



A Data Library Management System for Midwest FreightView and its Data Repository

Final Report

By

Peter Lindquist, Ph.D.

Associate Professor

Department of Geography and Planning

The University of Toledo

**Prepared for The University of Toledo University Transportation Center
and the U.S. Department of Transportation**

March 2011

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Executive Summary

Midwest FreightView (MWFV) and its associated data repository is part of a large multi-faceted effort to promote regional economic development throughout the Great Lakes system. The main objective for the system is to promote sustainable maritime transportation in the region by serving as an information clearinghouse for public policy decision making and for drawing the link between maritime freight movements and economic impacts throughout the region. In effort to maintain consistency over time within the MWFV system this project was devoted to the development of a set of formalized data library functions that include a complete set of user guides, technical manuals, and training modules for using the system, a detailed inventory of data repository contents, and complete documentation of repository contents according to national metadata standards. In addition, the project established a formally-defined directory structure for data storage, backup and security protocols, and the development of formal procedures for any updates, additions, or edits to the repository contents.

Previous phases of MWFV emphasized data collection, development of analytical tools, and customization of the geographic information system (GIS) platform. The bulk of this work has been devoted to assembling, reconciling, and integrating diverse data from a multitude of sources into a single comprehensive repository that provides access to millions of records within a wide range of datasets. This phase ensured that a sense of continuity is maintained in the project by building an “institutional memory” for the repository and delivery system. Specific outcomes resulting from this project are summarized as follows:

- Comprehensive Users Guide – for users to gain the necessary expertise to effectively use the system and to familiarize themselves with the contents of the repository;
- Complete Data Management Plan – for the project team to have a clearly defined, systematic approach for integrating new data into the system, for documenting updates, for maintaining data security, formalizing the repository directory structure, and for backing up data within the system to remote platforms;
- Contents Directory – for the project team to identify gaps in the database through the inventory process.

By providing current, detailed data coupled with analytical tools, MWFV contributes to the mission of the UT-UTC by providing a comprehensive resource to transportation analysts, regional stakeholders, public officials, and other interested parties in both the public and private sectors. In turn, this phase of the project provides greater utility to transportation professionals by organizing, documenting, and setting maintenance standards for this resource. The development of this Data Library Management System will thus enable users to take advantage of the strengths of MWFV.

Introduction

This project is a continuation of a long-term endeavor to develop and manage a comprehensive data repository and information delivery system for intermodal freight transportation in the Great Lakes Region. This system, entitled ***Midwest FreightView*** (MWFV), is designed to serve as a central focus for studying freight movements among all major modes in the Midwest. In contrast to previous work carried out on the repository, this project focused on developing a set of formalized data library functions that include a detailed data repository contents inventory, complete documentation of the repository contents according to national metadata standards, and a complete set of user guides, technical manuals, and training modules for using the system. In addition to these documentation efforts, the project includes the establishment of formalized directory structures for data storage, formal backup and security protocols, and the development of documentation procedures for any updates, additions, or edits to the repository contents.

The rationale for this project is based on several factors, the most important of which is to maintain an “institutional memory” for the repository and delivery system. This project has now evolved to the point where the volume of data and the range of data obtained from diverse sources exceeds the capacity of any one project team member to retain a complete knowledge of the contents of data files, the location(s) of files, and status of the data with regard to currentness, completeness, and compatibility with other data sets in the repository. In the literature, Ferrari and Shankaranarayanan reflect on the importance of comprehensive data and metadata management systems [1, 2]. Furthermore, the project team experiences significant turnover of graduate assistants from year to year, which requires significant startup costs when new students join the project team. Formal documentation of the database contents provide an effective resource for students to rapidly familiarize themselves with the system. Finally, the project team’s experience in training workshops has revealed gaps in the current documentation on the repository contents and in the operation of the system. As this system gains wider recognition over the region and its usage increases, the development of this Data Library Management System is a vital step toward its long-term viability. The development of these detailed, effective training manuals and help functions are essential for this system to be utilized by regional transportation professionals over the long-term.

Background

MWFV and its related web resource, The Great Lakes Maritime Research Information Clearinghouse (GLMRIC), were designed to furnish freight professionals with accurate and current data concerning the status of transportation infrastructure, vessel and commodity flows, patterns of economic activity and the location and size of markets for commodities and goods shipped through the system. The data within the repository provides a comprehensive picture of vessel traffic, commodity flows, intermodal connectivity and provides a linkage between the regional freight transportation system and the regional economy.

The concept for the original MWFV system was developed by the University of Toledo team in partnership with the Upper Midwest Freight Coalition (currently Mid American Freight Coalition). Additional contributing partners include the Great Lakes Maritime Research Institute at the University of Wisconsin-Superior, the UT-UTC, and the Center for Freight Infrastructure Research and Education (CFIRE) at the University of Wisconsin-Madison. In parallel efforts, the Toledo team is also contracted with the U.S. Army Corps of Engineers (Corps) to assist in the Federal Initiative for Navigation Data Enhancement (FINDE). In addition, the project team has recently finished a contract with the Corps to develop a web-based prototype data collection system for docks and facilities in the Great Lakes.

The MWFV system has been under development over the past five years with an emphasis on data collection and customization of the GIS platform. Up to this phase, the main emphasis of the project has been to assemble, reconcile, and integrate diverse data from a wide range of sources into a single comprehensive repository that provides access to millions of records within a host of datasets. In turn, the project also focused on the acquisition, management and exchange of data for dissemination to analysts within the industry. To this end, the project team developed GLMRIC – a central web location for dissemination of this information to users.

Presently, efforts have extended to developing a new MWFV dataViewer and a complete set of multifunctional analysis tools to be integrated into the system. It is anticipated that this system will provide a foundation for modeling flows between modes and for evaluating transportation alternatives according to their environmental, economic, and transportation system load impacts. Specific analysis tools envisioned for the system will include time-sensitive vehicle routing, intermodal routing, facility location and site selection, market identification and delineation of catchment areas, economic indicators, flow mapping, and spatial econometric analyses. Thus the value of this resource for transportation professionals increases significantly with the effective documentation of a contents directory, users' guide and standardized data update protocols to ensure consistency over the long-term.

Data Library Management System

Objectives

The objective of this project was successful to develop a formal data library management system for Midwest FreightView and its associated data repository. As previously mentioned, the outcomes from this project are threefold:

- for users to gain the necessary expertise to effectively use the system and to familiarize themselves with the contents of the repository;
- for the project team to have a clearly defined, systematic approach for integrating new data into the system, for documenting updates, for maintaining data security, formalizing the directory structure of the repository, and for backing up data within the system to remote platforms; and
- for the project team to identify gaps in the database by the inventory process.

The goal of this project – to provide a complete library of documents that can be readily downloaded by users from the project website – was attained. These documents include a detailed documentation of database contents and a users’ guide ranging from beginning users to more advanced users. In the area of data management, new requirements for data protocols have provided effective guidelines for the project team to follow, particularly with respect to training new graduate students and familiarizing them with the system. This is necessary for long-term consistency in the MWFV system for the project team and any future providers. Finally, data gaps have been reviewed and new data sets have been identified and incorporated into the system to fill some of these gaps.

In addition to meeting project objectives, this project provided an additional educational opportunity to a graduate student who worked under the direction of senior staff members. Time spent on this project enhanced the student’s educational experience in transportation and goods movement, as well as in data management. The student gained a broad perspective of the entire system by cataloging the data within it. It offered the student a hands-on opportunity to not only learn about available public and proprietary datasets, but also how they can be applied for transportation infrastructure system analysis. Excellent employment opportunities exist for students with experience and skills in these areas.

Methodology

The following tasks established in the project proposal were followed systematically to complete this project successfully:

- Task 1 - Create a users’ guide for the database that includes a detailed documentation of database contents and technique enhancement tips.
- Task 2 - Conduct inventory of contents for the current database.
- Task 3 - Develop complete metadata documentation for database contents.
- Task 4 - Identify data gaps in the repository.
- Task 5 - Develop protocols for managing and updating the database.
- Task 6 - Establish priorities for integrating data into the MWFV online GIS system.
- Task 7 - Set up procedures for data backup and updates to be carried out in a routine fashion.

A complete users’ guide was created to direct the user through step-by-step instructions on how to use the MWFV system. The users’ guide provides helpful insight to enhance the users’ ability to utilize the system effectively. It is available for download from the project website (pictured below) at www.maritime.utoledo.edu.

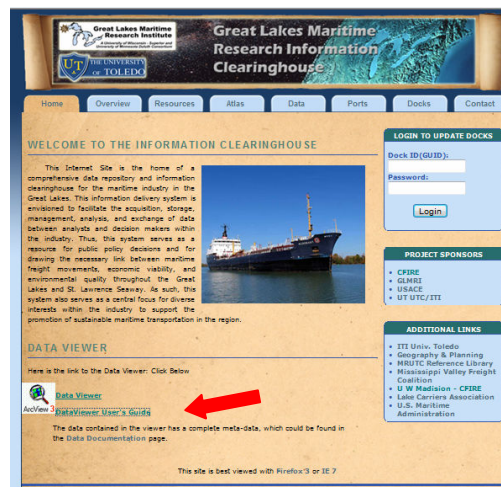


Figure 1. GLMRIC Home Page.
(link to User's Guide)

As new analytical modules are developed for MWFV, the necessary documentation will be added to the library. The graduate student and research staff working on this project carefully detailed the contents of the current database. This inventory was laid out and indexed for users to quickly locate datasets of interest. Metadata documents were created describing the exact contents of each dataset and linked to the database inventory. These features are now published on the project website. Upon completion of the first half of this project, the project team was able to identify gaps in the data. In other words, the team identified areas of the repository that either had very sparse datasets available or sections that had missing datasets that would be valuable to the MWFV system in order to provide a comprehensive analysis framework. The project team was able to find new public data sources for some of the data gaps identified during this process (see Table 1). The second half of the project timeframe was spent developing protocols for systematic data updates and backup procedures. These are documented for the use of future project team members or MWFV system providers. In addition, standardized procedures were set up for integrating new data into the repository. The final result is a systematic process that can be duplicated indefinitely to provide seamless integration of current datasets and to assure long-term system continuity. The electronic manual created here is a comprehensive 43 page document that details contents of the repository and metadata along with step-by-step procedures and responsibilities of student Research Assistants (RA) and the Data Manager. A copy is submitted along with this report.

Data Gaps	Hyperlink to Identified Sources
Highway Traffic Safety Data	National Center for Statistics & Analysis Crash*Stats Highway Performance Monitoring System Marine Transportation System Data inventory Annual Energy Review Railroad Safety North American Transborder Freight Data Truck Transportation, Messenger Services, Warehousing
Annual Service Survey	Agricultural Trade Data Transportation Energy Data National Transportation Stats American Recovery & Reinvestment Act
Transportation Investment	Border Crossing Data (BTS) North American Transportation Statistics Database Turner Fairbank Highway Research Center Operation Lifesaver (rail safety)
Oak Ridge National Laboratory	Center for Transportation Analysis
Environmental Impacts	EPA- transportation tools Clearinghouse

Table 1 – Identified data gaps with hyperlinks to new public data sources

Dissemination of Results

Technical transfer is of limited utility unless proper documentation is furnished to potential users of the system. The project team's experience in workshop training sessions for the MWFV system has revealed that users require clear, detailed, instructions for using the system. Users also tend to learn more effectively on their own, with tutorial sets, rather than interactive instructional sessions. As a result, the documentation function completed here is absolutely essential for effective technical transfer. This final product will contribute to the longevity of the MWFV system through a systematic approach for data updates and backups as well as integration of new data into the system. The new users' guide will enhance knowledge and applications for MWFV system users; this will ultimately assure effective technology transfer to stakeholders and decision makers for the intermodal transportation system. Given that the focus of this project is on documentation of the contents of the database and instructions regarding the use of MWFV, dissemination of results are based on making the documentation as accessible to users as possible. All data documentation and user's guide are thus available to users through the project web site. The project Data Management Plan is available electronically to all team members on a shared RA drive in addition to a printed copy in the GISAG lab.

Conclusion

To summarize, the overall purpose of the MWFV online GIS system is to promote more efficient, safe, and environmentally sustainable use of the freight transportation infrastructure in the Midwest region. This resource provides valuable information for sustainable economic development where users can study freight movements and commodity flows through the transportation network and relate these flows to the economic viability of the region. By providing current, detailed data coupled with analytical tools, the GLMRIC system contributes to the mission of the UT-UTC by providing a comprehensive resource to transportation analysts, regional stakeholders, public officials, and other interested parties in both the public and private sectors. Furthermore, the online format of the system provides an efficient framework for immediate dissemination of information.

This project produced a complete user's guide for the MWFV system available in a PDF format for download on the project web site. In addition, data documentation as well as metadata documentation are available for MWFV system users. In addition, a formal data management plan for data handling protocols has been created for the project team to proliferate system sustainability. As such, the project team was able to establish priorities for integrating new data into the MWFV online GIS system. Finally, completion of these processes provided the project team with additional insight to identify data gaps within the system. An index of these data gaps will be listed as a scope for further research within another project. The utility and sustainability of this resource to the community of transportation professionals is improved with properly organized, documented, and maintained data. The development of a Data Library Management System has strengthened the system by providing users with a clear understanding of its capabilities.

References

- [1] Ferrari, E. Access Control in Data Management Systems. San Rafael, Calif.: Morgan & Claypool, 2010.
- [2] Shankaranarayanan, G. 2004. "Managing Metadata in Data Warehouses: Pitfalls and Possibilities," *Communications of the Association of Information Systems*, 14: 247-274.