



Region 2
UNIVERSITY TRANSPORTATION RESEARCH CENTER

Final Report

Improvements on NYMTC Data Products

Prepared by

Dr. Kaan Ozbay

(Principal Investigator)

Department of Civil and Environmental Engineering

Rutgers University

Dr. Dilruba Ozmen-Ertekin

(Co-Principal Investigator)

Department of Engineering

Hofstra University

&

Dr. Cynthia Chen

Department of Civil Engineering

City College of New York

November 11, 2009

Disclaimer

The contents of this report reflect the views of those interviewed as interpreted by the authors except where specified. The authors are responsible for the facts, data, and analyses presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration, the Federal Transit Administration, nor the collective membership of the New York Metropolitan Transportation Council (NYMTC). This report does not constitute a standard, specification, or regulation. Acceptance of this report as evidence of fulfillment, in part or whole, of the objectives of the planning study PTCS.07.P12 does not constitute endorsement or approval of the need for any recommended improvement or additional study. It is being sponsored by the NYMTC for the benefit of its members and all interested stakeholders. Other legislation, regulation, executive order(s), official policy, and/or standard practice may supersede the recommendations or advice provided within.

1. Report No.		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Improvements on NYMTC Data Products				5. Report Date November 11, 2009.	
7. Author(s) Kaan Ozbay, Rutgers University, Dilruba Ozmen-Ertekin, Hofstra University, Cynthia Chen, City College of New York				6. Performing Organization Code	
9. Performing Organization Name and Address Rutgers, The State University of New York 33 Livingston Avenue New Brunswick, NJ 08901				8. Performing Organization Report No. 49777-14-19	
The City College of New York 135 th Convent Avenue New York, NY 10031				10. Work Unit No.	
12. Sponsoring Agency Name and Address University Transportation Research Center City College of New York New York, NY 10031				11. Contract or Grant No.	
				13. Type of Report and Period Covered Final Report 1/1/08 – 12/31/09	
				14. Sponsoring Agency Code	
15. Supplementary Notes					
16. Abstract <p>Just like any other scientific research field, the value of data quality is undisputed in the field of transportation. From policy planning to performance evaluation, from model development to impact studies, good quality data is essential to generate ideas and clear-cut solutions to be implemented by transportation professionals and decision makers.</p> <p>In order to improve scientific data quality and function within a continuous quality assessment and management framework, The New York Metropolitan Transportation Council (NYMTC) embarked on this research project in an effort to look for the latest methodologies and technological tools of data management to modernize the existing data products and improve the communication between the agency and the public.</p> <p>In this research project, first, a thorough review and examination of NYMTC data products was performed to identify specific issues about the existing data products, website and the current data maintenance process at NYMTC. Then, appropriate specific solutions, both data and website, and process oriented, were recommended for any issues identified, while taking into consideration the latest technological tools, results of the NYMTC's recent stakeholder survey, the best practices in various fields in the USA and around the world, and NYMTC's current data maintenance practices. Finally, an overall evaluation of the project's potential impacts was performed.</p>					
17. Key Words Data Quality, Data Life Cycle, Dynamic Data Maintenance			18. Distribution Statement		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No of Pages 101	22. Price

Improvements on NYMTC Data Products

FINAL REPORT

November 11, 2009

Submitted By

Dr. Kaan Ozbay
(Principal Investigator)

Associate Professor
Department of Civil and Environmental Engineering
Rutgers University
Piscataway, New Jersey 08854

Dr. Dilruba Ozmen-Ertekin
(Co-Principal Investigator)

Assistant Professor
Department of Engineering
Hofstra University
Hempstead, New York 11549

Dr. Cynthia Chen

Assistant Professor
Department of Civil Engineering
City College of New York
New York, NY 10031



NYMTC Research Project Manager
Mr. Salvatore Mamone

DISCLAIMER STATEMENT

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the New York Metropolitan Transportation Council (NYMTC). This report does not constitute a standard, specification, or regulation.

ACKNOWLEDGEMENTS

The project team would like to acknowledge the significant contributions made by the NYMTC staff. In every task of this project, NYMTC research project manager Mr. Salvatore Mamone, director of technical services Mr. Kuo-Ann Chiao, and from the NYMTC's technical group Mr. Munesh Patel and Ms. Leokadia Glogowski provided the project team with timely advice, and sharp, insightful and constructive criticism that significantly enhanced the quality of the final product.

This project was funded by a joint grant from and administered through NYMTC and the Region 2 University Transportation Research Center.

Their support is both acknowledged and appreciated.

TABLE OF CONTENTS

EXECUTIVE SUMMARY.....	1
CHAPTER 1. INTRODUCTION.....	4
CHAPTER 2. BACKGROUND REVIEW.....	6
2.1. Review and Summary of Best Practices.....	6
2.2. Review and Summary of Latest Technological Advances.....	13
2.3. Review and Summary of NYMTC Stakeholder Survey.....	18
CHAPTER 3. IDENTIFICATION OF ISSUES.....	20
3.1. Comparison of NYMTC to Best Practices.....	20
3.2. Evaluation of NYMTC Data Products Using Stakeholder Survey	26
3.2.1. <i>Customer Satisfaction Index</i>	26
3.2.2. <i>Importance Index</i>	28
3.3. NYMTC’s Current Data Maintenance Process.....	29
3.4. Final List of Issues.....	30
3.4.1. <i>Data and Website Oriented Issues</i>	31
3.4.2. <i>Process Oriented Issues</i>	32
CHAPTER 4. SUGGESTION OF SOLUTIONS.....	34
4.1. Data and Website Oriented Solutions.....	34
4.2. Process Oriented Solutions.....	48
CHAPTER 5. CONTEMPLATING THE NEAR FUTURE.....	52
5.1. Draft Timeline for the Implementation of the Project.....	52
5.2. Overall Evaluation of the Project’s Impacts.....	52
REFERENCES.....	55
APPENDICES.....	56
Appendix 1. Tabular Review of Best Practices.....	56
Appendix 2. Tabular Recommendations for NYMTC Data Products.....	74
Appendix 3. Screenshots from Best Practices.....	81

LIST OF FIGURES

Figure 2.1. NYMTC Stakeholder Survey Fact Sheet.....	19
Figure 3.1. Importance Index (ImpI).....	29
Figure 3.2. NYMTC’s Current Data Maintenance Procedure.....	33
Figure 4.1. Suggested Navigation Path for Data Access Using the Improved NYMTC Website.....	35
Figure 4.2. Current Look of a Data Product Page (County Profiles in this case) at NYMTC Website.....	41
Figure 4.3. Suggested Look for a Data Product Page (County Profiles in this case) at NYMTC Website (Only Center Panel is Shown).....	41
Figure 4.4. Suggested Look for Regional Transportation Statistics Report Page at NYMTC Website (Only Center Panel is Shown).....	42
Figure 4.5. Current NYMTC Home Page.....	44
Figure 4.6. Suggested Page Re-organization for Left, Center and Right Panels of the NYMTC Home Page.....	44
Figure 4.7. Current Data and Model Page at NYMTC Website.....	47
Figure 4.8. Suggested Data & Model Page on the NYMTC Website (Only Center Panel is Shown).....	48
Figure 4.9. Suggested Dynamic Data Maintenance Procedure.....	49
Figure 5.1. A Sample Scatter Diagram Depicting the Effectiveness of the Dynamic Data Maintenance Procedure Over Time.....	54
Figure A1. FHWA’s NHTS Website – User Forum.....	81
Figure A2. BTS Website – A Customizable Table with Radio Buttons (For US Air Carrier Traffic Statistics).....	81
Figure A3. BTS Website – A Customizable Table with Drop-Down Lists (For North American Transborder Freight Data).....	82
Figure A4. EEA Website – Data Viewer with Predefined or Custom Tables/Charts and Metadata.....	83
Figure A5. FARS Website - Query Pages with Commonly Selected Fields Indicated.....	83
Figure A6. FHWA’s NHTS Website – Table Designer Feature with Statistical Analysis and Different Output Options.....	84
Figure A7. BTS Website – A Customized Table with Data Sorting Option.....	84
Figure A8. BTS Website – A Data Table in HTML with Download Option in XLS or CSV Formats.....	85
Figure A9. FARS Website – Data Customization Through Simple Query with Query Exercises for Novice Users.....	85
Figure A10. NYSDOT Website - Home Page.....	86
Figure A11. IMF Website - Home Page.....	86
Figure A12. NCHS Website - Home Page.....	87
Figure A13. NJTPA Website - Helpful Links and Tools.....	87
Figure A14. NCHS Website – Subscription to ListServ (Email List).....	88
Figure A15. WHO-ROE Website – Subscription to Email List.....	88
Figure A16. EEA Website – Subscription to Email List and RSS Feeds.....	88
Figure A17. US Census Bureau Website – Help Page.....	89
Figure A18. CDOT Website – Contact and Feedback Page.....	89

Figure A19. FHWA’s LTPP Website – Feedback Form on the Main Page.....	90
Figure A20. FHWA’s LTPP Website – Customer Satisfaction Survey.....	90
Figure A21. US Census Bureau Website – American Fact Finder- Feedback Web Form.....	91
Figure A22. IMF Website – A More Detailed User Feedback Tool.....	91
Figure A23. IMF Website with Both Browse and Search Options.....	92
Figure A24. NJDOT Website- Research Report Search Boxes.....	92
Figure A25. DFT Website- Advanced Search Page.....	93
Figure A26. NCES Website – Table/Figure Search Page.....	93
Figure A27. NCES Website – Table/Figure Search Results Page.....	94

LIST OF TABLES

Table 1.1. Dimensions of Data Quality.....	4
Table 2.1. Review of Latest Technologies.....	16
Table 3.1. Comparison of NYMTC Data Products to Best Practices.....	21
Table 3.2. Combined Customer Satisfaction Index (CCSI).....	27
Table 3.3. Importance Index (ImpI).....	28
Table 5.1. Draft Timeline for Project Implementation	52
Table 5.2. Potential Pros and Cons of Recommendations.....	53
Table A1. Review of Best Practices.....	56
Table A2. Recommendations for the NYMTC Data Products.....	75

EXECUTIVE SUMMARY

In 2006, Washington State Department of Transportation (WSDOT) Secretary Doug MacDonald created and announced the following one-time \$1,000 Doug MacDonald Challenge in an effort to find new and innovative ways of connecting with the public in regards to transportation demand management techniques:

“Experience has shown that techniques to maximize the use of existing highway capacity can keep traffic moving. Approaches such as congestion pricing and incident management can increase overall system efficiency. Washington State and others are working hard to bring this information to the general public. However, despite our best efforts, opinion research shows we face a skeptical public. Despite our graphs and charts, we haven’t yet found the messages that work; messages that show the relationship between maximizing the use of existing system capacity, increasing efficiency, and keeping highway traffic moving.

We need fresh and creative ideas. How can the public and press – not just transportation performance geeks – get the picture? What are the messages or tools that can leave [ordinary] graphs [and charts] in the dust? Here’s your chance to help make the breakthrough” (WSDOT 2006).

The challenge, which was conducted by Transportation Research Board (TRB) and funded by Secretary MacDonald’s private donation to TRB, offered a “*one-time \$1,000 cash prize to the individual or group that proposes the most effective public communication tool.*” In TRB’s 86th Annual Meeting last year, the winner and the finalists were announced.

In embarking on this research project, NYMTC allied itself with the leading agency in this country in their efforts to modernize the existing data products and improve the communication between the agency and the public.

NYMTC is an association of governments, transportation providers and environmental agencies that is the metropolitan planning organization for New York City, Long Island and the lower Hudson Valley. NYMTC’s mission is 1)to serve as the collaborative forum to address transportation-related issues from a regional perspective; 2)to facilitate informed decision-making within the Council by providing sound technical analyses; 3)to ensure the region is positioned to capture the maximum federal funds available to achieve the goals of the Unified Planning Work Program, Regional Transportation Plan and Transportation Improvement Program; and 4) to focus the collective planning activities of all Council members to achieve a shared regional vision (Source: NYMTC website).

In order to fulfill its mission, NYMTC produces an array of information in support of the planning and policy decisions that shape the region. This information is produced in the form of various reports, brochures and databases, and “*is used by many transportation specialists and academics both in the NYMTC region and outside the region as well as people from other countries*” as explained in the Request for Proposal (RFP) for this project. Thus, NYMTC wants to ensure that its data products are “*current, accurate, consistent, and contain the information that can be used easily by users and serve their needs properly*”; i.e., an **overall quality of the**

data products is desirable. Therefore, it is essential to carry out a project that examines the data products and makes recommendations for further improvements.

The main goal of this research project is to perform a thorough review and examination of NYMTC data products (reports) from data quality point of view in order to ensure that these documents are current, professional looking, consistent, and “*contain the information that can be used easily by the users and serve their needs properly*” as explained in the RFP. Specifically, the research agenda focuses on:

1. Identification of specific issues and problems about the existing data products, website and the current data maintenance practices at NYMTC
2. Suggestion of appropriate solutions, both data and process oriented, for any issues identified,

while taking into consideration the latest technological tools, results of the NYMTC’s stakeholder survey, and best practices and NYMTC’s current data maintenance practices.

A careful investigation and analysis of the results from NYMTC’s Stakeholder Survey revealed important information about the stakeholders’ data usage, familiarity with the data products, preferences regarding formats, general concerns and suggestions for further improvements. The web-based survey carried out by NYMTC in April 2007 had about a 20% response rate. The results indicated that Travel Patterns in the New York Metropolitan Area Brochure and Regional Transportation Statistical Report are the top two most popular and mostly used data products. Freight Facilities and System Inventory turned out to be a data product that the users are least familiar with. The survey also revealed that the users’ most preferred data formats are MS Excel spreadsheets and full-color downloadable PDFs. Data insufficiency and timeliness issues were among the biggest concerns of the users. These invaluable pieces of information obtained from the survey were used as inputs in identifying and recommending solutions for the data, website and data maintenance process related issues.

The customer satisfaction index calculated for NYMTC’s data products turned out to be about 80%, which indicates a relatively high level of overall quality. However, it should be noted that the survey had only 32 responses in total, hence the accuracy of this result should remain rather questionable. A better and far more accurate evaluation would be possible with a larger sample size.

The average Importance Index for all 11 of the data products was calculated and found as 0.685, indicating that the stakeholders value the data products with a relatively high level of importance overall. Stakeholder survey results and detailed analysis can be found in Chapters 2.3 and 3.2 in this report.

Poor data organization and occasional lack of integration, insufficient data archives, quite long lag times in releasing NYMTC’s data products (ranging between 2-4 or even 7 years in some cases), insufficient and in some cases nonexistent metadata, variable dictionary, too few format choices, lack of data generation and extraction features and user feedback mechanisms were among the weaknesses of NYMTC’s data products when compared to the best practices reviewed in this project. Greater details can be found in Chapter 3.1.

A review of the report titled “Data Group Cookbook for Reports and Brochures” (NYMTC) revealed some minor issues with NYMTC’s current data maintenance process; *‘Monitoring Effectiveness’* and *‘Improving the Process’* components are not formal parts of the current process while their existence is significant in closing the loop in the life cycle of data products. Details can be found in Chapter 3.4.2.

Specific solutions recommended to all of the abovementioned issues are presented in great detail in Chapter 4. All of the prioritized recommendations made there aim to offer to the wide range of NYMTC website visitors and users better integrated, interactive, and easily accessible data that comes with sufficient metadata, is GIS-based and user friendly.

In addition to specific solutions to data and website related issues, a dynamic data maintenance procedure is also proposed, which might potentially be applied on a regular basis (bi-annually or whenever a new data is produced and required to be disseminated). This proposed procedure is built around three main functional components (review, monitor/assess, and modify/improve). The suggested procedure also incorporates the elements from the lifecycle of data (create, distribute, access, update) as explained in detail in Chapter 4.2.

A very crucial element in the proposed dynamic data maintenance process is creation of quality data first time every time by the data producing agencies. Thus, the research team proposes to organize and conduct a series of workshops (at least two) where all the contributing agencies are brought together to set up a more formal and reliable mechanism for NYMTC to work with these external agencies to obtain necessary data. Some details of the proposed workshop are also explained in this report; however, a complete workshop proposal will be submitted to NYMTC as a separate document.

CHAPTER 1

INTRODUCTION

Enterprise data quality is crucial in organizations across all industries in order to increase productivity, reduce the time spent on detecting and fixing errors, and meet end-users' expectations properly, thus increasing the organization's reputation in the competitive business environments.

The quality of data has various dimensions defining the characteristics of data in measurable forms. A data quality dimension is defined as “*a set of data quality attributes that most data consumers react to in a fairly consistent way*” (Wang et al. 2001). Table 1.1 below provides different categories and dimensions of data quality (Wang et al. 1996). These dimensions can be used as a starting point in determining the attributes related to data quality that are most suitable and relevant to specific industries.

Table 1.1 Dimensions of Data Quality (Wang et al. 1996; Lee et al. 2002)

Category	Dimension	Description
Intrinsic	Accuracy	The level to which stored data agree with accepted sources of “correct” information
	Objectivity	Whether the information was objectively collected and is based on facts
	Believability	The degree of credibility or trustworthiness of the information
	Reputation	Whether the information has a good reputation for quality
Accessibility	Accessibility	The ease and breadth of access to information
	Access Security	Whether the information is protected against unauthorized access
Contextual	Relevancy	Whether the information is useful / relevant / appropriate / applicable to intended purpose
	Value-Added	Overall quality so as to add value
	Timeliness	Freshness and up-to-date state of information
	Completeness	The expectation that certain attributes are expected to have assigned values in a data set
	Amount of Data	Sufficiency/Insufficiency of information
Representational	Interpretability	Enables the users to make accurate inferences
	Ease of Understanding	Format and presentation that is easy to understand
	Concise Representation	Whether the information is formatted compactly and presented concisely
	Consistent Representation	Whether instances of data are represented in a format that is consistent with the domains of values as well as consistent with other similar attribute values

Just as important as the data quality, is to have a well-established dynamic data maintenance process in place and to adopt an organizational approach in which all the agency's technical staff, agencies providing data, and other stakeholders make improving the quality of data products a priority.

Loshin (2007) argues that “*information follows a ‘lifecycle’ consisting of creation, distribution, access, updating, and retirement stages, and that it is necessary that a data quality framework*

provide protocols for measuring the quality of information at the various stages of that life cycle". That is why one of the main goals of this project is to propose an agency-wide, systematic, easily understandable and dynamic data maintenance procedure to guide the NYMTC staff in moving forward with effective data collection and dissemination efforts.

Another goal of this project is to perform a thorough review and examination of NYMTC data products (reports) from data quality point of view in order to ensure that these documents are current, professional looking, consistent, and "*contain the information that can be used easily by the users and serve their needs properly*" as explained in the RFP.

Thus, this project aims to recommend solutions to both data oriented and process related issues which will be identified in light of a thorough background review and a careful analysis of the status quo regarding the NYMTC data products, their quality control and maintenance processes.

This report is the result of the hard work of all individuals involved as staff members of NYMTC and the research team. More often than not, the authors produced a draft document that was dramatically changed on the basis of the feedback received from the NYMTC staff, whose comments significantly improved the quality of the document.

This report is comprised of five major chapters, including this introduction. A brief description of the chapters and the corresponding authors and contributors are shown next.

Chapter 1: Introduction (Author: Professor Dilruba Ozmen-Ertekin. Reviewed and Commented by: Professor Kaan Ozbay).

Chapter 2: Background Review (Author: Professor Dilruba Ozmen-Ertekin, in cooperation with Cynthia Chen of City College in the review of best practices section. Reviewed and Commented by: Professor Kaan Ozbay). This chapter presents the review of the best practices and the latest technological advances in the field of data quality and maintenance, as well as the results from NYMTC's Stakeholder Survey.

Chapter 3: Identification of Issues (Author: Professor Dilruba Ozmen-Ertekin. Reviewed and Commented by: Professor Kaan Ozbay). In this chapter, a comparison between NYMTC and best practices regarding data quality and maintenance efforts is made, an evaluation using the results of the NYMTC Stakeholder Survey is done, NYMTC's current data maintenance practices are summarized and a comprehensive list of data and process oriented issues are presented.

Chapter 4: Suggestion of Solutions (Author: Professor Dilruba Ozmen-Ertekin, in cooperation with Professor Kaan Ozbay). In this chapter, specific solutions are recommended for the issues identified.

Chapter 5: Contemplating the Near Future (Author: Professor Dilruba Ozmen-Ertekin. Reviewed and Commented by: Professor Kaan Ozbay). In this chapter, a draft time line is suggested for the implementation of this project, and potential impacts of the project, if implemented, are discussed.

CHAPTER 2

BACKGROUND REVIEW

The research team first embarked on an effort to collect and review the data products and reports published and made available to the general public by other leading and reputable agencies throughout the country and around the world. This way, knowledge on best practices regarding different dimensions of data quality was acquired. Practices were chosen from within the transportation community or from other areas. A complete review of the literature on best practices is presented in the following sections.

Next, a search on the latest technological advances related to data dissemination methods was conducted; to help the research team decide on the best ways of improving the presentation formats for NYMTC's data products. The results from the extensive review of the latest technological tools and methodologies available for data dissemination are presented in the following sections as well.

Finally, a copy of the results of NYMTC's Stakeholder Survey was obtained from NYMTC and examined thoroughly to understand stakeholders' specific concerns and recommendations about the data products. The stakeholder survey was crucial in identifying problems related to representational data quality especially, since it is the most subjective dimension and requires extensive user input. The results from the NYMTC's Stakeholder Survey are presented in the last section of this chapter.

2.1. Review and Summary of Best Practices

The research team's review of the best practices for database management and dissemination involved two major steps. In the first step, a broad but less detailed search was conducted to identify a relatively long list of websites that require further probing in the second step. The U.S. government website (http://www.usa.gov/Site_Index/D.shtml), which lists every department in the U.S. government was visited. Each departmental section was inspected to identify industries that have various and large amounts of data for the public. Then, for each identified industry, the institutions and agencies with the highest number of citations as data providers were selected using two search engines (Google and Academic Search Premier). This way, an initial list of 60+ websites was generated.. Then, a review of the over 60 websites was conducted in the second step to select the final list of best practice websites while keeping in mind three main selection criterion: the total number of databases available, the length of the data coverage, and accessibility to the datasets. For transportation related websites, in addition to the websites provided by the federal governments, other state DOTs' and MPOs' websites were also considered, as well as other English-speaking countries' transportation websites. As a result, a final list of 16 best practices was obtained. These 16 selected websites cover different industries (including transportation, weather, real estate, finance, health and forestry), both public agencies and private firms located in the U.S. or overseas.

Table A1 in Appendix 1 gives a concise summary of the 16 best practices reviewed in detail. Additional descriptions of the 16 best practices reviewed are presented in the following subsections.

2.1.1. Bureau of Transportation Statistics (<http://www.bts.gov/>)

BTS provides a variety of information through its website. Searching the available data items is possible using the search box on the main page. Data Finder feature allows access to databases grouped by different subject areas. Databases feature takes the user to a new webpage that contains an alphabetical list of all the databases and all the data tables in each database.

Some of the data tables are selectively downloadable in XLS format (i.e., the user can choose the data fields to be displayed in the tables), while some of them are downloadable as is, in HTML, XLS, or CSV formats.

Data Profiles provided within the Databases feature provides a description of each data table. However, variable definitions and explanations are not always available and not consistently located throughout the website. For example, the data dictionaries of different databases are often located in different places and not always easy to find.

The website lists data sources (database collector) in each database and provides a link to the original collector's website. It also offers links to other transportation related database/service websites (National Transportation Library, etc.).

Almost all the databases provided on the BTS website have table format, however, the provision of maps/charts is extremely limited.

The BTS website is limited in user interaction as it has almost no support or feedback tools. It is also difficult to find information on how to purchase or request a report.

2.1.2. New Jersey Department of Transportation (<http://www.nj.gov/transportation/>)

NJDOT provides various reports and publications to its users. Crash Records, for instance, are presented in HTML and PDF formats. Regarding GIS data and maps, maps are given in PDF format, whereas several databases associated with layers (e.g. road network) are given as ZIP files. Metadata for the GIS data regarding NJDOT roads is also provided. Various other reports are provided in PDF formats and are available through an efficient 'search' feature.

There are some data products (traffic counts, soil borings) that utilize interactive screens.

The NJDOT website also provides reports and publications prepared by other transportation agencies, for example, Federal Highway Administration (FHWA).

2.1.3. North Jersey Transportation Planning Authority (<http://www.njtpa.org/>)

NJTPA provides various data products as well as links to some data and maps from other sources (e.g. Environmental Systems Research Institute, NJ Department of Environmental Protection, US Census Bureau, NJ Image Warehouse) on its website.

The NJTPA website provides users a quick and easy access to databases of interest, because of its clear data descriptions and variable dictionary.

Many of the reports are presented in PDF format. Demographic and employment forecast tables are also given in PDF format. GIS databases and shape files are provided as ZIP files. Metadata

for the GIS data is also provided as ZIP files. Crash data are given in DBF format, which is quite easy to manage by the user after download.

Maps are available at different geographical levels (e.g., state, county). Different functional layers (e.g., surface, roadway network, etc.) are provided with images, statistics, and downloadable GIS shape files.

The website interface design is clear; it is easy for the user to find support tools.

The main limitation is that, relatively few databases are available and they date back only a few years. In addition, access to publications is not as transparent as the access to datasets. The publications and reports are listed under different categories (projects and programs, for example) and thus, are absent in publications list.

2.1.4. Ministry of Forests and Range, BC (<http://www.for.gov.bc.ca/mof/reports.htm>)

The British Columbia MFR (BCMFR) website provides various reports and publications to its users. Reports such as Annual Reports, Service Plans, and Ministry Publications are provided in the form of a list in both HTML and PDF formats. For the Ministry Publications item, there is a search box.

BCMFR has also recently begun to develop a data quality framework to encourage higher quality data within the Ministry as well as by the Ministry's various partners and stakeholders. The framework will utilize data quality report cards to identify problem areas within BCMFR data products.

Some of BCMFR's reports cover a very long time period and use various databases. For these reports, the website lets the user choose either downloading the full report or only the charts/maps/tables inside the report.

In the e-FM page of this website, a set of data quality survey and improvement feedback tools and education programs is available, which can be considered a pioneer practice in this format of services among its sister agencies' websites.

2.1.5. Colorado Department of Transportation (<http://www.dot.state.co.us/>)

CDOT provides a variety of data products on its website. It has a clear user interface for all databases that are highly accessible. After selecting a data category from the main page, the user gets presented with additional options and explanations. Maps are provided in PDF format whereas statistics are given in HTML or XLS formats. Shape files containing geographic data are given in the form of downloadable ZIP files. A link to the metadata describing GIS files is also available. The map products provided cover multiple aspects, which include both city/county maps with highway/railroads information and travel/traffic count maps for different regions. Reporting and displaying traffic count data in map format is more obvious and easy to visualize than the traditional table format, thus, more user-friendly.

On the Maps page, a statement indicating how to request paper copies of map products is also provided.

On the Straight Line Diagrams page, the user is presented with various choices regarding what information to be displayed in the final map (in PDF format).

2.1.6. Washington State Department of Transportation (<http://www.wsdot.wa.gov/>)

WSDOT provides a great variety of data products on its website and each one is clearly described. Directly on its main page, there is local traffic information, travel alert of accident/construction, the field picture/map of each highway, and the 95% confidence estimation of travel time. All these are key information for local residents. A search box on the main page facilitates locating desired maps and data products. A link showing the most requested items is also available. Reports, statistics, and maps are given in PDF format. Shape files containing geographic data are given in the form of downloadable ZIP files with a file preview option.

The website also provides a blog, where users can enter their opinions about the website design.

The limitation is that most databases are provided in a single file format. Another is that some of the data products (e.g. aerial photos) are only available for purchasing.

2.1.7. Fatality Analysis Reporting System, NHTSA (<http://www-fars.nhtsa.dot.gov/>)

The FARS website provides highway crash data and safety related reports/publications. The query tools for generating downloadable crash data are easy to operate and highly accessible, with detailed variable choice options. The user can choose to generate crash data tables, or choose to obtain analyzable cross-tables.

Another highlight of this website is its anonymous FTP service that requires no password. SAS/DBF formatted databases and variable dictionaries can be directly downloaded from this FTP server. Publications and reports are also well organized and thus easy to find.

2.1.8. US Census Bureau (<http://www.census.gov/>)

The US Census Bureau website has the most number of databases available and it offers a highly interactive data query service. In its geographic data service page, users can make selections for a query, save the selection on a local hard disk, and reload the saved selection later. This feature will save time for the user.

For some databases, basic statistical analysis tools like ranking is also available.

The website also provides related map links for some databases as well as links to the FTP server.

The main limitation is that the FTP service is not as informative as the FARS website's FTP service, since the databases are not properly labeled and lack necessary descriptions.

2.1.9. National Center for Education Statistics (<http://nces.ed.gov/>)

The NCES website provides information on local schools and national survey results. There are two ways to search databases; one through a list of data tools (subject specified) and another through the keyword search tool provided on the website. For each search result, the users can

choose a preferred format for downloading. Related information on the same geographical region can also be downloaded at the same time.

For each element in its national survey program, the website provides explanations on the program as well as its data outputs. In associated databases, publications and reports are also available for download on the same page.

Also on its main page, there is a section called “fast facts”, in which a list of popular topics is presented. For each fact, answers are given first, followed by a link to the source database and the links to other related reports/databases. This is a much more informative format designed for both novice and advanced users, and a much more preferred format than traditional, static FAQ page.

2.1.10. National Center for Health Statistics (<http://www.cdc.gov/nchs/>)

The NCHS website provides a list of data/statistics as well as highlights and related resources. The “getting started” tool on the main page acts as an advanced search engine where users can find information on what they need from an alphabetical listing of statistics, a list of important health topics, and links to the key annual reports.

The data/statistics service is a highly interactive service, as one can find facts, highlights, databases, or updated information on this page quickly. It allows various ways of selecting and accessing databases and publications at the same time. The user can also choose a particular topic. Once a topic (for example, health topic) is chosen, all related databases are listed and the user can choose from the list. For a specific database chosen, the user is first provided with a PDF which lists all the related tables. Each table has an associated hyperlink and clicking on it will direct the user to the table. If the user is not interested in specific tables, the same PDF contains a basic description and a brief report of the database.

2.1.11. National Climatic Data Center (<http://www.ncdc.noaa.gov/oa/ncdc.html>)

The NCDC website provides databases in various areas (climate, weather, surface condition) both in U.S and internationally. Ways of access to its databases is various. The user can choose to access by its collection method (e.g., satellite data, monitoring data, etc), by an alphabetical list of all the data products (which includes option to purchase requested data) or most popular databases lists, or search through a global map. For the latter option, the user is prompted to choose a specific database, its format, and variables of interest before downloading. This map-based search engine has a number of unique features. First, users can access all the free databases related to a specific location by utilizing the data tool on the map. Second, through the use of the graph tool on the map, the website will generate charts of specific data based on the selection of location, or even allow the user to compare data from two databases, conduct some basic analyses or draw charts of the results.

The website also provides an FTP service. The user can select both data and metadata from multiple databases in a single query. In other words, the user can ask for weather data of one county and coast data of another county (with metadata for both) and request the website to generate them into a single downloadable folder. The site will send all the requested items into a

folder in its FTP server and then allow the user to access and download them by using their email address as identification key and password.

2.1.12. National Association of Realtors (<http://www.realtor.org/>)

NAR is an organization of real estate agencies in the U.S with more than 1 million members. It provides a large amount of free data/report, covering almost all geographical areas in the U.S. For each topic (for example, sales), a list of sub-group topics are listed and its statistics are provided. The user can select detailed information for download through this sub-group. The site, however, discriminates between users. Many of the full reports are only accessible to member agency employees and require user login and password. The site, of course, still offers much useful information, despite the unavailability of many reports.

The user interface is friendly, providing a quick access. On the website, there is a “quick statistics service”, which lists basic (national level, for example) and key (most requested, for example) statistics on selected topics. Moreover, there is an online chat service available for users who may need any quick facts immediately.

The research team observed that the NAR website has the shortest time lag between data date and publication date among the 16 websites reviewed; it is generally less than 6 months and some of the publications only take 15 days from data date to the time they become available online.

2.1.13. International Monetary Fund (<http://www.imf.org/external/index.htm>)

IMF is one of the largest international financial/economics organizations and provides a great amount of databases, reports, and statistics online. The website has several interesting features. It dynamically updates the most frequently requested dataset and its related information. The “start here” tool provides additional accessibility to appropriate databases. Each row of the table is one topic of interest (payment balance, for example) and all databases are listed as single columns, which makes data selection easier.

For most databases, raw data is given in CSV format and charts are available only in PDF format.

Another interesting feature is that IMF can group its analysis and forecasting results, including tables and charts, on certain databases and create a new database.

For publications, there is a selection box on the website. The users can either run a search by key words or select a report from the full list of publications.

2.1.14. World Health Organization Regional Office for Europe (<http://www.euro.who.int/>)

The WHO-ROE website provides a great amount of health related data in Europe at country level. The “health for all databases” selection tool box on the website is very unique supposedly. With this tool box, users are supposed to easily obtain health related information collected by multiple agencies, with a single query. The research team, however, was not able to test this unique feature, as certain parts of the selection page always read “not available”.

Some statistical analyses like the histogram are available when choosing the output of the selected data.

The user can choose one health topic (for example, cancer) on the website and the site will automatically list all the related topics (news, data, publications, etc.) as well as associated search results within each of these topics.

In the search page for publications, in addition to the search by key words, users can either read the article list (sorted in health topics as well) or view the most recent publications.

2.1.15. European Environment Agency (<http://www.eea.europa.eu/>)

The EEA website provides environmental information for Europe and provides free access to various environmental databases. The website contains several useful items, including weekly highlights, frequently visited datasets and maps, and calendar of events. Associated links to each item are also provided (air pollution, energy, etc).

The website provides a highly interactive user interface for map customization. Users can choose from the themes of interest (household consumption, industry emission, etc.) and all the related maps inside the site will be listed. Each will link to a new page which contains a save-friendly version of the map with specification of data source and detailed metadata information. The research team considers this website having the most advanced map customization tools among all the 16 websites reviewed.

Another advanced service provided by the EEA website is called “live map and data”. By clicking on the database of interest, a map and data selection window will be generated, where users can interactively select display of data at a different geographic level by using the zoom tool and the data filter tool. These databases are actually collected by different agencies and even located at different websites; however the “live map and data” service provides a unified access to them.

2.1.16. Department for Transport, UK (<http://www.dft.gov.uk/>)

DFT is the transportation authority in the UK and its website contains traffic information for all modes and for all areas. On this website, the user can select between air, rail, road and sea; each selection will provide users with a list of related items. A well-designed quick search box is also available.

The latest releases are highlighted on the website. Moreover, the forthcoming release schedule is available. This service is unique, convenient and valuable to the users.

This website arranges its databases by a description of the purpose of the database, the link to the data provider (if not DFT itself), followed by all the documents related to this database in a descending sequence of publication date. The majority of these statistics is provided in PDF format.

Even if the latest releases are not available online yet, the website provides a bulletin of the key points.

The advanced search window allows the users to choose from similarly spelled words. This feature may improve the efficiency in search and provides more information to the user.

The website also contains a link to ‘popular pages’.

2.2. Review and Summary of Latest Technological Advances

New technologies are emerging all the time in the field of data dissemination which is a crucial step in data life cycle. These new technologies provide the users/consumers with more and different choices in retrieving quality data easily and in the desirable format. From data provider’s point of view, new technologies give way to wider circulation of data products.

The research team’s review of the latest technological advances involved three steps. In the first step, technologies in the area of data creation were researched. Next, the technologies that facilitate better data distribution and user feedback were examined. Finally, in the third step, technologies in data manipulation area were reviewed. In total, 14 technologies were examined in detail as presented in Table 2.1.

Additional descriptions for some of the technologies are given in the following subsections.

2.2.1. Web Modeling

Web modeling addresses the specific issues related to design and development of large-scale web applications. In particular, it focuses on the design notations and visual languages that can be used for the realization of robust, well-structured, usable and maintainable web applications. Designing a data-intensive website amounts to specifying its characteristics in terms of various orthogonal abstractions. The main orthogonal models that are involved in complex web application design are: data structure, content composition, navigation paths, and presentation model.

Several languages and notations have been devised for web application modeling. These include HDM-W2000, RMM , OOHDM, ARANEUS, STRUDEL, and WebML.

2.2.2. Data Mining

Data mining is the process of sorting through large amounts of data and picking out relevant information. It is increasingly being used to extract information from the enormous data sets generated by modern experimental and observational methods. It has been described as “*the nontrivial extraction of implicit, previously unknown, and potentially useful information from data*” (Frawley et. Al 1992) and “*the science of extracting useful information from large data sets or databases*” (Hand et. Al 2001).

Data mining identifies trends within data that go beyond simple analysis. Through the use of sophisticated algorithms, the users have the opportunity to identify key attributes of databases.

Although data mining is a relatively new term, the technology is not. For many years, businesses have used powerful computers to sift through volumes of data such as supermarket scanner data to produce market research reports (although reporting is not considered to be data mining).

Continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy and usefulness of data analysis.

The term data mining is often used to apply to the two separate processes of knowledge discovery and prediction. Knowledge discovery provides explicit information that has a readable form and can be understood by a user. Forecasting, or predictive modeling provides predictions of future events and may be transparent and readable in some approaches (e.g., rule-based systems) and opaque in others such as neural networks. Moreover, some data-mining systems such as neural networks are inherently geared towards prediction and pattern recognition, rather than knowledge discovery.

“An unavoidable weakness of data mining is that the critical data that may expose any relationship might have never been observed. Alternative approaches using an experiment-based approach such as Choice Modeling for human-generated data may be used. Inherent correlations are either controlled for or removed altogether through the construction of an experimental design.

Recently, there were some efforts to define a standard for data mining, for example the CRISP-DM standard for analysis processes or the Java Data-Mining Standard. Independent of these standardization efforts, freely available open-source software systems like RapidMiner and Weka have become an informal standard for defining data-mining processes”. (Taken from the online source <http://wikipedia.org>).

2.2.3. Really Simple Syndication (RSS) Feeds

RSS is a family of web feed formats used to publish frequently updated content such as blog entries, news headlines, and podcasts in a standardized format. An RSS document (which is called a "feed" contains either a summary of content from an associated web site or the full text. RSS makes it possible for people to keep up with web sites in an automated manner that can be piped into special programs or filtered displays.

RSS content can be read using software called an "RSS reader", which can be web-based or desktop-based. A standardized XML file format allows the information to be published once and viewed by many different programs. The user subscribes to a feed by entering the feed's link into the reader or by clicking an RSS icon in a browser that initiates the subscription process. The RSS reader checks the user's subscribed feeds regularly for new content, downloads any updates that it finds, and provides a user interface to monitor and read the feeds. (Taken from the online sources <http://guardian.co.uk> , <http://wikipedia.org>).

2.2.4. Wiki

A wiki is a collection of webpages designed to enable anyone who accesses it to contribute or modify content, using a simplified markup language. Wikis are often used to create collaborative websites and to power community websites. Wikis are used in business to provide intranets and Knowledge Management systems. Ward Cunningham, developer of the first wiki software, WikiWikiWeb, originally described it as “the simplest online database that could possibly work”.

“A defining characteristic of wiki technology is the ease with which pages can be created and updated. Generally, there is no review before modifications are accepted. Many wikis are open to alteration by the general public without requiring them to register user accounts. Sometimes logging in for a session is recommended, to create a "wiki-signature" cookie for signing edits automatically. Many edits, however, can be made in real-time and appear almost instantly online. This can facilitate abuse of the system. Private wiki servers require user authentication to edit pages, and sometimes even to read them. Many implementations allow users to supply an "edit summary" when they edit a page. This is a short piece of text summarizing the changes. It is not inserted into the article, but is stored along with that revision of the page, allowing users to explain what has been done and why; this is similar to a log message when committing changes to a revision control system. Most wikis keep a record of changes made to wiki pages; often every version of the page is stored. This means that authors can revert to an older version of the page, should it be necessary because a mistake has been made”. (Taken from the online source <http://wikipedia.org>).

2.2.5. Online Data Extraction/Web-Based Query

Online data extraction is facilitated by interactive websites that allow users to create custom tables/charts/maps using a graphical user interface (GUI)-based questionnaire.

Most of the information on the web today is in the form of Hypertext Markup Language (HTML) documents which are viewed by humans with a browser (Myllymaki 2001).

Pages in data-intensive sites are usually automatically generated: data are stored in a back-end Database Management System, and HTML pages are produced using scripts (i.e., programs) from the content of the database. Data extraction from HTML sites is usually performed by software modules called “wrappers” that translate a database query to a web request and parse the resulting HTML page (Crescenzi et. al 2001).

Online data extraction tools are available extensively on various websites such as the US Census Bureau's American Fact Finder and Data Ferret features, Long Term Pavement Performance of FHWA, and Table Designer of National Household Travel Survey, to name a few.

2.2.6. Online Data Analysis

Online data analysis is a feature of interactive websites that allow users to perform various statistical analyses and create related tables/charts.

“Not all data is suitable for analysis and presentation with online data analysis. The data should be rectangular in nature; in other words, it should be representable using a row-and-column structure. A common way of obtaining such data could be by querying relational databases or by reading from spreadsheets or other file-based data” (Moore and Clark 2004).

A popular tool for developing dynamic websites is JavaServer Pages (JSP) which is a Java technology that allows software developers to dynamically generate HTML/ XML/other types of documents in response to a web client request.

Table 2.1. Review of Latest Technologies

Technology Category	Technology Name	Description	Example Applications	Popularity
Data Creation	Web Modeling	High-level design of data-intensive, dynamic websites. Involves different models for data organization, redundant data definition, page composition, links, presentation	Business, Finance, Engineering	Moderate
	Data Mining	The process of sorting through large amounts of data and picking out relevant information	Business, Science, Engineering, Economics, National Security	Moderate
	Metadata Generation/Extraction	Generating/extracting a set of attributes, or elements, necessary to describe the data/resource in question	Extensive	High
Data Distribution/User Feedback	File Transfer Protocol (FTP)	A network protocol used to transfer data from one computer to another over the Internet	Various	Moderate
	Really Simple Syndication (RSS) Feeds	Summary of latest content from an associated website	Extensive	High
	Podcasts	A series of digital-media files distributed over the Internet using syndication feeds for playback on portable media players and computers	Extensive	High
	Electronic Mailing Lists	Collection of e-mail addresses used by an individual/organization to send material to multiple recipients	Extensive	High
	Audio/Video	-	Extensive	High
	Web Logs (Blogs)	A website with regular entries of commentary, descriptions of events, or other material such as graphics or video	Extensive	High
	Email a Friend/ Get&Share	A radio button that lets the users to email the web page that they are visiting to a friend	Extensive	High
	Webforms	A standard interface that can be downloaded from the Internet, and that contains text boxes for a user to enter data	Extensive	High
	Wiki	A World Wide Web application designed to allow multiple users to add, remove, and edit content), to create a common community environment for data providers as well as users	Extensive	High

Table 2.1. Review of Latest Technologies (Cont'd)

Technology Category	Technology Name	Level of Effort/Time/Money Required to Adopt/Use this Technology		Potential Benefits
		NYMTC Staff	User	
Data Creation	Web Modeling	High - knowledge about a web modeling language (e.g. WebML) and software tool (e.g. WebRatio, which is a popular, licensed software) and familiarity with XML required	N/A	Help NYMTC staff manage develop an exploratory, browsing-oriented website
	Data Mining	Moderate - familiarity with data mining techniques and software required	N/A	Help NYMTC staff analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified
	Metadata Generation/Extraction	Moderate - familiarity with metadata generation/extraction language (e.g. Dublin Core) and HTML required	N/A	Facilitates the understanding, characteristics, and management usage of data
Data Distribution/User Feedback	File Transfer Protocol (FTP)	Low - simply requires setting up and maintaining an FTP server	Low - simply requires the eligible user to follow the FTP link to connect and download the files	Facilitates user access to further data that is not readily available on the website
	Really Simple Syndication (RSS) Feeds	Moderate - requires familiarity with XML to build, edit, and publish feeds, but freeware/software also available	Low - simply requires user to copy-paste feed URL to an RSS feeder (such as My Yahoo!)	Allow users to keep up with website in an automated manner, Retain users
	Podcasts	Moderate - requires familiarity with XML to build, edit, and publish feeds, but freeware/software also available	Low - simply requires user to download and install appropriate software like iTunes, iPodder, My Yahoo that will subscribe to the podcast list and stay updated, show the available content and automatically retrieve the most recent files. These programs can also play the file on the computer	
	Electronic Mailing Lists	Moderate - requires familiarity with a programming language to set up list, but freeware/software also available	Low - simply requires user to subscribe to list through a link on website	
	Audio/Video	Low - simply requires uploading the recorded audio/video on to the website	Low - simply requires user to click on audio/video button on website (if automatic pop-up is not set up)	Improve data/metadata representation
	Web Logs (Blogs)	Low - requires NYMTC to develop the medium and maintain the logs	Low - requires user to participate in the logs	Increase user awareness of data
	Email a Friend/ Get&Share	Low - requires opying a simple HTML code on to the existing website to create the appropriate link on the page	Low - simply requires user to click the given link and fill in related fields	
	Webforms			Allow users to enter data about themselves or comment on the website
	Wiki	Low - same as creating a website	Low - simply requires registered (free) user to click the given button (edit) to add or update content	

Table 2.1. Review of Latest Technologies (Cont'd)

Technology Category	Technology Name	Description	Example Applications	Popularity
Data Manipulation	Online Data Extraction/Web-Based Query	Interactive websites that allow users to create custom tables/charts/maps using a GUI-based questionnaire	Various (e.g. US Census Bureau's American Fact Finder and Data Ferret features, LTPP of FHWA, Table Designer of NHTS)	Moderate
	Online Data Analysis	Interactive websites that allow users to perform various statistical analysis and create related tables/charts	Various	Moderate

Table 2.1. Review of Latest Technologies (Cont'd)

Technology Category	Technology Name	Level of Effort/Time/Money Required to Adopt/Use this Technology		Potential Benefits
		NYMTC Staff	User	
Data Manipulation	Online Data Extraction/Web-Based Query	Moderate - <i>familiarity with data extraction software and XML required</i>	Low to Moderate - <i>familiarity with SQL to perform queries may be required depending on the user interface of the website</i>	Lets users customize the way the raw data is presented
	Online Data Analysis	High - <i>familiarity with developing dynamic websites (e.g., JavaServer Pages (JSP) which is a Java technology that allows software developers to dynamically generate HTML/XML/other types of documents in response to a web client request. It allows Java code and certain pre-defined actions to be embedded into static content), Data must be in row-column format</i>	Low to Moderate - <i>familiarity with data analysis and statistical concepts</i>	Lets users do further analysis using the already presented raw data

2.3. Review and Summary of NYMTC Stakeholder Survey

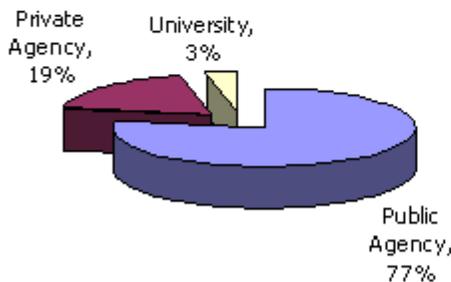
In this section, the research team has reviewed the completed survey and prepared a 1-page 'survey fact sheet' as given in Figure 2.1 in the following page by summarizing the overall survey results in a compact form along with pie charts/graphs so that everyone can get a quick idea about the research team's findings.

IMPROVEMENTS ON NYMTC DATA PRODUCTS: WEB-BASED USER SURVEY

Overview

NYMTC carried out a web-based survey in April 2007 consisting of 7 questions aiming to find out the most popular data products and formats, and to hear people’s suggestions about various ways to improve data formats and contents. Surveys were sent to 162 participants of public and private transportation agencies in New York Metropolitan area and suburbs, including NYMTC member agencies, via e-mail which included a link to the survey hosted by SurveyMonkey.com. In the end, there were 32 responses (about 20% response rate).

Profile of Survey Respondents



The following are some notable observations from the survey results:

Usage and Familiarity

Top 5 Mostly Used Data Products

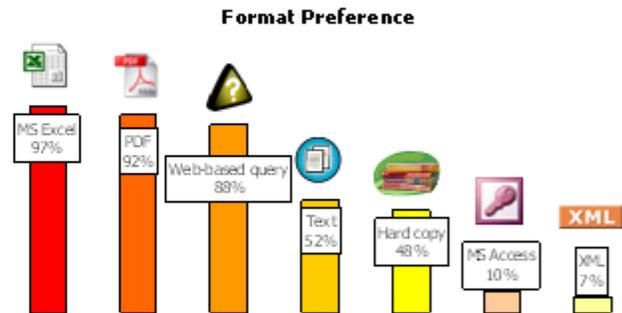
- Travel Patterns in the New York Metropolitan Area Brochure-Quarterly, and Regional Transportation Statistical Report-Annually (69% each),
- Regional Transportation at a Glance Brochure-Annually (62%),
- Hub-Bound Travel Report-Annually (59%),
- Regional Demographics at a Glance Brochure-Annually, Truck Toll Volume Report-Annually, and County Profiles (55% each),
- Hub-Bound Travel at a Glance Brochure-Annually (52%).

Top 5 Data Products Users are Unfamiliar With



Preferred Formats

- 97% preferred MS Excel spreadsheet format of data tabulations,
- 92% preferred full-color downloadable PDF,
- 88% would like web-based queries to be able to generate interactive tables, charts, maps.



Biggest Concerns/Suggestions for Improvements

- **Accuracy (5%)**
 - Cross-check with other agencies.
- **Consistency (11%)**
 - Consistency check with other agencies (5%) (specifically, Transit Bureau was mentioned),
 - Concurrence check with other agencies (5%).
- **Amount of Data (42%)**
 - Expand existing data (21%) (e.g. with more details on time/day of travel, vehicle class, County Profiles for the entire region),
 - Provide new data (21%) (e.g. longitude/latitude to facilitate map visualization by user, provide data at disaggregate geographic levels such as town, multiple zip code, crash data, airline boardings, Amtrak boardings).
- **Timeliness (32%)**
 - Keep at most 1-year lag between data and release times (specifically hub-bound travel and regional transportation statistical reports were mentioned).
- **Appropriateness of Format and Ease of Understanding (11%)**
 - Provide better metadata (5%) (specifically hub-bound reports were mentioned),
 - Provide better information on data sources (5%) (specifically hub-bound reports were mentioned),
 - Provide data in single MS Excel file formats for easier use, rather than PDF only (5%),
 - Enhance the data website by including more graphics (5%).
- **User Feedback (5%)**
 - Seek and *implement* suggestions for improvements from agencies (Mid-Hudson South Transportation Coordinating Committee was mentioned).

Figure 2.1. NYMTC Stakeholder Survey Fact Sheet

CHAPTER 3

IDENTIFICATION OF ISSUES

3.1. Comparison of NYMTC to Best Practices

NYMTC's online data products were reviewed to reveal their strengths and weaknesses (if any) and to compare them to the best practices that were reviewed in Chapter 2.1 earlier. Specifically, the following data products were the focus of the research team's review process, as specified in the RFP for this project:

- 1.** County Profiles (one for each County)-every 10 years
- 2.** Hub Bound Travel at a Glance Brochure-annual
- 3.** Hub Bound Travel Report-annual
- 4.** Regional Demographics at a Glance Brochure-annual
- 5.** Regional Transportation at a Glance Brochure-annual
- 6.** Regional Transportation Statistical Report-annual
- 7.** Transportation Safety at a Glance Brochure-annual
- 8.** Transportation Safety Report- annual
- 9.** Travel Patterns in the New York Metro Area brochure-quarterly
- 10.** Truck Toll Volume Trends at a Glance Brochure-annual
- 11.** Truck Toll Volumes Report-annual

A review of the NYMTC's 11 data products from different aspects was done as shown in Table 3.1 below, in comparison to the 16 best practices collectively, which were reviewed earlier in Chapter 2.1.

Table 3.1. Comparison of NYMTC Data Products to Best Practices

Data Provider	Agency Name	New York Metropolitan Transportation Council (NYMTC)*	Best Practices**
	Website Address	http://www.nymtc.org	Various
Data Description	Data Subjects	Transportation (safety, social/demographic, transportation statistics, traffic volumes, mode choice)	Transportation (freight, passenger travel, aviation, transit, infrastructure, roadway information,safety, traffic counts, straight line diagrams, economic/financial, social/demographic, energy, environment, national security) and various non-transport related data
	Dataset Organization	Alphabetically	Mostly by subject and/or alphabetically, some by mode, function (reports, plans, manuals), collection method
Data Quantity and Its Property	Newest Data	2008	Ranges between 2006-2008, but mostly 2007
	Oldest Data	1995	Ranges between 1953-2005, but mostly 70s and 90s
	Data Continuity	Continuous	Most (75%) provide continuous data, some (25%) do not
	Availability Period	Annual, quarterly, periodic	Mostly annual and/or monthly, some quarterly or periodic
	Smallest Geographic Level	County, Facility, Mode	Mostly County, some National or State
	Time Lag between Data Collection and Publish Date	1 month to more than a year	Ranges between 3 months to more than 1 year, but mostly 6 months to 1year
Data Source Information Availability	Data Collector	Collected by NYMTC from multiple agencies	Themselves or multiple agencies, but mostly themselves
	Collection Methodology	Survey/count	Survey/count
	Purpose of Collection (commercial, non-profit, etc.)	Some available	Some or not mentioned at all, but mostly not mentioned
User Accessibility to Data	Access Restrictions	None	None on transportation-related sites
	User Identification other than Automatic IP Address Recognition	No	No, but 2 non-transportation-related sites ask for employer
	Incentive of Use (require user provide incentive of use or not)	Not required	Mostly not required
	Privilege Discrimination	None	None on transportation-related sites

* Only the 11 data products mentioned in the RFP were considered in preparing this table

**16 agencies reviewed in Chapter 2.1 were considered collectively in preparing this table

Table 3.1. (Cont'd)

Data Provider	Agency Name	New York Metropolitan Transportation Council (NYMTC)*	Best Practices**
Advanced Accessibility	Further Data Availability Online	No	Mostly no, some yes, 2 transportation-related sites provide upon email request
	Further Data Availability by Hardcopy	Yes	Yes
Metadata Availability	Dataset Description (sufficient or not)	No	Most (75%) sufficient, some (25%) not
	Data Gaps / Overlaps	Yes	Mostly no
	Variable Dictionary	Few are available, but hard to find	Mostly available (63% of the sites) though some are not easily accessible, and not available at all for remaining 37%
Data File Format	Available Formats	pdf	Mostly pdf, zip, xls, some txt, html, very few csv, jpeg, png, doc, dbf
	Multiple Formats Available (for the same data)	No	Most (63%) no, some (37%) yes
	Format Choice Consistency (choice is fully/partially available)	-	When choice is available, some fully, some partially available
	Mostly Used Format	.pdf	pdf
Output Generation	Automatic Web Data Generation (in response to user request)	No	Yes for 4 transportation-related sites and 5 non-transportation-related sites, remaining no
	Availability of Online Data Extraction (by select options: time, location,etc.)	No	Most (81%) yes, some (19%) no
	Table/Chart Availability	Yes	Most (81%) yes, some (19%) no
	Map Availability	Yes	Most (81%) yes, some (19%) no
	Table/Chart/Map Customization	No	Some no, some yes
Ease of Output	Data Print/ Download Friendly	Yes	All, except 1, yes
	Table/Chart/Map Saving/Printing	Yes	Most (69%) yes, some (31%) no

* Only the 11 data products mentioned in the RFP were considered in preparing this table

**16 agencies reviewed in Chapter 2.1 were considered collectively in preparing this table

Table 3.1. (Cont'd)

Data Provider	Agency Name	New York Metropolitan Transportation Council (NYMTC)*	Best Practices**
Advanced Services	Both Raw Data and Statistics Available	No	All, except 1, yes
	Statistical Analysis Availability for Chosen Data	No	Most (75%) no, some (25%) yes
	Analysis Tools (if available)	-	For the 25% of the sites that do allow analysis: Count, Sort, Cross tab, Graph generation on map, Scattergram, Histogram, Time-series trend analysis in data, Spatial hotspot analysis on map
User Support and Feedback	FAQ Availability	No	Most (81%) yes, some (19%) no
	Most Requested Data (specified or not)	No	No for some, yes for some (and in those sites "Most Recent", "Popular Keywords", "Forthcoming Data" specified for some of them)
	Related Topics/Data (people who download this data are also interested in)	No	Most (75%) yes, some (25%) no
	Contact Information	Email/phone/person	Mostly Email/phone/person
	Feedback Request on Strategies and Content	No	Most (75%) yes, some (25%) no
	Newsletter/RSS/Podcast/Email Update Availability	NYMTC Newsletter available	3 sites have none, 1 has all, for the rest mostly Newsletter available, for a few RSS, Podcasts available
Final Remarks	Most Notable Strengths of the Service	Good data coverage, Data are presented in chart or table format, Contact information is available for every single report	-
	Most Notable Weaknesses of the Service	Datasets are not well organized, Little to no user support and feedback tools, Data/output customization does not exist, Charts and maps can not be saved conveniently, Long lag between data date and publish date, Limited metadata, Data provided in PDF only	-

* Only the 11 data products mentioned in the RFP were considered in preparing this table

**16 agencies reviewed in Chapter 2.1 were considered collectively in preparing this table

The 11 data products provided on NYMTC's website and listed above are accessible through the 'Data and Model' link on the homepage. However, duplicative links to some of these data products (specifically, County Profiles, Hub Bound Travel Reports, Regional Demographics at a Glance, Regional Transportation Statistical Reports, Transportation Safety Reports, Truck Toll Volumes Reports) are provided in the 'NYMTC Library' section of the website, along with metadata for these duplicative data products. This section also provides various reference materials that can only be viewed at NYMTC's physical library located inside NYMTC. The NYMTC Library section has an advanced search tool, which allows the users to search for data, and a What's New tool, which links to a list of recently posted or updated databases. However, the advanced search tool has limited functionality; the available datasets are divided into five categories and the users are required to search data within a single category at a time and thus may have to perform multiple searches using the same key words. Little information is provided about each category. The NYMTC Library section also includes a page called 'Data Depot'. This page is under construction and currently includes only the County Profiles data (duplicate of the one provided under Data and Model section) which is view and print-friendly only, but not downloadable, a digital version of NYMTC's famous Info Wheel which is view-friendly only, but not printable or downloadable, and links to various data at other agencies' websites.

Now turning to the 'Data and Model' section of the website which is the main section containing the data products that are the focus of this study; using Table 3.1 above, a comparative review of these products can be done here. The smaller sized bullets in the list that follows indicate a weakness in NYMTC's data products compared to the best practices, whereas bigger sized bullets indicate a strength in NYMTC's data products compared to the best practices. Bullets that are in the form of a small dash indicate an equality.

- As far as the Data Subjects are concerned, NYMTC's data coverage is quite good in comparison to the best practices in general. However, as mentioned in the previous paragraphs, NYMTC's data is not well-integrated and scattered through its website, under different links, making it difficult for the user to navigate efficiently and locate the needed data.
- NYMTC's data products are organized alphabetically, but the best practices give the data users more choices in accessing data by subject and/or alphabetically, by mode, function (reports, plans, manuals), as well as collection method.
 - The last release date for all of the annual data products of NYMTC is quite recent, which is also the case for the best practices reviewed.
- Archived data for the case of NYMTC dates back to mid 90's or early 2000s in most cases, but County Profiles are for year 2000 only, as this data is updated every 10 years. Most of the best practices reviewed had archived material dating back more than 50 years.
 - All 11 of NYMTC's data products are published on a continuous basis without any breaks in time, which is also the case for majority (75%) of the best practices reviewed.
 - The frequency of data product release is very similar for both NYMTC and the best practices reviewed, mostly annual or monthly.
- Depending on the type of data, level of detail for the NYMTC data products ranges from county level to facility-based data or data by mode of travel, which can be considered as quite good compared to the best practices which are mostly county based.

- Eight of the 11 data products have a time lag (between the data year and year of publication) longer than 1 year (ranging between 2-4 or even 7 years in some cases). Compared to the best practices reviewed, which have lag times ranging from 6 months to a year mostly, lag time in releasing NYMTC's data products is quite long.
- Regarding data collection methodology, cost and purpose, only some data products, namely, Hub Bound Travel reports, Regional Transportation Statistical Reports, Travel Patterns, and Truck Toll Volumes, provide information to the users, which is also the case for almost all of the best practices reviewed.
- Neither NYMTC nor the non-transportation related sites reviewed impose any restrictions on the users; access is free. But some of the non-transportation-related best practices make their data available to the registered users only.
- Neither NYMTC nor the non-transportation related sites reviewed require the data users to identify themselves before accessing the databases. But for usage monitoring purposes, some kind of user identification would work better to know what the users do with the data, what the most in-demand products are, etc.
- Regarding whether additional datasets can be delivered to the user as hardcopies or electronically if the user requests them, both NYMTC and most of the websites reviewed provide hardcopies and/or CDs in response to user request. Only a few websites, however, can provide electronic databases upon user request, and most (including NYMTC) do not.
- Metadata is not available for all of the data products of NYMTC, and for the ones that metadata is available; it is accessible through a separate link called NYMTC Library on the homepage. Most of the best practices reviewed have sufficient metadata though.
- Variable dictionary is not available for all the data products of NYMTC, and when it is available, it is again accessible through a separate link and thus, hard to locate by the user. For the best practices reviewed, data dictionaries are provided by most (63% of the sites) though some are not easily accessible, and not available at all for the remaining 37%.
- NYMTC's 11 data products are all in PDF only, whereas most of the best practices offer data in multiple formats. However, PDF is also the mostly used format in all the best practices reviewed.
- Data generation and extraction features do not exist for NYMTC's 11 data products studied here, whereas, they do exist for the majority of the best practices reviewed. For the case of NYMTC, the users are generally presented with a prepared set of data (in PDF), and not able to do any customized data generation/extraction. Data generation/extraction feature does exist however for NYMTC's GIS data files, provided on the homepage under a separate link called 'GIS'. Users can select or extract data by different data categories from multiple databases including vehicle counts, O-D survey, household survey, socioeconomic and land use data, and travel time and speed data. Users can view their selections in either a map format or a table style. This function, however, is not available for household survey database. The GIS section also contains some detailed metadata information. Variable dictionaries are only available for some databases (for example, the vehicle counts, O-D survey, and socioeconomic and land use data) and these dictionaries tend to be confusing and not very easy to understand, as they include a lot of extra information used for internal filing purposes only.
- Both NYMTC's and best practices' data products include tables, maps, charts.

- NYMTC’s data products, being in PDF, do not allow the user to generate customized tables/maps/charts, whereas more than 50% of the best practices reviewed do allow for such customization.
- Both NYMTC’s and best practices’ data products are print and save-friendly.
- All, except one, of the best practices present the user with both raw data and readily available statistics. However, NYMTC’s data products do not allow the user to work with raw data.
- Neither NYMTC nor most (75%) of the best practices allow the user to do statistical analysis with the chosen data. For the 25% of the sites that do allow analysis; Count, Sort, Cross tab, Graph generation on map, Scattergram, Histogram, Time-series trend analysis in data, and Spatial hotspot analysis on map are allowed.
- The User Support and Feedback aspect of the NYMTC website is quite limited. Popular features that exist on the websites of the best practices reviewed such as Frequently Asked Questions (FAQ), Help, Advanced Search, links to the Most Requested items, Newly Released data or updates to existing databases, online Feedback Forms, are not available. NYMTC publishes a biweekly electronic newsletter called NYMTC-Notes, to which users can subscribe. However other popular services offered by some of the websites of the best practices reviewed such as RSS feeds, Podcasts, E-mail updates, Online Chat are not available.

3.2. Evaluation of NYMTC Data Products Using Stakeholder Survey

Using the results of the NYMTC stakeholder survey, some simple but informative data evaluations were made which involved the calculation of the following:

- Combined Customer Satisfaction Index (CCSI)
- Importance Index (ImpI)

3.2.1. Customer Satisfaction Index

Customer satisfaction indices are widely used in measuring customer satisfaction in a variety of fields both from public and private sector. For example, the American Customer Satisfaction Index (ACSI) reports scores on a 0-100 scale at the national level and produces indexes for numerous economic sectors, industries, companies, and federal or local government agencies using a set of causal equations that link customer expectations, perceived quality, and perceived value to customer satisfaction (Source: ACSI Website).

The Minnesota Department of Employment and Economic Development (DEED) primarily uses two indices to express customer satisfaction with its services; the Minnesota Customer Satisfaction Index (MnCSI) and ACSI. The MnCSI uses responses to three questions about satisfaction, answered on a scale of 1 to 10, where “1” is the lowest (least satisfied) and “10” is the highest (most satisfied). Then, a single number, which varies from 0 to 100 is generated using Equation 1 below (Source: DEED Website).

$$MnCSI = \left(\frac{(Question1-1)}{9} * 33.3 \right) + \left(\frac{(Question2-1)}{9} * 33.3 \right) + \left(\frac{(Question3-1)}{9} * 33.3 \right) \quad (1)$$

where Question1 is the user’s rating on a scale of 1 to 10 to Question 1
 Question2 is the user’s rating on a scale of 1 to 10 to Question 2
 Question3 is the user’s rating on a scale of 1 to 10 to Question 3

A simple generalized formula shown in Equation 2 can be written based on the idea behind Equation 1 in order to obtain a formula that can be used for calculating the Combined CSI (CCSI) that varies between 0 and 100 (0 representing lowest quality, 100 representing highest quality) for each of the NYMTC’s data products using the responses to Question 4 in the stakeholder survey, which asked:

“Do you have any suggestions to improve these products?”

After reviewing the survey responses to this open-ended question, the research team calculated percentages of stakeholders who had concerns regarding 6 criteria (specifically, Accuracy, Consistency, Amount of Data, Timeliness, Appropriateness of Format and Ease of Understanding, and User Feedback) and then scaled those percentages on a 1-10 scale. After that, CCSI was computed simply using Equation 2.

$$CCSI = \sum_{i=1}^n \left(\frac{(C_i - 1) * 100}{9 * n} \right) \tag{2}$$

where C_i: User’s rating to Criteria ‘i’ on a scale of 1 to 10
 n: Total number of criteria (6 in this case)

Table 3.2 below shows the results from the CCSI calculations. As can be seen, the customer satisfaction index for NYMTC’s data products was computed as about 80%, which indicates a relatively high level of overall quality. However, it should be noted that, considering that the survey had only 32 responses in total, the accuracy of this result should remain rather questionable. A better and far more accurate evaluation would be possible with a larger sample size.

Table 3.2. Combined Customer Satisfaction Index (CCSI)

No	Criteria	User’s Rating*	Combined Customer Satisfaction Index (CCSI)**
1	Accuracy	9.5	15.741
2	Consistency	8.9	14.630
3	Amount of Data	5.8	8.889
4	Timeliness	6.8	10.741
5	Appropriateness of Format and Ease of Understanding	8.9	14.630
6	User Feedback	9.5	15.741

CCSI= 80.370

*On a scale of 1 to 10 where 1 represents least satisfied and 10 represents most satisfied

**On a scale of 0 to 100 where 0 represents least satisfied and 100 represents most satisfied

3.2.2.Importance Index

Next, using the responses to Question 3 in the stakeholder survey, a simple Importance Index (ImpI) that varies between 0 and 1 (0 representing low importance, 1 representing high importance) for each of the NYMTC’s data products was calculated by weighing the responses as shown in Equation 3. Question 3 asked:

“Please check one of the three options for each product whether you have used the product or not”

- Important
- Not important
- Not familiar with

$$\text{ImpI} = [0 * (\% \text{ responding "not familiar with"}) + 1 * (\% \text{ responding "not important"}) + 2 * (\% \text{ responding "important"})] / 200 \quad (3)$$

where ‘0’, ‘1’, and ‘2’ are the numerical values assigned to the three degrees of importance and ‘200’ is the value that normalizes the ImpI value to range between 0-1.

The results for ImpI calculations for each of the 11 data products are presented in Table 3.3 and Figure 3.1. As can be seen, stakeholders rated the Regional Transportation at a Glance brochure, Travel Patterns report, and the Regional Transportation Statistical report as the three most important data products. Average ImpI for all 11 of the data products is 0.685, indicating that the stakeholders value the data products with a relatively high level of importance overall.

Again, it should be noted that, considering that the survey had only 32 responses in total, the accuracy of this result should remain rather questionable. A better and far more accurate evaluation would be possible with a larger sample size.

Table 3.3. Importance Index (ImpI)

Data Product	# Respondents			% Respondents			Importance Index
	Saying Important	Saying Not Important	Who are Unfamiliar	Saying Important	Saying Not Important	Who are Unfamiliar	
Reg. Transp. at a Glance	22	2	5	75.9	6.9	17.2	0.793
Travel Patterns	22	0	6	78.6	0.0	21.4	0.786
Reg. Transp. Stat.	21	0	7	75.0	0.0	25.0	0.750
Truck Toll Volumes	18	2	7	66.7	7.4	25.9	0.704
Hub-Bound Travel	18	3	7	64.3	10.7	25.0	0.696
Hub-Bound Travel at a Glance	17	4	7	60.7	14.3	25.0	0.679
Truck Toll Volume Trends at a Glance	16	3	7	61.5	11.5	26.9	0.673
County Profiles	16	3	7	61.5	11.5	26.9	0.673
Reg. Demographics	17	2	8	63.0	7.4	29.6	0.667
Transp. Safety Stat.	16	1	9	61.5	3.8	34.6	0.635
Transp. Safety at a Glance	11	2	12	44.0	8.0	48.0	0.480
						Avg:	0.685

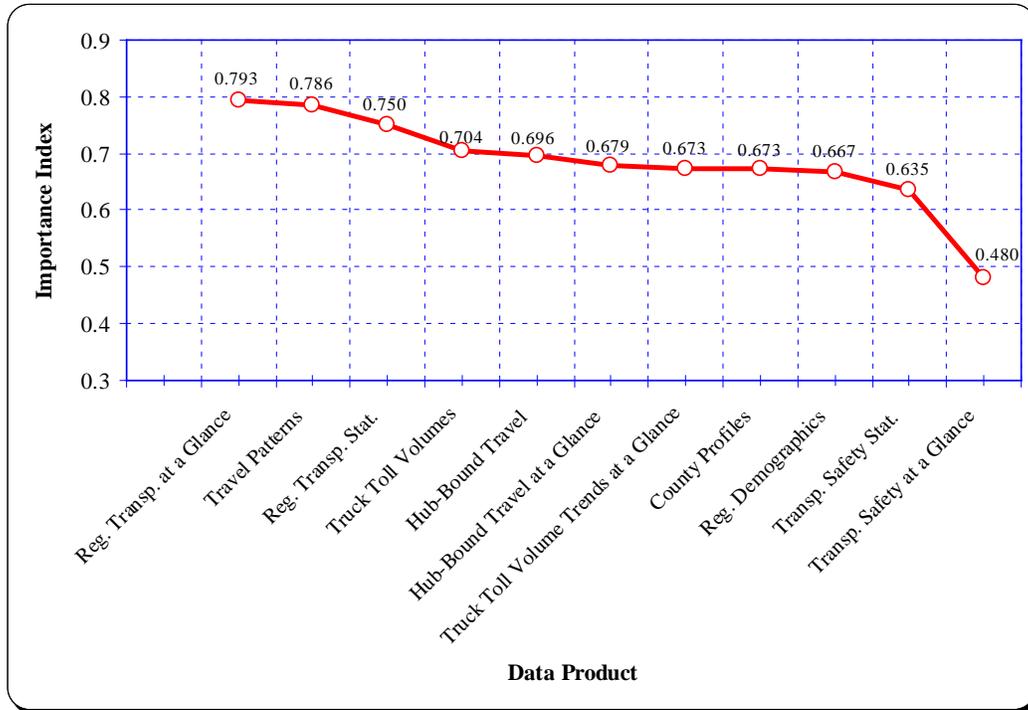


Figure 3.1. Importance Index (ImpI)

3.3. NYMTC’s Current Data Maintenance Process

The current practice of data maintenance and quality control within NYMTC is summarized in this section based on the *Data Group Cookbook for Reports and Brochures* (NYMTC) received as a PDF file from NYMTC as an input to this project, as well as the information gathered from the NYMTC technical staff during various project progress meetings.

Accordingly, NYMTC’s current data maintenance and quality control process starts with ‘*Creating the Data Product*’. For that purpose, emails are sent to all parties that are required to provide NYMTC with variety of data. Contact information and website addresses for data provider agencies for different data subjects are specified in detail in the “Data Group Cookbook for Reports and Brochures” (Source: NYMTC).

The next step is ‘*Conducting the Reviews*’. All the reports and brochures that are generated by NYMTC using the raw data provided by the source agencies go through an across-the-board review process. First the NYMTC data group manager performs the initial internal review and his/her comments are applied. Then, a person from the Public Relations Department at NYMTC checks the data product for quality and political sensitivity. Next, internal review is carried out by the NYMTC staff and any comments that may arise are then applied appropriately. The next step is the external review of the data product by the corresponding external data source and their comments are also applied. Finally, the data product is reviewed one more time by the data group manager, as well as the PR person. Each review takes about 2 weeks, and the total review process takes about 2 months

The next step is *‘Publishing the Data Product’*. The corrected and completed data product is published on the NYMTC website and an article is written for the NYMTC-Notes to announce the availability of the new data product highlighting any interesting findings or changes.

3.4. Final List of Issues

In identifying the issues with NYMTC’s data products listed in the previous section, the research team took into consideration four factors.

First, the results from the ‘comparison to best practices’ (Chapter 3.1) were used to shed light on the strengths and weaknesses of the NYMTC’s data products and help the research team identify problem areas.

Second, information on the ‘expectations of the NYMTC stakeholders’ from the stakeholder survey reviewed previously in Chapter 2.3, allowed the research team to obtain the desired end-user input on the subjective quality criteria. Compliance/noncompliance to some of the quality criteria (e.g. accuracy of information, whether or not data is out-dated) can be evaluated objectively using simple techniques such as visual inspection or side-by-side comparison with the original data source, to name a few. But evaluating the data on subjective criteria (e.g. the format in which the data should be presented, the way it should be represented) and concluding whether or not data is of certain quality is no simple task unless extensive end-user input is available. Therefore, results of the NYMTC stakeholder survey allowed the research team to pinpoint the most important problem areas by looking at the responses to the open-ended questions.

Third, understanding the ‘NYMTC’s point of view as the Metropolitan Planning Organization (MPO)’ serving the New York City, Long Island and the lower Hudson Valley region gave the research team a clear direction to go to in identifying the problem areas that require further improvements. *As the largest MPO in the nation, NYMTC provides a collaborative planning forum to address transportation-related issues from a regional perspective and plans and makes decisions on the use of federal transportation funds* (Source: NYMTC Website). Thus, the availability, utility, and quality of data is very important for the MPO to properly address a wide range of issues, whether through policy planning, performance evaluation, model development or impact studies, to generate ideas and clear-cut solutions for informed decision making. In fact, NYMTC’s new endeavor called *iNYMTC*, taken as part of the Strategic Data Coordination in the NYMTC Region, aims to offer integrated, interactive, easily accessible data to its wide range of users that comes with sufficient metadata, is GIS-based and user friendly.

Finally, the current data maintenance process at NYMTC was taken into consideration to identify the problem areas.

With the abovementioned four factors in mind, two main types of issues were identified: First, the issues related to the 11 data products data and the NYMTC website through which these data products are provided to the users online were prepared, which are presented in the next section. As can be seen in the next section, these issues were further categorized with respect to several dimensions of data quality. Unless stated otherwise, the issues presented apply to all 11 of the data products. Second, issues related to the data management process were identified.

Please note that specific solutions to these issues will be described and explained in great detail later in Chapter 4.

3.4.1. Data and Website Oriented Issues

Issues Related to Amount of Data

- As evident by the NYMTC stakeholder survey, users would like an expansion of the existing data (e.g. with more details on time/day of travel, vehicle class, County Profiles for the entire region). *County profiles* for the entire region does exist but under a separate link called Data Depot which can be accessed through NYMTC Library link. The county profiles provided in Data Depot section are poorly presented, and are not downloadable
- Data not integrated (Duplicative links to *County Profiles, Hub Bound Travel Reports, Regional Demographics at a Glance, Regional Transportation Statistical Reports, Transportation Safety Reports, Truck Toll Volumes Reports* given in ‘NYMTC Library’ section of the website creating confusion for the user)
- According to the NYMTC stakeholder survey, users would like to see new data (e.g. longitude/latitude to facilitate map visualization by user, data at more disaggregate geographic levels such as towns, multiple zip codes, and airline and Amtrak boardings)
- Insufficient Archives (Not dating back far enough)
- Raw data not available

Issues Related to Appropriateness of Format and Ease of Understanding

- Single file format, PDF available, thus not possible to customize/do analysis with. According to NYMTC stakeholder survey, users prefer MS Excel and TEXT formats in addition to PDF
- Metadata not available for *Travel Patterns in the New York Metro Area brochure*, and for other data products, it is available under a separate link called NYMTC Library on homepage
- Variable dictionary not available for all the data products

Issues Related to Timeliness, Accuracy and Consistency

- Long time lag between the date when the data is created and the date when the data is published
- According to NYMTC stakeholder survey, some users have concerns about the accuracy of data since it comes from multiple agencies, and users suggest periodic cross-checks with other agencies
- According to NYMTC stakeholder survey, some users have concerns about the consistency of data with other agencies and suggest periodic consistency checks with other agencies

Issues Related to Accessibility

- Data products are organized alphabetically on the website, and users are not given different choices in accessing data by subject, mode, function (reports, plans, manuals), as well as collection method, for example
- Search for specific charts or maps in multiple documents not possible

- Search tool that allows users to search through multiple databases using the same keywords once not available

Issues Related to Advanced Services and User-Friendliness

- Neither the user nor the NYMTC staff are able to generate reports automatically using different data sets. According to NYMTC stakeholder survey, users prefer web-based queries to be able to generate interactive tables, charts, maps
- Analysis such as descriptive statistics, sort, graph not available
- Data extraction from multiple databases using a set of criteria variables not possible
- Data not interactive or GIS-based

Issues Related to User Support and Feedback

- No monitoring mechanism to identify the users, what the users do with the data, what the most in-demand products are
- No RSS, Podcast, E-mail updates available, which would increase user awareness about data products and retain users. According to NYMTC stakeholder survey, users are least familiar with the *Transportation Safety reports*
- No help, advanced search features, FAQ, Online Chat, Most Requested, Recently Updated/Released Data, feedback forms available

3.4.2. Process Oriented Issues

The logical next step in the data maintenance process, based on the data life cycle concept suggested by Loshin (2007), would be to monitor the effectiveness and popularity of the data product from the user's point of view, and then to improve the overall process by taking into consideration user feedback. Even though these '*Monitoring Effectiveness*' and '*Improving the Process*' components are not mentioned in NYMTC's Data Group Cookbook for Reports and Brochures, the fact that NYMTC conducted a website hosting evaluation in 2007, conducted the NYMTC Stakeholder Survey in 2007, and initiated this "Improvements on NYMTC Data Products" project in 2008 are clear indications of NYMTC's efforts to monitor, modify and further improve the existing data products, thereby closing the loop in the data maintenance process in a way.

In Figure 3.2, NYMTC's current data maintenance process is shown in the form of a flowchart developed by the research team for ease of understanding and also to facilitate comparison with the dynamic process that will be proposed later in Chapter 4. In Figure 3.2, the components that are not formally mentioned in the Data Group Cookbook for Reports and Brochures, and thus are not part of the officially acknowledged current practice are shown with dashed lines and boxes.

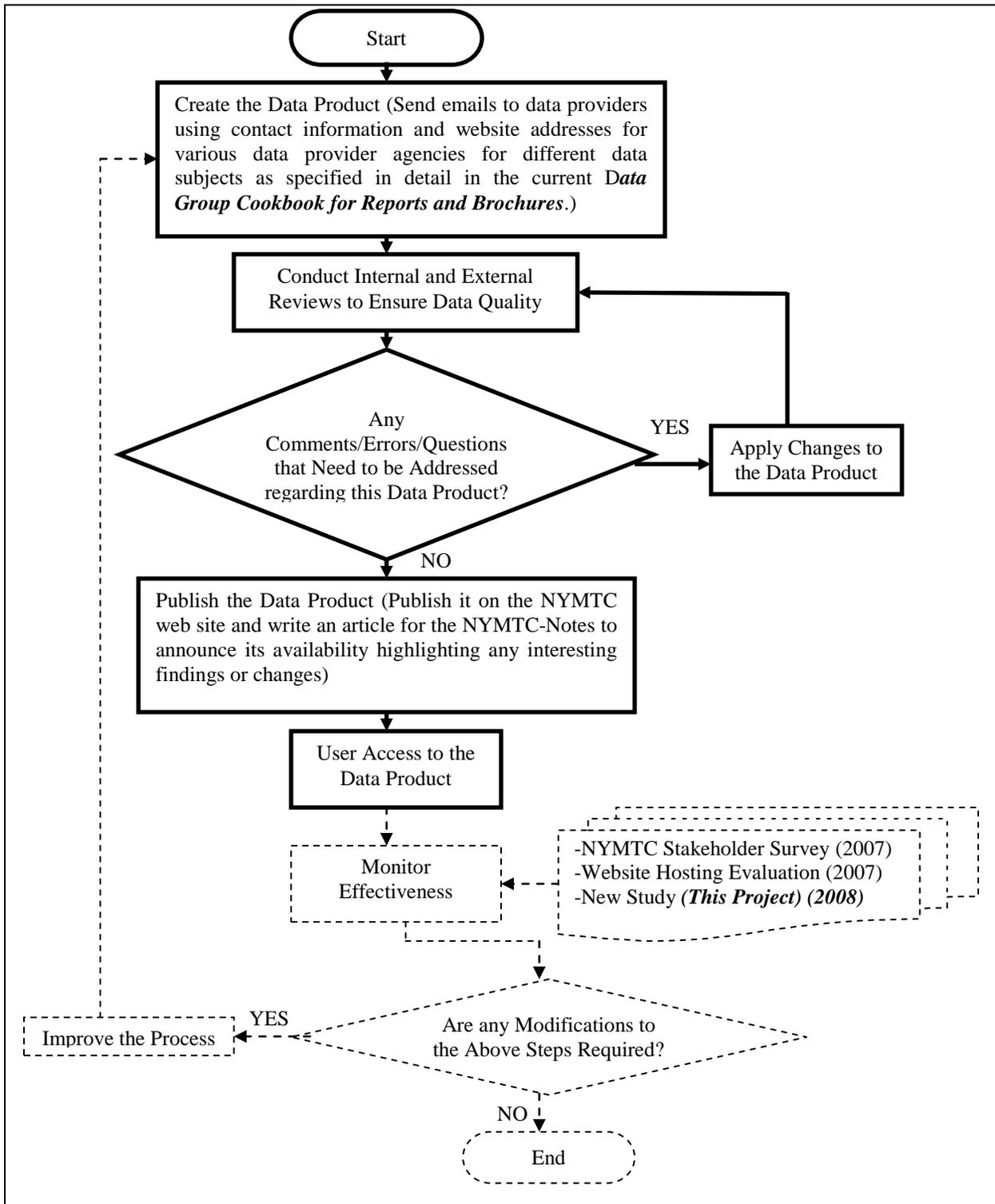


Figure 3.2. NYMTC's Current Data Maintenance Procedure (Put in Flowchart Form to Facilitate Comparison with Figure 2. Dashed lines and boxes are not official parts of the current process)

CHAPTER 4

SUGGESTION OF SOLUTIONS

In this chapter, specific recommendations are made for the issues identified in Chapter 3 related to the 11 NYMTC data products, the NYMTC website with special emphasis on the recommendations for the 'Data and Model' page, where the 11 data products are located (recommended to be renamed as 'Data & Model' as explained later in this chapter), and related to the NYMTC's current data maintenance process.

All of the recommendations made here aim to offer to the wide range of NYMTC website visitors and users better integrated, interactive, and easily accessible data that comes with sufficient metadata, is GIS-based and user friendly.

As can be seen in the following sections, it is indicated next to the heading of each recommendation whether it is considered to be a low, medium or high priority recommendation in terms of implementation.

4.1. Data and Website Oriented Solutions

In order to give a better idea about the big picture regarding the overall recommendations, a flowchart shown in Figure 4.1 is included. It summarizes the suggested navigation path for data access and also shows how the 'NYMTC Home Page', pages for the individual data products, and 'Data & Model' page should be linked to each other. Details on how each of these three elements (shown in bold letters in Figure 4.1) are recommended to be laid out and function as individual web pages are given in the following sections of this chapter.

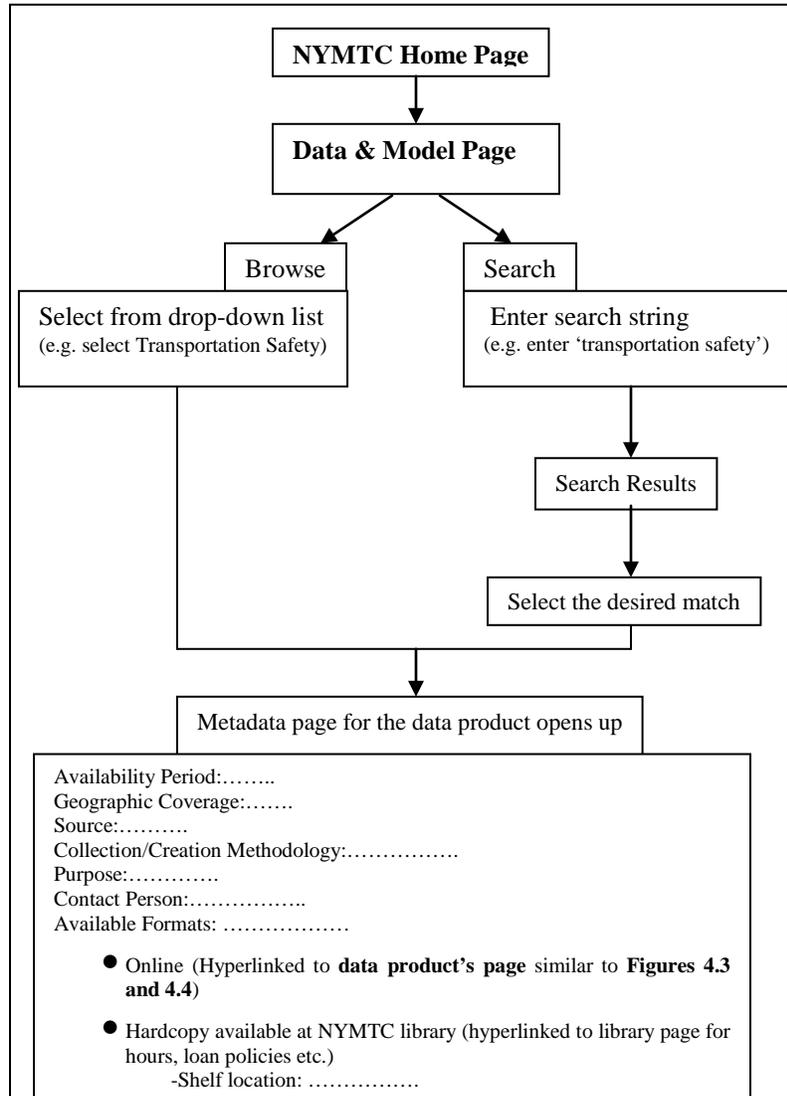


Figure 4.1. Suggested Navigation Path for Data Access Using the Improved NYMTC Website

4.1.1. Recommendations for the Eleven NYMTC Data Products

Recommendations for the individual NYMTC data products are summarized in Table A2 in the Appendix.

Below, each recommendation for the 11 data products is explained in more detail, along with some information on how each recommendation can be implemented and which ones of the 11 data products are in need of that particular solution. In addition, examples of best practices (with screen shots and URLs) that have already implemented the recommended solution(s) are given whenever possible.

Expand Data (High Priority)

As mentioned in the previous chapters of this project report, NYMTC stakeholders made several suggestions through the Stakeholder survey regarding amount of data presented at NYMTC website. Thus, the recommendations about “expanding data” are made in the light of stakeholders’ suggestions.

More details on time/day of travel are already available in the “Vehicle Classification and Occupancy” reports (not in the scope of this project). The availability of this detailed data should be clearly indicated in the metadata files for the **Hub Bound Travel reports and brochures**, **Regional Transportation Statistics reports and brochures**, **Truck Toll Volumes reports and brochures**, and **Travel Patterns in the New York Metro Area brochures**.

Similarly, detailed vehicle classification data (13 classes) is currently available in the “Vehicle Classification and Occupancy” reports, and this fact should be clearly indicated in the metadata files for the **Hub Bound Travel reports and brochures**, **Regional Transportation Statistics reports and brochures**, and **Travel Patterns in the New York Metro Area brochures** for the users who would prefer the travel data by detailed vehicle classes, rather than collective counts for autos, taxis, vans, and trucks.

County Profiles for the entire region can also be provided. It is already available in the ‘Data Depot’ section which can currently be accessed through ‘NYMTC Library’ link on the homepage, so it should simply be moved from there to the ‘County Profiles’ page.

Longitude/latitude data for each product should be provided as much as possible to facilitate map visualization by the user (recommended for the **data products where this suggestion is applicable**). That is, not only the databases, but also the associated maps (i.e., geocoded data) should be made available to the users.

Data related to airline and Amtrak boardings can also be provided.

Provide Latest Data (High Priority)

Most recent data, from the previous year or earlier should be made available to the public if possible. This will undoubtedly improve the usefulness and the value of the information provided. Currently, for the **Hub Bound Travel reports and brochures**, **Regional Demographics brochures**, **Regional Transportation Statistics reports and brochures**, **Transportation Safety reports and brochures**, and **Truck Toll Volumes reports and brochures**, the most recent available data is more than a year-old.

Provide Larger Data Archives (Low Priority)

In most of the cases, archived data do not cover long enough time periods, such as a 20-year old archive, meaning data from mid 1980s. Archives should be expanded as much as possible to include data starting from mid 1980s to provide a 20-year old data archive to the user. This recommendation is in line with the best practices which in general provide about 20-year old archives on the average as explained in detail in Table 3.1.

For the **Regional Transportation Statistical Report**, the **Hub Bound Travel Report**, and the **Truck Toll Volume Report**, hardcopies available at NYMTC Library already cover a sufficiently wide time span (data from 1970s, 1924, and 1979 respectively), so those materials can be digitized to provide the users an appropriate online data archive for these two products.

For the **remaining 8 data products**, however, older data is not provided neither digitally nor in hard copy form.

Provide Data Continuity (Medium Priority)

Data for missing years/quarters/decades etc., whichever time period applies to a particular data product, should be made available to provide data continuity. Specifically, the **Transportation Safety reports**, **Regional Demographics brochures**, **Regional Transportation brochures**, and **Hub Bound Travel brochures** have continuity issues.

Decrease Time Lags (High Priority)

Currently, except for the “Truck Toll Volumes at a Glance” brochure, all of the data products have lag times longer than one year and this is mostly due to the tragic events of September 11, 2001. For future data updates and/or uploads, it would be best to keep the time lag between data collection and publish date at 1 year or less as much as possible, for **all of the data products**.

Ensure Accuracy (High Priority)

It should be kept in mind that the data provided to the users should not only be accurate, but also that there are no typos or outdated information. For data accuracy and consistency, periodic cross-checks with other agencies should be performed. In addition, the products should be reviewed closely to remove any typos or to find out if any information needs updating.

Even though this recommendation is made for **all 11 of the data products**, specifically two errors were found in the **County Profiles**. First, the address that appears at the bottom of the “County Profiles” brochures needs to be updated, as it currently shows NYMTC to be located in Long Island. This is due to the fact that the address at the time of data product creation is shown, but then this fact should be pointed out to the data users. Second, on the cover it is indicated that the brochure is published on an annual basis, but in fact it is published every 10 years, so this should be corrected.

Integrate Data and Links to Data (High Priority)

Duplicative links for the same data product at several different pages throughout the website should be integrated under a common link.

Specifically, the **County Profiles** data appears at three different locations at NYMTC website; first at the ‘Data Depot’ section, second under ‘Census’ data on the current “Data and Model” page, and third on the current “Data and Model” page. The first two should be moved and combined with the third to provide data integrity and decrease confusion while navigating the site.

As for the **Transportation Safety reports**, **Regional Transportation Statistical reports**, **Hub Bound Travel reports**, **Truck Toll Volume reports**, and the **Travel Patterns in the NY Metro**

Area brochures, the duplicative link issue does not exist, however hardcopies of these data products (for some years) are available at NYMTC library. This should be indicated clearly in the metadata files for these data products. For example, “Transportation Safety” reports are available online for years 1997-2005, but hardcopies for years 1997-2004 can also be obtained from the NYMTC library.

Provide Better Data Access and Retrieval Options (High Priority)

The user should be able to access the data in an interactive manner by means of various filtering options at a desired level of detail and generate customized data in various output formats. Most of the users only need some specific data and they need to have an opportunity for processing the raw data to acquire only what they need.

The web pages for all 11 of the data products are currently in a simple static form and not interactive as shown in Figure 4.2 below. Figure 4.3 shows a suggested look for the ‘County Profiles’ data page, but **all 11 of the data product pages** require similar improvement. Please note that, as indicated in Figure 4.1 earlier, it is recommended here that the user gets directed to the webpage for a data product similar to the one shown in Figure 4.3, after selecting the ‘online’ option from the available formats in the Metadata.

As can be seen in Figure 4.3, with the suggested new format for individual data products, before proceeding with viewing/customizing/downloading the data itself, the user can view the Glossary of terms and acronyms, and Metadata. Metadata, as shown in the lower part of Figure 4.1 earlier, should provide information related to the availability period, geographic coverage, data source along with a link to the website of original owner/collector of this data if NYMTC is not the collector/creator of this data, data collection/creation methodology, clear statement of the purpose for collecting this data, contact person’s email address, phone and fax numbers, and available formats (electronic formats as well as hardcopy. For a product that is available as a hardcopy, shelf location at NYMTC library along with a link to the library webpage for hours etc. should be given. For an electronic data product, a link to the data page should be given).

As shown in Figure 4.3, a link to “User Support” tools can also be provided on each data product’s page to ensure that the user can:

- request a hardcopy,
- get more data through FTP,
- join a user forum (online multi-user collaboration),
- view the most frequently asked questions (FAQ) about a particular data product.

Users should be given access to a web form to submit requests for hardcopies.

Free FTP service should be provided to facilitate user access to further data that is not readily available on the website. It simply requires the NYMTC staff to set up and maintain a secure FTP server, and the user to follow the FTP link to connect to this server and download the files they need.

Online multi-user collaboration tools such as wiki (a World Wide Web application designed to allow multiple users to add, remove, and edit content), to create a common community environment for data providers as well as users should also be considered as important ways to enhance access to and use of NYMTC data. A very good example of online multi-user collaboration effort is the Federal Highway Administration's Next Generation Simulation (NGSIM) program which provides information sharing and collaboration on microscopic traffic simulation modeling, among the stakeholders of the broad transportation community (<http://ngsim.fhwa.dot.gov/>). Another good example is FHWA's National Household Travel Survey "Community of Practice" (CoP). The purpose of CoP is to provide a forum where an open exchange of information and knowledge of transportation issues can take place. Users are encouraged to share their "best practices", information about transportation trends, or transportation issues that face their communities on a national, state or local level. Clicking on the 'Add Discussion Item' link takes the user to a web form (Figure A1) where s/he can freely participate in the forum and send in a discussion topic. Tools that improve collaborations are found to be extremely useful by software users and developers of some of the popular products such as Arc-Info (ESRI) and Paramics (Quadstone) by directly allowing the members of the community to share their practical experience with others.

A Frequently Asked Questions (FAQ) link would complement the information provided in the Metadata of the data product. The following is the suggested list of FAQs. It can be made a living source of information by continuously adding new questions raised by the users.

- What information does this data product provide/ What does this data product tell us?
- How can this data product be used?
- How do I read or interpret a data entry in this data product?
- How is the data collected and reviewed?
- How is (*a certain value/variable*) calculated?
- What are the data sources that NYMTC utilizes?
- How often is the data published?
- How far back in time do your records cover?
- How can I view the data?
- How can I obtain further data?
- Who can I contact for questions/comments about this data product?

As it is shown in Figure 4.3, the user can either view the standard PDF version of the data product, or can generate customized data tables with chart/graph options with a simple query utilizing drop-down menus as suggested in Figure 4.3 or with radio buttons for selecting different data options. Figures A2 through A4 in the Appendix show table customization pages from three best practices. It is also a good practice to denote commonly selected fields in the query as done on the FARS website as shown in Figure A5.

If the user wishes to add graphs/charts options to his/her data customization process by checking the box that says "check here to create associated graphs/charts", then s/he is provided with a separate interactive window for selecting chart type, data series etc.

For the selected data set, users should be allowed to generate basic statistics like count, sort or cross-tab to name a few. A good example is FHWA's NHTS website which allows the users to design custom tables and also generate basic descriptive statistics as shown in Figure A6. Another good example is shown in Figure A7.

Data can be output in various formats such as HTML display with Print option, or XLS, TXT, CSV, PDF etc. exports, as well as DBF export whenever possible (it allows data mining and is preferable by most researchers). A good example is the Table Designer feature on the FHWA's NHTS website, which, as shown in Figure A6, allows the users to build customized data tabulations quickly and easily and get the output in both HTML and XLS formats. The users need to provide an email address to be able to use this feature. Another good example is given in Figure A8.

Implementing the idea of interactive data pages as explained in a simple way above and allowing generation of HTML/ XML/other types of documents dynamically in response to a web client's request, of course requires some expertise on the part of the NYMTC staff or consultants. For example, familiarity with the Java technology would make it possible to embed a Java code and certain pre-defined actions into the static web content. On the part of the users, however, some familiarity with data analysis and statistical concepts and running simple queries would be necessary. Still, for novice users there should be clear guidelines documenting the overall process. Figure A9 is an example.

Furthermore, achieving the interactivity described here for the pages of other data products, not just the 'County Profiles', is possible, with some simple modifications on the page illustrated in Figure 4.3. However, the level of time and effort required to create interactive content would depend on the data product and how comprehensive it is. So, to provide another example, an interactive look is also suggested for 'Regional Transportation Statistics Report' page as shown in Figure 4.4. As can be seen, if the users choose to customize data, they will be asked to:

- Select either the entire tri-state area or the individual States of NY, NJ, or CT
- Select County for the previously selected State
- Select a transportation mode of either Public, Highway or Air transportation
- Select from a wide range of variables including revenues, ridership (if Public transportation is selected previously), river crossings, crashes, etc.
- Select year(s).

Additionally, website maintenance issues need to be taken into consideration when designing the new website.

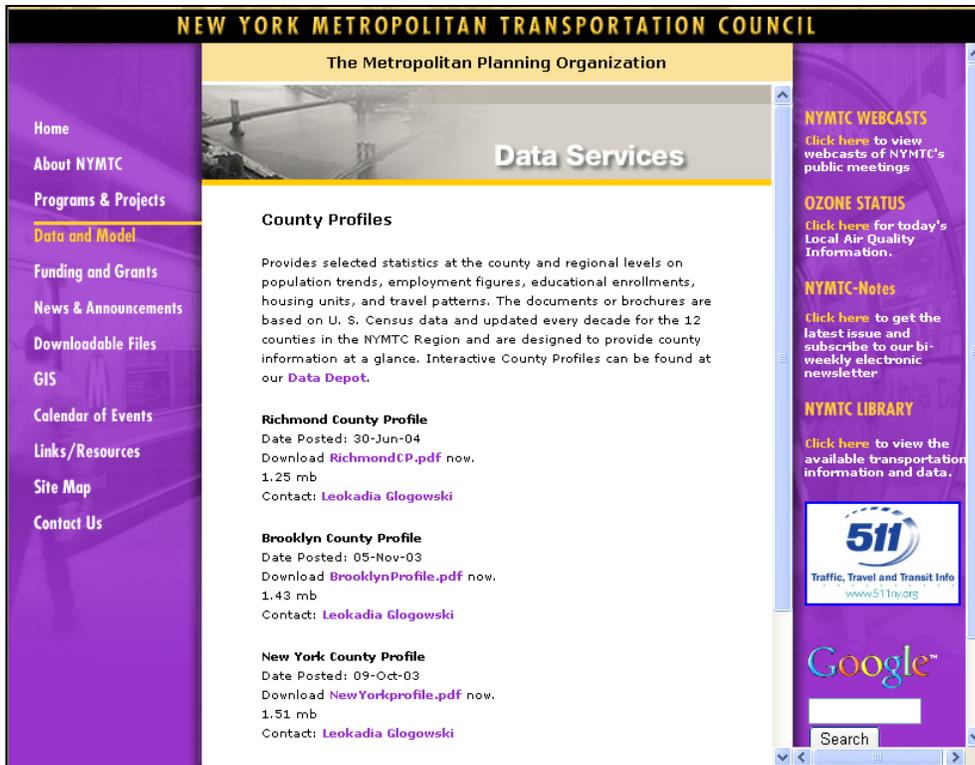


Figure 4.2. Current Look of a Data Product Page (County Profiles in this case) at NYMTC Website

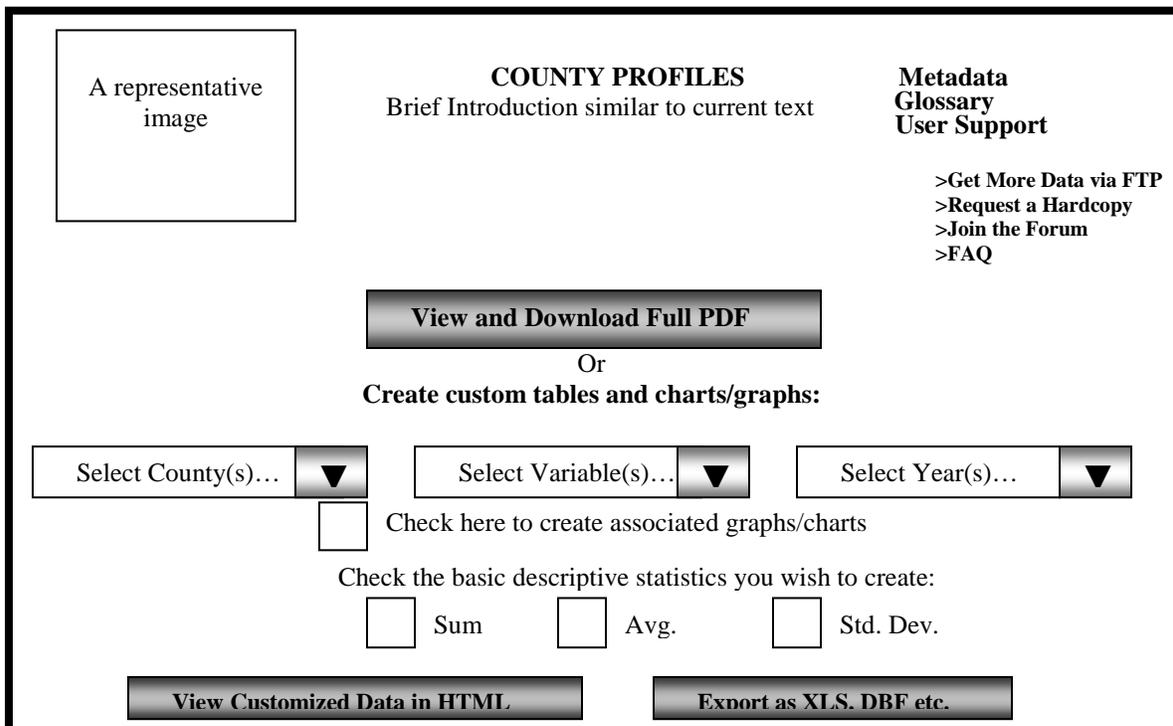


Figure 4.3. Suggested Look for a Data Product Page (County Profiles in this case) at NYMTC Website (Only Center Panel is Shown)

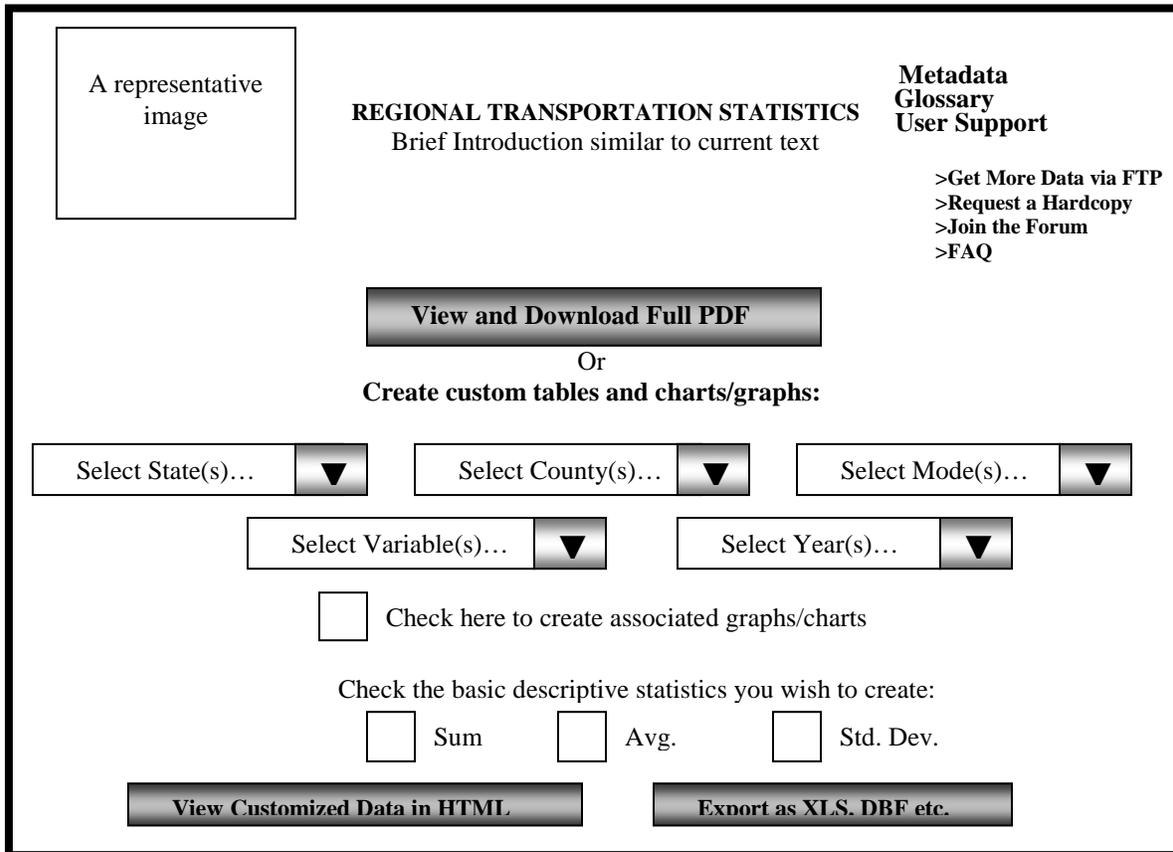


Figure 4.4. Suggested Look for Regional Transportation Statistics Report Page at NYMTC Website (Only Center Panel is Shown)

4.1.2. Recommendations for the NYMTC Website

This section presents some of the major recommendations for the NYMTC website in general in terms of site layout, appearance, organization, and content. Examples of best practices (with screenshots and their URLs) are given whenever possible.

General Recommendations (High Priority)

NYMTC's current site resolution of 800x600 causes only two thirds of the whole page to be utilized and, as a result, the text looks compressed and tiny. A scroll bar is located in the center panel (frame) that is not preferred for modern website designs. Most of the websites today use 1024x768 or higher resolutions, which facilitate utilization of the space more effectively and also, eliminate the need for a scroll bar in most cases.

The content on the site is not updated regularly and frequently. New reports, projects, data releases, etc. should be made available online as soon as possible.

A website monitoring mechanism such as a hit counter would make it possible to identify the users and the most in-demand products, allow NYMTC to keep track of successful data retrievals and returning users, and the popularity of pages and data products.

Recommendations for the NYMTC Home Page (High Priority)

The current NYMTC Home Page is shown in Figure 4.5, and suggested improvements are presented in Figure 4.6 below. Please note in Figure 4.6 that, for some of the categories, sub items are also shown. However, from an implementation perspective, those secondary and tertiary sub items should appear only when the website visitor clicks on the main category item. Figures A10 through A13 show several good examples from best practices regarding home page design and organization. As depicted in Figure 6, the major recommendations for the NYMTC Home Page are the following:

- NYMTC logo should be provided right at the top
- Currently a series of news and data updates are listed in the center panel. Enhancements can be made by organizing them under a proper heading such as 'Latest News' or 'What's New at NYMTC' and categorizing them by subject.
- The links on the left panel that have similar contents should be merged into a common link. Specifically, it is recommended that the 'News & Announcements' and 'Calendar of Events' links be combined to create a new link named 'News & Events' and a sub level link to NYMTC Webcasts be included there as well. Additionally, it is recommended that the data files, which are currently accessible through different links, namely 'Data and Model' and 'GIS', be centrally located under a new link named 'Data & Model'.
- 'Downloadable Files' should be renamed as 'Reports & Publications' as this page contains various research reports and agency publications/brochures.
- 'Links/Resources' can be renamed as 'External Links/Resources'.
- 'Site Map' and 'Contact Us' links can be moved to the right panel, and 'NYMTC Library' link to the left.
- The links 'NYMTC Webcasts', 'Ozone Status', 'NYMTC-Notes', and '511' should be combined together under the common heading 'Highlights...' on the right panel.
- Several user support links such as 'Help', 'RSS Feeds', and 'Email Sign Up' can be added to the right panel to provide the users help on how to best navigate the NYMTC website and access information, allow them to keep up with the website in an automated manner through RSS subscription, and get email notifications about latest programs, projects, data releases and everything in between that is happening at NYMTC by simply signing up to the electronic mailing list. Familiarity with XML or if not, utilization of a freeware/software would be required on the part of the NYMTC staff or the consultants to build, edit, and publish feeds, and the users would need to simply copy-paste feed URL to an RSS feeder (such as My Yahoo!). Figures A14 through A17 show several good examples from best practices related to Help, Email Lists, and RSS Feeds.
- The current 'Contact Us' page should be improved to include a web form for user feedback; comments and questions. Figures A18 through A22 show good examples from best practices.
- The unnecessarily prominent Google logo next to the search box in the right panel should be removed.



Figure 4.5. Current NYMTC Home Page

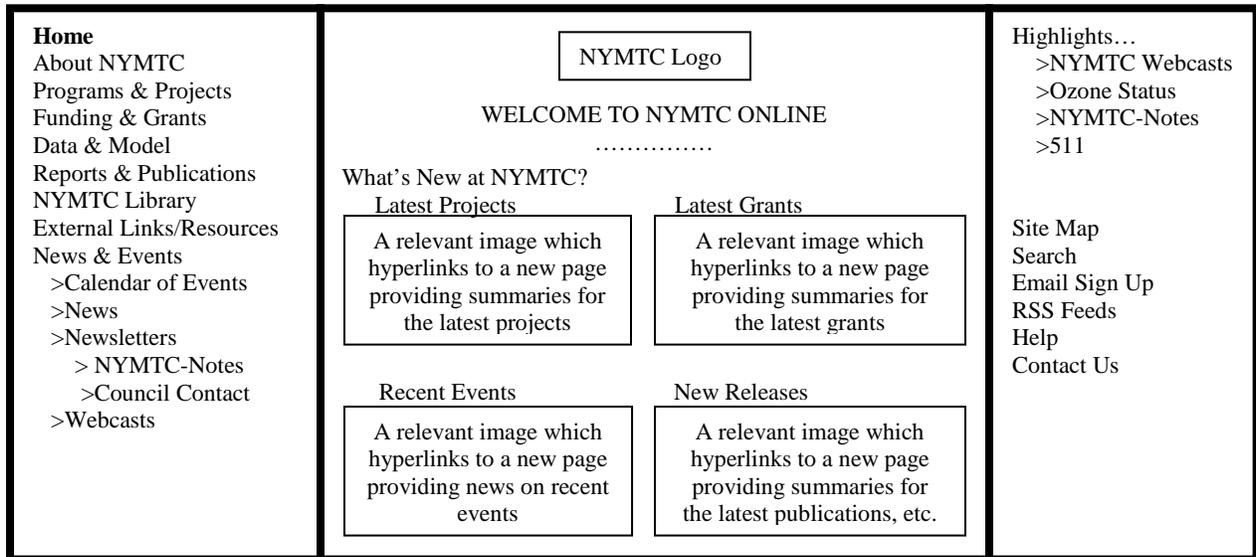


Figure 4.6. Suggested Page Re-organization for Left, Center and Right Panels of the NYMTC Home Page

Recommendations for the ‘About NYMTC’ Page (Medium Priority)

The major recommendations are given below:

- The first paragraph is a repetition of the mission statement, so should be deleted and the page should start with the mission statement.
- On the UPWP page, the four documents at the bottom should be titled ‘Recent Documents’ followed by a link to the complete list of UPWP documents.
- For a complete list of UPWP, TIP, and RTP documents, the user should be referred to (through a hyperlink) the Reports & Publications page.
- The Staff section should provide bios, responsibilities, as well as phone, fax numbers and email addresses.

Recommendations for the ‘Reports & Publications’ Page (High Priority)

Please note once again that it is recommended that the current ‘Downloadable Files’ page be renamed as ‘Reports & Publications’. Specific recommendations for this page include the following:

- There should be a statement at the opening page saying that the page only contains publications and reports that aim to provide information on various issues from Accounting Advisories to list of RFPs, etc., and that for databases and reports/files containing data, the users should refer to the ‘Data & Model’ page.
- The item named Census, which provides Journey-to-Work data, should be integrated into the Census item located at the ‘Data & Model’ page.
- The items Conformity, Major Investment Studies, Sustainable Development Studies should be moved to the ‘Data & Model’ page.
- NYMTC-Notes should be moved and integrated into the NYMTC-Notes section located on the ‘News & Events’ page.
- In addition to the current browsing capability by selecting files from the drop-down list, provide a search box for easier access to reports and publications similar to the search section shown in Figure 4.8.

Recommendations for the ‘NYMTC Library’ Page (High Priority)

Recommendations include the following:

- There should be a statement at the opening page saying that the page only contains publications and reports that are in hardcopy/CD/tapes formats, and that for a comprehensive list of data available in various formats the users should refer to the ‘Data & Model’ page.
- The current drop-down list style browsing should be kept. However, the current Advanced Search feature should be modified as it requires the users to perform five different searches to find what they need using the same key word(s) repeatedly. Instead, there should be a single search box and the search domain should consist of only the data products that are available in hardcopy or on CD and tapes. Once the user types a keyword and performs the search, the results page should bring up the metadata for the matching data products.

- The ‘Data Depot’ section, which is under construction and currently includes only the “County Profiles” data (duplicate of the one provided under the current ‘Data and Model’ section) should be removed entirely.
- The digital version of NYMTC’s famous “Info Wheel” should be moved to the new ‘Data & Model’ page.
- In addition to the current location and hours information provided for the library, loan policies for employees/partners vs. general public, and the name and contact details of a librarian should be provided.
- The links to various external libraries and resources shown on the page should be moved to the ‘External Links & Resources’ page to prevent repetition.

Recommendations for the ‘GIS’ Page (To be Accessible through Proposed ‘Data & Model’ Page) (Medium Priority)

Please note once again that it is recommended that the link to the GIS page is removed from the home page and instead, it is made accessible through the ‘Data & Model’ page. Specific recommendations for the ‘GIS’ page include the following:

- For the household survey database, currently only the data download function is available, so access to GIS map and the associated database should be provided.
- Metadata page should be provided.
- Variable dictionaries should be made available for all the databases, as they are currently available for the vehicle counts, O-D survey, and socioeconomic and land use data only.
- Variable dictionaries should be put into a more understandable form, as they currently tend to be confusing and include a lot of extra information used for internal filing purposes only.
- Shape files should be made downloadable.
- More data subjects and data for more recent years should be provided.
- The GIS interactive application should be removed because it does not work as advertised.

Recommendations for the ‘Data & Model’ Page (High Priority)

Since the ‘Data & Model’ page (currently called ‘Data and Model’) is where the 11 data products are located, which is the subject matter of this project, related recommendations are presented in detail. Examples of best practices (with screenshots and their URLs) that already use the recommended solution are given whenever possible.

The current ‘Data and Model’ page is shown in Figure 4.7, and suggested improvements for the proposed ‘Data & Model’ page are presented in Figure 4.8 and also listed as follows:

- The page should provide both browsing and search options to the user; there should be a drop-down list which allows browsing by topic, source, alphabetically, by type (At a Glance Brochures vs. Full Reports). Also, an advanced search tool that allows the users to search for the data products as well as for individual graphs/charts/maps should be provided. The search domain should be limited to the data files only and the search function should look for the typed keyword both in the contents and the titles of the data files. In the search results page, clicking on a result should display the relevant metadata in a new window, and from there,

the user would proceed by selecting the desired format from the Available Formats section in Metadata as shown in Figure 4.1 earlier. A few good examples of data pages providing both browsing and advanced search capabilities are shown in Figures A23 through A27.

- A regional map should be provided for better visualization by the user.
- A “What’s New” button which opens up a new page should highlight data updates, new releases, upcoming data products etc.
- Most requested items can be indicated.
- A link to various user support tools should be provided as explained in detail in Figures 4.3 and 4.4.
- It may also be useful to ask the user identify themselves (organization type (e.g. university, industry, government) and their country, which data product they are interested in, why they need that data).

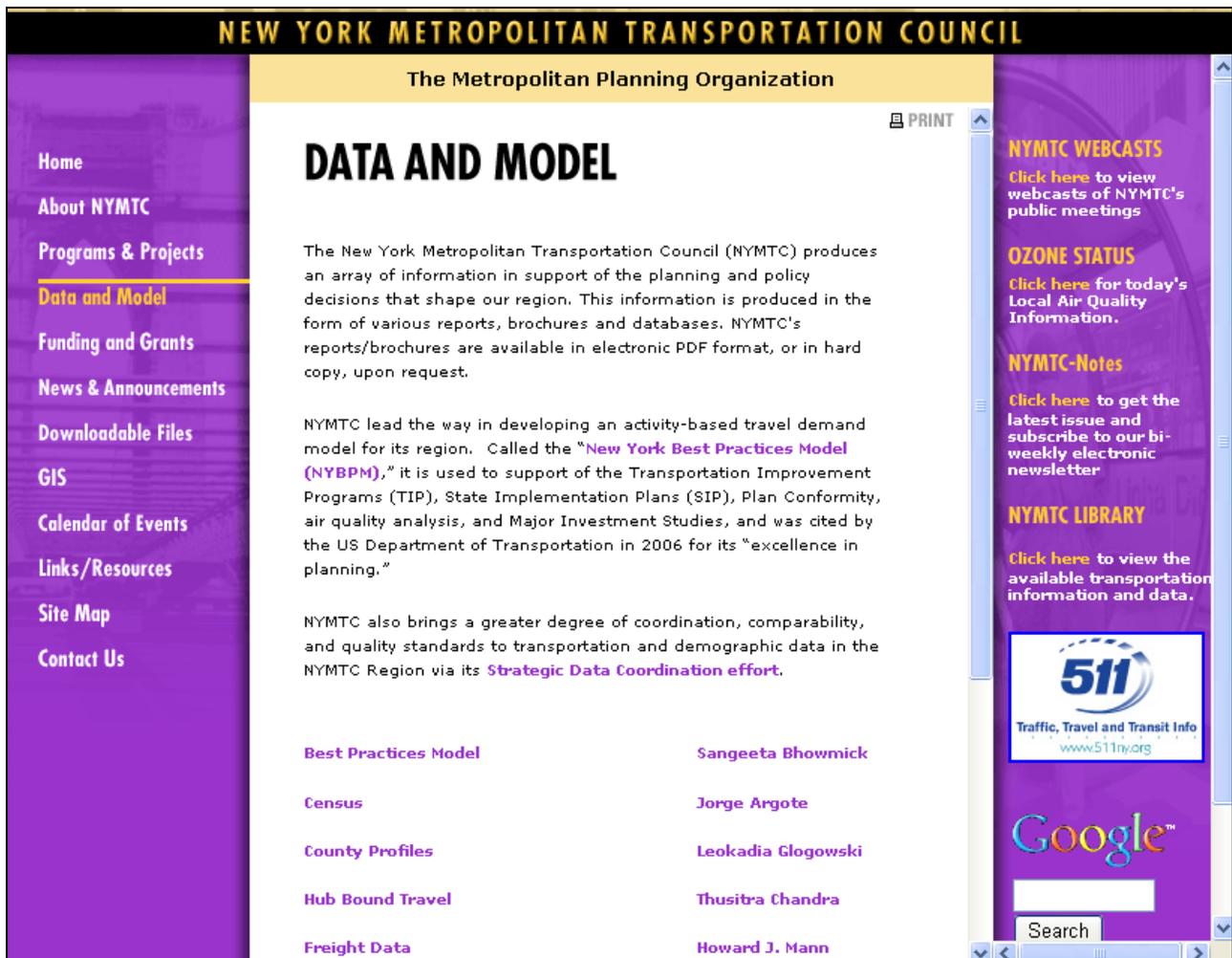


Figure 4.7. Current Data and Model Page at NYMTC Website

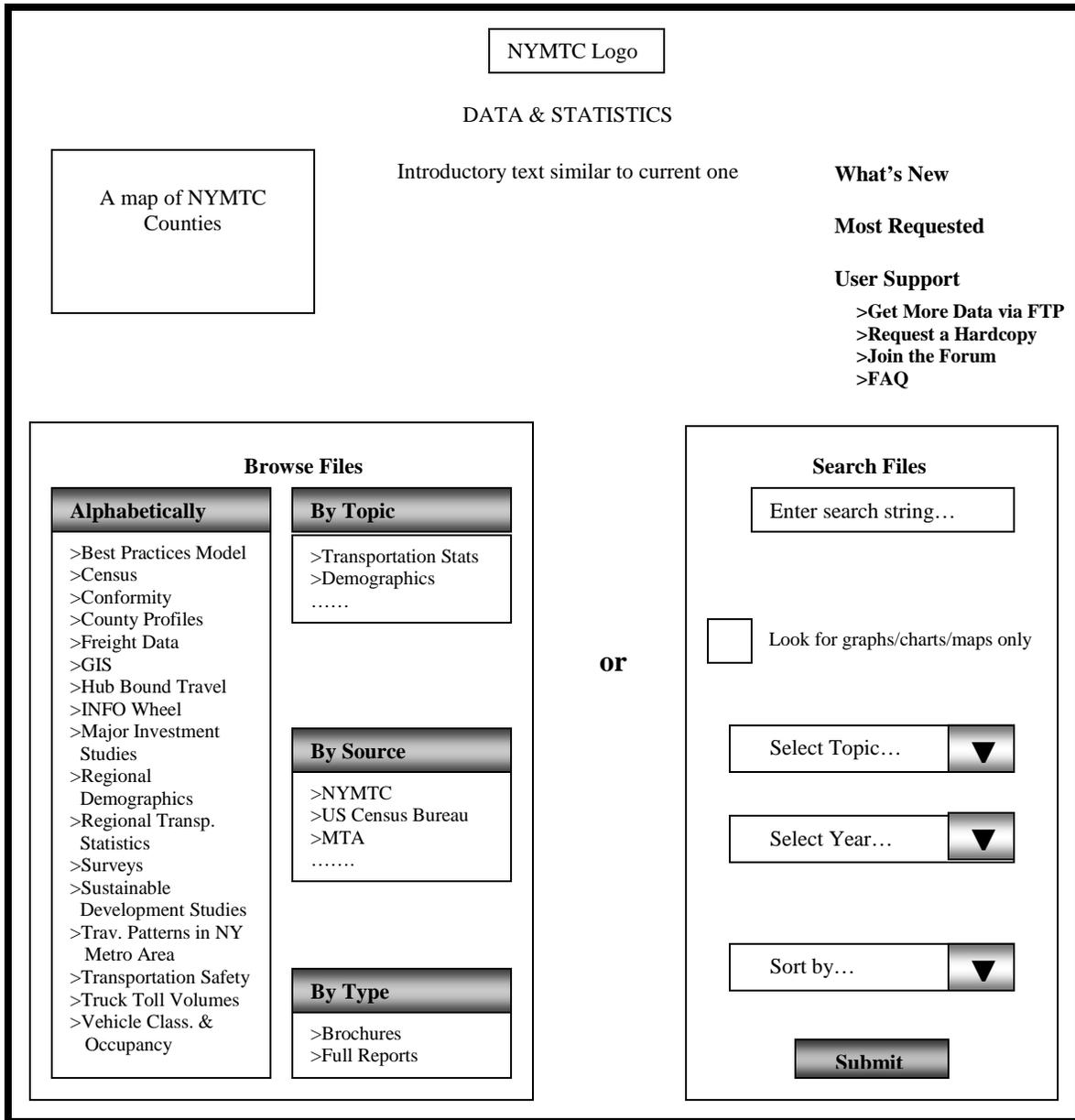


Figure 4.8. Suggested Data & Model Page on the NYMTC Website (Only Center Panel is Shown)

4.2. Process Oriented Solutions (High Priority)

Given Loshin's (2007) argument that "information follows a 'lifecycle' consisting of creation, distribution, access, updating, and retirement stages, and that it is necessary that a data quality framework provide protocols for measuring the quality of information at the various stages of that life cycle", a dynamic data maintenance procedure is proposed in this section, which might potentially be applied on a regular basis (bi-annually or whenever a new data is produced and required to be disseminated).

This proposed procedure is built around three main functional components (review, monitor/assess, and modify/improve). The suggested procedure also incorporates the elements from the lifecycle of data (create, distribute, access, update) as shown in Figure 4.9.

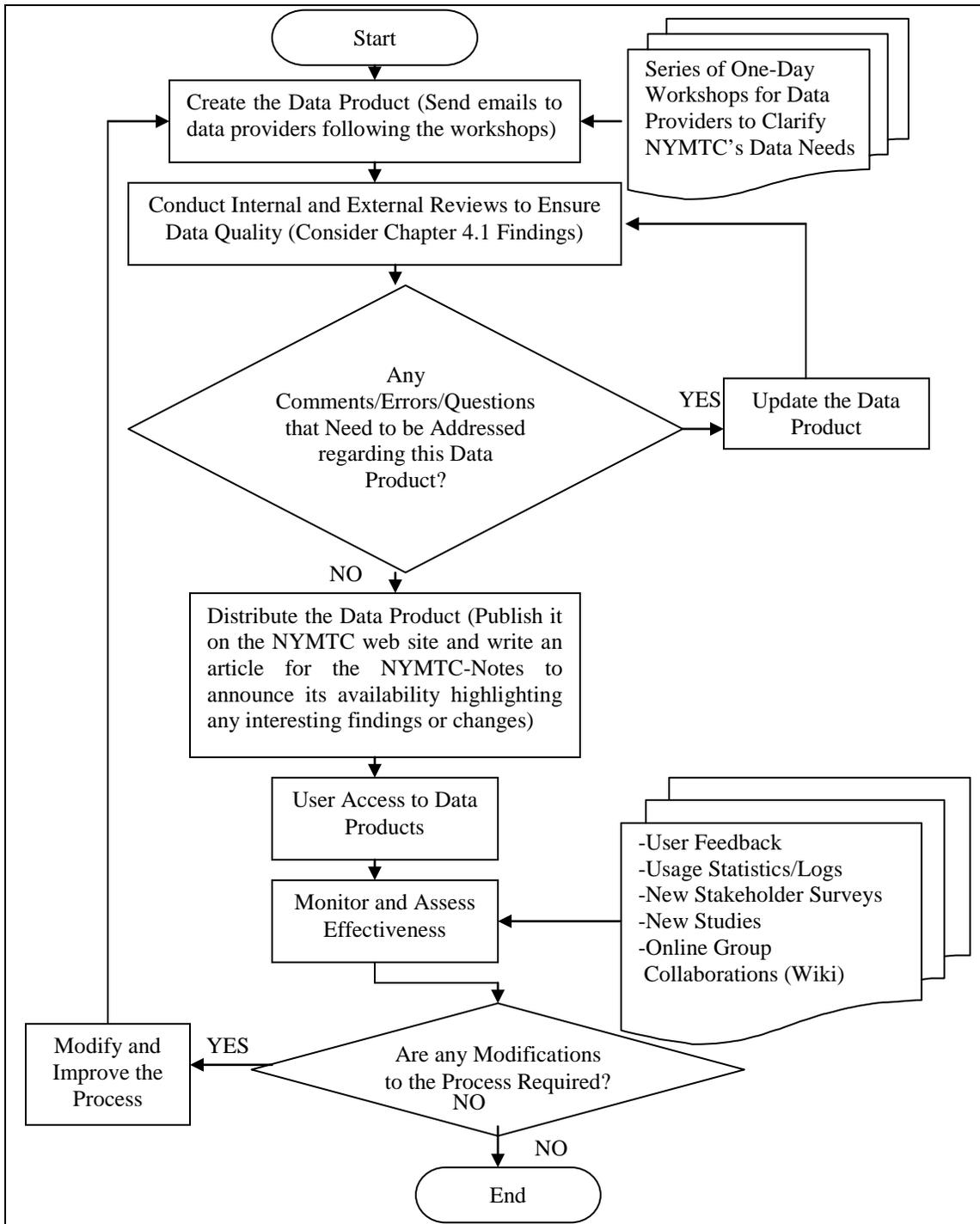


Figure 4.9. Suggested Dynamic Data Maintenance Procedure

4.2.1. Create

The suggested process starts with *creating the data product*. A very important input and feature to this step is a *series of one-day workshops* recommended as part of this dynamic data maintenance framework for the purposes of letting data providing agencies get a better idea about NYMTC's data needs.

Many agencies contribute various data needed by NYMTC. The availability, quality, and timeliness of this data generally depend on establishing a long-term relationship with each agency. However, there is not one institutional approach that is applicable to every external agency. Since the overall quality of data is crucial to the success of NYMTC's goals, it is important to establish long-term institutional procedures with all the contributing agencies. Moreover, it is also important to better explain and sometimes provide incentives to ensure that the data flow is consistent and satisfactory. Based on various meetings with NYMTC, the major time delay related to having latest data is caused by the delays in obtaining this data from individual agencies. There is little room in reducing time required to publish this data since NYMTC's internal process is already quite efficient given the rigorous internal and external requirements of the data to be published.

Thus, to determine novel ways to improve the efficiency of the data acquisition process from participating agencies, the research team proposes to **organize and conduct a series of workshops** (at least two) where all the contributing agencies are brought together to set up a more formal and reliable mechanism for NYMTC to work with these external agencies to obtain necessary data. This mechanism will be based on the following ideas:

1. Creation of an agency-specific formal protocol that ensures timely acquisition of high quality data. This protocol will be developed by emphasizing the mutual benefits of this process and will establish an institutional collaboration in addition to the current informal process that is in place.
2. Selection and adaption of a number of technologies that will reduce the burden of requesting and submitting the data will be explored. For example, a simple web-based data depositing system that is accessible to the external agencies can be created. The same system can be set up in such a way that it sends warnings to the agencies that are late in submitting the data.
3. Identification of incentives that are most likely to improve the current process of data acquisition.

A detailed proposal for the series of one-day workshops for the agencies providing data to NYMTC will be submitted to NYMTC as a separate document.

4.2.2. Review

The next step in the suggested data maintenance process, namely the *data product reviews*, uses the findings from Chapter 4.1 when conducting the reviews and looking for any issues in terms of various dimensions of data quality such as coverage, appropriateness of format, availability and content of metadata, accuracy, integration, accessibility and so on as discussed in detail previously in Chapter 4.1. If the data product presents conformity in terms of these various aspects of data quality, it will be distributed to the end-users. If not, it shall undergo necessary

modifications. This part of the procedure should be repeated until desired conformity to the desired level of data quality is achieved.

4.2.3. Distribute, Monitor and Assess

Once the *data product is distributed* by publishing it on the NYMTC website and the *users can access it*, another functional component of the maintenance process, *monitoring and assessment*, will come into play. It will basically involve utilization of the following tools and methods in order to determine if any modifications should be applied to the existing data quality dimensions.

1. Online user feedback forms
2. Website traffic monitoring tools such as hit counters to keep track of successful data retrievals and returning users, the popularity of pages and data products, and common entry points
3. Conducting new stakeholder surveys
4. Reviewing new studies on the latest technological trends and best practices
5. Use of new online multi-user collaboration tools such as wiki (a World Wide Web application designed to allow multiple users to add, remove, and edit content), to create a common community environment for data providers as well as users, as explained in Chapter 4.1.1 (“Provide Better Data Access and Retrieval Options” section) in detail along with a few examples
6. Implementing emerging web-based technologies that facilitate exchange of information while creating on-line communities beyond the tools described above. For example, “elgg (<http://elgg.org>)” an open source social networking tool is used to create professional and scientific communities such as, <http://scispace.net/>. These sites are easy to maintain and thrive on the contribution of its members. Elgg provides fine grained access control that can be easily customized to allow different contributors to have different access privileges. For example, some content that is published on the webpage can be accessible to everybody subscribing to this network while some other content such as agency specific data repositories can be made accessible to the corresponding agencies only.

4.2.4. Modify and Improve

If it is agreed by NYMTC that a number of modifications to the current data maintenance and quality control process are necessary, then, the third functional component, *modification and improvement*, comes into play and acts as a feedback mechanism and allows the incorporation of the results from the monitoring and assessment component back into the initial step of the process to guarantee the continuing success of the agency-wide data maintenance procedure.

CHAPTER 5

CONTEMPLATING THE NEAR FUTURE

5.1. Draft Timeline for the Implementation of the Project

A draft timeline for the implementation of this project in its entirety is given in Figure 3 below. As can be seen, the roles and responsibilities on NYMTC's part as well as the research team's part are shown.

Table 5.1. Draft Timeline for Project Implementation

NYMTC/New Consultant	Date	Research Team (Ozmen-Ertekin and Ozbay)
	Dec 09'	Complete the final report for the project
Start redesigning NYMTC website (home page, Data & Model page, data pages) (as described in Chapter 4)	Jan 10'	Provide feedback and get actively involved
First Workshop (will be described in a separate document)	Mar 10'	Organize
Second Workshop (will be described in a separate document)	May 10'	Organize
Perform monitoring and assessment (new survey, calculation of Customer Satisfaction index etc.) (as described in Chapter 5)	Aug 10'	(This will be an internal decision to be made by NYMTC)

5.2. Overall Evaluation of the Project's Impacts

In an effort to make an overall evaluation of this project's impacts, this section aims to answer the following two questions:

- What will be the overall value-added by this project? (pros)
- Is it worth doing? (cons)

Table 5.2 summarizes the anticipated pros and cons of the recommendations made in this project for both data and website related issues, and data maintenance related issues.

Table 5.2. Potential Pros and Cons of Recommendations

Type of Recommendation	Pros	Cons
Data and Website Oriented	<ul style="list-style-type: none"> +The time, money and effort spent cleaning poor quality data may be reduced significantly, thus increasing overall productivity +Data users get high quality, accurate, highly accessible, better integrated and customizable data in a timely manner + Increased stakeholder and data user satisfaction +Help NYMTC better project an image of quality to other agencies, stakeholders and data users +Promote website and other types of data user attraction and retention 	<ul style="list-style-type: none"> - Long-term commitment of the time and effort of NYMTC staff and the consultant implementing the project -Implementing and maintaining some technology bound solutions requires expertise and money -Technology changes constantly
Data Maintenance Process Oriented	<ul style="list-style-type: none"> +Interagency workshops will reduce delays for NYMTC in getting data from external agencies and will provide good input +Improvement regarding the notion of quality data and its maintenance over time so that NYMTC staff, agencies providing data, all the stakeholders and data users view data quality as key to success and making valid decisions +Provide broader networking and connectivity and create an atmosphere of cooperation among NYMTC, agencies providing data, stakeholders and data users 	<ul style="list-style-type: none"> -Workshops are costly and difficult to establish -Expenditures for the stakeholder surveys, and other monitoring and assessment tools

The overall success and impact of the recommendations made in this project can be assessed over time by calculating several quantifiable measures of performance such as, “Combined Customer Satisfaction Index (CCSI)” and “Importance Index (ImpI)” given by Equations 1 and 2 in Chapter 3, after re-conducting stakeholder surveys using the same sample of stakeholders. This way, a simple but informative chart as shown in Figure 5.1 can be obtained depicting the potential effectiveness and success of the suggestions made in this project.

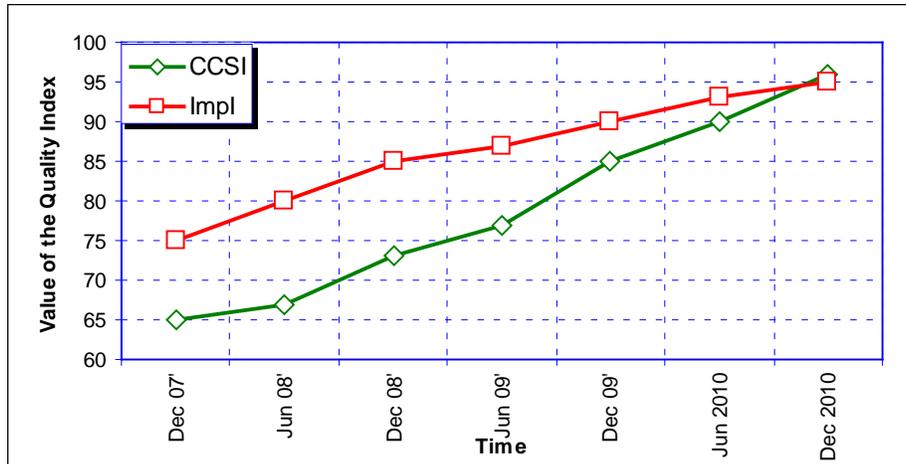


Figure 5.1. A Sample Scatter Diagram Depicting the Effectiveness of the Dynamic Data Maintenance Procedure Over Time

REFERENCES

- American Customer Satisfaction Index (ACSI) Website. < <http://www.theacsi.org/>> Accessed on June 26, 2007.
- Crescenzi, V., Mecca, G., and Merialdo, P. (2001). "RoadRunner: towards automatic data extraction from large websites", Proceedings of the 27th VLDB Conference, Roma, Italy.
- Frawley, W., Piatetsky-Shapiro, G., and Matheus, C. (1992). "Knowledge discovery in databases: an overview", *AI Magazine*, Fall, pp. 213–228.
- Hand, D. Mannila, H. and Smyth, P. (2001). *Principles of Data Mining*. MIT Press, Cambridge, MA. ISBN 0-262-08290-X.
- Lee, Y.W., Strong, D.M., Kahn, B.K., and Wang, R.Y. (2002). "AIMQ: a methodology for information quality assessment", *Journal of Information and Management*, 40, 133-146.
- Loshin, D. (2007). "Developing a data quality strategy", On-Line Article Published by *Business Intelligence Network*, January 18. < <http://www.b-eye-network.com/view/3816>> Accessed on June 1, 2007.
- Minnesota Department of Employment and Economic Development (DEED) Website. < <http://www.deed.state.mn.us/customersurvey/csi.htm>> Accessed on June 26, 2007.
- Moore, S. and Clark, D. (2004). "Dynamic data analysis on the web: a design approach", Online article. Available from <http://devshed.com>
- Myllymaki, J. (2001). "Effective web data extraction with standard XML technologies", *Association for Computing Machinery*, pp.689-696.
- New York Metropolitan Transportation Council (NYMTC) Website. <<http://nymtc.org>> Accessed on May 23, 2007.
- New York Metropolitan Transportation Council (NYMTC). *Data Group Cookbook for Reports and Brochures*.
- Wang, R.Y. and Strong, D.M. (1996). "Beyond accuracy: what data quality means to data consumers", *Journal of Management Information Systems*, 12(4), 5-33.
- Wang, R.Y., Ziad, M., and Lee, Y.W. (2001). *Data Quality*. Massachusetts, Kluwer Academic Publishers.
- Washington State Department of Transportation (WSDOT) Website (2006). <<http://www.wsdot.wa.gov/Traffic/Congestion/Rice/announcement.htm>> Accessed on June 24, 2007.

APPENDICES

Appendix 1. Tabular Review of Best Practices

Table A1. Review of Best Practices

No	Agency Name	Data Provider	Data Description		Data Quantity and Its Property					
		Website Address	Data Subjects	Dataset Organization	Newest Data	Oldest Data	Data Continuity	Availability Period	Smallest Geographic Level	Time Lag between Data Collection and Publish Date
1	Bureau of Transportation Statistics (BTS)	http://www.bts.gov	Transportation (safety, freight, passenger travel, infrastructure, economic/financial, social/demographic, energy, environment, national security)	By mode, by subject, alphabetically	Feb-08	1974	Certain proportion is missing, but not mentioned	Annual/monthly	Mostly state level, down to county	3 months to 1 year for new release; up to 5 years for updates
2	New Jersey Department of Transportation (NJDOT)	http://www.nj.gov/transportation	Transportation (crash records, sidewalk inventory, GIS, roadway information, traffic counts, soil borings, straight line diagrams)	Alphabetically	2007	1968	Small proportion is missing, not explained	Mostly annual	County	6 months to more than a year
3	North Jersey Transportation Planning Authority (NJTPA)	http://www.njtpa.org	Transportation (home interviews, journey-to-work, crash data, demographics, GIS)	By subject	2007	1998	Continuous	Annual/monthly	County	Around 1 year
4	British Columbia Ministry of Forests and Range (BCMFR)	http://www.for.gov.bc.ca/mof/reports.htm	Forestry (annual reports, service plans, maps)	By function (reports, plans, manuals)	2007	1956	Continuous	Annual/monthly	County	6 months to 1 year
5	Colorado Department of Transportation (CDOT)	http://www.dot.state.co.us	Transportation (geographic, highway, maps, straight line diagrams, statistics, traffic data)	By subject	2007	1997	Continuous	Annual	County	Around 1 year

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	Data Source Information Availability			User Accessibility to Data				Advanced Accessibility	
		Data Collector	Collection Methodology	Purpose of Collection (commercial, non-profit, etc.)	Access Restrictions	User Identification other than Automatic IP Address Recognition	Incentive of Use (require user provide incentive of use or not)	Privilege Discrimination	Further Data Availability Online	Further Data Availability by Hardcopy
1	Bureau of Transportation Statistics (BTS)	Multiple agencies	Survey/count	Not mentioned	None	No	Not required	None	Upon email request	Not mentioned
2	New Jersey Department of Transportation (NJDOT)	Itself	Count data	Not mentioned	None	No	Not required	None	No	Yes
3	North Jersey Transportation Planning Authority (NJTPA)	Multiple agencies	Count data	Some available	None	Email for further request	Not required	None	No	Yes
4	British Columbia Ministry of Forests and Range (BCMFR)	Itself	Not mentioned	Not mentioned	Registration required for some publications	Employer	Not required	Government user/ business user/residences	No	Yes
5	Colorado Department of Transportation (CDOT)	Itself	Count data	Some available	None	No	Not required	None	No	Yes

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	Metadata Availability			Data File Format				Output Generation				
		Dataset Description (sufficient or not)	Data Caps / Overlaps	Variable Dictionary	Available Formats	Multiple Formats Available (for the same data)	Format Choice Consistency (choice is fully/partially available)	Mostly Used Format	Automatic Web Data Generation (in response to user request)	Availability of Online Data Extraction (by select options: time, location, etc.)	Table/Chart Availability	Map Availability	Table/Chart/Map Customization
1	Bureau of Transportation Statistics (BTS)	Yes	Exists, but not mentioned	Available, but access not consistent at different levels	.xls, .html, .csv	Mostly not	Partial	.csv	Yes	Yes	Charts not available for most datasets	No	No
2	New Jersey Department of Transportation (NJDOT)	No	No	Available, hard to find	.xls, .doc, .pdf, .txt, .zip	No	-	.pdf	No	Yes	No	Yes	No
3	North Jersey Transportation Planning Authority (NJTPA)	Yes	No	Highly accessible	.xls, .pdf, .zip, .dbf	No	-	.pdf, .zip	No	Yes	Charts not available	Yes	No
4	British Columbia Ministry of Forests and Range (BCMFR)	No	No	Yes	.pdf, .html	Yes	Full	.pdf	Yes	Yes	Yes	Yes	No
5	Colorado Department of Transportation (CDOT)	Yes	No	No	.pdf, .xls, .html, .zip	No	-	.pdf	Yes	Yes	Yes	Yes	Yes

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	Ease of Output		Advanced Services		
		Data Print/ Download Friendly	Table/Chart/Map Saving/Printing	Both Raw Data and Statistics Available	Statistical Analysis Availability for Chosen Data	Analysis Tools (if available)
1	Bureau of Transportation Statistics (BTS)	Yes	Yes	Yes	No	-
2	New Jersey Department of Transportation (NJDOT)	Yes	Yes	Yes	No	-
3	North Jersey Transportation Planning Authority (NJTPA)	Yes	Yes	Yes	No	-
4	British Columbia Ministry of Forests and Range (BCMFR)	Yes	Yes	Yes	No	-
5	Colorado Department of Transportation (CDOT)	Yes	Yes	Yes	No	-

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	User Support and Feedback						Final Remarks	
		FAQs Availability	Most Requested Data (specified or not)	Related Topics/Data (people who download this data are also interested in)	Contact Information	Feedback Request on Strategies and Content	Newsletter/RSS/Podcast/Email Update Availability	Most Notable Strengths of the Service	Most Notable Weaknesses of the Service
1	Bureau of Transportation Statistics (BTS)	Yes	No	Related links available	Email/phone	No	Newsletter available for some databases	Good data coverage, Short time lags, Online data extraction possible	Data gaps, Variable definitions not always available and not consistently located, Provision of maps/charts extremely limited, Almost no user support and feedback tools
2	New Jersey Department of Transportation (NJDOT)	Yes	No	Related links available	Phone, personal contact available for certain datasets	No	No	Good data coverage, Online data extraction possible	Output format choice not available, No data customization, Metadata limited
3	North Jersey Transportation Planning Authority (NJTPA)	Yes	No	Yes	Person	Yes	Newsletter available	Map customization, Best user support and feedback tools among transportation related websites reviewed	Short data coverage, Lack of data customization and table/chart output, Access to publications not as transparent as the access to datasets
4	British Columbia Ministry of Forests and Range (BCMFR)	Yes	No	No	Person/phone/fax/email	About data quality	Podcasts	Good data coverage, Data quality survey and improvement feedback are collected from business/govemer users	Lack of table/chart output customization, Limited user support and feedback tools
5	Colorado Department of Transportation (CDOT)	No	No	Yes	Email/phone	Yes	Newsletter available	Data covers variety of subjects at county level, Strength in data/output customization, especially highway data	Output format choice not available

Table A1. Review of Best Practices (Cont'd)

No	Data Provider		Data Description		Data Quantity and Its Property					
	Agency Name	Website Address	Data Subjects	Dataset Organization	Newest Data	Oldest Data	Data Continuity	Availability Period	Smallest Geographic Level	Time Lag between Data Collection and Publish Date
6	Washington State Department of Transportation (WSDOT)	http://www.wsdot.wa.gov	Transportation (maps, geospatial, collision reports, traffic data)	By subject, alphabetically	2007	1960	Continuous except oldest data	Annual/monthly	County	6 months to 1 year
7	National Highway Traffic Safety Administration's Fatality Analysis Reporting System (FARS)	http://www-fars.nhtsa.dot.gov	Transportation (detailed crash data)	By subject	2006	1994	Continuous	Annual	State	More than 1 year
8	US Census Bureau	http://www.census.gov	Demography (people and households, business and industry, geography)	By subject	2007	1970	Continuous	Periodic	County	6 months to 2 years
9	National Center for Education Statistics (NCES)	http://nces.ed.gov	Education (variety of education topics)	By subject	2008	1980	Continuous	Periodic	County	Less than 1 year
10	National Center for Health Statistics (NCHS)	http://www.edc.gov/nchs	Health (variety of health topics)	Alphabetically	2006	1985	Continuous	Periodic	State	Around 1 year
11	National Climatic Data Center (NCDC)	http://www.ncdc.noaa.gov/oa/ncdc.html	Climate (variety of climate related data)	By collection method, alphabetically	2007	1990	Continuous	Annual	National	Around 1 year

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	Data Source Information Availability			User Accessibility to Data				Advanced Accessibility	
		Data Collector	Collection Methodology	Purpose of Collection (commercial, non-profit, etc.)	Access Restrictions	User Identification other than Automatic IP Address Recognition	Incentive of Use (require user provide incentive of use or not)	Privilege Discrimination	Further Data Availability Online	Further Data Availability by Hardcopy
6	Washington State Department of Transportation (WSDOT)	Itself	Survey/count	Some available	None	No	Not required	None	No	Yes
7	National Highway Traffic Safety Administration's Fatality Analysis Reporting System (FARS)	NHTSA	Count data	Not mentioned	None	No	Not required	None	Upon request, requires personal information	Yes
8	US Census Bureau	Itself and other government agencies	Survey/count	Not mentioned	None	No	Not required	None	No	Yes
9	National Center for Education Statistics (NCES)	Itself	Survey/self-report	Some available	None	No	Not required	None	Yes	Yes
10	National Center for Health Statistics (NCHS)	Itself	Survey/count	Not mentioned	None	No	Not required	Business buyer	No	CD-ROM
11	National Climatic Data Center (NCDC)	Itself	Survey	Not mentioned	None	No	Not required	None	Yes	Yes

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	Metadata Availability			Data File Format				Output Generation				
		Dataset Description (sufficient or not)	Data Caps / Overlaps	Variable Dictionary	Available Formats	Multiple Formats Available (for the same data)	Format Choice Consistency (choice is fully/partially available)	Mostly Used Format	Automatic Web Data Generation (in response to user request)	Availability of Online Data Extraction (by select options: time, location, etc.)	Table/Chart Availability	Map Availability	Table/Chart/Map Customization
6	Washington State Department of Transportation (WSDOT)	Yes	No	No	.pdf, zip	No	-	.pdf	Yes	Yes	Yes	Yes	Yes
7	National Highway Traffic Safety Administration's Fatality Analysis Reporting System (FARS)	No	No	Yes	.txt, .xls, .pdf	Yes	Full	.xls	Yes	Yes	Yes	No	Yes
8	US Census Bureau	Yes	Exist, and explained	Accessible	.pdf, .xls, .zip	Yes	Partial	.pdf, .zip	Yes	Yes	Yes	Yes	Yes
9	National Center for Education Statistics (NCES)	Yes	Not mentioned	No	.pdf, .html	No	-	.pdf	No	Yes	Yes	Yes	No
10	National Center for Health Statistics (NCHS)	Yes	Not mentioned	Available for some databases	.pdf, .jpeg, .html	No	-	.pdf	No	No	Yes	Yes	Yes
11	National Climatic Data Center (NCDC)	Yes	Not mentioned	Yes	.csv, .txt, .html	Yes	Full	.html	Yes	Yes	Yes	Yes	Yes

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	Ease of Output		Advanced Services		
		Data Print/ Download Friendly	Table/Chart/Map Saving/Printing	Both Raw Data and Statistics Available	Statistical Analysis Availability for Chosen Data	Analysis Tools (if available)
6	Washington State Department of Transportation (WSDOT)	Yes	Yes	Yes	No	-
7	National Highway Traffic Safety Administration's Fatality Analysis Reporting System (FARS)	Yes	No	Yes	Yes	Count, sort, cross tab
8	US Census Bureau	Yes	Yes	Yes	No	-
9	National Center for Education Statistics (NCES)	No	No	Yes	No	-
10	National Center for Health Statistics (NCHS)	Yes	Yes	Yes	No	-
11	National Climatic Data Center (NCDC)	Yes	No	Yes	Yes	Graph generation on map

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	User Support and Feedback						Final Remarks	
		FAQs Availability	Most Requested Data (specified or not)	Related Topics/Data (people who download this data are also interested in)	Contact Information	Feedback Request on Strategies and Content	Newsletter/RSS/Podcast/Email Update Availability	Most Notable Strengths of the Service	Most Notable Weaknesses of the Service
6	Washington State Department of Transportation (WSDOT)	Yes	Yes	Yes	Phone/person	Hard to find	No	Good data coverage, Data/output customization, Clear and highly accessible metadata; highlights most requested data to users	Output format choice not available, Varieties of data/maps not available (only available by hard copy); Limited support and feedback tools
7	National Highway Traffic Safety Administration's Fatality Analysis Reporting System (FARS)	No	No	No	Phone/fax	No	No	Provides additional analysis tools, Highly interactive data customization, Provides open access of ftp to users	Long lags between data date and publish date; Report/Charts not saving friendly; Lack of user support tools
8	US Census Bureau	Yes	"Most recent" specified	Yes	Email/phone/person	Yes	Newsletter available	Good data coverage, highly interactive data customization, sufficient metadata, ftp link available	Most are period data, ftp service not very user-friendly and informative
9	National Center for Education Statistics (NCES)	Yes	Popular keywords, new items specified	No	Phone	Available in "contact" section	Newsletter, RSS available	"Fsat facts" tool provides various key information about datasets, Relatively short lags between data date and publish date, RSS feeds available	Limited data/output customization, Not download/saving friendly
10	National Center for Health Statistics (NCHS)	Yes	Yes	Yes	Email/phone	Easy to find	All available	Map/chart generation and customization, easily accessible and various user support and feedback tools	Online data extraction not available, Data format choice not available
11	National Climatic Data Center (NCDC)	Yes	Yes	No	Email/phone	Yes	Newsletter available	Good data coverage, Interactive chart/map generation, Spatial analysis tool available, ftp link available	Output map/charts not saving friendly

Table A1. Review of Best Practices (Cont'd)

No	Data Provider		Data Description		Data Quantity and Its Property					
	Agency Name	Website Address	Data Subjects	Dataset Organization	Newest Data	Oldest Data	Data Continuity	Availability Period	Smallest Geographic Level	Time Lag between Data Collection and Publish Date
12	National Association of Realtors (NAR)	http://www.realtor.org	Real Estate	By subject	Mar-08	2005	Continuous	Annual/monthly	State	Less than 3 months
13	International Monetary Fund (IMF)	http://www.imf.org	Finance (statistics, exchange rates)	By database title and subject	Apr-08	1980	Continuous	Annual/monthly	National	Less than 3 months
14	World Health Organization Regional Office for Europe (WHO-ROE)	http://www.euro.who.int	Health (variety of health topics)	By databases	2006	1970	Continuous	Periodic	National	Around 1 year
15	European Environment Agency (EEA)	http://www.eea.europa.eu	Environment (variety of environmental data)	Both by function and by subject	2007	1970	Certain proportion is missing online	Mostly annual	National	6 months to 1 year
16	Department for Transport, UK (DFT)	http://www.dft.gov.uk	Transportation (aviation, crime and public transport, economics, freight, railways, safety, security)	By mode and by function	Mar-08	1953	Continuous	Annual/quarterly	Mostly national	Less than 1 year

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	Data Source Information Availability			User Accessibility to Data				Advanced Accessibility	
		Data Collector	Collection Methodology	Purpose of Collection (commercial, non-profit, etc.)	Access Restrictions	User Identification other than Automatic IP Address Recognition	Incentive of Use (require user provide incentive of use or not)	Privilege Discrimination	Further Data Availability Online	Further Data Availability by Hardcopy
12	National Association of Realtors (NAR)	Itself	Survey/count	Available	Registration required for most products	Employer	Not required	Member/ non-member org. employee	Yes	Yes
13	International Monetary Fund (IMF)	Multiple agencies/companies	Count/self-report	Some available	Registration required for some products	No	Not required	None	Yes	Yes
14	World Health Organization Regional Office for Europe (WHO-ROE)	Partner organizations	Survey/count	Not mentioned	None	No	Required for quote	Member org. users	Yes	Yes
15	European Environment Agency (EEA)	Member/partner agencies/organizations	Survey/count	Some available	None	No	Not required	None	Yes	Yes
16	Department for Transport, UK (DFT)	Itself	Survey/count	Some available	None	No	Required if further information is needed by user	None	No	Yes

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	Metadata Availability			Data File Format				Output Generation				
		Dataset Description (sufficient or not)	Data Gaps / Overlaps	Variable Dictionary	Available Formats	Multiple Formats Available (for the same data)	Format Choice Consistency (choice is fully/partially available)	Mostly Used Format	Automatic Web Data Generation (in response to user request)	Availability of Online Data Extraction (by select options: time, location, etc.)	Table/Chart Availability	Map Availability	Table/Chart/Map Customization
12	National Association of Realtors (NAR)	Yes	Exist, and highlighted	No	.pdf, .xls, .txt	Yes	Partial	.pdf	No	No	Yes	No	No
13	International Monetary Fund (IMF)	No	Exist, and explained	Available, not easily accessible	.pdf, .xls, .txt, .png, .csv	No	-	.xls	No	Yes	Yes	Yes	Yes
14	World Health Organization Regional Office for Europe (WHO-ROE)	Yes	No	Accessible	.pdf, .xls, .txt, .zip	Yes	Full	.zip	Yes	Yes	Yes	Yes	Yes
15	European Environment Agency (EEA)	Yes	Exist, not mentioned	Yes	.pdf, .xls, .zip	No	-	.xls	Yes	Yes	Yes	Yes	Yes
16	Department for Transport, UK (DFT)	Yes	No	No	.pdf, .zip	No	-	.pdf	No	No	Yes	Yes	No

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	Ease of Output		Advanced Services		
		Data Print/ Download Friendly	Table/Chart/Map Saving/Printing	Both Raw Data and Statistics Available	Statistical Analysis Availability for Chosen Data	Analysis Tools (if available)
12	National Association of Realtors (NAR)	Yes	Yes	Only statistics are available for free	No	-
13	International Monetary Fund (IMF)	Yes	Yes	Yes	No	-
14	World Health Organization Regional Office for Europe (WHO-ROE)	Yes	Yes	Yes	Yes	Scattergram, histogram, rank
15	European Environment Agency (EEA)	Yes	No	Yes	Yes	Time-series trend analysis in data and spatial hotspot analysis in map
16	Department for Transport, UK (DFT)	Yes	No	Yes	No	-

Table A1. Review of Best Practices (Cont'd)

No	Agency Name	User Support and Feedback						Final Remarks	
		FAQs Availability	Most Requested Data (specified or not)	Related Topics/Data (people who download this data are also interested in)	Contact Information	Feedback Request on Strategies and Content	Newsletter/ RSS/Podcast /Email Update Availability	Most Notable Strengths of the Service	Most Notable Weaknesses of the Service
12	National Association of Realtors (NAR)	No	Yes	Yes	Email/phone	Yes	Newsletter, RSS	Very short lag between data date and publish date, web interface highly interactive and contains efficient support tools (online chat); various news release and update services	Data customization only available for member org. users, Data gaps/overlaps exist, Limited metadata
13	International Monetary Fund (IMF)	Yes	Yes	Yes	Email/phone	Yes	Newsletter available	Variety of datasets, Very short lag between data date and publish date, Provides full data/chart/map customization, Contains efficient support tools	Limited metadata, Data gaps exist
14	World Health Organization Regional Office for Europe (WHO-ROE)	Yes	No	Yes	Email/phone	No	Newsletter available	Provide various databases collected by multiple agencies, Sufficient metadata, Highly interactive data customization, Same window for both data customization and further analysis functions	Requires incentive of use for quote; Most databases are periodic
15	European Environment Agency (EEA)	Yes	Yes	Yes	Email/phone/ person	Yes	Newsletter, RSS available	Provides access to various databases collected by multiple agencies; Database organization can be customized; Strength in "live map" services: GIS-based map customization and location/spatial analysis available; sufficient user support and feedback tools (e.g. RSS)	Data gaps for certain databases, Format choice for data query not available, Some outputs not saving friendly
16	Department for Transport, UK (DFT)	Yes	Yes and 'most recent' and forthcoming data also	Yes	Email/phone	Yes	Newsletter available	Good data coverage, Short lag between data date and publish date in transportation websites, Metadata provided	Data/output customization not fully available, Maps not saving friendly

Some of the column headings in Table A1 require detailed explanations as given below.

“Data Description”

This category describes the subject of the datasets disseminated through the websites of the best practices reviewed. The research team examined two attributes: subject of the data and the organization of the data. 7 out of 16 websites reviewed are transportation related and the rest pertains to health, education, demography, environment, climate, finance, real estate, and forestry. The datasets in most of the websites reviewed were organized in several different ways: by mode, by subject, by alphabetical order, by function, by database title, and by data collection method. Some agencies use multiple ways to organize their datasets on their websites.

“Data Quantity and Its Property”

This category reports the continuity of the datasets in time as well as their geographical availability. Reporting interval generally varies from monthly to annually. Geographical availability varies from national to country level. The time lag between the date of data collection and its publication date varies from a couple of months to more than 1 year. The research team considers no longer than 1 year as the appropriate lag time.

“Data Source Information Availability”

This category describes the entity that collects the data, the method of data collection, and the purpose. In some cases, for example, in North Jersey Transportation Planning Authority’s (NJTPA) website, multiple agencies are involved in data collection. Some websites, such as Bureau of Transportation Statistics’ (BTS) website, provide extra links to the database of other agencies with a page that includes a warning message about leaving the site, so that users can choose to go to the other sources or go back to the original site. When it comes to the data collection methodology, not all the websites reviewed reveal that information.

“User Accessibility to Data”

This category examines the ease of access to the datasets by a user. The research team looked at the following issues in this category: a) whether registration (i.e., signing up for a username and password) is required for a user to access the datasets; b) whether a user needs to be identified; c) whether the user needs to provide a reason for being interested in a particular dataset; d) whether the website distinguishes between different types of users to provide priority for usage. The review of the 16 selected websites revealed that, in most cases, users do not need to register, identify their employer or provide a purpose to use the datasets. Most websites do not distinguish between different types of users.

“Advanced Accessibility”

This category examines whether additional datasets can be delivered to the user as hardcopies or electronically if the user requests them. Most of the websites reviewed provide hardcopies and/or CDs in response to user request. Only a few websites (BTS, National Center for Education Statistics (NCES), National Climatic Data Center (NCDC), National Association of Realtors (NAR), International Monetary Fund (IMF), World Health Organization Regional Office for Europe (WHO-ROE), European Environment Agency (EEA)) can provide electronic databases upon user request.

“Metadata Availability”

This category looks at the availability of metadata. The research team looked at three issues: a) whether the website provides a data description; b) how complete the metadata is; c) whether a variable dictionary is available or not, and if it is, how easy it is to access it. Most of the websites reviewed provide a data dictionary. Half of them have complete metadata information, and most have a variable dictionary, but the ease of access varies. Metadata information for GIS-based maps are highly accessible since it is usually downloadable in ZIP format and mostly contains a variable dictionary, whereas other metadata are usually available only on the webpage and thus, is less accessible when the variable dictionary is located on a separate page (i.e., not the data download page).

“Data File Format”

This category looks at the available format provided, availability of multiple formats, consistency in formatting, and the most frequently used format. All 16 websites offer more than one format, for example, PDF, CSV, XLS, HTML, DOC, TXT, or ZIP. The most frequently used format seems to be the PDF format.

“Output Generation”

This category looks at various issues relating to output generation. Majority of the websites reviewed can allow automatic web data generation in response to users' requests. Most websites allow users to perform web-based queries. Almost all the websites provide table/chart capability. Majority of the websites can produce maps for the users. Some also offers customized tables and charts.

“Ease of Output”

This category describes whether the websites' data or other information can be downloaded or saved in an easy and user-friendly way or not. The two basic types of output considered here are data and associated maps/charts. A popular strategy implemented in the 16 websites reviewed is to provide a “Download” button on the data output page. This is especially true for those websites that allow data customization and output selection. Most websites use the ‘pop-up’ feature for download, which may cause inconvenience for some users. The ‘print’ feature, which allows the user to print the data without saving on the hard drive, is however not often used. The research team observed that the websites that provide TXT output format provide this feature, while those with PDF or XLS format output do not.

The ease of output can also be measured by other means. The most popular way of outputting a map, table or chart is to provide them inside a PDF or HTML file, so the user can save the whole report instead of the map/table. For maps, it is useful to let the users download the whole GIS shape file and metadata as a single ZIP file. One website (EEA) allows users to search maps within PDF files and provides “Save” or “Print” buttons for each single map.

Another interesting feature relates to the output of the webpage. One website (Department for Transport, UK (DFT)) provides a download button on each of its webpages. By clicking on this download button, the user tells the website to generate the context on the webpage as a downloadable PDF.

“Advanced Services”

This category includes several attributes. “Both Raw Data and Statistics Available” column examines whether both or only one type is provided. All but one of the websites reviewed provide both types.

Only a few websites provide users with tools to analyze the data. Typically available are basic statistical tools, such as count, sort, rank, cross-tab, and chart/table output generation (histogram, scattergram, etc.) Such websites often have built-in map tools which allow the users to conduct some analysis online without installing the professional GIS software. Some websites (National Highway Traffic Safety Administration’s Fatality Analysis Reporting System (FARS), NCDC, WHO-ROE, EEA) have advanced analysis tools; for example, the user can generate hotspots from highway cross-section data or detect trends in a time-series dataset.

Some websites (NCDC, WHO-ROE, EEA) offer map-based search engines and provide supplementary radio buttons on the map. This allows the users to select data on the map, save/download the data, or even conduct table/charts analysis.

“User Support and Feedback”

This category examines whether the users can find help if they need to. In particular, it looks at whether the website provides a user guide on how to find specific information/data, and who or where to contact if further help or assistance is needed. A website can be considered as user-friendly in this aspect if the “Contact Us” page lists both the email and phone/fax contact information for each branch or section of the agency/company. In addition, frequently asked questions (FAQ) needs to be available and the website needs to publish newsletters and provide users with a feedback dialog box. Some websites also provide a survey or evaluation button so the user can enter comments about the website.

In addition to publishing contact information and providing feedback entries, it is useful to provide information that can help the users locate databases/maps/charts easily. Some of the websites reviewed lists the most requested data, provide a summary page of all the databases, or provide related topics/data links for a particular database. Some websites also provide information on other people who are interested in a database, links to other relevant webpages, and links to external webpages that the users might be also interested in. The Washington State DOT (WSDOT) website contains a survey button asking the users enter their level of satisfaction about their website. Clicking on the button will take the users to a blog page, where they can write down their opinions or simply rate the website.

Appendix 2. Tabular Recommendations for NYMTC Data Products

The cells in Tale A2 given below are shaded whenever a recommendation applies to a specific data product (i.e., if a cell corresponding to a data product is shaded, then it means the recommendation in the corresponding column should be applied, if not then it means the data product can be considered acceptable with respect to that particular recommendation.) In addition, explanatory remarks are included in parentheses in some cases, no matter if the cell is shaded or not.

Table A2. Recommendations for the NYMTC Data Products

No	Data Product			Suggested Solutions						
	Name	Frequency	Smallest Geographic Level	Expand Data	Provide Latest Data	Provide Larger Data Archives	Provide Data Continuity	Decrease Time Lags	Ensure Accuracy	Integrate Data and Links to Data
1	County Profiles	Every decade	County	Profiles for counties in entire NYMTC region requested by stakeholders. Already available at Data Depot, should be moved from there to current page. Long./lat. data.	Most recent data year is 2000. Since published every 10 years, next up is 2010.	The oldest data is from 1990		3, 4-year lags currently	On the cover it should be indicated that the brochure is published every 10 years.	County profiles for entire region exist under Data Depot. They also exist under Census on Data and Modal page. All should be combined.
2	Hub Bound Travel at a Glance Brochure	Annual	Facility	Time/day of travel, detailed veh. class. requested by stakeholders. Already exists in Veh. Class. and Occupancy reports, so should be indicated in metadata file. Long./lat. data.	Most recent data year is 2007. Provide 2008.	The oldest data is from 1996	1998 data should be provided	1,2,3-year lags currently		
3	Hub Bound Travel Report	Annual	Facility	Time/day of travel, detailed veh. class. requested by stakeholders. Already exists in Veh. Class. and Occupancy reports, so should be indicated in metadata file. Long./lat. data.	Most recent data year is 2007. Provide 2008.	The oldest data is from 1996. Hardcopy versions from NYMTC Library cover a much wider time span, so they should be digitized.		1, 2, 3, 4-year lags currently		Hardcopy available at NYMTC Library for years 1924-1961, 1963, 1971, 1973-2001. This should be indicated in metadata.

Table A2. Recommendations for the NYMTC Data Products (Cont'd)

No	Data Product			Suggested Solutions						
	Name	Frequency	Smallest Geographic Level	Expand Data	Provide Latest Data	Provide Larger Data Archives	Provide Data Continuity	Decrease Time Lags	Ensure Accuracy	Integrate Data and Links to Data
4	Reg. Demo. at a Glance Brochure	Annual	County	Long./lat. data	Most recent data year is 2006. Provide 2007-2008.	The oldest data is from 1995	1997-1998 data should be provided	1, 2, 3-year lags currently		
5	Reg. Transp. at a Glance Brochure	Annual	Facility	Time/day of travel, detailed veh. class. requested by stakeholders. Already exists in Veh. Class. and Occupancy reports, so should be indicated in metadata file. Long./lat. data.	Most recent data year is 2006. Provide 2007-2008.	The oldest data is from 1995	1997-1999 data should be provided	1, 2, 3-year lags currently		
6	Reg. Transp. Stat. Report	Annual	Facility	Time/day of travel, detailed veh. class. requested by stakeholders. Already exists in Veh. Class. and Occupancy reports, so should be indicated in metadata file. Long./lat. data.	Most recent data year is 2006. Provide 2007-2008.	The oldest data is from years 2000-2001. Hardcopy versions from NYMTC Library cover a much wider time span, so they should be digitized.		2, 3, 5, 7-year lags currently		Hardcopy available at NYMTC Library for years 1995-2005. There are also older ones from 1986-94 period and from 1970s with a different name "Regional Transportation Status Report". This should be indicated in metadata.
7	Transp. Safety at a Glance Brochure	Annual	County	Long./lat. data	Most recent data year is 2006. Provide 2007-2008.	The oldest data is from 2002		2, 3-year lags currently		

Table A2. Recommendations for the NYMTC Data Products (Cont'd)

No	Data Product			Suggested Solutions						
	Name	Frequency	Smallest Geographic Level	Expand Data	Provide Latest Data	Provide Larger Data Archives	Provide Data Continuity	Decrease Time Lags	Ensure Accuracy	Integrate Data and Links to Data
8	Transp. Safety Report	Annual	County	Long./lat.data to facilitate map visualization	Most recent data year is 2005. Provide 2006-2008.	The oldest data is from 1985	1999 data should be provided	2, 3, 4-year lags currently		Hardcopy available at NYMTC Library for years 1997-1998, 2000-2004. This should be indicated in metadata.
9	Trav.Patterns in NY Metro Area Brochure	Quarterly	Facility	Time/day of travel, detailed veh. class. requested by stakeholders. Already exists in Veh. Class. and Occupancy reports, so should be indicated in metadata file. Long./lat. data.	Most recent data year is 2008 3rd and 4th Quarters	The oldest data is from 2002		0-5-year lags currently		Hardcopy available at NYMTC Library for 2002. This should be indicated in metadata.
10	Truck Toll Vol. Trends at a Glance Brochure	Annual	Facility	Time/day of travel requested by stakeholders. Already exists in Veh. Class. and Occupancy reports, so should be indicated in metadata file. Long./lat. data.	Most recent data year is 2007. Provide 2008.	The oldest data is from years 2000-2001		No lags or only 1-year lags		
11	Truck Toll Vol. Report	Annual	Facility	Time/day of travel requested by stakeholders. Already exists in Veh. Class. and Occupancy reports, so should be indicated in metadata file. Long./lat. data.	Most recent data year is 2007. Provide 2008.	The oldest data is from years 1998-1999		0, 1, 2-year lags currently		Hardcopy available at NYMTC Library for years 1997-2004 and also from 1979. This should be indicated in metadata.

Table A2. Recommendations for the NYMTC Data Products (Cont'd)

No	Data Product			Suggested Solutions							
	Name	Frequency	Smallest Geographic Level	Provide Better Data Access and Retrieval Options							
				Allow Automatic Web Data Generation and Extraction	Provide Metadata	Include Variable Dictionary (Glossary)	Provide Interactive Charts/Graphs Customization	Allow Statistical Analysis for Chosen Data	Provide Multiple Formats	Make All Data Print/Download Friendly	Provide User Support and Feedback Tools
1	County Profiles	Every decade	County	At Data Depot interactive tables provided. Should be moved to current page and improved.	Should be created		At Data Depot interactive charts provided. Should be moved to current page and improved.		HTML display in Data Depot and at Census. PDF at current page. More variety needed.		
2	Hub Bound Travel at a Glance Brochure	Annual	Facility		Should be created						
3	Hub Bound Travel Report	Annual	Facility		Available at NYMTC Library page. Should be hyperlinked there from data page. Contents need to be expanded				XLS format available for 2007, in addition to PDF		

Table A2. Recommendations for the NYMTC Data Products (Cont'd)

No	Data Product			Suggested Solutions							
	Name	Frequency	Smallest Geographic Level	Allow Automatic Web Data Generation and Extraction	Provide Metadata	Include Variable Dictionary (Glossary)	Provide Interactive Charts/Graphs Customization	Allow Statistical Analysis for Chosen Data	Provide Multiple Formats	Make All Data Print/Download Friendly	Provide User Support and Feedback Tools
4	Reg. Demo. at a Glance Brochure	Annual	County		Should be created						
5	Reg. Transp. at a Glance Brochure	Annual	Facility		Should be created						
6	Reg. Transp. Stat. Report	Annual	Facility		Available at NYMTC Library page. Should be hyperlinked there from data page. Contents need to be expanded						
7	Transp. Safety at a Glance Brochure	Annual	County		Should be created						

Table A2. Recommendations for the NYMTC Data Products (Cont'd)

No	Data Product			Suggested Solutions							
	Name	Frequency	Smallest Geographic Level	Provide Better Data Access and Retrieval Options							
				Allow Automatic Web Data Generation and Extraction	Provide Metadata	Include Variable Dictionary (Glossary)	Provide Interactive Charts/Graphs Customization	Allow Statistical Analysis for Chosen Data	Provide Multiple Formats	Make All Data Print/Download Friendly	Provide User Support and Feedback Tools
8	Transp. Safety Report	Annual	County		Available at NYMTC Library page. Should be hyperlinked there from data page. Contents need to be expanded						
9	Trav. Patterns in NY Metro Area Brochure	Quarterly	Facility		Available at NYMTC Library page. Should be hyperlinked there from data page. Contents need to be expanded				XLS format available for 2004-2008, in addition to PDF		
10	Truck Toll Vol. Trends at a Glance Brochure	Annual	Facility		Should be created						
11	Truck Toll Vol. Report	Annual	Facility		Available at NYMTC Library page. Should be hyperlinked there from data page. Contents need to be expanded				XLS format available for 2007, in addition to PDF		

Appendix 3. Screenshots from Best Practices

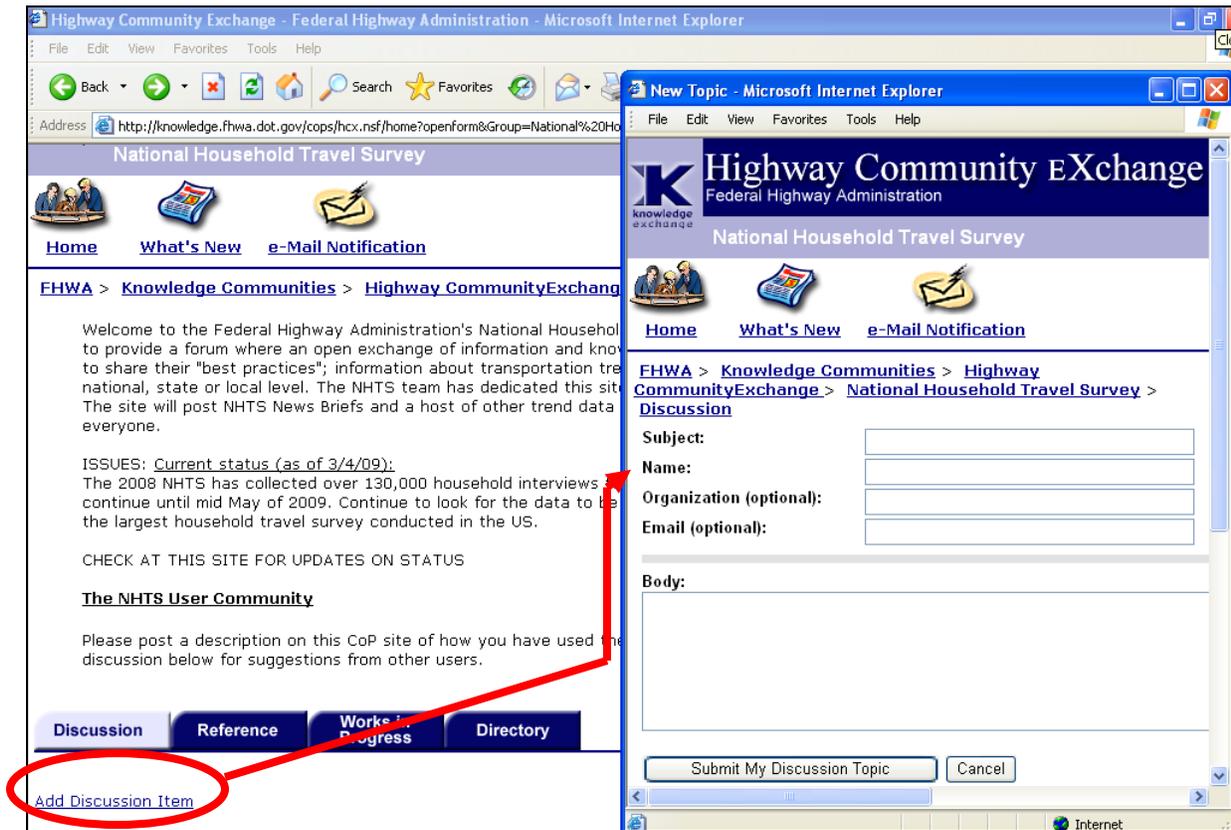


Figure A1. FHWA’s NHTS Website – User Forum

<http://knowledge.fhwa.dot.gov/cops/hcx.nsf/home?openform&Group=National%20Household%20Travel%20Survey>

Customize Table

(Includes U.S. Air Carrier Passenger and Cargo Services)

Geographic Area: Domestic International System (domestic and international)

Schedule Type: Scheduled Non-Scheduled Total

Service Class: Passenger Cargo

Operating Statistics:

<input type="radio"/> Revenue Passenger Enplanements	<input type="radio"/> Revenue Passenger Miles	<input type="radio"/> Available Seat Miles
<input type="radio"/> Revenue Freight Ton Miles	<input type="radio"/> Total Revenue Ton Miles	<input type="radio"/> Available Ton Miles
<input type="radio"/> Revenue Departures Performed	<input type="radio"/> Revenue Aircraft Miles Flown	<input type="radio"/> Revenue Aircraft Hours (Airborne)

From: Jan 1996 To: Nov 2008

Figure A2. BTS Website – A Customizable Table with Radio Buttons (For US Air Carrier Traffic Statistics) http://www.bts.gov/xml/air_traffic/src/index.xml#CustomizeTable

<p>US State</p> <ul style="list-style-type: none"> Aggregate all US States All US States Detail Alabama (AL) Alaska (AK) Arkansas (AR) Arizona (AZ) California (CA) Colorado (CO) 	<p>Trading Partner</p> <ul style="list-style-type: none"> NAFTA (Canada and Mexico) Canada Mexico All Canada Provinces Detail All Mexico States Detail Alberta (CAN) British Columbia (CAN) Manitoba (CAN) 	<p>Year</p> <ul style="list-style-type: none"> All Years Detail 2008 2007 2006 2005 2004 2003 2002 	<p>Month</p> <ul style="list-style-type: none"> Annual Summary All 12 Months Detail January February March April May June
<p>Customs Port</p> <ul style="list-style-type: none"> Aggregate all Ports Ports by Name Land Ports Air ports Sea Ports Ports by Code Northern Border Ports Southern Border Ports Ports by State Order Alaska (AK) Alabama (AL) Arkansas (AR) 	<p>Mode</p> <ul style="list-style-type: none"> Aggregate all Modes Aggregate all Land Modes --Mail --Truck --Rail --Pipeline --Other --Foreign Trade Zones Aggregate all AV Modes --Air --Vessel 	<p>Measure</p> <ul style="list-style-type: none"> Value (in US Dollars) Weight (in Pounds) Weight (in Kilograms) Weight (in US Short Tons) Weight (in Metric Tons) <p>Trade Type</p> <ul style="list-style-type: none"> Total Trade --Exports --Imports 	<p>Data Display</p> <ul style="list-style-type: none"> Actual In Thousands In Millions In Billions <p>Output Option</p> <ul style="list-style-type: none"> Table w/o % Change Table w/ % Change <p>Other Options</p> <ul style="list-style-type: none"> Suppress Zero and NA Show Commodity Desc
<p>Port Name, Code and Type</p> <ul style="list-style-type: none"> --No Selection Here-- 		<p>Commodity</p> <ul style="list-style-type: none"> Aggregate all Commodities All Commodities Detail 01----Live animals 02----Meat and edible meat offal 03----Fish and crustaceans, mollusks and other aquatic invertebrate 04----Dairy produce; Birds eggs; Natural honey; Edible products of e 05----Products of animal origin, not elsewhere specified or included 06----Live trees and other plants; Bulbs, roots and the like; Cut flow 07----Edible vegetables and certain roots and tubers 08----Edible fruit and nuts; Peel of citrus fruit or melons 	
<p>Submit Reset</p>			

Figure A3. BTS Website – A Customizable Table with Drop-Down Lists (For North American Transborder Freight Data)

(http://www.bts.gov/programs/international/transborder/TBDR_QA.html#)

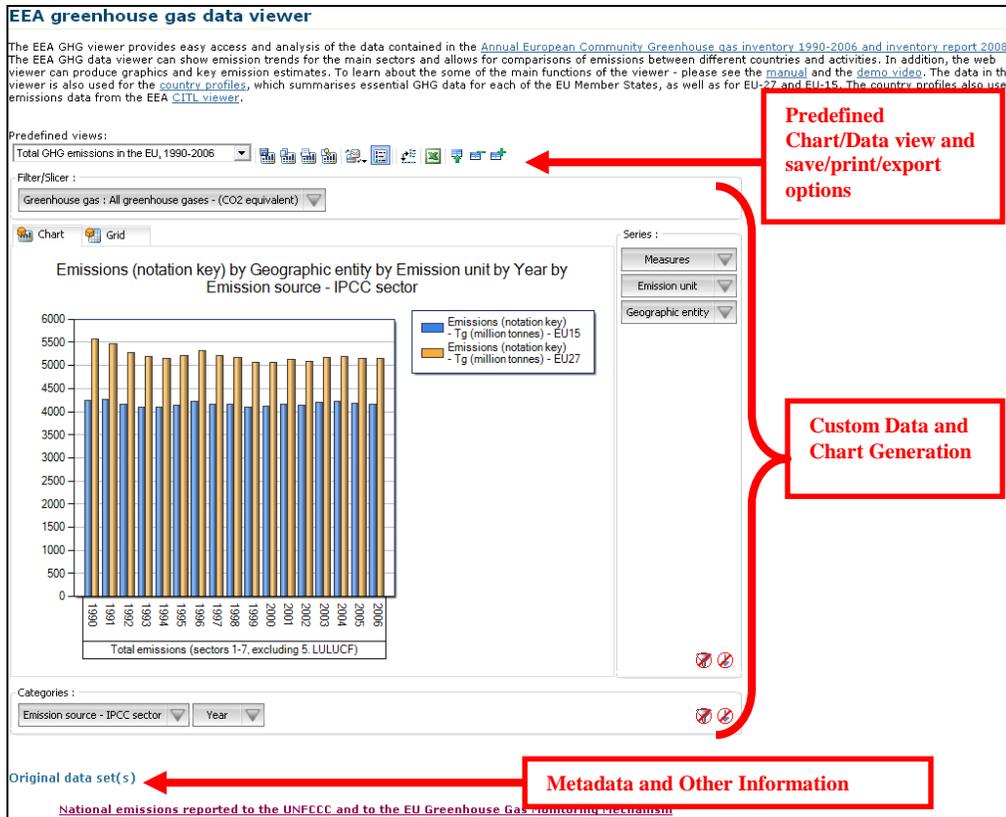


Figure A4. EEA Website – Data Viewer with Predefined or Custom Tables/Charts and Metadata (<http://dataservice.eea.europa.eu/PivotApp/pivot.aspx?pivotid=455>)

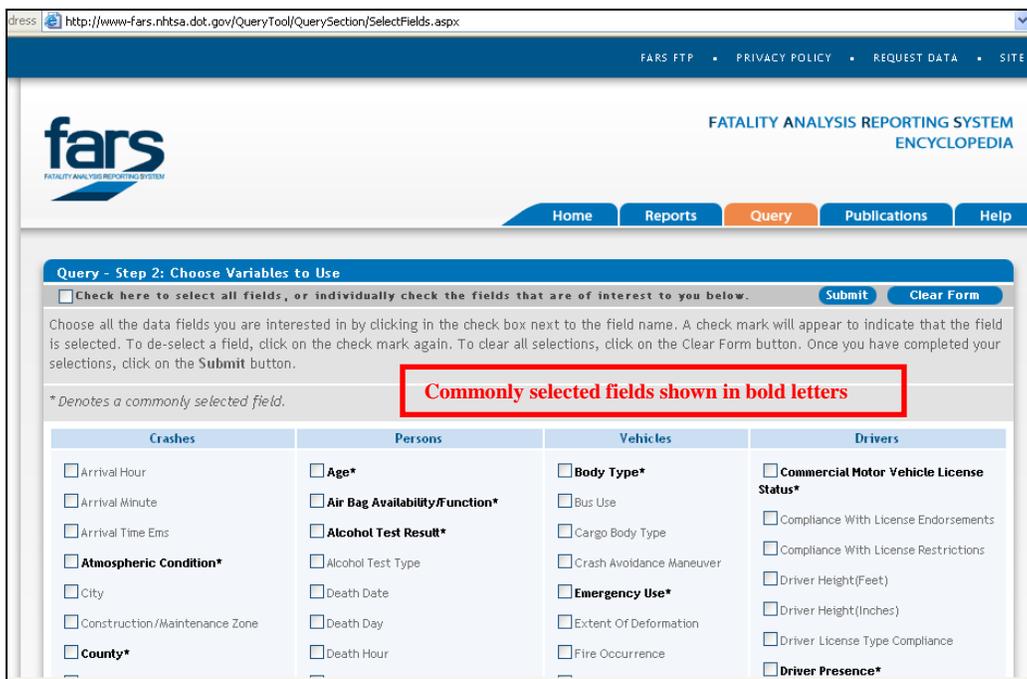


Figure A5. FARS Website - Query Pages with Commonly Selected Fields Indicated (<http://www.fars.nhtsa.dot.gov/QueryTool/QuerySection/SelectFields.aspx>)

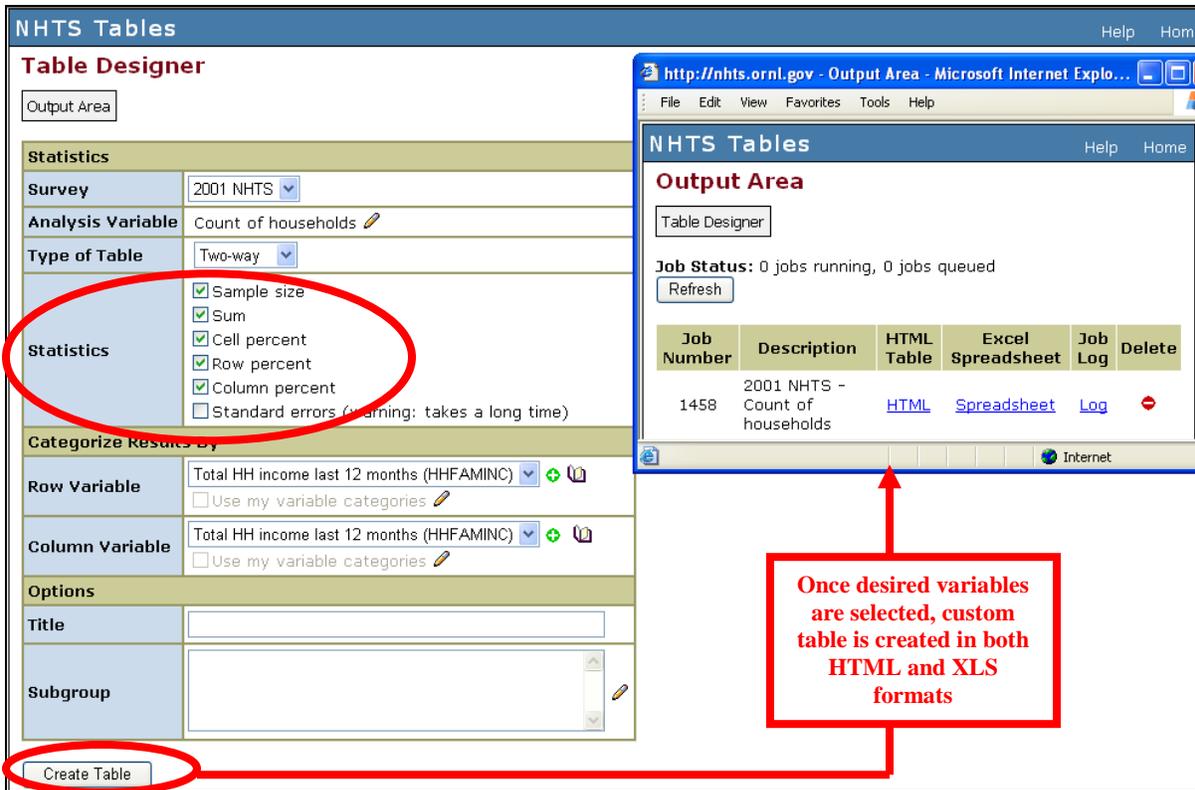


Figure A6. FHWA's NHTS Website – Table Designer Feature with Statistical Analysis and Different Output Options (<http://nhts.ornl.gov/tables/ae/TableDesigner.aspx>)

Table 1: Trade between USA and NAFTA (Monthly)

[Back to top](#)
[Suppress zero and NA values](#)
[Rotate table\(s\)](#)
[Go to download or print](#)

Value (in Actual of US Dollars) Click any column header to sort

Trade Type and Mode	Trader	Partner	Commodity Code	01/2006	02/2006	03/2006	04/2006	05/2006	06/2006	07/2006	08/2006	09/2006
Total Trade Value by All Modes	USA	NAFTA	01	213,220,421	220,886,861	267,820,022	199,732,594	150,823,668	145,010,259	134,581,426	169,495,950	200,524,000
Total Trade Value by All Modes	USA	NAFTA	02	382,669,152	338,334,872	386,476,781	344,126,719	404,204,667	397,539,095	381,801,660	394,956,249	389,966,000
Total Trade Value by All Modes	USA	NAFTA	03	174,576,681	156,420,361	188,736,276	200,928,412	334,124,087	270,574,556	273,005,032	283,085,668	261,636,000
Total Trade Value by All Modes	USA	NAFTA	04	56,046,066	66,239,971	56,461,367	50,076,302	55,107,047	60,619,927	63,159,679	65,129,317	62,777,000
Total Trade Value by All Modes	USA	NAFTA	05	43,378,067	34,434,927	40,749,747	34,696,340	42,432,175	46,077,326	45,664,990	47,737,439	50,256,000

Figure A7. BTS Website – A Customized Table with Data Sorting Option

Twelve Months - System

(Excludes all-cargo services. Includes domestic and international)

[Excel](#) | [CSV](#)

	December 2007 - November 2008			December 2006 - November 2007		
	Scheduled	Non-Scheduled	Total	Scheduled	Non-Scheduled	Total
Revenue Passenger Enplanements (000)	744,907	5,977	750,884	769,520	6,281	775,801
Revenue Passenger Miles (000)	815,473,989	11,599,278	827,073,267	827,814,272	12,385,831	840,200,103
Available Seat Miles (000)	1,027,583,453	19,747,988	1,047,331,441	1,034,922,210	22,012,381	1,056,934,591
Passenger Load Factor (%)	79.36	58.74	78.97	79.99	56.27	79.49
Revenue Freight Ton Miles (000)	6,684,980	100,070	6,785,050	6,810,087	88,016	6,898,103
Total Revenue Ton Miles (000)	89,147,141	1,260,034	90,407,175	90,418,855	1,326,609	91,745,464
Available Ton Miles (000)	145,717,902	3,128,089	148,845,991	149,170,402	3,328,719	152,499,121
Ton Mile Load Factor (%)	61.18	40.28	60.74	60.61	39.85	60.16
Revenue Departures Performed	10,258,992	143,001	10,401,993	10,699,013	156,479	10,855,492
Revenue Aircraft Miles Flown (000)	7,390,953	97,001	7,487,954	7,535,265	107,581	7,642,846
Revenue Aircraft Hours (Airborne)	17,725,957	253,060	17,979,017	18,079,271	275,612	18,354,883

SOURCE: Bureau of Transportation Statistics, T-100 Market and Segment

Figure A8. BTS Website – A Data Table in HTML with Download Option in XLS or CSV Formats

Query - Step 1: Choose a Year

Select a Query Year: 2007

To begin building a query, select a year from the drop-down list above, and click the 'Submit' button.

The Fatality Analysis Reporting System (FARS) contains data on all vehicle crashes in the United States that occur on a public roadway and involve a fatality. This FARS Query System provides interactive public access to fatality data through this web interface. Due to the complexities within the FARS data, users cannot query across multiple years. If you are interested in trend or cross year information, check out the Trend Reports.

The exercises and final report links listed below point to PDF documents, and require the use of the free Adobe Reader. Links will open in a new browser window or tab.

👉 If you would like additional training on using this query system, try some of the exercises below.

-  [Univariate Report Exercises](#)
-  [Cross Tab Report Exercises](#)
-  [Case Listing Report Exercises](#)

👉 If you would like to see the final reports, click links below.

-  [Univariate Tabulation Reports](#)
-  [Cross Tabulation Reports](#)
-  [Case Listing Reports](#)

Figure A9. FARS Website – Data Customization Through Simple Query with Query Exercises for Novice Users (<http://www-fars.nhtsa.dot.gov/QueryTool/QuerySection/SelectYear.aspx>)

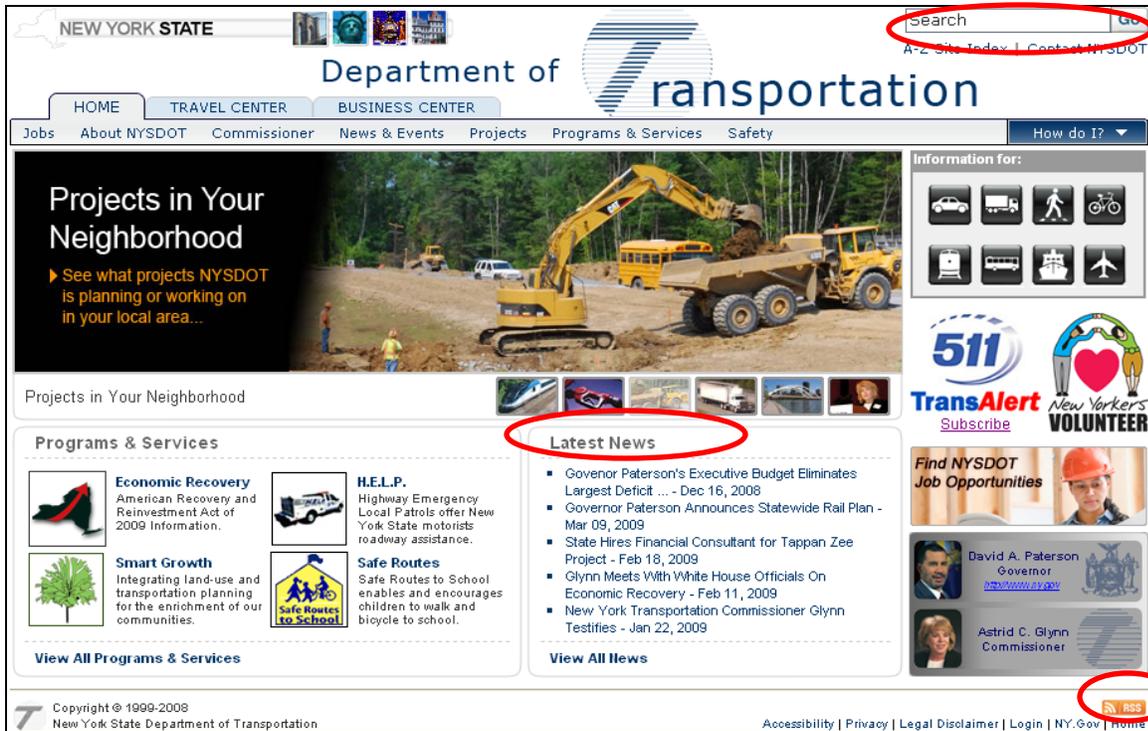


Figure A10. NYSDOT Website - Home Page (<https://www.nysdot.gov/index>)

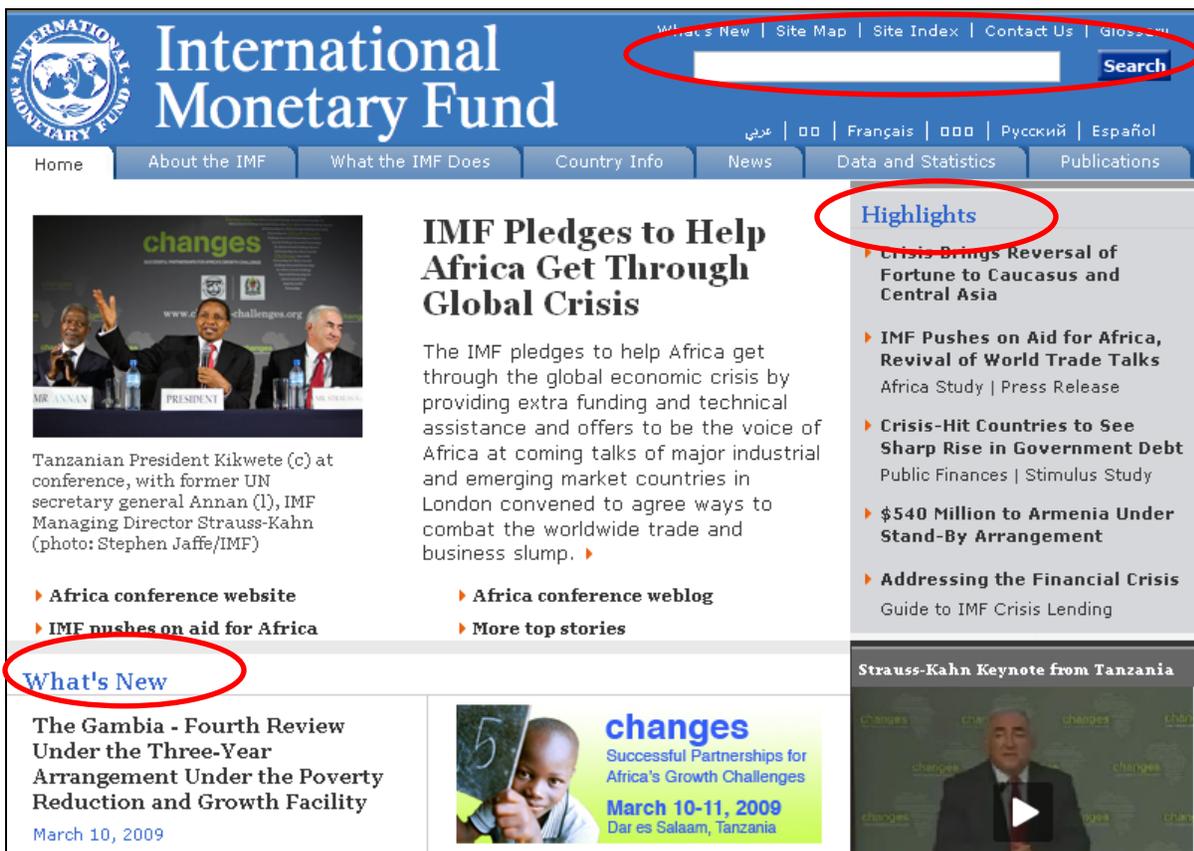


Figure A11. IMF Website - Home Page (<http://www.imf.org/external/index.htm>)

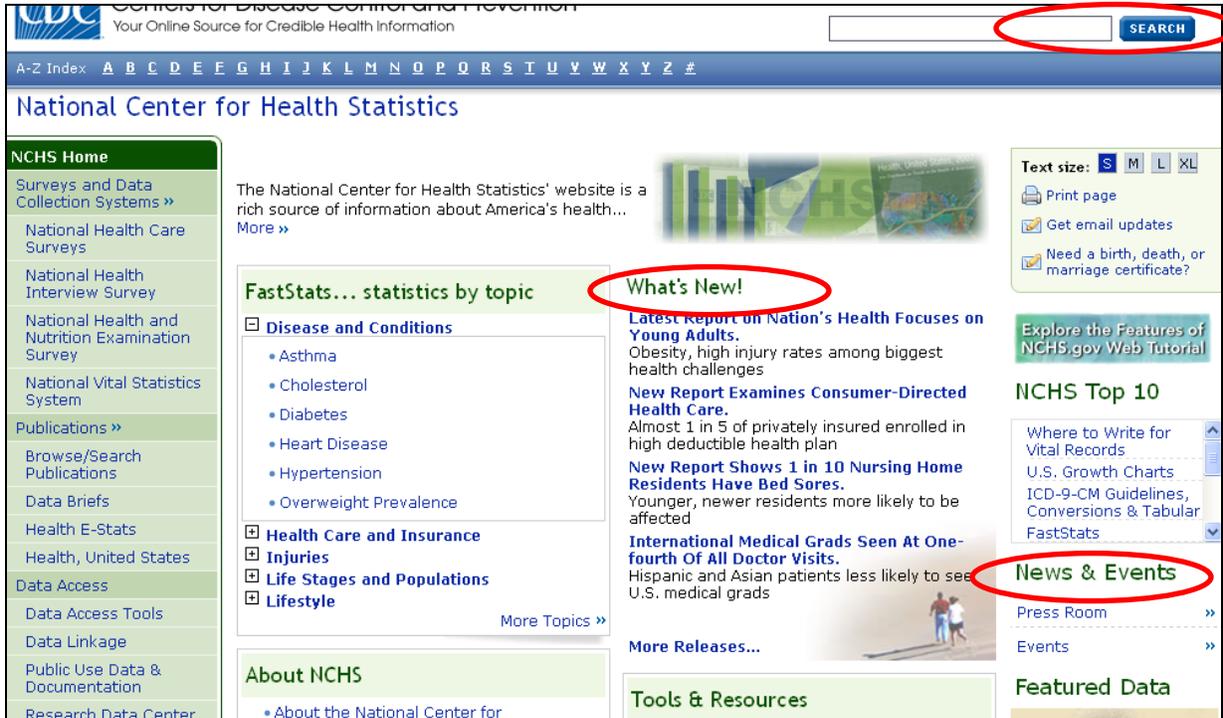


Figure A12. NCHS Website - Home Page (<http://www.imf.org/external/index.htm>)



Figure A13. NJTPA Website - Helpful Links and Tools (<http://njtpa.org/>)

How to subscribe (be added) to the RDC List

- Click on the envelope on the right and in the **body** of the e-mail message (leave the subject line blank), type or paste: **subscribe Research_Data_Center (your name)**
- If your browser does not invoke an e-mail form send this message to: listserv@cdc.gov and in the **body** of an e-mail message (leave the subject line blank), type or paste: **subscribe Research_Data_Center (your name)**

Example subscription to the RDC List

How to unsubscribe (be removed) from the RDC List

- Click on the envelope on the right and in the **body** of the e-mail message (leave the subject line blank), type or paste: **signoff Research_Data_Center (your name)**
- If your browser does not invoke an e-mail form send this message to: listserv@cdc.gov and in the body of an e-mail message (leave the subject line blank), type or paste: **signoff Research_Data_Center (your name)**
- If your browser does not invoke an e-mail form send this message to: listserv@cdc.gov and in the **body** of an e-mail message (leave the subject line blank), type or paste: **signoff Research_Data_Center (your name)**

Figure A14. NCHS Website – Subscription to ListServ (Email List)
<http://www.cdc.gov/nchs/r&d/rdclist.htm>

World Health Organization
Regional Office for Europe

Country information

Data and publications > Publications

Electronic mailing list

To sign up, please send a blank e-mail to: join-whopublications@pr4.netatlantic.com

Data and publications

Figure A15. WHO-ROE Website – Subscription to Email List
http://www.euro.who.int/InformationSources/Publications/20080118_1

» Home » EEA subscriptions » News feeds subscription (RSS)

News feeds subscription (RSS)

Subscribe to EEA's RSS feeds in order to get announcements delivered directly to your desktop or mobile device!

Main EEA RSS feed – all environmental news and products

- <http://www.eea.europa.eu/highlights.rdf>, subscribe by:
 - E-mail
 - RSS 2 PDF
 - Subscribe

sign up now!
Get notifications on new reports and products.
Enter email

RSS Feeds » Archive »

Figure A16. EEA Website – Subscription to Email List and RSS Feeds
<http://www.eea.europa.eu/subscription/news-feeds>

Figure A17. US Census Bureau Website – Help Page (<http://www.census.gov/main/www/help.html>)

Figure A18. CDOT Website – Contact and Feedback Page (<http://www.dot.state.co.us/Contacts/form.htm>)

The screenshot shows the main page of the FHWA LTPP website. On the left is a navigation menu with items like 'Feedback Reports' and 'LTPP Customer Survey'. The main content area includes links to a forum, contact information for LTPP Customer Support Service, technical services contractor, regional links, newsletter, and a list of publications. On the right, there are sections for 'Events', 'Contacts', and a 'We Want Your Feedback' form. The feedback form is highlighted with a red border and contains two questions with radio button options and text input fields.

Figure A19. FHWA’s LTPP Website – Feedback Form on the Main Page
[\(http://www.fhwa.dot.gov/pavement/ltppl/\)](http://www.fhwa.dot.gov/pavement/ltppl/)

The screenshot shows the 'Customer Satisfaction Questionnaire' page on the FHWA LTPP website. The top navigation bar includes 'Research', 'Design', 'Construction', 'Preservation', 'Maintenance', 'Management', and 'Rehabilitation'. The breadcrumb trail is 'FHWA > Engineering > Pavements > Research > LTPP'. The main content area is titled 'FHWA LTPP Customer Support Service Customer Satisfaction Questionnaire' and includes an introduction, a note about confidentiality, and a list of five statements to be rated from 1 to 5. A sidebar on the left contains a menu with 'LTPP Customer Survey' circled in red. On the right, there are sections for 'Features', 'More Information', 'Events', and 'Contact'.

Figure A20. FHWA’s LTPP Website – Customer Satisfaction Survey
<http://www.fhwa.dot.gov/pavement/ltppl/survey2.cfm>

American FactFinder Feedback

Enter your feedback and click 'OK'.

- If this feedback is about a problem, please include a detailed description of the steps that led to the error. (what you were trying to do; the last two steps you took before encountering the error...)
- If your problem involves a slow response, also include the type of network connection you have (telephone modem, cable modem, DSL, LAN...)

Feedback:

Is this a Comment Problem

Name (optional):

E-mail (optional):

Phone (optional):

If you would like a response, please provide your name, e-mail address, and/or phone number. Name, phone and e-mail information will only be used to contact you regarding your feedback and not for any other purpose.

Search for answers to frequently asked questions by clicking the [FAQ](#) link displayed at the top of every page in American FactFinder.

Figure A21. US Census Bureau Website – American Fact Finder- Feedback Web Form

April 2008 World Economic Outlook (WEO) Feedback

1. Which aspect of the WEO would you like to comment on?

Text
 Data
 Text and Data

2. Do you find the WEO Text useful?

3. Do you find the WEO Data useful?

4. How would you characterize the WEO on the IMF website as an information resource for your work?

5. What is your area of employment or affiliation?

6. Where do you live?

If you would like to add any additional comments, please do so in the space below.

Thank you for taking the time to complete this survey. Select Submit Survey now to send your response to us.

Figure A22. IMF Website – A More Detailed User Feedback Tool
<http://www.imf.org/external/pubs/ft/weo/2008/02/surv/index.asp>

The screenshot shows the IMF Publications website interface. At the top, there is a 'Browse by' section with links for 'Title', 'Author', and 'Subject'. Below this is a 'Search by' section with three text input fields: 'Title:', 'Author/Editor:', and 'Subject / Keyword:'. The 'Author/Editor' field has a note below it: '(last name or last name, first name)'. There is also a 'Series' dropdown menu with options: 'All', 'Economic Issues', and 'IMF Country Reports'. Below the search fields are three more dropdown menus: 'Date' (with options 'During' and 'Year'), 'Language' (with options 'All', 'English', and 'French'), and 'Sort by' (with option 'Date'). At the bottom of the search section are 'Search' and 'Clear' buttons.

Figure A23. IMF Website with Both Browse and Search Options

<http://www.imf.org/external/pubind.htm>

The screenshot shows the NJDOT website's 'Research' section. The URL in the browser is 'http://www.state.nj.us/transportation/refdata/research/ReportsDB.shtm'. The page features the NJDOT logo and navigation links. A left sidebar contains a menu with items like 'Reference Data', 'Research Overview', 'News and Events', 'Research Library', 'Ongoing Research', 'Final Reports', 'Guidelines, Formats and Forms', 'Research Partners', 'Useful Links', and 'Contact Us'. The main content area is titled 'Research' and 'Final Reports'. It contains a paragraph: 'NJDOT Bureau of Research maintains a database of completed research projects. The database is periodically populated with additional reports as they become available.' Below this is a search form with several dropdown menus: '- Select keyword -', '- Select Year -', '- Select School -', '- Select Organization -', '- Select Manager -', '- Select Principal Investigator -', '- Select Report Number -', and '- Sort By -'. A 'SUBMIT' button is located at the bottom of the search form.

Figure A24. NJDOT Website- Research Report Search Boxes

<http://www.nj.gov/transportation/refdata/research/ReportsDB.shtm>

DfT home [Advanced search](#)

Advanced search

[\[Show help \]](#) Print 

Enter search term

 Include similarly spelled words

Results must contain

All words in search term

Any words in search term

The exact search term

Publication date

From date 

To date 

Mode of transport
[\[Hide options \]](#)

Air Rail Road Sea and water

Subject (e.g. Safety)
[\[Show options \]](#)

Audience (e.g. Parents)
[\[Show options \]](#)

Region
[\[Show options \]](#)

Content type (e.g. Annual report)
[\[Show options \]](#)

Figure A25. DfT Website- Advanced Search Page (<http://www.dft.gov.uk/advsearch>)

Tables & Figures Search

Enter a word or phrase to search for tables/figures.

Refine your search:

Select year(s) of data.

2007	▲
2006	▲
2005	▲
2004	▼

Select a topic area.

Usage Tips

This search tool lets you locate all tables/figures/charts published in the inventory of NCES' National Education Data Resource Center (NEDRC) [Postsecondary Tables Library](#); the [Condition of Education](#); the [Digest of Education Statistics](#); [Indicators of School Crime and Safety](#) and other NCES publications. Tables are constantly being added (**thousands of tables, graphs & figures are now available**). State tables can be created by visiting [State Education Data Profiles](#).

- Enter a word or phrase to search for tables/figures. Searches are conducted on the actual textual information provided in the tables and figures.
- Further refine your search by selecting a topic area and the year(s) you are interested in. By holding down the control key you can search for multiple years. After finalizing your criteria, click on the Search Now button for your results.

Popular Keywords are also available to help with your search. By selecting a phrase you will get results from all the topic areas and years for the selected term.

The Results Page will list your selection criteria, a count of matches, and descriptive information for each table/figure matching your criteria including: Title, Topic Area, Survey Source, and Year of Data. View the Figure/Table by clicking on the "View Table/Figure" link. The results are presented in alphabetical order but can be resorted.

Figure A26. NCES Website – Table/Figure Search Page (<http://nces.ed.gov/quicktables/index.asp>)

The search returned **5050** matches:

Figure/Table Title	Topic Area	Source	Year
1. Percentage of private schools with teaching vacancies in this school year with one or more vacancies in various fields, by private school typology and NAIS membership: 1999–2000 (View Table/Figure)	Elementary/Secondary	Schools and Staffing Survey (SASS)	2000
2. Percentage of public schools with a formal school improvement plan and percentage of those schools that used various methods to assess the improvement plan, by state: 1999–2000 (View Table/Figure)	Elementary/Secondary	Schools and Staffing Survey (SASS)	2000
3. Students transported at public expense and current expenditures for transportation: Selected years, 1929-30 through 2004-05 (View Table/Figure)	Resources/Other	Digest of Education Statistics	2005
4. Percentage of private schools with teaching vacancies in this school year and percentage of those schools that used various ways of filling or eliminating the vacancy, by school level and selected school characteristics: 1999–2000 (View Table/Figure)	Elementary/Secondary	Schools and Staffing Survey (SASS)	2000

Figure A27. NCES Website – Table/Figure Search Results Page