Online California Freight Data Repository for Freight Modeling and Analysis

Yeow Chern Andre Tok, Miyuan Zhao, Joseph Y.J. Chow, Stephen G. Ritchie, Dimitri Arkhipov, Institute of Transportation Studies, University of California, Irvine, CA 92697-3600. Corresponding author: Yeow Chern Andre Tok, ytok@uci.edu.

Background

The importance of understanding freight demand is increasing with the appreciation of freight transportation’s multi-faceted impact on our economy. There is a significant need to access a wide array of data sources for freight modeling and analysis. However, current data sources are not always easily accessible or obtainable even with the availability of the Internet. There are several reasons for this:

- First, freight related data is collected and stored by different agencies such as U.S. Census Bureau, Federal Highway Administration (FHWA), Bureau of Transportation Statistics (BTS), State Department of Transportation (DOTs), Metropolitan Planning Organizations (MPOs). This results in different types of web user interfaces, which organize the data in various ways. Hence, users who are unfamiliar with a particular web-based resource may expend significant effort in locating the desired data.

- Second, the definition of data is critical for robust model design. Data dictionaries help to define the types of measurements, units of measure, scope of the measurement and geographical association of the data source. However, data dictionaries are not always explicitly provided. Furthermore, data sources of different origins with similar fields may have differing underlying definitions that may not be apparent to the user.

- Third, a variety of data formats are adopted by various data sources. Not all formats are easily convertible from one to another. This creates potential incompatibilities across data sources for modeling purposes which, to be corrected, may require significant effort by the modeler.

- Fourth, the data provided is usually text-based. Text representation does not effectively convey the scope and the usefulness of the information. However, data visualization is typically a protracted task which requires acquiring the data and setting it up using charting software. If geographical representations are desired, further effort is needed to display the data in geographical information systems-based (GIS-based) software and to acquire the associated geometric data.

The Online California Freight Data Repository (http://moon.its.uci.edu/FDR/) is an example of a freight data assessment and compilation framework that addresses the needs identified above for the State of California. It is a user-centered online tool that provides convenient access, standardized interface and
a centralized location for obtaining freight data from a variety of public sources for purposes of regional and statewide freight transportation planning and analysis. Data dictionaries and lookups associated with each data source are provided and data tables associated with each source can be downloaded in a variety of data formats to accommodate user needs. Analysis tools are also available to provide a convenient evaluation and visual assessment of data sources. Data sources that are relevant but not stored in the repository can be accessed via external links which are also provided within the repository. Lastly, a built-in Google-based freight related custom search engine allows users to easily locate other freight-related information.

System Architecture

The data repository system consists of two main components: the Freight Repository Database Server and Freight Repository Web Server as shown in Figure 1. Communications between these components facilitates the exchange of information between users and the Freight Data Repository Server to provide the requested data through user-friendly inputs.

The web-based architecture was chosen as it offers the highest end-user compatibility and the least user-system requirements as it operates through existing web browsers such as Mozilla Firefox and Google Chrome and does not require any application download or installation. The web-based data repository was designed with an advanced user interface which features the Asynchronous JavaScript and XML (AJAX) technology, enabling asynchronous data retrieval from the server in the background. This design approach offers improved interactivity and dynamism of the user web interface.
**User Interface and Modules**

The Online California Freight Data Repository user interface was designed with an emphasis on enhanced user interactivity as well as intuitive graphical displays for ease of use. It consists of five modules: Data Source Query, Time Series Analysis, Spatial Analysis, External Data Sources and Freight Data Search. Each module is accessible via a tab located near the top of the interface.

A unique feature of the web-interface is the ability to access all modules without navigating away from a single web URL. This allows the user to return to the current states of other modules by simply clicking on the navigation tabs near the top of the page (see Figure 2).

![Freight Data Repository Home](image)

**Figure 2**  
Freight Data Repository Home

**Data Source Query Module**

The Data Source Query module (shown in Figure 3) allows the user to access all the raw data tables for a core set of data sources found to be most relevant for California regional and statewide freight modeling and analysis. The data ranges from national to local sources, for any year that is available (mostly ranging between 2002 and 2007—the two most recent benchmark years for several economic data sources).

![Data Source Query interface](image)

**Figure 3**  
Data Source Query interface
Eighteen freight-related data sources are available for download. The data can be organized into a variety of formats for software compatibility such as tab-delimited, comma separated values (csv) and hypertext markup language (html). Data dictionaries and lookup tables, which provide definitions for fields and data found in each data source, can also be downloaded through the Data Source Query.

**Time Series Analysis Module**
The Time-Series Analysis module allows users to observe trends in data using graphical representations. The data fields available for time-series analysis are hierarchically organized by geographical aggregation and data source in the user input panel. Results are displayed when all four selections – Aggregation, Data Source, Field and Location – are filled, as shown in Figure 4a.

**Spatial Analysis Module**
The Spatial Analysis module allows a user to visualize the data using color scales at a spatial level by associating spatial data with their corresponding GIS shape data overlaid on the Google Maps interface. Upon selecting the desired geographical aggregation, data source, data field and time period, the spatial analysis tool uses the Google Map interface to generate geometric polygons of each region, thematically shading each from white to vivid red reflecting the size of values from the selected data field as shown in Figure 4b. Regions or locations with no data values are shaded in grey.

![Figure 4](image1.png)  
**a. Time Series Analysis** 
**b. Spatial Analysis**

**Data Analysis Tools Sample Results**

**External Data Sources Module**
The External Data Sources Module provides information on data sources that are not available directly from the Freight Data Repository. The data sources are divided into two categories: public and private data sources. External public data sources are organized by their geographical scope into four categories: National Data Sources, California Data Sources, MPO Data Sources and Intermodal Facility Data Sources. Each data set is linked to an external URL where the data can be obtained. A list of private data sources affiliated with freight is also provided in this module.
Freight Data Search
The Freight Data Search consists of a custom search engine, called FREDDY, developed using Google technology. It currently incorporates 45 websites from around the country related to freight and goods movement, providing users with the convenience of locating other freight-related information.

Applications
The Online California Freight Data Repository serves a wide array of applications related to regional and statewide freight modeling and analysis, such as freight facility investment planning, land use development, environmental impacts, economic development, security and hazardous material planning, supply chain design, and safety impacts. Several freight applications were used as test cases to see if the Online California Freight Data Repository could address their data inputs. In particular, the following four applications representing different types of models or analytical methods were considered, summarized by the data input requirements that were met with the repository.

1) Freight network optimization: the Cal-GIFT model
a. Highway, railway, waterway networks, transfer facility locations
b. Truck/rail/vessel emissions, energy consumption from EPA
c. Port/rail terminal volumes, goods types transferred, handling delay
d. PeMS data
e. Commodity flow OD’s (FAF2)

2) Freight data analysis: Commodity Flow Disaggregation
a. FAF2
b. Socioeconomic data: County Business Patterns, Census Populations, Quarterly Census of Employment and Wages

3) Metropolitan Planning Organization Goods Movement Studies (SCAG)
a. PeMS data
b. Truck VMT/HPMS
c. FAF2
d. VIUS

4) Land use model integration (Caltrans PECAS)
a. Highway, railway, waterway networks, transfer facility locations
b. FAF2
c. Socioeconomic data: County Business Patterns, Census Populations, Quarterly Census of Employment and Wages

Acknowledgement
The Online California Freight Data Repository was developed under contract to the California Department of Transportation, Division of Transportation System Information. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the state of California. This paper does not constitute a standard, specification, or regulation.