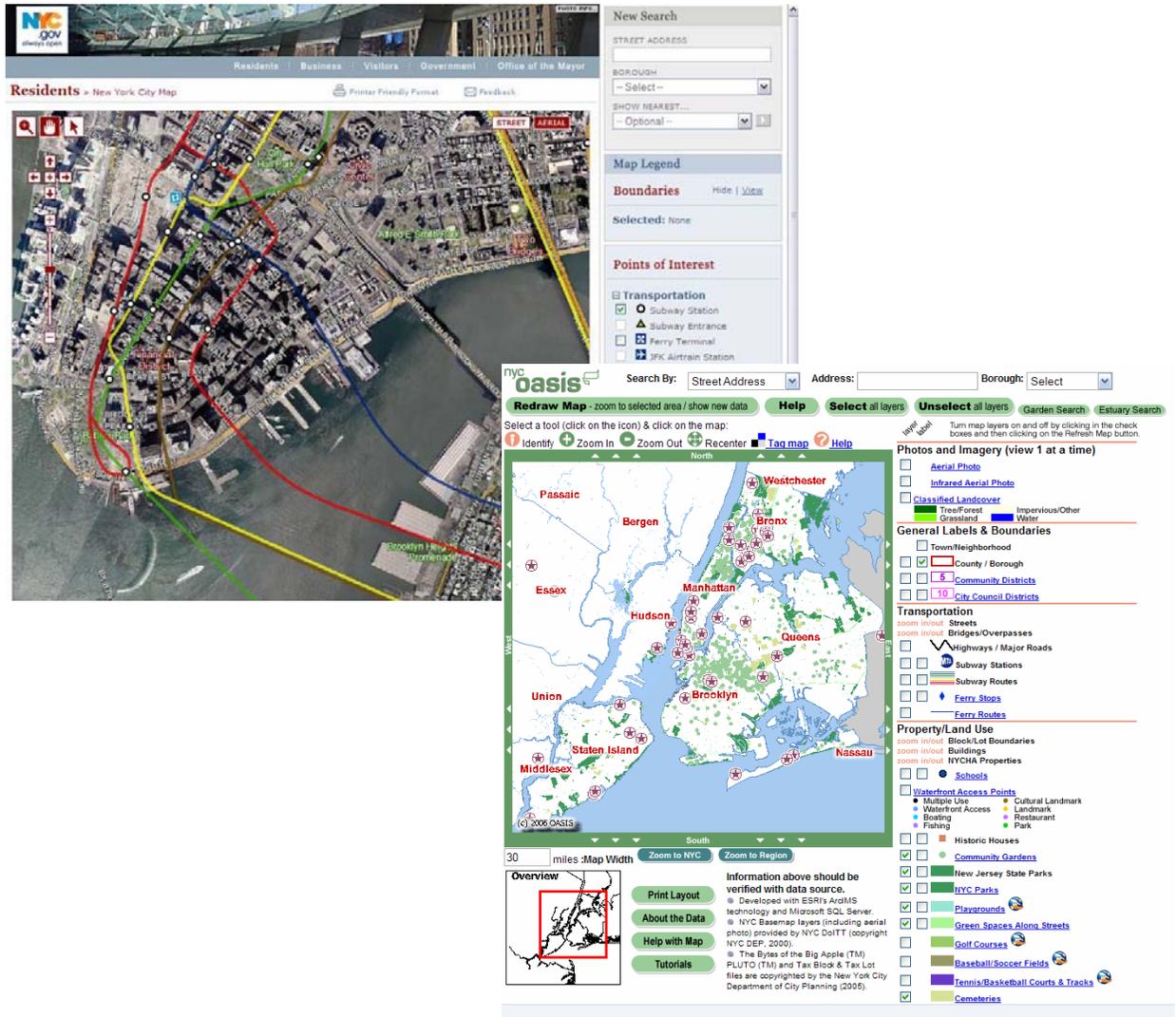


Figure 2: NYC: OASIS

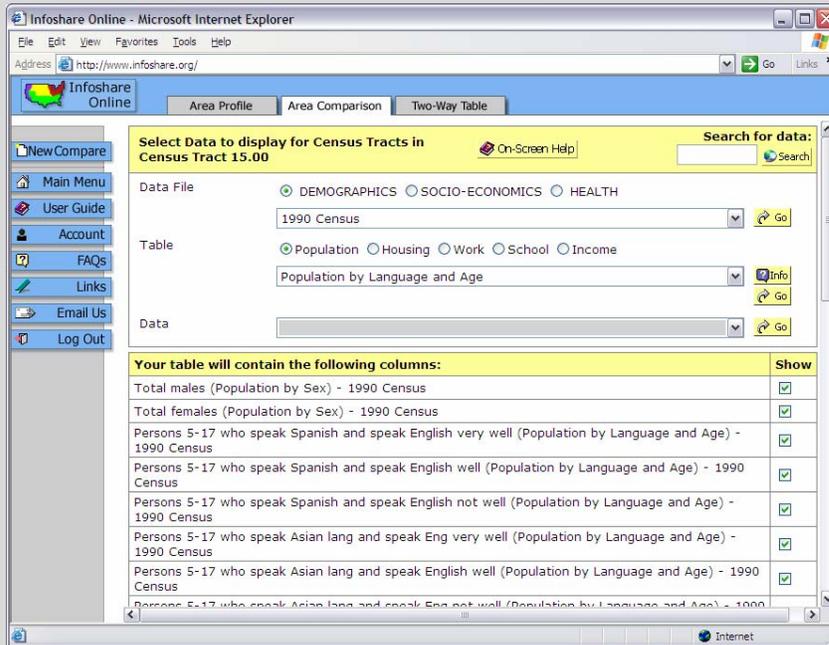
Non-GIS Web Application Development

The NYMTC must develop a browser based non-GIS web application with predefined and interactive tabulations, charts, and maps that will allow users to view, query, analyze, print, and download the data and information. These application(s) will allow us to export our results in different formats and visually present the data in various ways such as customized tables and charts. The below section provides a design reference to better illustrate our vision.

Design Reference = INFOSHARE (<http://www.infoshare.org/>)

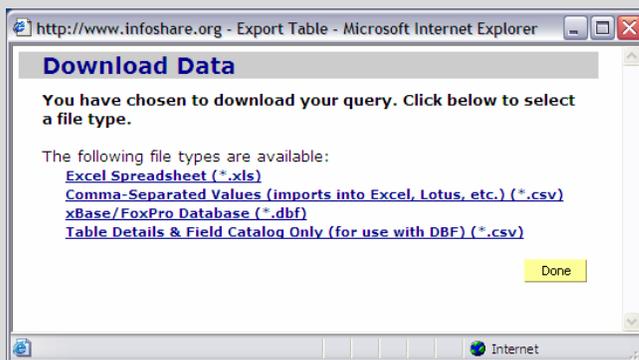
This site allows a user to organize census data in a tabular format. Dynamically selecting how to aggregate the data (based on regions, city, census tract, block groups or blocks).

The user can also include multiple columns of data in order to create a rich dataset for analysis (see below interface).



Building a table

INFOSHARE also gives the user the option to download the tabular output as different file types (see below interface).



Ability to download output in different formats

The detailed five strategies are the first steps towards creating a more efficient and effective data management practice. In order for the overall strategy to be successful it is critical to initiate all five. In order to provide the final product of data accessibility to our staff, members, and the public, it is critical to build off the foundation that is set forth in the earlier steps. At the same time, this process is not a static one and will require constant vigilance and reevaluation based on performance and changes in technologies that normally occur over time.

Implemented Technologies and Future Steps

NYMTC is in the process of implementing a number of the strategies set forth in the previous section. Many of the projects were initiated prior to the official articulation of a data management strategy. The impetus for some of these projects was due to the additional data collection and visualization requirements mandated in SAFETEA-LU. The nine projects detailed below are either well ahead in the planning process, currently underway, or completed.

Project Title: UPWP WebAccess: Online Application

Status: **Completed**

Reason: Council & Internal Collaborative Process Improvement

Category: Data Management & Reporting

Group(s): Information Technology Management Group



Description: A menu driven collection of tools created to assist Council members in developing sponsor activities for inclusion in the NYMTC UPWP. The system utilizes Cold Fusion and a Microsoft SQL Server database on the backend. The tool also allows for instant report generation using a Crystal Reports module. The final UPWP document is constructed from the data collected through this application.

Project Title: NYMTC Bicycle Data Collection Program

Status: **Completed**

Reason: SAFETEA-LU & Council

Category: Visualization (Web Site/Google)

Group(s): Planning Group

Description: From 2002 to 2005 a total of 385 locations have been surveyed to create a database of bike paths and facilities. Of the 385 total study locations, 226 are in the five boroughs of New York City, including 82 multi-use path locations and 144 on-street locations. One-hundred twenty-six of the locations are in Nassau, Suffolk, Westchester, Putnam or Rockland counties, including 43 multi-use paths and 83 on-street locations. The remaining 33 locations were train stations. The data from these surveys were mapped to their locations, utilizing image mapping to link the detail to the data point. Future revisions of the website are planned, potentially mapping the data using Google maps for visualization instead of a static image.

Project Title: Project Information Management System

Status: **Completed** (Version 1)

Reason: Council & Internal Collaborative Process Improvement

Category: Data Management, Reporting, & Modeling



Group(s): Modeling Unit

Description: PIMS allows NYMTC's members to enter TIP information directly into a database that then feeds the BPM model. The PIM also associates a spatial reference to the project information that is entered, creating a modified network to be used in the BPM forecasting process, as well as a table with the specific variables/factors as well. Currently, NYMTC is the process of refining the system and will be releasing the next version of PIMS in approximately one year.

Project Title: Development of the Graphical User Interface and Improvement of NYMTC's Best Practice Model

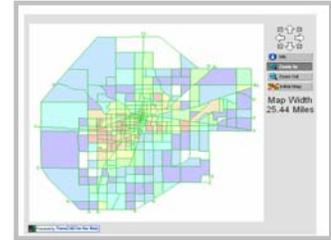
Status: **Underway**

Reason: SAFETEA-LU

Category: Visualization (GIS)

Group(s): Modeling Unit

Description: Caliper will develop a web application that will provide model-based data to the public through an interactive dynamic web server. Using TransCAD for the WEB, Caliper will design and implement a web application that will provide input data and forecasts to the public in a form that is easy to access. This will include maps, charts, and tables that clearly illustrate the forecasts and permit downloading of some key data of interest.



Project Title: Demographic and Socioeconomic Forecasting

Status: **Underway**

Reason: SAFETEA-LU & BPM

Category: Visualization (GIS)

Group(s): SED Unit

Description: Urbanomics will develop a web application for the display and review of forecasts to the county level. Given the rapid advances since the development of the prototype application (back n 2002 the consultant developed a static web-based application utilizing GIS output [images] within a dynamic web site framework, akin to the NYSDOT travelinfo) in web-based GIS and mapping (including new third-party interactive client-side mapping tools), the consultant will evaluate current technologies to determine the most suitable and efficient means to accomplish the goals of this task. It is expected that purchase of a license for a web service such as ArcWeb GIS will be necessary in order to implement the application's interactive mapping capability.

Project Title: NYMTC: New Technology Scan

Status: **Underway**

Reason: Council & SAFETEA-LU

Category: Visualization, Data Management, Organization, & Improved Public Interaction/Participation

Group(s): Planning Group

Description: This initiative is designed to inform the staff and council of the latest technologies/methods available in the marketplace and academic communities that might assist in the regional transportation planning process. The University Transportation Research Center, Region 2 in conjunction with Rensselaer Polytechnic Institute are the two groups who are performing this survey for NYMTC.

Project Title: Integrating Technology in Regional Planning Processes: A Review of the State-of-the-Art

Status: **Underway**

Reason: SAFETEA-LU

Category: Visualization, Data Management, Organization, & Improved Public Interaction/Participation

Group(s): Planning Group

Description: The goal of this process is to familiarize NYMTC staff and council members with the latest technologies that will help them to organize and visualize their work. This initiative is being organized by Hunter College, a part of the City University of New York, and consists of a number of work sessions that will be ongoing throughout 2007. The data collection will be based on reviews of published materials, targeted telephone interviews with selected individuals who are expert users of these applications, and additional conversations with software developers and technical specialists.

Project Title: NYMTC Transportation Information Gateway

Status: **Planned**

Reason: SAFETEA-LU

Category: Data Storage & Visualization (GIS)

Group(s): Data Unit & Information Technology Management Group

Description: The primary objective of TIG will be to create a data warehouse to store all raw data collected by NYMTC staff. Additionally, we will create a predefined browser based Geographic Information System (GIS) and other non-GIS applications to enable staff and our members to retrieve the data in order to create reports or other types of outputs.



Project Title: NYMTC Website Redesign: Utilizing a Content and Document Management Solution

Status: **Early Planning Stages**

Reason: SAFETEA-LU

*tegor*y: Data Management, Organization, & Improved Public Interaction/Participation

Group(s): Technical & Information Technology Management Groups

Description: SAFETEA-LU, along with other internal initiatives aimed towards greater transparency and visualization of our data and planning products, will make it necessary for NYMTC staff to update the site daily. In order to accomplish this effectively and efficiently we will need a Content Management System (CMS) to spread the workload throughout the organization. NYMTC can derive benefits from a CMS on many different levels, from creating work flow processes to structuring the way organizational data is stored. Utilizing a CMS as the foundation for the website will enable direct authoring of page content and easier site management while providing us with editing controls (work flows) and versioning that is associated with many internal Document Management Systems. This solution is an excellent fit for NYMTC due to the public nature and vast amount of published the materials we produce.

Summary: Benefits for the Region

Reviewing the partial listing of projects provided, it is clear that NYMTC is taking steps to address the data management deficiencies that were raised earlier. Projects like PIMS and the UPWP Online enable the membership to collaborate and communicate more effectively with the staff, without the requirement of being tied to a specific location. While the Caliper BPM and Bike Paths projects enable the visualization of both collected and generated transportation data. Finally, projects like TIG and the website redesign will enable the organization to communicate more directly with all the stakeholders in the region. Through these current initiatives, and future projects that incorporate the articulated five-point strategy, NYMTC is not only fulfilling its legislative obligations, but is also poised to become the preeminent transportation data clearinghouse for the entire region.