Appendix U3 – EGIS Quality Assurance program

Table of Contents

TABLES

FIGURES

1. INTRODUCTION

The North Jersey Transportation Planning Authority (NJTPA) has undertaken the Enterprise Geographic Information System database (EGIS) project, contained in the agency's Unified Planning Work Program (UPWP) for 2009. The goal of the project is to consolidate the agency's data into a single, web-enabled platform and to improve information sharing among the project EGIS participants. The document *System Requirements*, developed in Task 1, details the overall design and strategy of the EGIS. This document summarizes the EGIS's policy, technology, and operating standards and describes its overall quality assurance and data management programs.

2. GOVERNANCE AND POLICY

Governance

The NJTPA EGIS is governed and overseen by a data management team referred to as the *NJTPA Data Management Group (*NJTPA-DMG). The NJTPA-DMG sets and enforces policies, guides decision-making related to technology, and oversees the agencies Quality Assurance, Data Inventory, and Data Exchange programs. At its core, the EGIS is a network of people, data, and technology resources. The NJTPA-DMG mission is to ensure that these resources are used to best serve the transportation goals of the region.

Policy

The NJTPA-DMG has implemented a set of comprehensive, high-level policies that serve to guide the overall administration of the NJTPA's EGIS.

- Commitment to a creating a framework for the open exchange of GIS and other digital products
- Commitment to providing user-friendly and internet-based public access to GIS and other digital products
- Commitment to the use of state-of-the-art data management tools to support data development and data exchange programs
- Enforcement of a Quality Assurance program to establish trust in the data products published by NJTPA

3. TECHNOLOGY PLATFORMS

Implementation of the EGIS, as a data sharing network, relies on several technology components. Data exchange standards among and within NJTPA partner agencies requires a basic understanding of the hardware and software platforms used to build the EGIS. These components make up the EGIS infrastructure and are part of the overall EGIS standards.

3.1 Hardware and Software Standards

The EGIS is built, operated, and maintained using an array of IT components. These components can be divided into three groups; 1) Server-Side: Database and Application, 2) Server-Side: Application, and 3) Client-Side: Application. These components are supported by NJTPA's local area network which consists of a Citrix – IP Centrex/Primary ISP run on a 50 mb per second back-bone. Figure 1 depicts the overall EGIS hardware configuration.

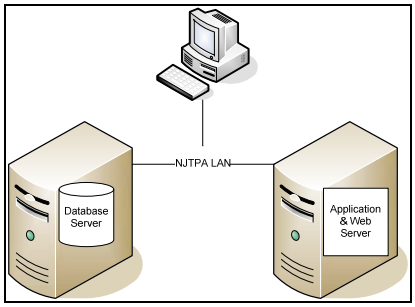

Figure 1 - EGIS Hardware

Tables 1, 2, and 3 enumerate the software elements that reside within in the hardware profile.

Table 1 - Server-Side: Database and Application

#	Software Element	Description
1	Operating System	Microsoft Windows Server Standard 2008
2	Database server	Relational Database Management System - Microsoft SQL Server 2008; Version 10.0.1600.22
3	GIS Gateway	ESRI ArcSDE 9.3.1; Release 93004
4	Data Management Applications	Microsoft SQL Server Business Intelligence Development Studio with ASP.NET. C## is the programming language. *Note that initial EGIS data management applications are run from the database server.*
5	Web Server for Data Management Applications	Microsoft IIS

Table 2 - Server-Side: Application

#	Software Element	Description
1	Operating System	Microsoft Windows Server Standard 2008
2	Server GIS	ESRI ArcGIS Server for the Microsoft .NET Framework, Version 9.3.1 R 3000
3	Web Server for ArcGIS Server	Microsoft IIS
4	GIS internet Metadata portal	ESRI ArcGIS Server Geoportal Extension Version 9.3.1

5	Web server Geoportal Extension	Apache, Version 6.0
6	Servlet for the Geoportal	Apache Tomcat Version 6.0 /Apache Directory Server Version 1.5, Apache Directory Studio Version 1.5
7	Geoportal Extension developer platform	Java SE Runtime Engine (JRE) 6 Update 16 / Java SE Development Kit (JDK) Update 16

Table 3 - Client-Side: Application

#	Software Element	Description
1	GIS connection	ESRI ArcGIS 9.3.1
2	Database connection	Microsoft SQL Server Management Studio and Access 2007

3.2 Geodatabases Standards

The EGIS is implemented through an ESRI ArcSDE geodatbase platform to support the development and distribution of quality data sets. A geodatabase is a database that has been spatially enabled through the ArcSDE software. For further detail related to its design, refer to EGIS documents 3.1 and 3.2. The contents of the EGIS geodatabases are made available through other ESRI software components, such as ArgGIS and ArcGIS Server. Together, these components make up the EGIS data-flow, available for use by EGIS participants. The data flow process for GIS oriented data is depicted in Figure 2.

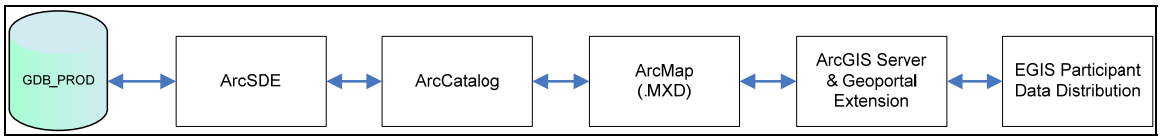

Figure 2 - EGIS Data-Flow Process

The geodabase is the cornerstone of the EGIS. A series of standards have been developed to govern its use and associated data distribution. These standards serve to guide all aspects of data development, submission, storage, and distribution. These standards, in their aggregate, are one part of the overall EGIS Quality Assurance Program. The program is based on a work-flow that develops spatial and non-spatial data through four stages. Each stage represents one physical geodatabase. Data sets are progressively developed and reviewed before they are checked into subsequent geodatabases. Figure 3 depicts the sequence of the geodatabase as data is developed through each stage.

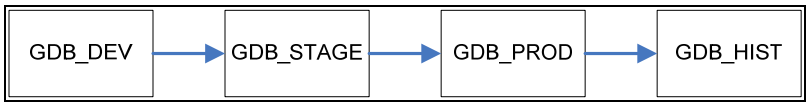

Figure 3 - EGIS Geodatabase Sequence

Table 4 describes the four geodatabases at the physical database level.

Table 4- EGIS Geodatabases Description

#	Database	Purpose
1	GDB_DEV	Development database that contains ArcSDE GIS layers and other data items that are in development and subject to a QA/QC process before they are checked into the GDB_STAGE database. This database is used to initiate datasets, develop their overall structure, and populate the structure with record values.
2	GDB_STAGE	Staging database that contains ArcSDE GIS layers and other data items. These items can come from the GDB_DEV database or potentially from other agencies. All data items in this geodatabsae are subject to NJTPA-DMG quality assurance guidelines. After the items are in conformance with the quality assurance program, they can be placed in the GDB_PROD database.
3	GDB_PROD	Production database that contains the active, day-to-day ArcSDE GIS layers and non-GIS database items. These items have been reviewed by the NJTPA_DMG group and its quality assurance program. No GIS editing will occur in this database; standard SQL data transactions on business tables will occur in this database.
4	GDB_HIST	Database that stores time-stamped, ArcSDE GIS layers. These items come from the GDB_PROD database.
5	GDB_REP_*AgencyX*	Geodatabases that will participate in a distributed data replication process, between NJTPA and other participants. *Note that the databases that will be set up for replication do not currently have the requisite details needed to support the replication process.*

Quality Assurance

As data is developed through the four geodatabases, standards are imposed on each item to ensure quality and conformace with the underlying technology. The process should begin at the GDB_DEV stage, but is essential at the GDB_STAGE level. Table 5 describes the quality assurance items for which EGIS data sets will be subject to during the data development process.

Table 5 - NJTPA Quality Assurance Program

#	QA/QC Item	Description
1	Each layer must have a responsible party from an NJTPA division to maintain the layer.	Each layer must have a custodian and a sponsoring department.
2	Each layer must use the NAD_1983_StatePlane_New_Jersey XY Coordinate System	Each layer must contain a proper spatial reference.
3	Each layer must conform to NJTPA naming conventions	Refer to the document EGIS – Logical Geodatabase Design for a an explanation of layer naming conventions and Table 6 for SQL Server 2008 and ArcSDE Table Naming Conventions
4	Each layer must contain a spatial index	Spatial indexes improve the efficiency of data queries and retrieval
5	Each layer must contain complete metadata	Meta data, to varying degrees, should be maintained for all EGIS data objects. Meta is catalogued in two places. 1. In ArcCatalog where it is stored in the SQL Server Geoportal931 database and available for wider use through the Geoportal web application 2. In the NJTPA Data Management application (Data Dictionary / Adaptive Data Inventory Template) where it is stored in the t_DataItems and t_SubjectArea tables. 3. NJTPA follows all FGDC and ISO standards related to information about the data sets for which it owns and is responsible.
6	Each layer must be checked for quality assurance by the NJTPA database team against the quality assurance items	NJTPA has established a Database Group whose tasks consists of, among other task, running through the agencies Quality Assurance Program check-list, before data is checked into the GDB_PROD geodatabase
7	Each NJTPA owned layer must contain the Create Date, CreateBy, Updatedate, UpdateBy fields	These fields enable NJTPA staff to track the changes of an ArcSDE layer.
8	Each layer must conform to the schema established in the NJTPA EGIS data model. Some important fields are: • Data sets that will grow over time must contain a **Date** field to facilitate data exchange processes.	Generally, layers and tables for which NJTPA is not the custodian, the layer/table schema will be established by the custodian agency. For layers/tables where NJTPA is the custodian, table definitions should be observed. Note that some tables or layers could be considered hybrid, in the sense that NJTPA has taken data

	• Each layer, where possible, should contain County and MPO fields with values. • Join fields must be consistent to maintain the integrity of underlying views and relationship classes.	from an outside source, and redesigned the schema to facilitate data storage.
9	Each layer must be aggregated, at a minimum, up to the regional level. Layers aggregated at the municipality and county level should be avoided.	Layers, if possible, should be entered into the EGIS as complete regional data sets. One notable exception is county-level cadastral data sets, denoted by lc_*County*.

The data items themselves are stored in the GDB_PROD geodatabase according to **Framework Categories**, depicted in Table 6.

Table 6 - NJTPA Framework Categories

#	Suffix	Framework Category	Description
1	c	Cadastral	Includes legal, block and lot data sets
2	d	Design Plans	Includes CAD-type data sets
3	e	Environmental	Includes data sets that are environmentally related
4	f	Facilities	Includes data sets that are related to buildings
5	g	General	Includes data sets that do not fit into any of the formal data categories
6	p	Planning	Includes data sets that relate to the planning process
7	pb	Political Boundaries	Includes data sets that denote artificial boundaries created by humans
8	t	Transportation	Includes data sets that are transportation related
9	u	Utilities	Includes data sets that are related to utilities

Data items are also subject to a set of nomenclature standards. Table 7 provides a list of guidelines to consider when naming database tables and GIS layers.

Table 7 - SQL Server 2008 and ArcSDE Table Naming Conventions

#	Convention	Example
1	Table names should, as clearly as possible, describe the data contained in the table, keeping in mind the 32 character limitation for SQL Server database object names.	
2	Words in table names should follow three basic conventions: • Base tables should be separated by a capital with no spaces or underscores. • Join tables should be capitalized, and separated with underscores • Where two capitalized words meet, they can be separated by underscores	• le_UrbanArea • t_PR_FACILITY_RAILLINE • t_HH_PUD
3	Table names should avoid the use of abbreviations.	Exceptions include: • CMS: Congestion Management System • SLD: Straight Line Diagram • SMS: Safety Management System •
4	Names of database objects (tables, feature classes, etc.) should be singular.	
5	Only use words like 'BOUNDARY, 'ZONE', 'LINE', 'POINT' when they are necessary to clarify the data contained by the table. Logical, non-technical	

	descriptors are preferred over technical jargon that may make sense to technical staff only.	
6	Column names should be singular where possible. Avoid using dates or years as part of the table name in a production database. Historical data sets, with their time-stamped suffix, should be stored in the GDB_HIST geodatabase. This process, in effect, represents NJTPA's change management system.	• lt_BusLine_12282009
7	In general, the first word in the table's name should reflect the data content while additional words will refine the description. This makes it easier for users to find data that is most often listed alphabetically or when using search engines.	• Streams_MOA instead of MOA_Stream • BuildingMyNeighborhood instead of MyNeighborhoodBuilding

3.2 Data Management

Data management is a fundamental part of the EGIS. In addition to the underlying ArcSDE and ArcGIS Server software platforms, the management of NJTPA's data resources is supported through two external facing environments.

Data Inventory

The EGIS contains a dynamic data inventory process, maintained within the Geoportal931 SQL Server database and on the NJTPA Geoportal site on the Inventory page. The process consists of SQL Server and .NET tools. Figure 4 depicts the data inventory process.

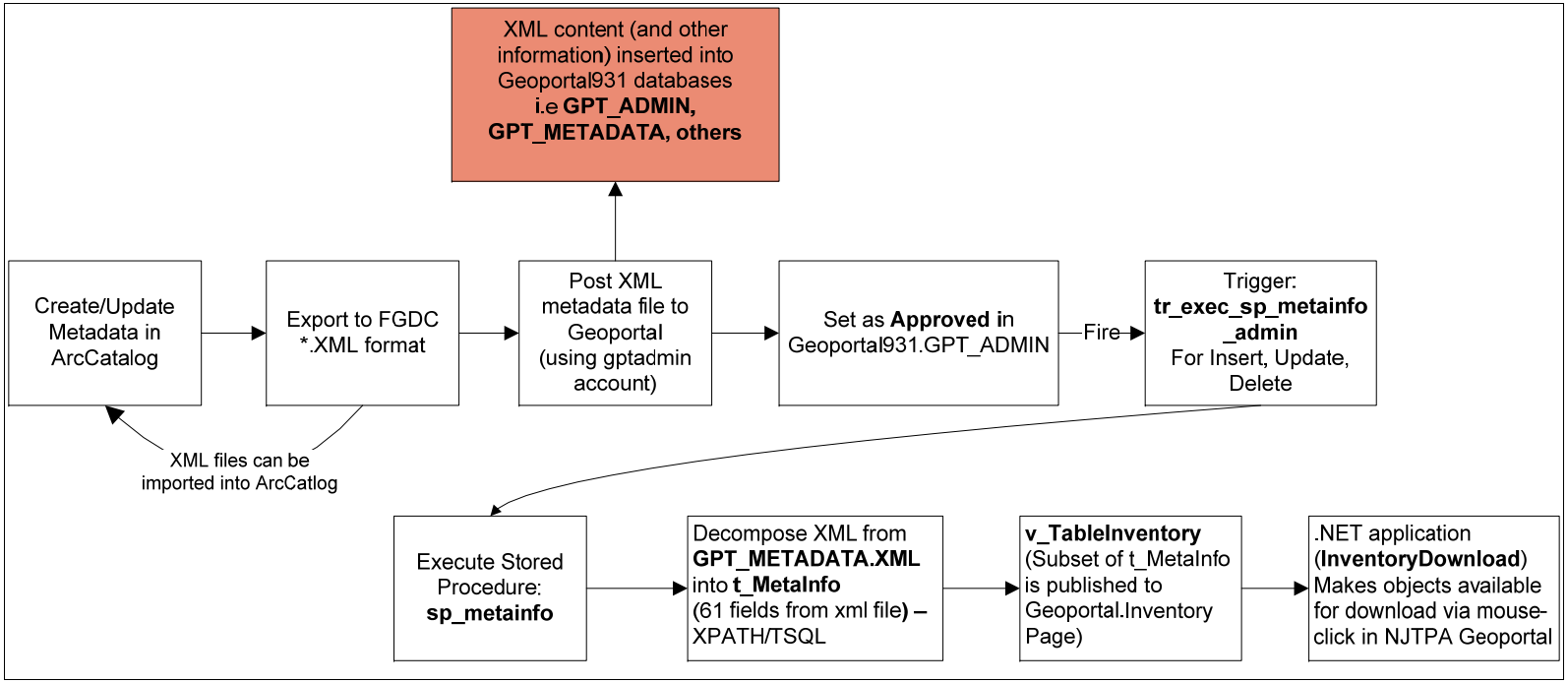

Figure 4 - Dynamic Data Inventory Process

The Data Inventory process starts with ArcCatalog where XML files are created that contain information about each EGIS data object. These xml files are than published to the Geoportal where they are processed and placed within the Geoportal/Inventory web page and made available for download to the public.

Geoportal

The NJTPA Geoportal provides a public interface to the EGIS by allowing users to search the agencies metadata (data about data). The Geoportal will allow NJTPA staff, subregions, state agencies, MPO's and the public to discover what data NJTPA and other agencies have available and how that data can be acquired. The Geoportal is a web browser based application that can be accessed by anyone over the internet to search metadata that has been uploaded to the site and approved as complete by the NJTPA Geoportal Administrator. In this way, the NJTPA Geoportal site is key component of the agencies overall data management and quality assurance program. Figure 5 present the home-page of the Geoportal. For a full discussion of the NJTPA's Geoportal, refer to the EGIS Admin and User manuals.

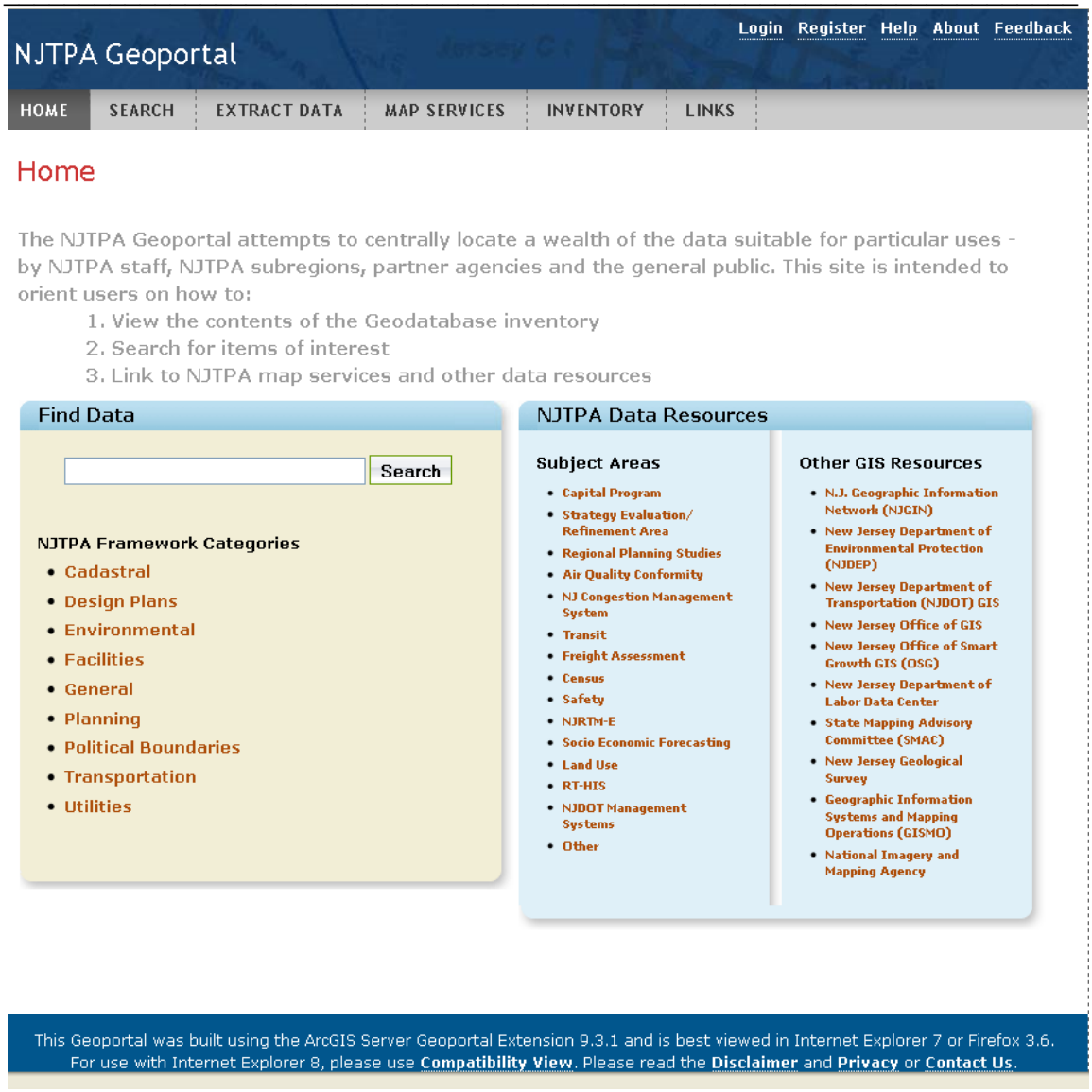

Figure 5 - NJTPA EGIS Geoportal

4. INTERAGENCY STANDARDS

The EGIS is a data repository for spatial and non-spatial data. Exchanging data in and out of the repository is also subject to a series of standards. Table 8 summarizes the access and exchange methods available to EGIS participating agencies.

Table 8 - EGIS Data Exchange Method

#	Method	Description	Agency
1	Direct access to geodatabase content	via credentialed login through ArcGIS Desktop	• NJTPA Central Staff • Subregions
2	Access to published	• via ArcGIS Server web	All map services can be

	ArcGIS Server map services	Applications (credentialed and non credentialed) • via ArcGIS Desktop (credentialed) • via NJTPA EGIS Data Management web page and link to geoportal (credentialed and non credentialed) • via NJTPA Geoportal (non credentialed)	exposed as either restricted or non restricted to any individual or organization • Map Services embedded in an ArcGIS Server application can be restricted at the login level • Map Service available through ArcGIS are restricted at the connection level • Map services available through the Data Inventory are restricted at the login level • Map Services via the NJTPA Geoportal are not restricted
3	Access to meta data and associated map services	via the EGIS Geoportal	• NJTPA Central Staff • Subregions • State Agencies • MPO's • Public
4	Replication to and from EGIS geodatabases	via ESRI geodatabase service: • Extraction – data can be copied from a geodatabase (Check-out/check-in) • Replication – Edits can be synchronized with another geodatabase (with ArcSDE only) – Both one-way and two-way	• NJTPA Central Staff • Any NJ agency
5	ftp	File transfer	• NJTPA Central Staff • Any NJ agency