

1. REPORT NUMBER TPF-5 (108) Transportation Pooled Fund	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
4. TITLE AND SUBTITLE Software Tools for Sharing and Integrating GIS Data		5. REPORT DATE 06/29/2011
7. AUTHOR Washington State Department of Transportation		6. PERFORMING ORGANIZATION CODE
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lead State: Washington State Department of Transportation 310 Maple Park Avenue SE P.O. Box 47300 Olympia, WA 98504-7300		8. PERFORMING ORGANIZATION REPORT NO.
12. SPONSORING AGENCY AND ADDRESS California Department of Transportation Division of Research and Innovation, Mail Stop 83 1227 O Street, 5th Floor Sacramento, CA 95814		10. WORK UNIT NUMBER 191
15. SUPPLEMENTARY NOTES		11. CONTRACT OR GRANT NUMBER Transportation Pooled Fund (TPF)-5(108)
		13. TYPE OF REPORT AND PERIOD COVERED Final: June 2011
		14. SPONSORING AGENCY CODE

16. ABSTRACT

Computer software tools that facilitate local government geo-spatial transportation data sharing and integration, for the creation and maintenance of State transportation networks.

There is a critical need to develop, implement and distribute a variety of tools and process for sharing and integrating statewide geo-spatial transportation data. There are various obstacles to development and maintenance of such a statewide network using data from multiple of sources, both organizational and technical. This study looked at both, but focused on research to help develop processes and software utilities to overcome the technical obstacles.

The basic research objectives included:

- Design and implementation of a core database, based on federal standards, within the state as a central repository of multi-jurisdictional location data.
- Translation of the data from provider's schema to a centralized database and from one database to another.
- Integration of data from disparate data sources into a seamless whole.
- Quality Assurance/Quality Control processes and software to monitor data quality, security, data entry and retrieval processes.

17. KEY WORDS software tools, database, data	18. DISTRIBUTION STATEMENT No restrictions.
19. SECURITY CLASSIFICATION (of this report) Unclassified	20. NUMBER OF PAGES 10
	21. COST OF REPORT CHARGED

TPF 5(108) Final Report

Software Tools for Sharing and Integrating GIS Data

Executive Summary

Development of a statewide transportation network, with roads and other transportation modes has become a critical need for many agencies and while there are various options for how this development is handled the most cost effective way over the long-term is to develop this network using data maintained and provided by local governments about the portions of the network they maintain. Thus data is developed and maintained at the location where the most information is available and then shared for the creation and maintenance of a statewide transportation network.

There are various obstacles to development and maintenance of such a statewide network using data from multiple of sources, both organizational and technical. This study, TPF-5(108) Software Tools for Sharing and Integrating GIS Data, looks at both, but focuses on research to help develop processes and software utilities to overcome the technical obstacles.

Organization and Oversight

A consortium of public and private entities was established for the purpose of researching, developing and implementing computer based tools that facilitate geo-spatial transportation data sharing and integration for a variety of purposes. The Geo-spatial Integration and Sharing Data Consortium (GISDC) is funded by a Transportation Pooled Fund arrangement and managed by the Washington Department of Transportation (WSDOT) through the Washington State Transportation Framework Project (WA-Trans). The states participating in this pooled fund include; California, Idaho, Nebraska, Ohio, Oregon, Tennessee and Washington.

Objectives

The goal of the consortium is to develop, implement and distribute a variety of tools and process for sharing and integrating geo-spatial transportation data. The basic objectives are:

- Design and implementation of a core database, based on federal standards, within the state as a central repository of multi-jurisdictional location data,
- Translation of the data from provider's schema to a centralized database and from one database to another,
- Integration of data from disparate data sources into a seamless whole,
- QA/QC processes and software to monitor data quality, security, data entry and retrieval processes,
- Change detection, and change management to support maintenance of the data over time,



-
- Documentation of a set of processes necessary to support data sharing from a variety of sources, e.g. data sharing agreements, agreement points,
 - Linear Referencing Integration.

Results

At the Annual TPF partners (GISDC), Meeting in September 2010 the software research was prioritized. The software and processes of most interest were identified for continued research. This prioritization was necessary due to the funding reality that this research project would run out of funds by June 30, 2011. The goal was to complete the software, identified as the top priority, and do as much as possible on the remaining software. There was a focus on delivery of this software to each partner to include documentation specific to each state DOT environment.

The prioritized software and processes identified for continued research met all of the original TPF study objectives. The research priorities formulated at the Annual GISDC Meeting in September 2010 were met and all process and software packages were delivered.



TPF-5(108): Software Tools for Sharing and Integrating GIS Data: Products

The WA-Trans Transportation Pooled Fund effort, TPF-5(108) study: Software Tools for Sharing and Integrating GIS Data is complete. The processes and software identified for research were delivered via a set of TPF Packages, to the TPF partners, California, Idaho, Nebraska, Ohio, Oregon, and Tennessee. Final deliver occurred on June 29, 2011. The Processes and Software delivered are listed with a brief description in this document.

TPF Package Contents:

All TPF packages are structured similarly.

All TPF packages include:

- 1.) A READ ME document describing the contents of the package. The basic contents of these “READ ME” documents are included in this document.
- 2.) A description document for each identifiable process that includes a table at the beginning of the document listing; referenced project documents, applicable software and scripts, any Quick Start Guides, and other items included within, or referenced in that document.
- 3.) Quick Start Guides. It is a goal of project delivery to enable all partners to create an environment enabling them to test the software. WA-Trans is hoping you will be able to implement all software, in a test mode, and make it work enough to better understand the processes and further to help understand the changes that may be necessary to implement this totally in your environment. The Quick Start Guides walk partners through detailed steps that will create this kind of environment. Except for the WA-Trans Global Quick Start Guide, the Quick Start Guides are located within related documents and not provided as separate independent documents.
- 4.) Opportunities for Improvement: All document will have a section that includes what we wanted to do, but did not; ways to possibly improve the process and or the software etc.
- 5.) A Folder structure that includes software, documents, scripts, data store templates and other files related to the processes described in the package documents.

Another goal of WA-Trans was to design processes that were as software agnostic as possible so as to make implementation if a variety of different environments as possible. The processes are at least as important as the software, and possibly more so.

The WA-Trans Global Quick Start Guide can be followed to implement the entire set of WA-Trans processes. This WA-Trans Global Quick Start Guide will reference other more specific Quick Start Guides included in all packages.



The “Acquire, Standardize, Load, and Evaluate Provider Data” Package:

Documents included in this package describe creating a relationship with data providers, the acquisition of data, connectivity between provider data at jurisdictional boundaries and loading of this data into a WA-Trans database. Documents located in the Document Folder include:

- 1.) Providing DataToWA-Trans.doc, Meeting a data provider and getting familiar with their data.
- 2.) WA-TransAttributeCrosswalk.xls: The crosswalk file a provider fills out to relate their data to WA-Trans attributes (discussed in the ProvidingDataToWA-Trans.doc).
- 3.) WA-TransAgreementPoints.doc: Discussion and case study for the agreement and disagreement point process (connectivity between data from different providers within and between jurisdictional boundaries).
- 4.) WA-TransProviderDataLoader.doc: Describes in detail the ProviderDataLoaderWorkspaceTemplate.fmw: (found in the FMEWorkspaces folder). Includes the Provider Data Loader Quick Start Guide.

NOTE: All documentation refers to related information in other documents. Some of these other documents may be included in other WA-Trans packages.

Database:

Database Models and create scripts are included in the separate WA-Trans Database Package. The Quick Start Guides for the Loading and Production database referred to in the Provider Data Loader Quick Start Guide are included in the WA-Trans Database Package.

FME Workspaces:

Safe Software Feature Manipulation Engine (FME) workspaces are included in this folder.

- 1.) **The ProviderDataLoaderWorkspaceTemplate.fmw:** This workspace can be used as a template to create provider data loaders for providers in your environment. This workspace is also designed to be used to test the loading process according to the Provider Data Loader Quick Start Guide.
- 2.) **Custom transformers:** Custom transformers are used in the ProviderDataLoaderWorkspaceTemplate.fmw and are separate files that need to be placed on our computer. Refer to the Provider Data Loader Quick Start Guide for details.
- 3.) **WATransAdditionalProcessesRealWorldExamples.fmw:** This is a workspace that includes examples of processes that could be used when transforming provider data. Refer to the WA-TransProviderDataLoader.doc, Appendix B: Additional Useful FME Process Examples.



The “WA-Trans Database” Package:

Data Definitions and Descriptions:

The data definitions are included in the “**WA-TransAttributeDefinitions.xls.**” Refer to the “**WA-TransDataMetadata.doc**” for explanations of how data is used and configured. NOTE: The “**WA-TransDataMetadata.doc**” document is referenced by almost every other WA-Trans TPF document. The “**WA-TransDataMetadata.doc**” can be considered a Global WA-Trans TPF Package document.

Database Structure:

The Washington State Transportation Framework (WA-Trans) database infrastructure and schema comprise three distinct databases. Refer to the “**WA-TransDataMetadata.doc,**”

Appendix A.

- 1.) The Loading database is where a provider’s data is initially placed after it is transformed from the provider’s data schema into the WA-Trans database schema using an FME translation workspace.
- 2.) The Staging database is where provider’s data is integrated with other data during the process of creating a statewide integrated transportation dataset for Washington State.
- 3.) The User/Production database is where the statewide integrated transportation dataset for Washington State was to be placed in order to be available to data users.

Data Models:

All WA-Trans databases are very similar, but there are differences related to purpose and functionality. Refer to the data model PDF documents located in the DataModel folders. Also, refer to the “**WA-TransTableListMatrix.xls**” for an overview of the attribute differences. The models and the “**WA-TransTableListMatrix.xls**” are designed easily to reference each other.

Database Scripts:

Located in the DatabaseScripts folders in each LoadingDatabase, StagingDatabase, and ProductionDatabase folders are scripts to create those databases. All the scripts to create the database Tables, Stored Procedures, Functions and Views are included. In addition, there are scripts to insert lookup table values. The creation of the database and relationship to your geospatial software is not included in this package. That process will need to be done by an ESRI SDE database administrator.

In order to create an environment to leverage the other WA-Trans Quick Start Guides do not make any major changes to the database scripts. NOTE: Due to your environment, slight changes may be necessary. These changes may not affect the Quick Start Guides but should be noted at the time for debugging purposes.



WA-Trans tested using a Microsoft SQL 2005 database.

Microsoft SQL Enterprise Edition (64bit)

Collation: SQL_Latin1_General_CP1_CI_AS

Is Clustered: False

This is a spatial database using ESRI SDE 9.2

For the Loading database only SegmentGeometry is a spatial table and **none** of the other tables are registered with the SDE. None of the Loading database views are spatial.

Refer to the “**WA-TransDataMetadata.doc**” sections 1.1 – 1.6 for spatial information.

The “Change Detection and Management” Package:

Change Detection:

The automated processes that compare newly received and loaded provider data with existing, integrated data from the same provider are described in the document “**WA-TransChangeDetection.doc**.” This document describes in detail the methods implemented by the accompanying FME workspace “**ChangeDetection.fmw**” as well as the SQL script “**ChangeDetection.sql**.” Usage instructions are also included in the document.

Change Management:

The propagation of Change-Detected data from the Loading database into the Staging database is described in the document “**WA-TransChangeManagement.doc**.” There is also a spreadsheet “**ChangeManagementFieldMapping.xls**” that describes the mapping of every field for every Change Detection code for each of the three tables of provider data. Finally, the three FME workspaces “**ChangeManagement_Geometry.fmw**,” “**ChangeManagement_Address.fmw**,” and “**ChangeManagement_Route.fmw**” are included to perform the database synchronization and their function and usage is fully described in the document.



The “Data Integration” Package:

Integration Pre-Processing:

The automated processes that initialize and prepare data for manual review and processing, the document **“WA-TransIntegrationPreProcessing.doc”** describes the context, the processes, and the usage. The accompanying FME workspace **“Create-Assign Reference Points.fmw”** is used to generate Reference Points and associate them with features and records from other core tables.

Integration Processing:

The detailed workflow and methodology of the manual processes by which a provider’s dataset is examined and edited, is described fully in the document **“WA-TransIntegrationProcess.doc.”**

Change Integration Pre-Processing:

To integrate updated data from providers who have already submitted data to WA-Trans, additional processes are required to prevent massive duplication of effort. These automated processes use Change Detection statuses, spatial analysis of proximity of a provider’s updated features to other providers’ data, and processing history of the provider’s existing data to automate the integration of as many records as possible and to flag the remaining records to ease the subsequent manual processing. The document **“WA-TransChangeIntegrationPreProcessing.doc”** describes the context, the processes, and their usage. The accompanying FME workspace **“ChangeIntegrationPreProcessing.fmw”** is used to generate identify and flag updated records to indicate their proximity to other providers’ data, and the SQL script **“ChangeIntegrationPreProcessing.sql”** uses those flags to finalize the processing of some records and to re-flag others.

Change Integration Processing:

The workflow and methodology of the manual processes by which a provider’s remaining, flagged, updated data are examined and edited are described in the document **“WA-TransChangeIntegrationProcess.doc.”**



The “Data Extraction Production Data” Package:

Documents:

The “**WA-TransDataExtraction_ProductionData.doc**”

Database Model and SQL Scripts:

In the “Database” folder there is a “DatabaseModel” folder and a “DatabaseScripts” folder. The model is the “**WA-TransStagingCoreERD.pdf**” also found in the WA-Trans Database package. The scripts are the same scripts found in the WA-Trans Database package, but only the ones you will need for creating the integration databases according to the “Create a WA-Trans SQLExpress Database Quick Start Guide” in the “**WA-TransDataExtraction_ProductionData.doc**”.

FME Data Extraction Workspaces:

In the “FMEWorkspaces” folder, you will find three FME Workspaces:

- 1.) All Pertinent Records Duplicate Geometries.fmw
- 2.) All Pertinent Records Many-To-One relationships.fmw
- 3.) Many-To-One with Points_FullAttributeNames.fmw

Directions for use are included in the comments within each workspace.

Data Store Templates:

The FME data extraction workspaces can load data in a variety of formats. WA-Trans only included the formats for data requested for the counties we processed. These templates are located in the “DataStoreTemplates” folder.

In the DataStoreTemplates folder, there are the following folders each with a database template and a metadata file template. These can be used with the FME Data Extraction Workspaces to store the extracted data.

- FileGeodatabase_MultiGeometries
- FileGeodatabase_No_Lookups
- FileGeodatabase_with_lookups
- PersonalGeodatabase
- Shapefile

Metadata files:

Various metadata files are included in this folder. Included are:

- 1.) FGDC standard HTML metadata file templates.
- 2.) Data Structure documents describing the data store structures.
- 3.) WA-Trans Processing and Processing Tasks Overview.doc, describing the processes the extracted data were subject to.



Pre-Processing Quality Assurance Quality Control Package:

Documents:

The “**WA-TransDataQAQC.doc**” details the Quality Assurance and Quality Control of data prior to data loading, after data loading, reference QA/Qc in the “**WA-TransIntegrataionProcess.doc**”, after data is processed and reviews address and route quality.

Data Loading Check:

A SQL Script to run right after data loading to make sure important attribute values are assigned correctly. These important attribute values are necessary for further process and successful data loading, to include necessary attribute values, can be affected by changes in the provider data loaders.

Final Event Placement and Analysis:

The “Final Event Placement and Analysis” folder contains FMEWorkspaces and SQL Scripts that check for accuracy of the Route and Address values on the production ready (processed) WA-Trans data. These processes perform the QA/QC on the entire dataset, not just a sample.

Network Connectivity:

The “NetworkConnectivity” folder contains FME workspaces and SQL scripts that create a network connectivity dataset that can be reviewed for connectivity and networking issues in the data.

Provider Reports:

The “Provider Reports” folder contains SQL scripts that compile data for a Provider data report. A Provider Report Template is also included.

Business Needs and Agreements Package:

Business Needs:

The “**WA-Trans Bus Needs Specs January 2008.doc**” details the WA-Trans Business needs. There was an extensive business needs process performed. Also included are spreadsheets detailing and prioritizing these needs prior to completing the business needs document.

Agreements:

Although no final agreement documents we approved, there were some drafts. The MOU is intended to precede a Data Sharing Agreement.