Submitted to:

Pennsylvania Geospatial Data Sharing Standards (PGDSS) V 2.5

Spatial Data Migration Tools Report

May 2, 2007

Table of Contents

1.0 Document Purpose

The Pennsylvania Mapping and Geographic Information Consortium's initiative to develop geospatial data sharing standards for Pennsylvania has recently culminated with the Pennsylvania Geospatial Data Sharing Standard (PGDSS) Version 2.5. In order to encourage and support use of this latest version PAMagic has commissioned an overview report of the primary tools available to perform geospatial data migration. This report will provide a reference for organizations who wish to migrate to, or provide their geospatial data in, the PGDSS format.

In order to comply with the PGDSS data model, spatial data extract, transform and load (ETL) techniques are required for the diverse spatial data owners and maintainers in Pennsylvania. Spatial ETL tools allow GIS users to extract spatial data from their native data model environment, map features and attributes to a new data model (PGDSS), transform the data to align with the new model (PGDSS) and load it into the new data model. This report provides an overview of the spatial ETL process and reviews several popular tools available to perform spatial ETL. Also included are the licensing options for the commercial spatial data migration tools.

2.0 Overview of Spatial Data Migration Tools

2.1 Spatial Data Migration Defined

The goal of spatial data migration, also known as GIS interoperability or Spatial ETL, is to move data from one datastore to another using three separate steps - extract, transform and load. Extracting spatial data involves reading specific data from a specified source datastore and extracting it. Transforming processes the acquired data and packages it for the data model of the target datastore. The Load function writes the resulting data to the target datastore. For PGDSS, the transformation processing may involve conversion of coordinate systems and projections, spatial transformations, topological transformation and migration, conflation, or combination of source spatial data attributes in order to map to the PGDSS data model. By using these spatial data migration techniques, disparate spatial datasets maintained by multiple organizations and agencies can be combined into a common data model providing a seamless view of Pennsylvania spatial data.

2.2 Spatial Data Migration Uses

Spatial data migration tools are generally applied to one or more of the following uses. **Data Cleanup** for removing errors within a dataset or during migration. **Data Merging** involves bringing multiple datasets together into a common framework. This is also known as conflation. **Data Verification** is the comparison of multiple datasets for quality assurance purposes. **Data Translation** provides for the conversion of spatial data from one format directly to another with no intended change in structure or schema. **Supporting Legacy Applications** involves maintaining two applications prior to replacing a legacy system. **Initial Data Loading** involves data from multiple systems into a new integrated system. **Multi-vendor Solutions** are used if multiple organizations use different spatial data systems and need to share data. **Data Sharing** is for the efficient and accurate sharing of data between multiple organizations. This is the primary spatial data migration use for PGDSS.

2.3 Spatial Data Migration Capabilities

Any spatial data migration tool that is chosen to support PGDSS compliance should be characterized by the following capabilities. First, multiple input support is needed to accommodate the extraction of data from at least one of the multiple source datastores used in Pennsylvania. Second, geoprocessing capabilities are needed for the transformation process so source data can be manipulated to match the target data model. Finally the tools should have multiple output support. While PGDSS has been defined as an ArcGIS data model, most of these tools are designed to output to multiple GIS formats.

3.0 Overview of Existing Spatial ETL Tools

3.1 Safe Software (Feature Manipulation Engine)

Safe Software's Feature Manipulation Engine (FME) is the industry's commercial product standard for spatial data migration. The ESRI spatial data migration module is based on FME technology. FME provides a "workbench" component that allows users to build the Geoprocessing steps and attribute mapping required to move data out of, and into different spatial data sources such as CADD and GIS. FME also contains a universal viewer used to preview transformations and a universal translator that can save spatial data to over 150 different formats. FME can be licensed in several methods: A full professional edition or specific to the following source formats: ESRI, Intergraph, Oracle, DB2 and Smallworld.

http://www.safe.com/products/fme/index.php

3.2 ESRI (Data Interoperability Module)

The ESRI Data Interoperability Module allows ESRI users to read from over 75 different spatial data formats and output to over 50. The tool leverages the ArcGIS Model Builder for spatial data format manipulations including Geoprocessing and attribute mapping. The Data Interoperability Module is built on Safe Software's Feature Manipulation Engine and thus utilizes and integrates the FME workbench functions to specify custom Geoprocessing, attribute translation and schema redefinitions.

http://www.esri.com/software/arcgis/extensions/datainteroperability/index.html

3.3 Intergraph Geomedia (Fusion)

GeoMedia Fusion provides tools for Geomedia users to integrate data through conflation, validation, queued editing and schema remodeling. Conflation rules are used for comparing two datasets containing multiple representations of the same features and selectively combining their geometries and attributes to create a singular representation. Validation rules are used in GeoMedia Fusion to check on the geometry of individual features, or Connectivity Validation that examines the geometries of features to other features. Queued Editing provides the capability to graphically present a list of items (data) in an environment for review and resolution. Lastly, GeoMedia Fusion is used to perform schema remodeling so that when data exists in a different model than the desired enterprise data base model, the data can be changed or remodeled to a new schema. This product is meant to migrate data into GeoMedia so it may be used for reading PGDSS data but GeoMedia users would need to use another product to migrate their data to the PGDSS format standard.

http://www.intergraph.com/gfusion/default.asp

3.4 Open Source (PostGIS / PostgreSQL)

PostgreSQL and PostGIS were developed by the University of California and is maintained by volunteer developers and runs on UNIX compatible platforms. PostgreSQL is an object-relational database system with most of the features of a commercial database system. PostGIS adds support for geographic objects to the PostgreSQL object-relational database. PostGIS essentially "spatially enables" the PostgreSQL server, allowing it to be used as a backend spatial database for geographic information systems (GIS), much like ESRI's SDE or Oracle's Spatial extension. When used together, these tools can transform geometry and restructure attribute data from disparate data sources into a common schema. These tools are currently being used by the North West PA Emergency Response Group to pull spatial data from multiple counties into shape files and the PGDSS V 2.0 data model

http://www.postgresql.org/docs/faqs.FAQ.html

http://postgis.refractions.net/

3.5 Open Source (Feature Data Objects)

Feature Data Objects (FDO) was developed by Autodesk and first distributed in 2004 with Autodesk Map 3D 2005. FDO was first released as open source along with Autodesks MapGuide open source web mapping product in 2006 and is part of the Open Source Geospatial Foundation. The open source version includes read/write ability for the following formats; SDF (Autodesk's spatial data format), SHP (ESRI shape file), MySQL, ArcSDE, ODBC, OGC WFS, and OGC WMS.

http://fdo.osgeo.org/index.html

4.0 Licensing and Pricing Options

4.1 SAFE SOFTWARE – Feature Manipulation Engine (FME)

FME Professional Edition Single Fixed Node License: $2,250 USD

FME Professional Edition Floating: $6,400 USD for the first user and $2,100 USD for each additional concurrent user

FME ESRI/Intergraph Edition Single Fixed Node License: $3,350 USD

FME ESRI/Intergraph Edition Floating: $8,700 USD for the first user and $2,900 USD for each additional concurrent user

An evaluation version and tutorial of FME is available from the following link: http://www.safe.com/evaluation/index.php

4.2 ArcGIS - Data Interoperability Module

ArcGIS Data Interoperability – Single Use License: $2,500 USD

4.3 Intergraph - GeoMedia Fusion

GeoMedia Fusion – Concurrent license: $9,000

GeoMedia Fusion – Node Locked license: $7,000

Intergraph does offer a Government & Education Discount on this product.

4.4 Open Source

All open source products are free to download. The drawbacks to open source tools include more limited documentation and support and less formal versioning.