Managing Spatial Data for

Effective Support of Business Processes

ESRI Paper UC1295

Jun Wu, Ph.D. and John Farley, GISP
IT-GIS, North Carolina Department of Transportation

Dana Nibby, GIS Database Analyst
ARCADIS

May, 2006

Abstract

Increasing demand for GIS technology has driven organizations of all sizes to produce and collect massive amounts of spatial data. To leverage the full benefits of that data, organizations must plan for and invest in spatial data storage and infrastructure. When managed effectively, spatial data can support business processes at a larger organizational scale.

Local and centralized spatial data will likely co-exist for years to come. In transitioning from ad hoc, single user mode GIS to enterprise GIS, organizations should reassess their spatial data management practices. Planning for and implementing a spatial data infrastructure can lower the total cost of ownership as well as provide a more robust environment for GIS applications.

This article will discuss the general principles of, and specific practices for, managing a diverse enterprise spatial data environment.
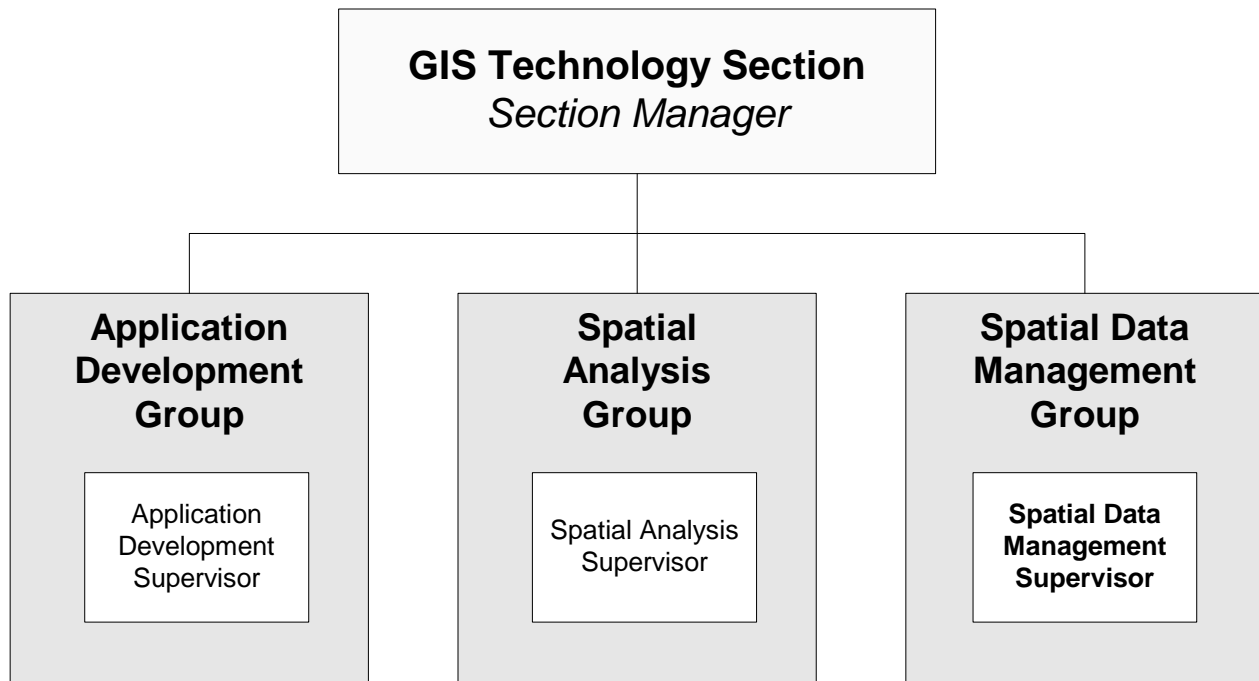

History and Vision

The North Carolina Department of Transportation's (NCDOT) GIS Unit was established in the late 1980's by combining the Road Inventory and Mapping sections within the Statewide Planning Branch. When demand for GIS services began to originate from NCDOT branches beyond Statewide Planning, the Unit became part of Information Technology.

The Engineering Transportation Systems Branch, as directed by the NCDOT Chief Information Officer, managed the Unit's reorganization. Two design goals guided the reorganization:

1) To enable the Unit to provide quality products and services

2) To better serve the Agency with an improved organizational structure and more effective business processes

During the reorganization, the Spatial Data Management Group was formed as a dedicated resource for spatial data administration. Tasks the Group is responsible for include:

Spatial Data Management
Spatial Metadata Management
Spatial Data Services
Database Management Support for GIS
Digital Spatial Data Dissemination

```
┌─────────────────────────────────┐
│       GIS Technology Section     │
│          Section Manager         │
└─────────────────────────────────┘
```

GIS Technology Section
Section Manager

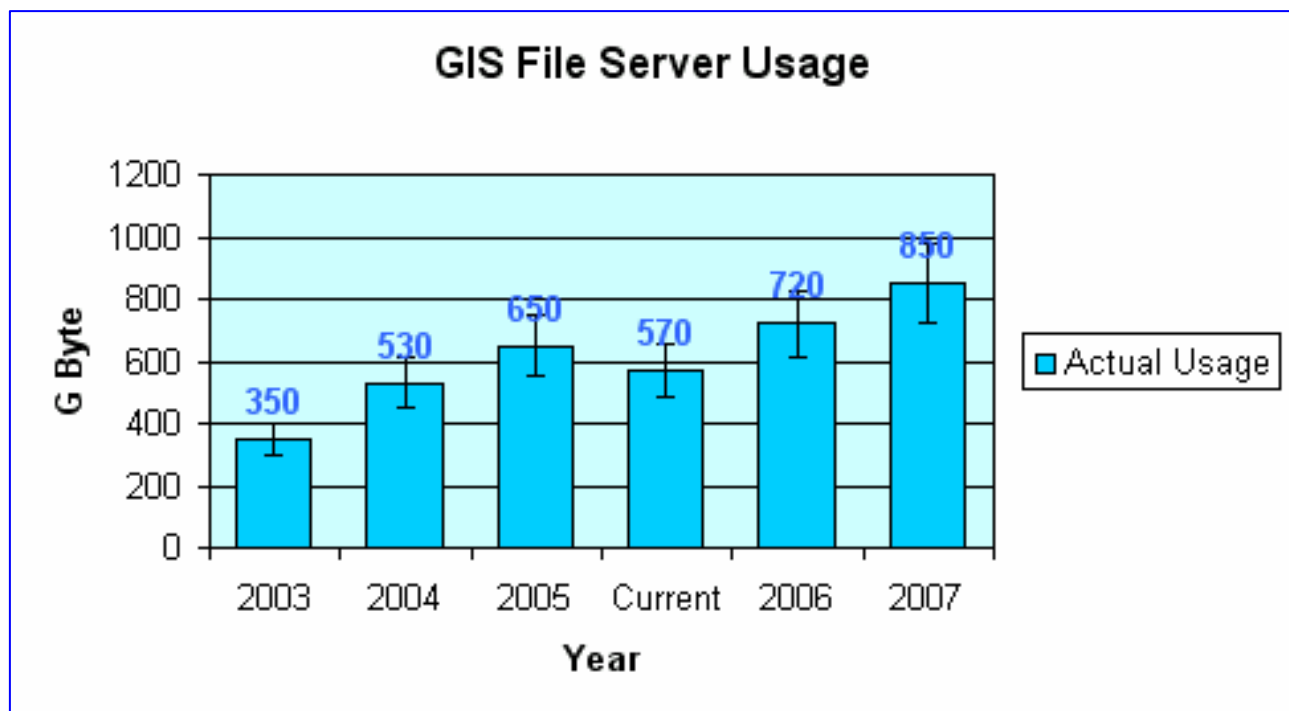| **Application Development Group**	**Spatial Analysis Group**	**Spatial Data Management Group**
Application Development Supervisor	Spatial Analysis Supervisor	**Spatial Data Management Supervisor**

Figure 1: GIS Technology Section Organization Chart

Introduction

GIS as a technology and an industry has experienced steady growth in the past two decades (USDOL, 2005). Massive conversion of paper maps into digital data began in the 1980s (David Sonnen, 2005). Such conversions are an ongoing effort for the entire industry.

NCDOT's GIS Unit has experienced rising demand for spatial data and GIS services. Additional data storage capacity is only one element required to satisfy this demand (Figure 2). Other requirements include having a sufficient Information Technology (IT) infrastructure, additional human resources to manage massive amounts of spatial data, effective business processes to streamline the data flow, and an appropriate organizational structure.

Figure 2: NCDOT GIS Unit Storage Usage

Goals of Spatial Data Management

Spatial data management activities should serve two purposes:

1) To ensure that spatial data of interest is available to the organization

2) To reduce the total cost of ownership of spatial data

The goals of data management in conjunction with the nature of spatial data itself suggest principles and best practices for spatial data management. In our business and organizational environment, we aim to follow these principles:

Principle 1: Institutionalize spatial data management with data policies

Principle 2: Comply with North Carolina's Statewide Technical Architecture

Principle 3: Adopt industry-wide IT standards and good practices

Principle 4: Balance current needs with future growth

Key Activities of Spatial Data Management

Establish a Data Policy

We formally established a data policy to cover many spatial data management related issues. The process of drafting the policy was educational for both the spatial data management team and GIS Management.
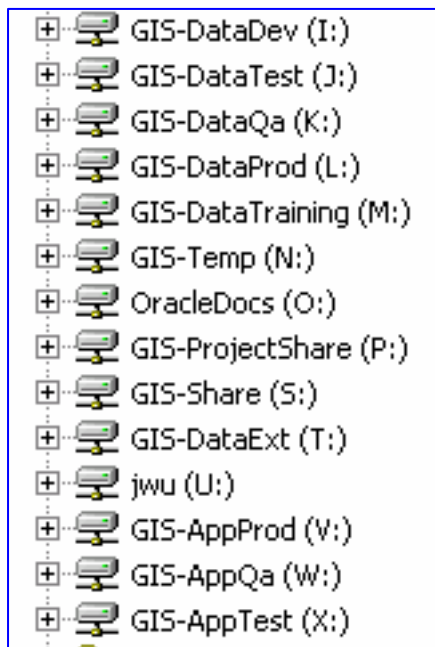

Figure 3: GIS Data Policy cover

We took the opportunity of a recent NCDOT Active Directory implementation to restructure the file server and deploy the policy. Most of the work was completed in about three months and the project was kept open to handle remaining issues for six additional months.

Restructuring the file server has resulted in the following benefits:

> Separated project documents from spatial data
> Separated the production instance from other instances
> Categorized users following the organizational chart and functional roles
> Standardized workstation drive mapping
> Defined security roles and access rights
> Defined a process of data promotion

Figure 4: Separation of production instance from other instances

Establish Spatial Metadata Requirements

We are adopting FGDC's Content Standard for Digital Geospatial Metadata (CSDGM) as the foundation for NCDOT's own spatial metadata standard. NCDOT's proposed metadata standard specifies what metadata elements are mandatory and optional for data sets. An NCDOT metadata standards document will provide guidance for data producers within NCDOT as well as contractors delivering data to NCDOT.

Additionally, we have developed standard metadata templates to simplify and encourage the production of metadata by data producers. Templates include information common to all NCDOT data sets, such as statements of liability and contact information.

We review metadata for each data set and provide feedback before we promote the data to production. The review process has two parts. Metadata is inspected manually for semantic correctness (are the descriptions meaningful?) and completeness. Structural and syntactic correctness (can the metadata be interpreted by software?) are checked by a program called MP (Metadata Parser) produced by Peter Schweitzer at USGS.

Producing metadata should not be optional. It is a necessary component of effective spatial data management.

Inventory Existing Spatial Data

Currently, the Unit manages approximately 600 gigabytes of spatial data. This data resides on a local file server dedicated primarily, but not exclusively, to spatial data. To assist with tracking changes made to data sets, the Unit has been compiling a data catalog; a list of existing spatial data sets. This has been a valuable, but time-intensive undertaking.

To make the process more efficient, the Unit plans to automate the production and maintenance of the data catalog.

```
Bridge Location Maps (Category of Datasets)

        Responsible:              Terry Norris
        Storage Location:         K:\Bridgemaps
        File Name:                Varies
        File Format:              TIF Image
        Status:                   QC
        Last Refresh:             Jan. 9, 2003


Color Infrared Image for 1998 (Category of Datasets)

        Responsible:              Terry Norris
        Storage Location:         L:\ColorIR1998
        File Name:                Named by Tiles
        File Format:              SID Images
        Status:                   Production
        Last Refresh:             Dec. 5, 2005
```

Figure 5: Data Catalog entry example

Manage the Life Cycle of Spatial Data

The life cycle of spatial data sets should be managed with defined processes.

The Unit tracks milestones for the full life cycle of each data set. These include data requirements gathering, metadata review, and the promotion of a data set from inventory or development to higher stages such as production. Particular attention is paid to storage and security requirements.
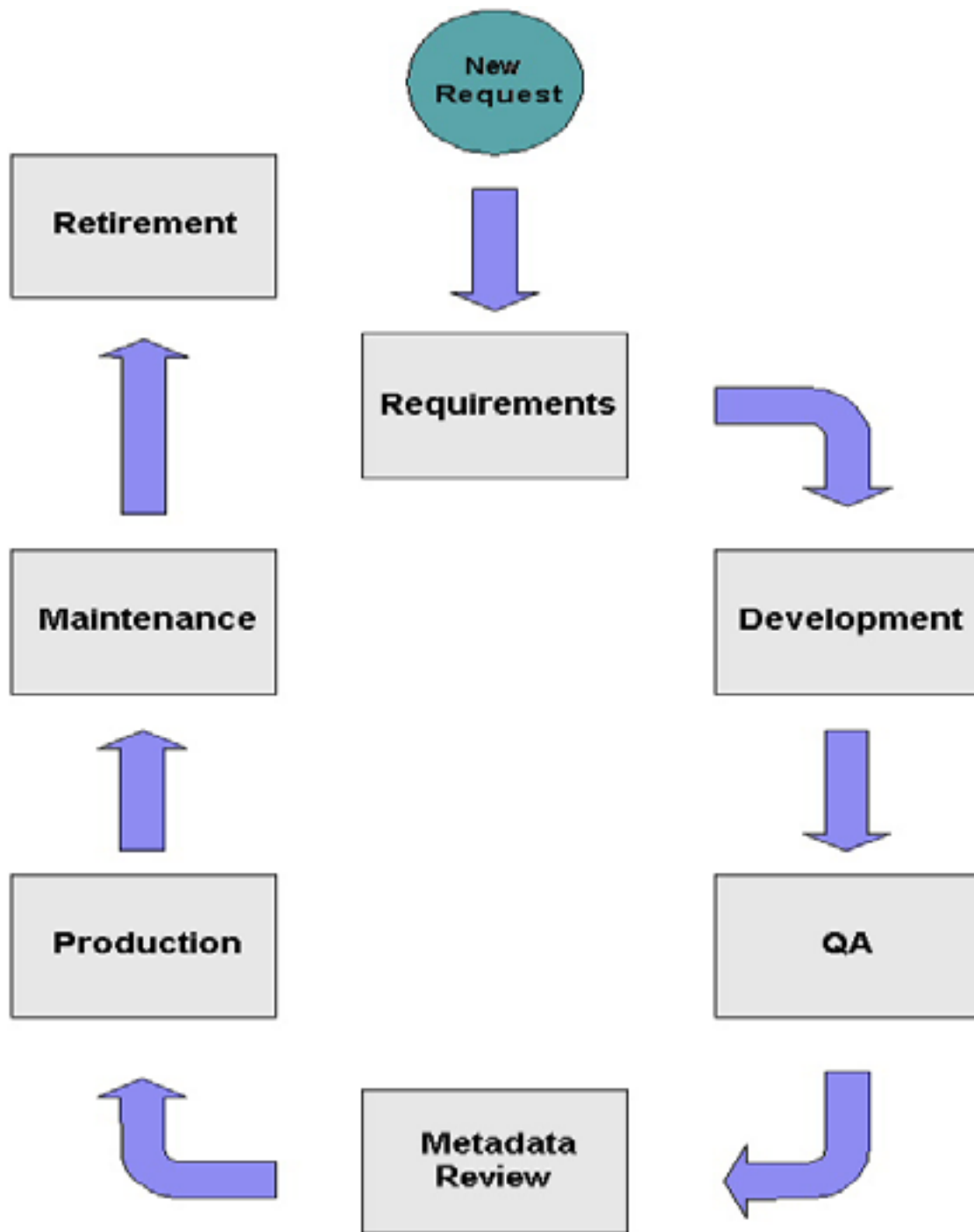
Figure 6: A typical data set life cycle model.

Manage Data Architecture

Data architecture defines how data is stored, managed, and used in a system. In particular, data architecture describes (Lewis, et al. Software Engineering Institute):

> How data is persistently stored
> How components and processes reference and manipulate this data
> How external/legacy systems access the data
> Interfaces to data managed by external/legacy systems
> Implementation of common data operations

While supporting the current infrastructure for daily operations, the Unit is moving toward ESRI's Spatial Database Engine server technology to meet the demand for GIS services across NCDOT.

The next generation of our spatial database environment will chiefly consist of a centralized spatial data repository, spatial database engine services, and Citrix Application Servers hosting GIS applications. A mix of thin and thick client technology will be utilized to meet different user needs within NCDOT. The addition of thin client options to the current infrastructure will enable us to push spatial data to multiple locations with varied network bandwidths.
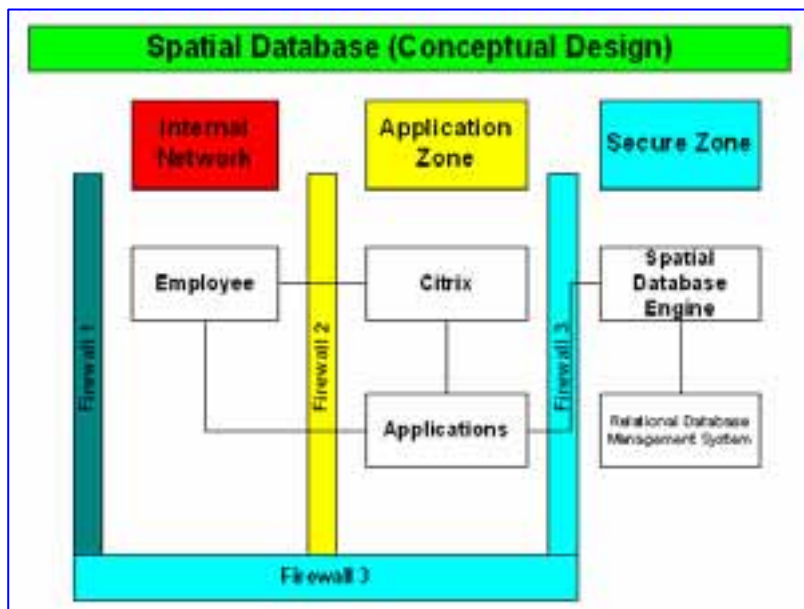

Figure 7: Enterprise Spatial Database Conceptual Design

Manage Spatial Data Availability

We perform daily incremental and monthly full backups of our spatial data sets. The recovery time span is six months for live data. A permanent archive is also made annually following our production cycle. Separate DVD and CD archives are also made for application or project specific data as required.

Summary

It is difficult to formulate universal rules of data management given vastly different business environments and organizational cultures. Every situation is different. However, similarities persist through the variation and should be considered by any organization. Practitioners in other organizations may find another balance of data management principles and practices appropriate.

We have adopted a specific series of principles and practices for spatial data management that have enabled us to more effectively support operations for NCDOT. Perhaps the most important principle one can follow is a commitment to formalizing processes, procedures and policies for spatial data management. It is always better to manage spatial data than be managed by it.

References

High Growth Industry Profile, U.S. Department of Labor, 2005

Spatial Information Management (SIM) - Then, Now, Next, David Sonnen, Directions Magazine, Sept. 2005

North Carolina Statewide Technical Architecture – Data Domain, Office of Enterprise Technology Strategies, Sept. 2005.

LRS Project Definition, ver 1.0, LRS Task Force, IT-GIS, NCDOT, April 2006.

An Enterprise Information System Data Architecture Guide, Grace Alexandra Lewis, et al., Software Engineering Institute

Author Information

Jun Wu, Ph.D.
Spatial Data Management Supervisor
IT-GIS Unit
North Carolina Department of Transportation
919-212-6031
jwu@dot.state.nc.us

John Farley, GISP
GIS Technology Section Manager
IT-GIS Unit
North Carolina Department of Transportation
919-212-6020
jcfarley@dot.state.nc.us

Dana Nibby
GIS Database Analyst
ARCADIS
919-212-6003
dmnibby@dot.state.nc.us

Appendix 1 - North Carolina Department of Transportation GIS Unit: File Server Access Policy Version 1.0

NC DEPARTMENT OF TRANSPORTATION

GIS UNIT:

File Server Access Policy

Version 1.0

Geographic Information Systems Unit
North Carolina Department of Transportation
April 28, 2006
http://www.ncdot.org/it/gis

File Server Access Policy

Revision History

Doc Type	Number	Action	Name	Edit Date	Comments
Draft	1	Check-In	DDJ	3/29/06	Submitted to John Farley for review.
Draft	1	Review	JF	3/30/06	Reviewed and routed to Cindy for revisions.
Draft	2	Revise	CM	3/30/06	Revised with Dave Johnson.
Draft	2	Commit	CM	3/31/06	Committed document to CVS.
Draft	2	Submit	CM	3/31/06	Submitted (4) copies to John Farley.
Draft	3	Submit after revision	Jwu	04/26/06	
Version	1.0	Approved by GIS Management	Jwu	04/28/06	GIS Director and Section Managers approved the policy.

File Server Access Policy

Background

Historically the GIS Unit has employed few conventions and standards with respect to storage of and security for spatial data sets and general files in the GIS File Server (csfs01). Recently, there have been changes to the GIS unit's organizational structure. In addition, workstations and file servers have been migrated to the DOT Domain. As a result, the file server has been reorganized to meet the GIS Unit's operational needs. The Spatial Data Management Group was established to handle data management related issues such as maintaining the GIS File Server directory structure and managing access to spatial data sets and general files for the Unit.

Purpose of Policy

The purpose of this policy is to clearly identify the structure and intended use of the GIS file server. Adherence to the following policy will allow the Unit to ensure security of data and a standardized data flow.

Terminology

Organizational Group: defined by the GIS Unit organizational chart. These groups are relatively static. Please refer to the attachment "GIS Unit Organizational Groups" for more details.

Functional Group: defined by the roles and functions across organizational boundaries. These groups are more dynamic than organizational groups and are subject to change based on GIS Unit priorities. The Spatial Data Management Group will notify users of their assignment to any functional groups. Please refer to the attachment "GIS Unit Functional Groups" for more details.

Policy Procedure

1. **Maintaining Active Directory inheritable permissions**

   All inheritable permissions initially defined and set by DOT IT Operations should be maintained unless otherwise justified by operational requirements.

2. **Setting file permissions**

   All file permission changes should be requested through the Spatial Data Management Group and documented by the DOT Help Desk. GIS employees should not change file permissions except for urgent or emergency situations. Under such circumstances, the Spatial Data Management Group is allowed to change permissions with written approval of a GIS Section Manager or the GIS Director. A notification should be sent to the DOT Help Desk after any file permissions are changed.

3. **User Drive (U:)**

Employees have 50M of disk space. The GIS Director and GIS Section Managers have 250M of disk space.

GIS employees may use the U: drive for storage of their personal work related files.

4. **Shared Working Drive (S:)**

   Provides storage space for working files based on GIS Organizational Groups. Everyone in the group has change access to all folders and files in their respective group directories.

   Some functional groups may store working files on the S: drive with the approval of GIS management.

5. **Shared Project Drive (P:)**

   The purpose of this drive is to support projects with members across organizational boundaries and to share unit-wide documents. Everyone in the GIS Unit has change access to all folders on this drive. This drive replaces the old GIS share directory.

   Please observe the following rules for using the shared project drive:

   - **No personal folders are allowed in the root of this drive**
   - **Each folder must be a DOT project, an approved GIS SOW, or a special folder designated by GIS management**
   - **Project and SOW folder names should follow the naming convention below:**

     ```
     <project or SOW>_<project lead user id>

     Example: DgnLinkNodeReport_ctilley
     ```
   - **Project and SOW folders should be used primarily to store project management documents**
   - **Project team members should coordinate the revision, ownership and responsibility of project documents stored on the P: drive**

6. **Data Working Area and Data Library Drives (I: through L:)**

   The data working area and data libraries shall be used for storage and dissemination of spatial data, including, but not limited to, GIS data layers, CAD data, and raster data.

The I: drive is designated as the Data Development Drive for spatial data. Initial development for all file-based spatial data sets should start from this drive.

The J: drive is designated as the Data Test Drive for spatial data. This data area accommodates the need for additional testing of data sets.

The K: drive is designated as the Data QA/QC Drive for spatial data. This data area accommodates the need for additional QA/QC activities of data sets.

The L: drive is designated as the Data Production Drive for spatial data. All data sets that are periodically updated or refreshed are stored on this drive, as well as snapshots of data sets that are modified daily.

The M: drive is designated as the Data Training Drive for spatial data. This data area accommodates data storage for training activities.

7. Shared Temporary Drive (N:)

The main purpose of the Shared Temporary drive (N:) is for temporary file transfer within the GIS Unit. Please use folder `N:\Transfer` to transfer files across organizational group boundaries. The contents of this folder will be cleaned-up periodically.

Other DOT employees may be granted special access to the Unit's spatial data through the N: drive.

8. Script Drives (V: through X:)

The V: drive is designated as the Script Production Drive. The V: drive will contain scripts / applications used internally by the GIS Unit. Change access is granted to employees responsible for maintaining the files.

The W: drive is designated as the Script QA/QC Drive. This area accommodates additional testing.

The X: drive is designated as the Script Development Drive. Initial development of scripts may start from this drive.

9. Use of (local) Workstation Hard Drive (C: and D:)

C: drive is reserved as the system drive (operating system and other required standard software) for the workstations. Do not use the C: drive to store working data/files.

It is strongly recommended that employees not use the D: drive to store working data/files. Local workstation hard drives are not backed up. There is no possibility of file recovery should a hard drive failure occur.

10. **Access to Managerial Files**

Managerial folders are designated for sensitive personnel information and management documents. Access to this area is granted only to GIS managers and others who are responsible for maintaining these files.

A list of the managerial folders:

```
S:\\GIS-Management
S:\\GIS-DataCompSpvsr
S:\\GIS-ProdSpvsr
S:\\GIS-TechSpvsr
```

11. **Promotion of Spatial Data Sets to Production Data Drive (L:)**

The Spatial Data Management Group is responsible for promoting all data sets from the Data Development Drives (I: through K:) to the Production Data Drive (L:).

Per workflow requirements, those who wish to store data on this volume should contact the Spatial Data Management Group. The Spatial Data Management Group will then create folders and request access permissions adherent to the user objectives and related policies.

12. **Promotion of Scripts and Applications through Script Development drives (W: and X:) to Script Production Drive (V:)**

The GIS-AppAdm group (please refer to the "Definition of GIS Functional Groups" for the members) is responsible for promoting all files through the Script Development drives (W: and X:) to the Script Production drive (V:).

Employees requesting that applications be stored on this volume will contact the GIS-AppAdm group members and inform them of their storage needs. The SDMG will then create folders and request access permissions adherent to the user objectives and related policies.

File Server Access and Usage Chart

File Server Access and Usage Chart			
Label	**Name**	**File Type**	**Access Limited To:**
I:	Data Development	Spatial Data	Organizational Functional Group
J:	Test	Spatial Data	Organizational Functional Group
K:	QA\QC	Spatial Data	Organizational Functional Group
L:	Production	Spatial Data	Organizational Group
M:	Training	Spatial Data	Functional Group
N:	Shared Temporary	Spatial Data Transfer	GIS-All and External Groups
P:	Project Share	Project Documentation	GIS All
S:	Share	Working Data\Files	Organizational Group
U:	User	Personal Working Data\Files	User Only
V:	Script Production	Script and Application	GIS-All
W:	Script QA\QC	Script and Application	GIS-All
X:	Script Test	Script and Application	GIS-All

Policy Enforcement

GIS employees are responsible for understanding and following this policy. Employees refusing to carry out this procedure are subject to the policy enforcement actions, up to and including dismissal.