SCOPING STUDY FOR A FREIGHT DATA EXCHANGE NETWORK

Prepared for:
American Association of State Highway and Transportation Officials (AASHTO)
Standing Committee on Planning

Prepared by:
Dr. Michael S. Bronzini
George Mason University
Fairfax, Virginia
and
Sashank Singuluri
David B. Roden
AECOM Consult, Inc.
3101 Wilson Blvd, Suite 400
Arlington, VA 22201

June, 2009

The information contained in this report was prepared as part of NCHRP Project 8-36, Task 79, National Cooperative Highway Research Program, Transportation Research Board.
Acknowledgements

This study was conducted for the American Association of State Highway and Transportation Officials (AASHTO), with funding provided through the National Cooperative Highway Research Program (NCHRP) Project 8-36. The NCHRP is supported by annual voluntary contributions from the state Departments of Transportation. Project 8-36 is intended to fund quick response studies on behalf of the AASHTO Standing Committee on Planning. The report was prepared by Dr. Michael S. Bronzini of George Mason University and Sashank Singuluri and David B. Roden of AECOM Consult, Inc. The work was guided by a task group chaired by Mary Lynn Tischer which included Roy Cornelius, Patricia Oliver-Wright, and Jienki Synn.

Disclaimer

The opinions and conclusions expressed or implied are those of the research agency that performed the research and are not necessarily those of the Transportation Research Board or its sponsoring agencies. This report has not been reviewed or accepted by the Transportation Research Board Executive Committee or the Governing Board of the National Research Council.
Table of Contents

1 Introduction .................................................................................................................................................. 1

2 Data Sharing Technology ............................................................................................................................. 1
  2.1 Freight Data Needs for Transportation Planning .................................................................................... 2
  2.2 Existing Freight Data Sources .................................................................................................................. 2
  2.3 Data Processing and Maintenance Needs ............................................................................................... 3
  2.4 Database Systems and Data Warehouses ............................................................................................... 4
  2.5 Data Portal .............................................................................................................................................. 5
    2.5.1 Data Providers ................................................................................................................................. 5
    2.5.2 Data Security ............................................................................................................................... 6
    2.5.3 Data Interchange Format .............................................................................................................. 6
    2.5.4 Data Size and Transfer Speed ..................................................................................................... 6

3 Views of Potential Participants: Data Users and Providers ...................................................................... 7
  3.1 Data User Interview Results ................................................................................................................... 7
    3.1.1 Experience in Using Freight Data ............................................................................................... 7
    3.1.2 Interest in the Proposed Freight Data Exchange Network ........................................................... 8
    3.1.3 Requested Features for the Freight Data Exchange Network ....................................................... 8
    3.1.4 User Fee Issues .......................................................................................................................... 8
    3.1.5 Obstacles to System Use .............................................................................................................. 9
  3.2 Data Provider Interview Results .......................................................................................................... 9
    3.2.1 Experience in Providing Freight Data .......................................................................................... 9
    3.2.2 Potential for Participation in the Freight Data Exchange Network ............................................. 9
    3.2.3 Obstacles to Participation .......................................................................................................... 10
    3.2.4 Overcoming Obstacles ................................................................................................................. 11

4 Views of Potential System Operators .......................................................................................................... 11
  4.1 System Implementation .......................................................................................................................... 11
  4.2 Costs and User Fees .............................................................................................................................. 12
  4.3 Handling Restricted Data ......................................................................................................................... 12
  4.4 Obstacles to System Implementation .................................................................................................... 12

5 Freight Data Exchange Concept .................................................................................................................. 13
  5.1 Metadata ................................................................................................................................................. 15
  5.2 Data Schema ......................................................................................................................................... 16
  5.3 Datasets and Potential use ..................................................................................................................... 16
    5.3.1 Freight Analysis Framework Data (FAF2) ..................................................................................... 16
    5.3.2 Commodity Flow Survey (CFS) ..................................................................................................... 17
    5.3.3 TranSearch Database .................................................................................................................... 18
    5.3.4 Highway Information Pertaining to Truck Vehicles ..................................................................... 20
    5.3.5 Local and Regional freight datasets ............................................................................................. 21
    5.3.6 Socio-economic Data from Regional Studies ........................................................................... 21
    5.3.7 Other data sources ....................................................................................................................... 22
    5.3.8 Data Translation .......................................................................................................................... 23

6 Summary and Conclusions .......................................................................................................................... 23
6.1 Design Concept Questions ................................................................. 23
6.1.1 Can freight data be shared efficiently and securely over the internet? ..................................................... 23
6.1.2 Are there sufficient markets of data users and providers? ........................................................................ 23
6.1.3 What are the incentives to participate for users and providers? ................................................................. 24
6.1.4 What are the concerns and potential impediments to participation for users and providers? .......... 24
6.1.5 What data would be desired and available? .................................................................................................. 24
6.1.6 How would metadata be made available in the network? .......................................................................... 24
6.1.7 How would the exchange network be implemented? .................................................................................. 24
6.1.8 What requirements and costs would be associated with establishing and maintaining the exchange network? ...................................................................................................................... 24
6.1.9 What would stewardship of the exchange network involve? Who might be willing and able to take this role? What are potential impediments to stewardship? ........................................................................... 25
6.2 Conclusions .................................................................................... 25
1 Introduction

Domestic intercity freight traffic has almost doubled in the last 30 years. Most of this growth has been experienced by the air and truck modes. In addition, the advent of internet-based shopping sites has increased the shipment of packages of all sizes and weight to a much larger set of pick-up and delivery locations. Individuals and businesses expect and demand more reliable and efficient freight services to support just-in-time delivery systems. In other words, an efficient freight transportation network is an essential component of the economic development and vitality of a region.

In July 2007, the Transportation Research Board conducted a workshop on Meeting Freight Data Challenges. An underlying theme of the workshop was reviewing and reemphasizing the many reasons that both the public and private sectors require good data in order to make sound, supportable decisions about freight transportation system investments, operations, and policies. One of the focal points of the workshop was identifying current and future sources of freight data. A white paper on this topic demonstrated that there is a wealth of data available covering many aspects of the universe of freight data. A second white paper reviewed possible business models and public-private partnerships that might be used to enable the sharing of freight data among its providers and users.

Through the workshop and white papers, it was observed that potential users of freight data have difficulty in identifying the data that are available as well as the quality and other important attributes that affect their ability to productively use the data at a reasonable cost. Conversely, data originators and providers lack systems to inform potential users about what data are available and how that data might be accessed. One suggested means of overcoming these limitations is to establish a cooperative data exchange network, modeled on the “wiki” concept, where freight data and analyses could be shared efficiently and securely over the internet.

This research investigated the feasibility of implementing a freight data exchange network to provide timely access to higher quality freight data. The exchange network is envisioned as a centralized data repository where data providers and user would have the ability to enter or access freight datasets, metadata, reports of quality or other data issues, and study reports documenting how data were used for specific applications. The system would increase both the use of data and statistics from all available sources and increase the demand for improvements in the completeness, quality, and timeliness of the data.

2 Data Sharing Technology

The first question that needs to be answered is can freight data be shared efficiently and securely over the internet? The simple answer from a technology perspective is yes. Internet technology and communications protocols have the ability to transfer encrypted data reliability and at relatively high speeds. This answer, of course, leads to another whole set of questions about:

- The types of data and the size of the datasets that are shared;
- The data processing services and tools that are required to maintain or use the data;
- The technical capabilities and skills of data users and data providers;
- The types and levels of data and access security that are required;
- The software and hardware required by network users;
- The level of effort and cost to develop the exchange network; and
The level of effort and cost to operate and host the exchange network.

The answers to questions such as these address the overall feasibility of the freight data exchange network from a cost-effectiveness and likelihood of success perspective.

2.1 Freight Data Needs for Transportation Planning

Though most Metropolitan Planning Organizations (MPOs) develop regional household demand models, very few develop freight demand models. Many MPOs would like to develop freight profiles and data driven frameworks to address regional freight mobility needs, but have difficulty acquiring the necessary data. A snapshot of a region’s freight demand is essential for developing plans that mitigate freight-related congestion and increase regional economic activity. The lack of access to disaggregate freight data to construct freight flow patterns is a major roadblock in the development of freight demand models. The creation of a freight data exchange network could provide MPOs with the data needed to analyze existing freight flow patterns and anticipate future problems as an integral part of regional planning and investment decisions.

The U.S. Environmental Protection Agency (EPA) is developing the Motor Vehicle Emission simulator (MOVES) to replace MOBILE6 for emissions analyses. A draft version of MOVES software is currently available and the final version is planned for late 2009. Once finalized, MPOs will need to evaluate pollutant and greenhouse gas (GHG) emissions using the MOVES software. This tool requires the MPO to input volumes and speeds by vehicle type and roadway functional classification to estimate emissions. It also encourages MPOs to use local distributions of vehicle types by age, fuel and power attributes as well as research on local emissions rates. Since trucks contribute substantially to regional emissions, it is important to estimate these emissions correctly to minimize regulatory constraints. A freight data exchange network could store truck emissions information and research that would help improve emissions estimates and refine emissions rates. This would help agencies assess their collected data, remove outliers and create more accurate emissions inventories.

2.2 Existing Freight Data Sources

This section provides general background information about several of the public and commercial freight data sources available for model development and planning studies.

*Commodity Flow Survey (CFS)* – The CFS database is part of the economic census. It is compiled every five years by the Bureau of Transportation Statistics (BTS). It contains commodity flow information by Standard Classification of Transported Goods (SCTG) code for most retail and all government shippers. Since only domestic shippers are surveyed, most import flows are not included in the database. This database is used by transportation planners to analyze freight flow patterns by commodity type, mode, shipment size and value. The CFS data for the years 1997, 2002 and 2007 are publicly available through the BTS website at [http://www.bts.gov/programs/commodity_flow_survey/](http://www.bts.gov/programs/commodity_flow_survey/). The website also presents a number of interactive tables and pre-aggregated summary reports by region and mode.

*Rail Waybill data* – The Rail Waybill master data file is a confidential database compiled annually that contains aggregate commodity flows by Bureau of Economic Analysis (BEA) regions and five-digit Standard Transportation Commodity Classifications (STCC). In order to maintain confidentiality, this database does not reveal the true origin, destination or shipment value of most shipments. The master data file provides routing information in the form of railroad and stations while the public use file provides routing information as interchange states.
This database is used by transportation planners to assess the demand on the rail network and analyze the potential for mode shift to rail. The public use file developed from the master file is available on the Surface Transportation Board (STB) website at http://www.stb.dot.gov/stb/industry/econ_waybill.html. The structure of the database is complex and the data needs to be processed before use. The website lacks options to interactively query and summarize data.

**Foreign Trade data** – This database is compiled every month from U.S. international merchandise import and export trade documents and contains information at the state or province level. Summary tables for this data can be obtained from http://www.census.gov/foreign-trade/www/. In addition, fine grained trade data can be ordered for a fee through this website. Transportation planners use this database to understand the movement of goods through ports; analyze the economic impact of the port and foreign trade within the region; and suggest changes to policies as deemed necessary.

**Freight Analysis Framework 2 (FAF2)** – FAF data contains freight flow data at an aggregate geographic level (114 domestic regions, 17 international gateways, and 7 foreign trade zones) by 43 commodity classes (SCTG) and seven modes. The data are available through the Federal Highway Administration (FHWA) website at http://ops.fhwa.dot.gov/freight/freight_analysis/faf/. This website also provides freight profiles for each state in the form of various summary tables. However, the aggregate nature of this database prevents it from being used by most counties and Metropolitan Planning Organizations (MPO).

**TranSearch Commodity Flows Database** – IHS Global Insight, Inc. compiles the TranSearch database from a variety of public and private sources including the commodity flow survey and interviews of various trucking firms. This information can be procured at various levels of spatial fidelity, industry sector and mode of transport of goods. The cost of appropriation varies by the level of detail needed and by the entity purchasing the data. Costs typically range from $10,000 to $100,000.

**Freight Databases from local and regional studies** - As part of corridor studies and other planning and design efforts, state DOTs and MPOs occasionally perform facility inventories, freight flow surveys, and vehicle classification counts. These datasets are rarely stored in formal databases or made available beyond the immediate needs of the study. The cost and effort of collecting the data for a given study could be more easily justified if the data were uploaded to a data repository and thereby made available to other studies within the state or by other states and urban areas.

### 2.3 Data Processing and Maintenance Needs

The creation of a regional freight demand model or a freight movement analysis toolkit requires good freight flow data. While a multitude of freight data sources exist, most end users feel that freight data is in-accessible at best. This feeling can be mostly attributed to the variety of sources, the differences in nomenclature, and the lack of a tool to translate the data to a format that is comprehensible to the end user. For example the Commodity Flow Survey and the Freight Analysis Framework classify commodities by SCTG code while the railway waybill data classifies commodities by STCC code. The need for normalized datasets or a set of tools that could normalize the various sets of data is rising. The lack of a unified data structure is also a concern. Modelers typically need to compile the necessary information from various sources into a consistent format prior to any model development or analysis. Methods or procedures that facilitate maintenance and access to freight data are important.
A freight data exchange network could provide the ability to transform the collected data, transfer it, and access pre-existing data as tabulations and summary reports. The provision of these data processing services would involve:

- The enumeration and re-organization of data tables from existing public agency and private sources;
- The development of a database schema and an entity relationship model or a data model to define and enumerate logical relationships between the data entities;
- The development of scripts to accommodate the conversion of data uploads to fit the database schema; and
- The development of an exchange file format and schema for data uploads and downloads.

This implies a data repository or data warehouse function that is more involved to develop and operate than a simple “wiki” like portal would imply. The costs and skills required to develop and operate this type of system are not insignificant. Given the typical capabilities and skills of MPOs and freight data providers, it is likely that considerable upfront investment in easy-to-use tools and user interfaces would be needed to make these capabilities useful to potential data users and providers.

### 2.4 Database Systems and Data Warehouses

The advent of the information superhighway has provided not only numerous new technologies, but also numerous data storage and exchange formats. Today, most large datasets are stored in a relational database management system (RDBMS) such as:

- Oracle
- Microsoft SQL Server
- Microsoft Access
- MySQL

Oracle and Microsoft SQL Server are complex and powerful software systems for managing massive databases. Microsoft Access and MySQL are more appropriate for smaller and less complex databases. In both cases custom user interfaces are typically created to present the end user with a friendly interface to input, search, summarize, and comprehend the data.

Unfortunately, existing freight datasets differ significantly in both structure and content. Bridges or translators will need to be built to effectively interpret the data and present it in a format comprehensible to humans. Open Database Connectivity (ODBC) provides a standard software application programming interface (API) to interpret and present the data, and could be used to link multiple data sources.

It is also essential that the data users and data providers develop a familiarity with the data and how the data is defined and organized. Each dataset needs to include metadata (i.e., a description of the structure and content of a dataset) to inform data users about how to read and use the dataset and instruct data providers about how to write and format data uploads.

Requirements such as these typically lead a system developer away from an isolated RDBMS to a data warehouse concept. A data warehouse is a repository of data with tools to analyze, extract, transform, and load data. It also provides tools to manage and retrieve metadata associated with each of the data elements. One potential vision is a fully integrated data warehouse that has the capability to provide the necessary tools needed to build the freight data exchange network. A data warehouse with an effective internet portal could serve the
analysis and reporting needs of the end-user of freight data. This data warehouse is envisioned to provide the tools needed to facilitate the following:

**Extracting data** – Since a multitude of data sources already exist, a set of tools could be developed to extract the necessary information from these sources and re-format the data to the data warehouse structure. These tools could automate the task of updating the information in the data warehouse when new information is available from one or more data sources.

**Transforming data** – A set of tools would need to be developed to integrate and reconcile data from multiple sources. This will reduce redundancy in the data and transform or link the data for cross-referencing, query building, and searching within the data warehouse. These transformation tools would apply a series of rules or procedures to conform the data to a standard format.

**Loading data** – Tools for loading new data sources in various formats into the database would be helpful. These tools would provide an effective procedure for uploading locally collected data into the data warehouse.

### 2.5 Data Portal

An internal-based portal or user interface needs to be established to enable the end-user to search for information and analyze, format, upload and download data. This portal is envisioned to provide the following services.

1. Information about and hyperlinks to publicly available datasets and commercial data vendors;
2. An interface to search for specific information or create database queries;
3. An interface to upload one or more data sets to the data warehouse;
4. An interface to choose data sets and aggregation levels for data downloads;
5. Documentation and metadata about database tables and data items;
6. A variety of tools to post-process some or all of the data sets available through the portal;
7. A wiki-like interface to develop and maintain process and methodology guidance and study reports;
8. A list-serve to keep interested parties informed about data updated or additions;
9. A help-desk or an online forum to facilitate dialogue and support among members;
10. Secure access to sensitive datasets by managing and validating user accounts and passwords; and
11. Mechanisms for tracking data requests and collecting use fees (if necessary).

The following sections discuss several of these requirements in more detail.

#### 2.5.1 Data Providers

Data providers are envisioned to encompass a wide range of agencies and firms related to freight transportation including State DOTs, MPOs, trucking and distribution firms. As multiple providers pool and share data, the benefits to them and the industry in general is magnified. They improve their knowledge about the freight transportation system in the region and the types of solutions that are likely to make a difference. In addition, the tools offered by the portal could enhance their analysis and processing of their own data.

Though smaller MPOs and Council of Governments (COGs) may not have any data to share, they are likely to find the information provided by others helpful in understanding the freight needs of their region and thereby be more receptive to improvements that benefit the industry. The availability of this data to smaller MPOs and
COGs could result in cascading benefits to the data providers. In other words, though some of the agencies would not be able to provide data, all agencies would benefit by sharing data. This suggests that fees for data should not be collected from MPOs, COGs and other public agencies. However, a fee could be charged to firms requesting data for commercial purposes, such as locating a distribution center or intermodal facility.

2.5.2 Data Security
Data sets such as the rail waybill data are confidential and cannot be publicly available. Data security measures would have to be incorporated into the data warehouse to ensure that confidential data are protected. If these data are not critical for planning and design applications, the confidential data elements should be stripped during the uploading process. If the data are necessary to make the information useful, data security measures will be applied. This is likely to include user accounts with different security clearance levels and data encryption and masking techniques to protect the data within the database and during all file transfer activities. Terms of Use agreements will need to be signed with both the data providers and the data users to ensure legal protection for all parties.

2.5.3 Data Interchange Format
One or more data interchange formats should be adopted for data transfers to and from the data warehouse. XML (eXtensible Markup Language) is one format option that has become increasingly popular and flexible in recent years. A complex relational data set can be converted into a hierarchical XML file with the use of standard converters. This makes it possible for a whole dataset can be transmitted from or to the data warehouse with ease. The XML format is a human readable text format that includes all of the information needed to read and interpret the data. The result is in a file that contains a significant amount of redundant information. The file size is much larger than other comparable binary formats. Confidential information can be transmitted by encrypting the data in some or all of the data tables before converting it to an XML format. In addition, the XML files can be compressed before transmitting to reduce the file size.

2.5.4 Data Size and Transfer Speed
The size of freight datasets such as commodity flow information is a concern for data transfer, analysis and reporting. Uploads of large datasets to the data warehouse should be pre-processed by the data providers in order to eliminate redundancy of data and compressed to ensure the data transfer is efficient. To assist with this process, a set of pre-processing tools could be provided to normalize the datasets and compress them for upload. The datasets requested by the data users could also be compressed to ensure that data transfer speeds are high. The file could include a self extraction tool to uncompress the data on the destination computer.

If the data warehouse receives a high number of requests for datasets that degrades the performance of the whole system, a set of tools to query, analyze, and develop reports from the downloaded data could be made available through the data portal. This would reduce the data processing load on the server and provide the data user with greater flexibility and improved responsiveness.

The set of software tools developed for these purposes need to be flexible, interoperable and user-friendly. It is recommended that these tools be written with a commonly available back-end database file system such as Microsoft Access or Excel in order to ensure that most of the data users experience a familiar “look and feel”. The analysis and reporting software itself should be developed using standard database interface technologies such as SQL scripts, Visual Basic and Crystal Reports.
3 Views of Potential Participants: Data Users and Providers

In order for the Freight Data Exchange Network to succeed there must be a set of willing participants of two types—data users and data providers. Data users are primarily state DOTs and local MPOs, and the contractors (consulting firms, universities) who support their transportation planning, design, and operations needs. Data providers are envisioned to be primarily transportation carriers (truck, rail, water, intermodal), shippers, and the various third parties who support freight transportation and logistics activities (3rd party logistics providers, freight bill audit firms, freight forwarders, etc.).

Data users might also serve as data providers if they make the results of their studies available to other users, as discussed in section 2.2 above. A premise of this study, however, was that a primary objective of the Freight Data Exchange Network is to provide a mechanism for private sector firms to readily make available data that are at present quite difficult for the public agencies to identify and access. In this context the aims and views of the data user and provider communities are quite different, hence firms and agencies interviewed were divided into these two groups.

This section presents the results of interviews conducted to obtain the views and requirements of these potential participants. Interviews were conducted either in person or by telephone, using the interview guides provided in Appendix A. In most cases the guide was sent to the respondent prior to the interview, with the directive that this was not intended to be a self-administered questionnaire, since it would be necessary for the interviewer to follow up initial answers with added in-depth questions.

3.1 Data User Interview Results

3.1.1 Experience in Using Freight Data

The first set of questions was designed to elicit information on the level of experience of the agency in conducting freight studies and using freight data. All of the states and large MPOs interviewed had conducted one or more freight transportation studies, covering truck, rail, intermodal, and occasionally the marine and air modes. Most of the data sources listed in section 2.2 have been used for such studies. The core dataset is usually local traffic and truck count data for the primary highway and intermodal connector system, supplemented by traffic data for the other modes of interest (e.g., Rail Waybill and Waterborne Commerce data). Many areas have purchased TranSearch commodity origin-destination data extracts. Special data collection has included freight facilities, and data from enforcement agencies on hazardous materials shipper/carrier compliance and truck oversize, overweight permits. They also use employer and property data provided by their state economic development office. One area noted for possible improvement is to make use of the data collected at roadside truck inspection stations, which has not commonly occurred; perhaps the proposed exchange network would facilitate this. These agencies all expressed a willingness to share their data, subject to observing any proprietary rights and privacy restrictions to which they had agreed.

Small MPOs and states newly embarked on freight studies have a slightly different outlook. They have usually conducted only limited scope and preliminary studies, with a focus on truck traffic and occasionally intermodal facilities, and with strong reliance on data provided by their state DOT. A typical application has been input to county profile databases that are used for economic development purposes. An emerging issue for some MPOs is relocation of logistics centers to less congested areas, which will require more freight data. These smaller MPOs recognize their relative lack of experience, and look to their state DOT and more experienced MPOs for guidance. Most have attempted little or no original freight data collection.
3.1.2 Interest in the Proposed Freight Data Exchange Network
Both large and small agencies expressed strong interest in the Freight Data Exchange Network. They particularly desired access to commodity origin-destination data, route and link-level freight vehicle counts, detailed truck data, and vehicle emissions data. It was noted that traffic engineering and traffic safety agencies already collect and provide accident data, and there may be a liability issue for MPOs to also handle this type of data.

This interest is hardly surprising, as any agency that has attempted a freight study is keenly aware of current data availability shortcomings, and the respondents were presented with a “shopping list” of potential data types that might be included. Nonetheless this expression of interest in the principle data types was consistent across all respondents, indicating that there is a potential market for the proposed service.

3.1.3 Requested Features for the Freight Data Exchange Network
Data users requested that the following features be considered for inclusion in the data system.

- The site should be easy to use. The Census Bureau website was cited as an example of an easy to use site.
- The system should process user queries, e.g., to search or sort the data.
- The system should provide information on data quality and information on or links to descriptions of how data are being collected and used.
- A “nice to have” feature would be methodology guides or reports, particularly drawn from the experience of others, such as:
  - examples of how to link commodity O-D data to actual vehicle trips;
  - case studies organized by size, geography.
- Provide links to government data sites, such as FHWA, BTS, Census Bureau, etc. The goal should be to provide easier access to government data, e.g., the Rail Waybill data.
- Consider providing separate sites or branches for neophytes, vs. experienced users.
- Legal protections should be used to ensure insulation from release of competitive or private data, in order to reduce some of the obstacles to collecting data from private firms. This is necessary because, as one respondent put it, “If one or two careless users ‘blow it’ the site would be shut down.”
  - MOUs should spell out the terms of use.
  - Use model agreements from other business sectors as a guide.
  - Potential users routinely execute such agreements with data providers, so requiring this should not be an obstacle to system development.

3.1.4 User Fee Issues
Fees that might be charged to download data from the site are definitely a concern of the potential users. All agencies were willing to pay reasonable fees to support the exchange network, with the caveat that reasonableness of cost would be judged with respect to data content, timeliness, and quality and value added for their operations. Subscription fees were mentioned as one way to spread out the access cost over time, rather than paying a high cost at the time of a freight study effort. However, some small MPOs suggest paying for what is downloaded, rather than a subscription fee, since that would tend to minimize their costs. It is reasonable to
suppose that current costs to use private data (such as TranSearch) provide an upper limit on user willingness to pay, and some respondents verified this.

One idea to encourage states and MPOs to upload their data and reports is to waive usage fees for those who make such uploads. In another variant uploaders would receive credits that could be used as partial payment for their future downloads. As might be expected this idea found favor with the larger MPOs. A small MPO pointed out, however, that that large MPOs and state DOTs are the likely uploaders of data, and they are the ones who are better able to pay fees. Design of an appropriate user fee system, or alternatively finding some other way to pay for the expenses associated with providing and maintaining the service, will be an important design consideration.

3.1.5 Obstacles to System Use
Large MPOs and state DOTs listed the lack of incentives to participate and share data, and institutional or inertia barriers, as obstacles to their use of the proposed Freight Data Exchange Network. Small MPOs had different concerns, including cost; the complexity and large size of the dataset, making it hard to find exactly what is needed; and lack of staff expertise. Lack of incentives to participate was a major issue, and was also cited by potential data providers, as discussed in the next section.

3.2 Data Provider Interview Results
There are many subgroups of potential private sector providers of freight data, so it was possible to interview only one or two representatives of each. The respondents have quite different experiences in collecting their own data (e.g., motor carriers vs. barge companies) and providing it to governmental entities. Hence there was some divergence in their views, so it is difficult to find any sort of consensus.

3.2.1 Experience in Providing Freight Data
There was one point of consensus—nearly all respondents had been asked to provide data to a state or metropolitan area for a study with specific consideration of freight traffic. Carriers are often asked for data, by freight consultants and all levels of government. Private sector value-added data providers sell detailed freight data to public agencies as a regular line of business, and invest considerable resources in compiling such data.

Carriers provide to federal agencies data required by law or regulation, e.g., the former motor carrier financial and operating data, safety, rail waybill, waterborne commerce, lock performance monitoring system, etc. The trucking industry has, on rare occasions, released sample O-D data. Detailed O-D and trip data provided by carriers have been used to develop freight facility investment and operating plans. Company-specific data has not been released.

3.2.2 Potential for Participation in the Freight Data Exchange Network
Transportation carriers collect and maintain data that would be useful for tracking freight transportation activity (e.g., freight vehicle movements, shipment origins and destinations, freight vehicle routes, empty vehicle movements, fuel consumption, transportation charges). Firms like UPS and FedEx, and virtually any trucking company with more than 150 trucks, and any large rail or marine carrier, have detailed data. Carrier trade associations also maintain relevant data. For example the trucking industry maintains a GPS-based truck position dataset for a sample of instrumented vehicles, up to three years, and has provided this data to FHWA, state DOTs, and MPOs. Carriers maintain detailed operational data indefinitely, and use it to develop P&L statements, etc. Such carrier data are typically in electronic formats.
Carriers might be willing to share their data with states and MPOs, if the prospective results of such cooperation would lead to a return on investment (ROI) within no longer than 24 months. Releases of scrubbed and aggregated data are most likely, rather than detailed proprietary data. Strict confidentiality assurances would be required. Some carriers did express interest in the proposed easy-to-use internet upload site as a means of facilitating data sharing. Carriers already use methods like FTP sites, so there are really no technical data sharing issues.

In contrast to the considerable reservations about providing data, nearly all respondents expressed some interest in using an internet data sharing site to download data provided by public sector agencies and other parties. Such data, if it is not too old, could be useful for market and performance analysis. However, they did feel that most of the data that would be on such a site is already available from other parties. Obtaining the data from an easy-to-use internet download site would be of value to them. Carriers would be willing to pay fees to use such a system, if the quality of the available data warrants this. Any usage fees should be waived for those who have uploaded data to the site. Firms already participate in marketing surveys under quid pro quo arrangements. They would require the user to sign license agreements or usage restrictions to use an internet data sharing site.

3.2.3 Obstacles to Participation
There are significant obstacles that would prevent carriers and other parties from sharing their data. The private sector and government agencies operate on vastly different timescales, which causes difficulty in meeting private sector ROI requirements. There is a perceived lack of “mission criticality” on the part of government. What can the government offer the private sector in the way of quantified benefits in the near-term to induce private companies to want to invest effort in providing freight data? Where’s the ROI for the private companies to be able to explain to their owners why they are expending effort to participate? The traditional government role in enforcement and regulation is also a big concern. The idea that an open source site would provide the necessary level of data confidentiality and security seems unrealistic.

The value-added freight data providers see significant problems with the proposed system, and would likely oppose its establishment. Private data firms maintain and estimate detailed freight activity data, and essentially operate a freight data exchange network for the benefit of their members and customers. Why should this be displaced by a free or low cost government-sponsored competitor? They also question the assertion that data providers now lack systems to inform potential users about what freight data are available. The web already serves this purpose well via search engines and guides to information that have been published by the Transportation Research Board and others. Freight data vendors are receiving frequent cold inquiries from new potential data users as a result of their existing web content.

The trucking industry rarely provides any detailed data to requesters due to concerns about:

- Civil litigation
- Release of proprietary operating and financial data
- Release to government could result in further releases

Other carriers have similar concerns. These concerns about release of proprietary data would likely mean low participation in the Freight Data Exchange Network by the private sector.
These findings are central to this study. While private companies collect and maintain considerable freight data for their own business uses, they have no incentive to share these data, and many serious concerns about adverse business consequences that could result from data sharing. The data maintained by these firms have unique formats and characteristics, which (even with use of carefully defined metadata standards) poses system design problems, and virtually requires some type of freight data exchange network to make the data routinely usable by third parties. But the privacy concerns and lack of incentives makes it very unlikely that the firms would participate. This absence of truly new data from the private sector makes the whole enterprise questionable, even if the public agencies are willing to invest the time and resources needed to upload their data.

### 3.2.4 Overcoming Obstacles

The most significant action that would overcome the obstacles to private sector participation would be to create or find some win-win examples. For example, the barge industry would probably provide detailed data to the U.S. Army Corps of Engineers even if not required, due to the broad industry benefits of Corps’ programs. Their industry association is engaged in a cooperative effort to improve accident and incident data, again due to broad industry benefits.

The only sure way to obtain industry involvement is to demonstrate that their concerns about subsidized competition (in the case of the freight data vendors) and release of proprietary data can be overcome, and that costs of participation will yield benefits in the near term. If real bottom line improvements accrue to carriers, this could be a real trust builder.

### 4 Views of Potential System Operators

In addition to data users and data providers, the third party needed to make the Freight Data Exchange Network a reality is someone to develop and operate the system. Potential system custodians include a variety of scientific, technical, and trade organizations, as well as selected federal government agencies. A subset of the potential operators was interviewed, and their views are presented here.

All of the parties interviewed were interested in operating the system, with some caveats. Since this is not a trivial undertaking a business case for providing this service would need to be made, as all organizations have an approval hierarchy to follow for approval of new ventures. There would need to be an identified community of interest, and a relatively assured revenue stream to support system development and long-term operation.

### 4.1 System Implementation

Prior to doing any technical work an oversight board or user group would be created; comprising representatives from the data user and data provider communities. This board would recommend policies that would guide system development and operation.

The system would be implemented using a combination of regular agency staff and contractors. The initial software development and other information technology (IT) tasks likely would be contracted to IT consultants, while operation would use full time and part time staff. The system could be operated within the organizational framework of the existing library, information, and software services provided by the host. It would not take a full time person to run the system, particularly in its early days, so part-time contract workers could be used. In fact once the system is up it could be operated from almost any location, so it lends itself to being operated by a telecommuter worker.
Quality assurance and quality control activities will be important. To facilitate this, the host might provide or review all metadata, and might provide an expert to assist with data input by the users. Some other tasks that the host organization might perform include publicizing the system at professional meetings by doing demonstrations, and providing manuals, training, and webinars.

### 4.2 Costs and User Fees

The costs to provide the system include one-time development costs, and ongoing maintenance and operating costs. If a sponsor could be found it might be possible to obtain a grant or contract to cover the initial costs. For example, interested states could conduct a pooled funds study, or an appropriate cooperative research programs project could be proposed. Most sponsors would be unwilling to cover long-term operations, so user fees would need to be imposed. As a minimum the fees would need to cover any IT costs, such as internet hosting, and data storage charges, and “service call” assistance, as well as the labor cost for the part-time staffer. There may be expenses related to periodic meetings of the oversight panel. Without an up-front development grant the fees would need to be high enough to recover development costs.

Potential system operators have various ways to collect user fees. They already sell publications, and sometimes software or data, to the public, either per item or under subscription arrangements. They also sponsor conferences and workshops and have business systems in place to account for costs and collect participant fees. They can discriminate between members and non-members, subscribers and others, etc., so it would be possible to implement various types of tiered user charges. As noted earlier one suggestion is for fees to be waived or reduced for those users who upload data to the system. Again, the systems to handle such exceptions are in place or could easily be provided.

### 4.3 Handling Restricted Data

Potential system operators do have in place procedures for handling restricted datasets, but this can be an expensive feature. For example one organization currently has a copyrighted database that is available to sponsors only. Access to this is controlled by inspecting the requesters IP address.

A related issue is how best to incorporate or take advantage of commercial data that is offered for sale by private vendors, without violating the related terms of use. It was suggested that the procedures used by the Federal Geographic Data Committee might provide some guidance in this matter, as they have found ways to utilize both public domain and private data.

### 4.4 Obstacles to System Implementation

A significant obstacle is the mindset among many data users that freight data should be free, and freely available from government agencies. So any user fees above media and distribution costs might be resisted, or at least would reduce the number of system users. While this stereotype may have some validity, it is noted that this does not agree with the sentiments reported by the data users in section 3.1.4 above. While fee levels are a concern, they are not a make-or-break concern; users will pay for value delivered.

By far the biggest obstacle is the reluctance of potential data providers, particularly in the private sector, to upload their data. There is also no real incentive for public agencies to upload reports and data. Real or perceived competition with private sector data vendors will also be a problem for the potential system operators.
There are some ways to alleviate private sector concerns. One is that the system might provide an index to private vendor data, which would tend to direct to them added customers. Another is that private sector data vendors might be persuaded to release or sell at a reduced price their outdated data. For example NOAA makes available old sea state data obtained from a vendor who provides the real-time data to fishing fleets.

5 Freight Data Exchange Concept

The freight data exchange network is envisioned to contain an access portal through which data providers can upload data while end-users can download data in the form of summary tables, reports and customized tabular data. Initially, the freight data exchange network would act as a bridge between the data provider and the user. It would effectively extract, transform and load the standard existing datasets to the data warehouse. In addition, it would provide a user interface for the data providers to peruse data standards and formats; and upload data, metadata and documentation. These uploaded datasets would be analyzed by a database administrator in order to develop routines to extract, transform and load the data into the data warehouse and eventually develop a set of scripts to automate these repetitive tasks.
Exhibit - 1: Freight Data Exchange Network conceptual design

Existing Databases
- FAF2 Database
- Foreign Trade Database
- CFS Database
- Railway Waybill Public use Database

Agency Data Files
- HPMS Data
- Commodity Flow Data
- Emissions Data
- Vehicle Classification Count and Speed Data

Internet Portal
- Data Provider Interface: Upload
- Extract | Transform | Load Tools

Database Warehouse
- Metadata and Data mapping Equivalencies
- Transformed Databases
- Pre-processed Summary Tables

Online Analytical Processing Tools

Internet Portal
- Helpdesk
- Wiki & User Forum
- Documentation & Information

Data User Interface: Search, filter, query, aggregate, download
- Summary Tables
- Custom Tabular Data
- Custom Reports

Offline Data Processing Applications for download
- Data Translator tools
The internet portal itself would contain documentation (wiki), metadata, a user forum, and static and dynamic information (hyperlinks). As the number of data sets uploaded to the network increases, the network’s engine would be enhanced and scaled to be able to effectively transform and reformat the data. Components such as a data processing applications and data translating tools would be added to the portal for download. User input from the forums could be used to debug, enhance the capabilities of these applications. Finally, a set of documents and tools could be added to the network to help the end-user not only develop a better understanding of the ways to effectively use multiple data sets but also expedite the process of analyzing the data through the use of a set a tools built to automate standard procedures. Exhibit - 1 presents a conceptual design of the freight data exchange network, key features, inputs and outputs.

One of the advantages of a data warehouse is that it integrates data from multiple sources into a common format. This section outlines approaches to effectively create and manage data schema, standards, formats, sources and metadata for various types of freight-related datasets.

5.1 Metadata

Metadata is a description of the structure and content of a dataset that informs data users and software applications about how to read and use the dataset. The information may be included as header records in the dataset, an associated text or metadata file, and/or written documentation. For generic or existing datasets, an XML file is frequently used to distribute metadata over the internet.

The creation of metadata standards or rules is essential to manage complex data warehouse. Metadata should include field specifications, lookup tables, enumerations and other descriptive data to help the end-user effectively decipher the data and utilize it effectively. Exhibit - 2 presents the contents of a simple XML-based metadata file.

Exhibit – 2: XML Metadata Sample

```xml
<Metadata>
  <Data Source Name="MPO Commodity flow data"/>
  <Use Restrictions Text="None"/>
  <Created Date="Dec 01, 2008"/>
  <Copyrights Text="Freight Data Exchange Network"/>

<Datasets>
  <Table name="Commodity Flow" description="Freight flow movement information"/>
  <Attributes
    Name="Origin" type="Integer" description="Origin Region – place of production of goods"
    Name="Destination" type="Integer" description="Destination Region – place of attraction of goods"
    Name="Commodity Code" type="Integer" description="Type of commodity by group such as food"
    Name="Total Tons" type="Integer" description="Total tonnage of goods moved"
    Name="Truck Tons" type="Integer" description="Tonnage moved by truck mode"
    Name="Rail Tons" type="Integer" description="Tonnage moved by rail mode"
    Name="Air Tons" type="Integer" description="Tonnage moved by air mode"
    Name="Truck Vehicles" type="Integer" description="Number of truck vehicles used"
    Name="Rail Containers" type="Integer" description="Number of rail containers used"
  />
</Datasets>
```
5.2 Data Schema

The creation of an expandable standard data dictionary or a data schema is essential to ensure this data schema could be further expanded to include data pertaining to intermodal traffic, water tons, etc. as needed.

The basic data schema for the freight data exchange network would focus on the following types of data tables and data relationships.

1. Origin-Destination Commodity flow data
2. Origin/Destination code descriptions
3. Commodity code relationships and cross references for STCC, STCG, etc.
4. Truck vehicle emission rates by region/zone.
5. Truck volume and speed distributions by facility type and time of day.

5.3 Datasets and Potential use

The following section describes various types of freight related data sets and suggests potential ways to utilize the datasets.

5.3.1 Freight Analysis Framework Data (FAF2)

The FAF2 dataset available in various formats including text, dbase, MS Access, contains domestic and border crossing commodity flow movements and provides reasonable estimates for national and multi-state corridor analyses. This data does not have enough detail to support local planning and policy. However, it provides an excellent source to compare and validate local data. Exhibit - 3 presents the commodity flow data structure for the FAF2 database and Exhibit - 4 presents the seven modes reported the FAF2 database. An organization such as a State performing regional freight demand analysis could effectively use this dataset to create a quick snapshot of the current freight productions and attractions for the region. If the organization were to perform a detailed analysis, the data in FAF2 would also provide a reasonable estimate of forecasts for comparison.
Since the FAF2 database is relatively small in size, this dataset could easily be hosted by the freight data exchange network. In addition, the portal could provide online and offline tools to query and analyze the data. The FAF2 database could be normalized by using numeric codes for regions, commodity types and modes. This re-format of the database would not only reduce the size of the database, but also expedite querying and reporting processes. In addition, the use of numeric codes for commodity grouping (SCTG for FAF2) would allow for easy translation to a different commodity grouping such as the STCC and in turn help organizations perform comparisons of their data to FAF2 with ease.

5.3.2 Commodity Flow Survey (CFS)
The CFS dataset provides data on shipments from select types of businesses in the United States by North American Industry Classification System (NAICS) code. Exhibit - 5 presents a partial data structure for the commodity flow survey. The data from the CFS which includes survey from auxiliary establishments such as warehouses is used by organizations to assess the demand for transportation facilities. The lack of a data analyzing tool is a roadblock for most of the organizations to effectively analyze and interpret this data. The CFS dataset could be processed and converted to a more user friendly schema. A proposed conversion schema is presented in Exhibit - 6. This conversion process would include developing look up tables to convert NAICS industry / employment based codes to a standard commodity code such as SCTG which would be interpreted easily by the end user. The inclusion of standard FIPS state and county codes for the regions in the CFS data would improve query and analysis capabilities. A set of tools could be developed to perform the necessary conversions and upload this data to the freight exchange network data warehouse. A set of online and offline
tools could be developed to query and analyze this data and hosted on the portal. This process along with the analysis tools available would greatly help organizations analyze the CFS data and utilize it effectively.

Exhibit - 5: Commodity Flow Survey data structure

<table>
<thead>
<tr>
<th>Short Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSA</td>
<td>FIPS Combined Statistical Area code</td>
</tr>
<tr>
<td>DESTGEO</td>
<td>Destination geographic area</td>
</tr>
<tr>
<td>GEOTYPE</td>
<td>Type of geography flag</td>
</tr>
<tr>
<td>GEO_ID</td>
<td>Origination geography</td>
</tr>
<tr>
<td>MD</td>
<td>FIPS Metropolitan Division code</td>
</tr>
<tr>
<td>MSA</td>
<td>FIPS Metropolitan Statistical Area or Micropolitan Statistical Area code</td>
</tr>
<tr>
<td>SECTOR</td>
<td>NAICS economic sector code</td>
</tr>
<tr>
<td>ST</td>
<td>FIPS state code</td>
</tr>
<tr>
<td>TMILE</td>
<td>Ton-miles (millions)</td>
</tr>
<tr>
<td>TON</td>
<td>Tons (thousands)</td>
</tr>
<tr>
<td>VAL</td>
<td>Value ($million)</td>
</tr>
<tr>
<td>YEAR</td>
<td>Year</td>
</tr>
</tbody>
</table>

Exhibit - 6: Commodity Flow Survey data conversion

<table>
<thead>
<tr>
<th>Short Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin</td>
<td>Origin Region</td>
</tr>
<tr>
<td>O_FIPS</td>
<td>Origin FIPS state /county code</td>
</tr>
<tr>
<td>Dest</td>
<td>Destination region</td>
</tr>
<tr>
<td>D_FIPS</td>
<td>Destination FIPS state /county code</td>
</tr>
<tr>
<td>SCTG</td>
<td>SCTG commodity classification</td>
</tr>
<tr>
<td>SECTOR</td>
<td>NAICS economic sector code</td>
</tr>
<tr>
<td>TMILE</td>
<td>Ton-miles (millions)</td>
</tr>
<tr>
<td>TON</td>
<td>Tons (thousands)</td>
</tr>
<tr>
<td>VAL</td>
<td>Value ($millions)</td>
</tr>
<tr>
<td>YEAR</td>
<td>Year</td>
</tr>
</tbody>
</table>

5.3.3 TranSearch Database

The TranSearch database is available for purchase from IHS Global Insight, Inc. It is a comprehensive dataset containing commodity flow information disaggregated at a geographic unit as small as a zip code. The dataset can be procured at a commodity level (STCC) disaggregation of over 500 commodities. Exhibit - 7 presents the data structure of a typical TranSearch database.
### Exhibit - 7: TranSearch Database structure

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geography: Origin</strong></td>
<td></td>
</tr>
<tr>
<td>Origin FIPS</td>
<td>Five Digit county code, independent city</td>
</tr>
<tr>
<td>Origin BEA</td>
<td>BEA region code (county grouping)</td>
</tr>
<tr>
<td>Origin BEA Name</td>
<td>BEA name</td>
</tr>
<tr>
<td>Origin State</td>
<td>Two digit state code</td>
</tr>
<tr>
<td>Origin State Name</td>
<td>State name</td>
</tr>
<tr>
<td>Origin Region</td>
<td>Finer zone such as a zip code</td>
</tr>
<tr>
<td><strong>Geography: Destination</strong></td>
<td></td>
</tr>
<tr>
<td>Destination FIPS</td>
<td>Five Digit county code, independent city</td>
</tr>
<tr>
<td>Destination BEA</td>
<td>BEA region code (county grouping)</td>
</tr>
<tr>
<td>Destination BEA Name</td>
<td>BEA name</td>
</tr>
<tr>
<td>Destination State</td>
<td>Two digit state code</td>
</tr>
<tr>
<td>Destination State Name</td>
<td>State name</td>
</tr>
<tr>
<td>Destination Region</td>
<td>Finer zone such as a zip code</td>
</tr>
<tr>
<td><strong>Commodity</strong></td>
<td></td>
</tr>
<tr>
<td>STCC</td>
<td>Standard Transportation Commodity Code</td>
</tr>
<tr>
<td>Commodity Name</td>
<td>Descriptive name of the commodity by STCC code</td>
</tr>
<tr>
<td><strong>Shipment Volumes: Motor</strong></td>
<td></td>
</tr>
<tr>
<td>Truckload Tons</td>
<td>The net annual tons carried by truckload common carriers.</td>
</tr>
<tr>
<td>LTL Tons</td>
<td>The net annual tons carried by less than truckload common carriers.</td>
</tr>
<tr>
<td>Private Truck Tons</td>
<td>The net annual tons carried by private truck fleets</td>
</tr>
<tr>
<td>Truckload Trucks</td>
<td>The net annual truckloads carried by truckload common carriers.</td>
</tr>
<tr>
<td>LTL Trucks</td>
<td>The net annual truckloads carried by less than truckload common carriers.</td>
</tr>
<tr>
<td>Private Trucks</td>
<td>The net annual truckloads carried by private truck fleets</td>
</tr>
<tr>
<td><strong>Shipment Volumes: Rail</strong></td>
<td></td>
</tr>
<tr>
<td>Rail Carload Tons</td>
<td>The number of annual short tons carried by rail (non intermodal)</td>
</tr>
<tr>
<td>Rail Carloads</td>
<td>The number of rail carloads carried by rail (non intermodal)</td>
</tr>
<tr>
<td>Intermodal Tons</td>
<td>The number of annual short tons carried in rail (intermodal)</td>
</tr>
<tr>
<td>Intermodal Units</td>
<td>The number of trailers and containers carried (intermodal)</td>
</tr>
<tr>
<td><strong>Shipment Volumes: Other</strong></td>
<td></td>
</tr>
<tr>
<td>Air Net Tons</td>
<td>The number of annual net short tons carried by cargo and combination air carriers</td>
</tr>
<tr>
<td>Water Net Tons</td>
<td>The number of annual net short tons carried by barge operators on the inland waterways</td>
</tr>
<tr>
<td><strong>Shipment Value</strong></td>
<td>Estimated Value of goods at origin in U.S. dollars</td>
</tr>
</tbody>
</table>

A database of this nature procured for a state could potentially contain over 2 million records making effective analysis of the data complicated and cumbersome. An organization after procuring the database typically spends a considerable amount of effort querying and analyzing the database. IHS Global Insight, Inc’s Terms-of-Use agreement for the data does not permit the data to be hosted on the freight data exchange network. However, an application could be developed to automate most of the standard querying and reporting functions performed by the organizations. This application would aggregate the data by a user-defined set of summary classifications related to:
Freight Data Exchange Network

- **Productions**: Aggregation by origin;
- **Attractions**: Aggregation by destination;
- **Modal Split**: Aggregation by mode.

Moreover, the aggregations could include input filters for geography and mode. The availability of such an application would help the organizations not only analyze the data quickly and effectively, but also compare the data to other sources such as FAF2.

In order to compare the TranSearch and FAF2 databases, the geographic units in both databases need to be the same and a commodity code mapping between TranSearch’s STCC and FAF2’s SCTG codes needs to be created. The application could contain the mapping data and a tool to convert the dataset's codes. In addition, forecast information contained in the FAF2 data could be used to extrapolate the TranSearch flow patterns to a future year.

### 5.3.4 Highway Information Pertaining to Truck Vehicles

Most state DOTs collect and maintain information pertaining to truck vehicles such as

- Vehicle classification counts
- Designated truck routes
- Bridge restrictions on trucks

These datasets are simple and could be converted to an XML format and uploaded to the freight data exchange network to be used by other state organizations. In addition, this co-operative exchange of information would also help neighboring states develop a better understanding of the demand on the existing transportation system and help them develop a plan to meet the expected data. Exhibit - 8 presents a sample vehicle classification count and speed information data structure.

**Exhibit - 8: Sample Vehicle Classification Count and Speed Data Structure**

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route</td>
<td>Highway Route Name</td>
</tr>
<tr>
<td>Mile Point</td>
<td>Mile point location</td>
</tr>
<tr>
<td>Count Date</td>
<td>Date of traffic count</td>
</tr>
<tr>
<td>Count Hour</td>
<td>Hour of the day of traffic count</td>
</tr>
<tr>
<td>Vehicle Type</td>
<td>Vehicle Classification</td>
</tr>
<tr>
<td>Count</td>
<td>Vehicle Class Traffic Count</td>
</tr>
<tr>
<td>Speed</td>
<td>Vehicle Class Observed Speed</td>
</tr>
</tbody>
</table>

Traffic count and observed speed information by time of day on a roadway section is essential for validating traffic simulation studies and as an input to an emissions calculating engines such as EPA’s MOVES. This data if uploaded by the organization collecting it could be used by other organizations with similar characteristics performing a traffic simulation study or an air quality analysis study. Though this regional data cannot be directly used by other organizations, this data could be used to derive percentages or ratios to be applied to models from other regions. In addition, the presence of a post-processing tool to allow the user to filter and reformat the data could be extremely beneficial.
Highway Performance Monitoring System (HPMS) data is a publicly available database that contains information pertaining to roadway characteristics including traffic counts and percentages for various classes of trucks. This data is collected for a sample of roadway sections and expanded to the universe. An organization that needs to use the information from the HPMS dataset might often be concerned that the data set is not well sampled in the region of interest. In such cases, the organization could use traffic count and vehicle percentage information from a local or regional study. Easy access to this data would help these organizations refine their dataset and remove outliers before using it for a transportation related study.

5.3.5 Local and Regional freight datasets.  
Freight data sets from local and regional studies could be uploaded to the freight data exchange network. The data from these studies could include:

- Commodity Flow information;
- Vehicle Classification Counts by time of day;
- Truck vehicle emission information by type of pollutant.

Truck vehicle emission information is not yet available easily. EPA’s MOVES tool is making progress towards calculating emissions by geography and vehicle type. An organization that has performed an elaborate air pollutant emissions study could share their data and findings through the freight data exchange network. Exhibit - 9 presents a data structure for uploading pollutant emission information. A dataset of this nature could be converted to an XML format and uploaded to the freight data exchange network to be used as a guide for other organizations with similar geographic and climatic characteristics.

Exhibit - 9: Truck Vehicle Pollutant Emission Rates Data Structure

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>Speed of travel (mph)</td>
</tr>
<tr>
<td>Pollutant</td>
<td>Pollutant type</td>
</tr>
<tr>
<td>Combination Long-haul Truck</td>
<td>Emission rates for combination long haul trucks</td>
</tr>
<tr>
<td>Combination Short-haul Truck</td>
<td>Emission rates for combination short haul trucks</td>
</tr>
<tr>
<td>Light Commercial Truck</td>
<td>Emission rates for light commercial trucks</td>
</tr>
</tbody>
</table>

Organization using EPAs MOVES tool to estimate emissions information need vehicle volume and speed information by roadway type.

5.3.6 Socio-economic Data from Regional Studies

Socio-economic data sets such as population, household units and employment forecasts are developed by organizations such as DOTs conducting regional planning level studies. These forecasts are the driving agents behind freight demand forecasts. These datasets if shared could be used by smaller organizations such as MPOs, COGs, counties etc. This population and employment forecast information at the census block, block group or county level could be used to forecast commodity flow data. Exhibit – 10 presents the structure of a socio-economic data file. A commodity flow forecast requires population data (commodity attractor) and employment data by industry sector (commodity producer). In addition, key indicators such as transportation and warehousing sector related employment would not only enhance an existing freight demand model but also help produce a more reasonable forecast.
Exhibit - 10: Socio-economic Data Structure

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>Census block ID, block group ID, county FIPS etc</td>
</tr>
<tr>
<td>Year</td>
<td>Forecast Year for Socio-Economic data</td>
</tr>
<tr>
<td>Population</td>
<td>Population forecast for the region</td>
</tr>
<tr>
<td>Household Units</td>
<td>Forecasted Household units</td>
</tr>
<tr>
<td>Employment – Total</td>
<td>Total Employment forecast</td>
</tr>
<tr>
<td>Employment - Farming</td>
<td>Farming Employment forecast</td>
</tr>
<tr>
<td>Employment – Mining</td>
<td>Mining Employment forecast</td>
</tr>
<tr>
<td>Employment – Construction</td>
<td>Construction Employment forecast</td>
</tr>
<tr>
<td>Employment – Manufacturing</td>
<td>Manufacturing Employment forecast</td>
</tr>
<tr>
<td>Employment – Transportation</td>
<td>Transportation related Employment forecast</td>
</tr>
<tr>
<td>Employment – Services</td>
<td>Services Employment forecast</td>
</tr>
<tr>
<td>Employment - Other</td>
<td>Other Employment forecast</td>
</tr>
</tbody>
</table>

5.3.7 Other data sources

Data collected by public agencies for freight studies could potentially be shared over the freight data exchange network. The following is a list of data sources that could be mined and summarized for distribution by the freight data exchange network.

Continental 1 Corridor: The Continental 1 (C-1) alliance is a multi-state public-private partnership that includes representatives from nine states along the Atlantic coast. Freight data procured as part of this corridor study including analysis output data such as trade and economic development opportunities could potentially be shared over the freight data exchange network.

Source: [http://www.continental1.org](http://www.continental1.org)

I-95 Corridor Coalition: The I-95 Corridor Coalition is an alliance of transportation agencies, toll authorities and related organizations of states along the Interstate 95 corridor. The Coalition provides a common forum for decision and policy makers to address transportation related issues. This alliance has conducted freight transportation and intelligent transportation system studies; the data from these studies could potentially be shared over the freight data exchange network.

Source: [http://www.i95coalition.org](http://www.i95coalition.org)

I-81 Corridor freight studies: The Virginia department of rail and public transportation has conducted various freight transportation studies including a corridor freight rail study on I-81. The data from these studies could potentially be shared over the freight data exchange network.


Advantage I-75: Advantage I-75 is a public private partnership to reduce congestion, increase efficiency and safety along Interstate 75 by implementing an intelligent transportation system to allow properly documented transponder equipped trucks to travel through the entire corridor including Canadian highway 401 with not more than one stop at an inspection center. The data and technology if shared through the freight data exchange network could potentially benefit other agencies performing similar studies.
Corridors of the future: Six Interstate routes (I-5, I-10, I-15, I-69, I-70, I-95) have been selected from among 38 applications received from public and private sector entities to develop super-regional approaches to reduce congestion and increase freight transportation efficiency. The data from this ongoing study could potentially be shared through the freight data exchange network.


5.3.8 Data Translation
Publicly available data sets such as FAF2, CFS etc could be cumbersome to use as part of a regional study due to the differences in coding of the various databases. The development of a set of tools to translate, aggregate a select few attributes in the data set to a different notation would be extremely helpful to the end-user. For instance the end-user might want to convert a SCTG commodity group to STCC, and aggregate the FAF2 database by state.

6 Summary and Conclusions
One purpose of this study was to provide a proposed concept design for the Freight Data Exchange Network. That was done in section 3 (see Exhibit 1). The rest of the desired information is summarized below, in the form of answers to the questions asked at the outset of the study.

6.1 Design Concept Questions

6.1.1 Can freight data be shared efficiently and securely over the internet?
The simple answer from a technology perspective is yes. Internet technology and communications protocols have the ability to transfer encrypted data reliability and at relatively high speeds. Sections 2 and 5 of this report go into considerable detail about the relevant technical design issues.

6.1.2 Are there sufficient markets of data users and providers?
This is a more difficult question, and the answer is a weak “yes.” The great variety of data users and providers, particularly the latter, made it difficult to achieve a reliable consensus view. There is more consensus among data users, as nearly all of those interviewed were at least interested in the Freight Data Exchange concept. Caution, however, requires interpretation of this in light of the usual types of responses expected in a marketing study for a product that does not exist. It is common for prospective customers to state that they would try or buy a new product, but those expressions of intent to not yield the same level of purchases. The same may be true here. Without knowing with some certainty the data quality, timeliness, and price, data users tend to assume that something (anything!) may help to solve their problems, and hence they express support for the proposed system. They also were more willing as a group to make their data available to the system, subject to observing the restrictions of any terms of use attached to the data that they had obtained.

Potential data providers in the private sector were considerably more reluctant to support the concept. Among carriers and other types of logistics firms the concern is release of proprietary data in ways that could later harm their business or violate confidentiality agreements. Among data vendors the concern is competition with their value added services. All potential private sector participants see real problems with the lack of incentives for them to expend the time and resources necessary for participation, with little apparent return on investment. Hence it is possible that initial participants in the Freight Data Exchange Network might all be state DOTs and
MPOs, who would be both data users and data providers. It will take some time to gain the buy-in of the private sector.

6.1.3 What are the incentives to participate for users and providers?
Lack of incentives was mentioned by many survey respondents, and is a major impediment to implementing the Freight Data Exchange Network. Providing a data upload, even if it is relatively easy, takes time and effort by the data provider, and there is little or no direct recompense for this effort. This is particularly true for private sector firms. Virtually all of them commented on the lack of identifiable benefits to them of cooperating with the public sector, and the different time scales at work in the public and private sectors.

6.1.4 What are the concerns and potential impediments to participation for users and providers?
As noted above, the principal private sector concerns are release of business sensitive data, and little or no identifiable return on investment. The principal public concerns are user fees and the level of effort required for participation.

6.1.5 What data would be desired and available?
Public sector agencies most often requested reliable and timely commodity origin-destination data by mode, preferably at a county-level or finer geographic scale. Links to government data sites should be included. In addition virtually any other data about freight would be welcome. The private sector was less specific, but would welcome the opportunity to use freight data for market and competitive analysis. What is not clear is if the requested data could actually be made available, given data use restrictions and the likely sparse private firm participation.

6.1.6 How would metadata be made available in the network?
In general the user group oversight board would promulgate metadata standards, metadata would be provided by the user as part of the input process, and would undergo quality control by the system operator. More details are provided in section 3.

6.1.7 How would the exchange network be implemented?
Since there is no organization empowered to set up such a Freight Data Exchange Network the initial implementing step is not clear. The early needs are to establish an oversight board and select a system custodian. Who will do this? This concept feasibility study was requested as a research task through the National Academies/Transportation Research Board Cooperative Research Programs process. It might take a full blown NCFRP or NCHRP project to take the next actions to actually implement the concept. A key charge to any such study should be to identify a willing system operator, and an initial set of system users.

6.1.8 What requirements and costs would be associated with establishing and maintaining the exchange network?
System requirements are detailed in section 5. While no specific cost estimate was made, cost categories were discussed in section 4.2. There will be a relatively high initial cost for system development, a moderate second stage cost to publicize the system and get some initial participants, and a fairly modest cost (a part-time staffer, perhaps at the level of one day per week) to operate the system.
6.1.9 What would stewardship of the exchange network involve? Who might be willing and able to take this role? What are potential impediments to stewardship?

A number of scientific and technical organizations have the capability to implement the Freight Data Exchange Network, and might be willing to do so. The principal impediments are securing stable funding, establishing a user base, and dealing with concerns about real or perceived competition with the private sector.

6.2 Conclusions

It is not clear at this point if the Freight Data Exchange Network would succeed, at least not in its originally proposed form. It might be that a less ambitious exchange network, involving only public sector agencies, would be a good starting point. This would allow state DOTs and MPOs to have an internet site for sharing their own freight data and best practices, which would eventually create a large and growing freight data user community actively engaged in assisting each other. Developing and operating such a system would also solve initial technical problems and instantiate a design concept. This would provide experience that would be valuable for future expansion to include the private sector. At the same time it would be valuable to use system resources to begin working on private sector participation. An early activity should be to identify some possible public-private data sharing partnership opportunities.
Appendix A: Interview Guides

Data User Interview Questions

1. Has your organization conducted a study with specific consideration of freight traffic?
   a. If yes, what modes were included?
      __Truck
      __Rail
      __Water
      __Air
      __Intermodal
      __Other (please identify)

2. Have you collected any freight data, other than routine vehicle classification counts?
   a. If yes, what did you collect?
   b. What did you do with it?
   c. Would you be willing to share your data with states and MPOs?
      i. Would you require confidentiality or other assurances?
      ii. Would an easy-to-use internet upload site be of value to you?

3. Do you use freight data that is maintained by another agency in your state or region?
   a. If yes what kind of data?

4. What freight data would you like to have?
   a. Commodity origin-destination, tons by vehicle type
      __Truck
      __Internal trips
      __External trips
      __Through trips
      __Rail
      __Other modes (please specify)
      Time dimensions: year, season, month, day of the week, time of day
   b. Route or link data, by mode and vehicle type
      __Tons
      __Loaded vehicles
      __Empty vehicles
   c. Hazardous materials traffic
   d. Cost of freight shipping
   e. Fuel consumption
   f. Emissions
   g. Accidents
   h. Other

5. Would you like to obtain freight data that have been collected by states and MPOs, or provided by third parties such as transportation companies or shippers?
   a. Would an easy-to-use internet download site be of value to you?
   b. Would you be willing to pay fees to use such an internet data-sharing site?
   c. Should usage fees be waived for those who have uploaded data to the site?
   d. Would you sign license agreements or usage restrictions to use an internet data sharing site?
6. Some other questions related to a possible internet-based freight data exchange:
   a. Should the site contain anything besides data (e.g., methodologies, case studies)?
   b. What obstacles do you see that would prevent you from using the site?
      i. How can they be overcome?
   c. What incentives would motivate you to use the site?
   d. What legal protections can be used to ensure insulation from release of competitive or self-incriminating data, which would reduce some of the obstacles to collecting data from private firms?
   e. What kinds of partnerships could enhance freight data availability?
   f. Are there benefits from freight data exchange partnerships beyond the actual data sharing?

7. Do you have any other suggestions about how to organize and operate a freight data exchange network?

THANK YOU
Potential Data Provider Interview Questions

1. Has your organization been asked to provide data to a state or metropolitan area for a study with specific consideration of freight traffic?
   a. **If yes**, did you provide any data?
      i. If yes, what data?
         1. Did you attach any usage restrictions?
            a. What were they?
            b. Was a formal agreement signed?
         2. Did you receive any feedback from the recipient on what they did with your data, and what resulted?
   b. **If no**, why not?

2. Do you routinely maintain data that would be useful for tracking freight transportation activity (e.g., freight vehicle movements, shipment origins and destinations, freight vehicle routes, empty vehicle movements, fuel consumption, transportation charges)?
   a. If yes, what do you collect?
   b. How long do you keep it?
   c. What do you do with it?
   d. Are the data readily accessible in electronic formats?

3. Would you be willing to share your data with states and MPOs?
   a. Would you require confidentiality or other assurances?

4. What obstacles do you see that would prevent you from sharing your data?
   a. How can they be overcome?
   b. What legal protections can be used to ensure insulation from release of competitive or self-incriminating data, which would reduce some of the obstacles to sharing your data?

5. Would an easy-to-use internet upload site make it easier for you to share your data?
   a. What incentives would motivate you to use the site?

6. What kinds of partnerships could enhance freight data availability?

7. Would you like to obtain freight data that have been collected by states and MPOs, or provided by third parties such as transportation companies or shippers?
   a. Would an easy-to-use internet download site be of value to you?
   b. Would you be willing to pay fees to use such an internet data-sharing site?
   c. Should usage fees be waived for those who have uploaded data to the site?
   d. Would you sign license agreements or usage restrictions to use an internet data sharing site?

8. Are there benefits from freight data exchange partnerships beyond the actual data sharing?

9. Do you have any other suggestions about how to organize and operate a freight data exchange network?

THANK YOU
Potential System Steward Interview Questions

Brief description of the freight data exchange network concept:

Potential users of freight data have difficulty identifying the data that are available as well as the quality and other important attributes that affect their ability to productively use the data at a reasonable cost. Conversely, data originators and providers lack systems to inform potential users about what data are available and how that data might be accessed. One suggested means of overcoming these limitations is to establish a cooperative data exchange network, modeled on the open source data repository (or “wiki”) concept, where freight data and analyses could be shared efficiently and securely over the internet.

1. Would your organization be interested in hosting and operating the freight data exchange network? Some typical tasks would include:
   - Web site design and maintenance
     - Facilitate chat rooms and community help options
     - Automatic email notification of data additions and updates
     - Wiki-based documentation, reports, and case studies
     - Links to proprietary data sites/vendors
   - Develop a data archive.
   - Develop a metadata system.
   - Accept data uploads.
   - Service data queries and download requests.
   - Monitor user comments and suggestions, implement improvements.
   - Maintain a user registry.
   - Administer any user pay features.
   - Administer data security and confidentiality requirements.

2. How would you implement the exchange network?

3. How would metadata be made available in the network?

4. What requirements and costs would be associated with establishing and maintaining the exchange network?

5. If user fees are required, do you have the systems in place to accept payments?
   a. Should usage fees be waived for those who have uploaded data to the site?
   b. Do you have the mechanisms in place to handle fee waivers?

6. If data usage restrictions are necessary do you have the means to implement them?

7. What other tasks would stewardship of the exchange network involve?

8. What are potential impediments to your taking on this stewardship role?
   a. How can they be overcome?

THANK YOU