GIS Architecture and Data Management Practices

Boone County GIS

**Created and Maintained by the
Boone County Planning Commission
GIS Services Division**

June 2008

Background

Boone County's GIS program has been in existence for fifteen years beginning with a formal implementation plan in 1993. During the remainder of the 1990s, the Boone County Planning Commission took on the daunting task of creating the 'core' GIS datasets (parcels, street centerlines, building footprints, contours, aerial photography, and a few others).

Once this task was completed around 2000, the GIS Division of the Planning Commission embarked on an aggressive initiative to create customized GIS computer applications and toolsets. The primary application resulting from these efforts became known as "BooneMap". Along with updated data, BooneMap is now deployed to more than 700 end users that work throughout Boone County in over thirty different agencies ranging from the Public School system to the Public Works department to Fire and Police officers. Nearly every public employee now has real-time access to the same quality GIS data as all other public employees, creating an environment that fosters improved communication, cooperation, collaboration and coordination of government services.

The focus of this document is to describe the technologies, methodologies and procedures that are in place to support the maintenance and dissemination of the GIS data.

Production and Publication Data

To understand the unique characteristics of Boone County's data architecture, one of the most important concepts to understand has to do with the relationship between production data and published data. Simply put, this concept promotes the idea that the data being actively maintained does not have to have the same schema as the data that is being provided to end users for analysis and consumption. A natural *parent-child* relationship exists between the two types of data in the sense that published (or child) data sets are produced from the production (or parent) data sets. The bullet points below attempt to describe the difference between production and publication data by outlining some rules for each.

- **Production** data is considered to be any data set that is designated as a working copy.

  RULES FOR PRODUCTION DATA SETS
  ➢ Only the person(s) responsible for editing a data set should have access to it in its production environment.
  ➢ Regardless of whether or not someone is the designated custodian, a production data set should only be accessed by users if the data set is in a state of being edited.
  ➢ Production data sets are to be exported periodically. The resulting *published* data sets are to be made available for consumption by end users.

- **Published** data is considered to be any data set that has been processed and prepared for consumption by end users.

  RULES FOR PUBLISHED DATA SETS
  ➢ All non-sensitive published data sets should be available to all end users.

> ➢ Published data sets should be read only. (Any edits that need to occur on a layer must be performed on the *production* version of the data set and should be performed by the designated custodian).
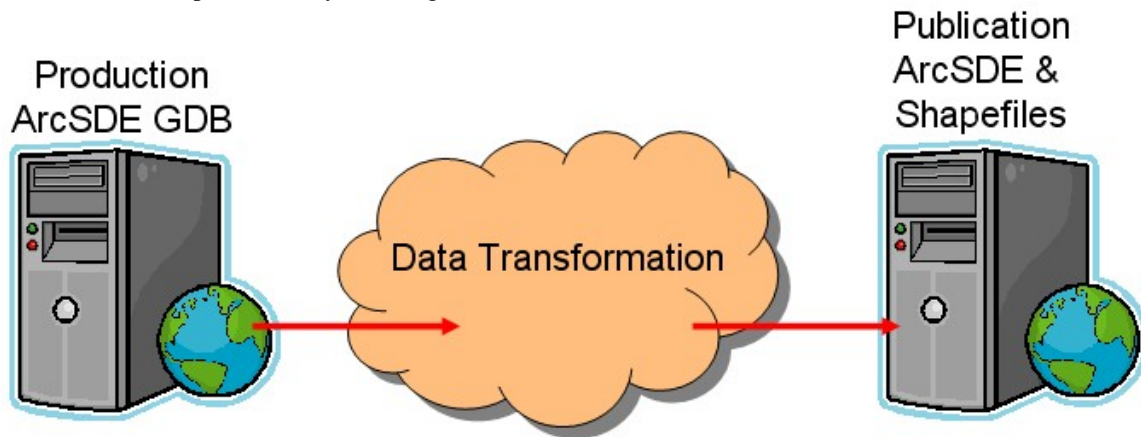

Figure 1: Production Publishing Paradigm

The *production environment* is a term that refers collectively to all of the hardware, software, data editors, procedures and production data layers that factor into the on-going maintenance of the GIS layers. Similarly, the *publishing environment* refers to the hardware, software, end users, procedures and published data layers that are necessary to support the consumption of GIS data.

In some instances, one person's published data may be another person's production data, thereby blurring the line between the two environments. An example of this can be seen in how the data custodian responsible for the maintenance of the zoning layer utilizes the parcel boundaries to enforce coincident geometry between the two layers. Since the zoning layer custodian is not making changes to the parcel layer, the rules outlined above dictate that the editor must use the published parcels for this function; yet because the published parcels are used in this fashion, they are inherently part of the zoning layer's *production environment*.

Boone County's Production Environment

Boone County utilizes a centralized data maintenance environment, whereby nearly every layer is maintained by the GIS Services Division of the Boone County Planning Commission. All production layers are stored in ArcSDE with a SQL Server relational database back-end. Staff members of the GIS Services Division are equipped with ArcInfo licenses of ArcGIS Desktop. ArcSDE editing is accomplished with out-of-the-box editing tools included in ArcMap.

The *Maintenance* geodatabase (aka "main") is the master repository for all feature classes. This geodatabase currently contains 142 feature classes organized within ten feature datasets. Feature classes within this geodatabase are currently supported by 243 domain tables. All edits that are expected to become permanent are posted to the Maintenance geodatabase. If a published layer ever gets corrupted or is otherwise in

need of being refreshed, the GIS Services Division can always re-export the layer from this geodatabase.

Each feature class in the Maintenance geodatabase contains certain minimum feature level metadata fields. Information stored within these fields indicate when the feature was created (CREATIONDT), when it was last edited (REVISIONDT), the technician who edited the feature (DATATECH) and other fields that help track changes and monitor progress.

In cases where non-GIS Services Division users are charged with the maintenance of data layers, production workflows that incorporate geodatabase versioning and disconnected editing protocols are in place to accommodate the need to centrally manage the production layers in the *Maintenance* geodatabase. Such users may perform their edits with other ArcView or ArcEditor licenses of ArcGIS Desktop or with ArcPad loaded on a GPS-equipped hand-held device. Check-out geodatabases are stored and accessed in a centrally located shared directory that is mapped as each custodian's 'V:' drive. Each sub-directory has permissions established to allow only the designated custodian access to the check-out geodatabase that they are responsible for editing.

Boone County's Publishing Environment

Boone County maintains two distinct publishing environments – one for users whose computers exist on the same network with the GIS Services Division, and another for those users who do not have direct network access to ArcSDE.

'Connected' Environment

The first publishing environment consists of a second ArcSDE *Publishing* geodatabase (aka "gis"). This geodatabase currently contains 173 feature classes. All publicly available feature classes within this geodatabase are stored at the root of SDE to maximize performance. Six feature datasets exist to store sensitive data layers and assist with implementing a strict permission scheme. Because most of the coded domain values that are utilized within the *Maintenance* geodatabase are converted to their corresponding textual values in the *Publishing* geodatabase, no domain tables are necessary.

Georeferenced raster data is served up to users via a third ArcSDE *Raster* geodatabase (aka "ras"). This geodatabase currently contains six years' of aerial imagery, USGS digital raster graphics, and several non-imagery grid layers.

'Disconnected' Environment

Boone County's standard publishing environment for users that do not exist on the same network as the GIS Services Division's *Publishing* geodatabase is often referred to as the 'X:' drive because that is the partition where most users will find their GIS data. Any GIS layer or supporting digital content that is intended for widespread use must be published on the X drive. For non-networked workstations and MDTs, GIS data and supporting content is stored locally on their internal hard drive's 'C:' partition. Because the X drive content is frequently updated and overwritten, any files added to the X drive

must be added by the GIS Services Division. All published data is read-only and only the GIS Services Division staff has write access to the contents of the X drive.
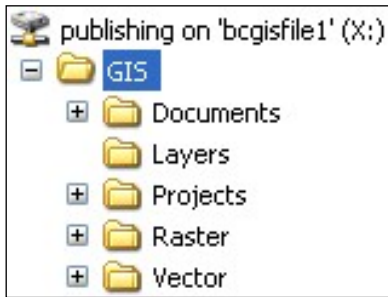
Immediately underneath the X: drive is a root folder named 'GIS'. The directory structure within the root folder is designed to help with organization of the GIS data and supporting content. The directory structure and specific location of GIS content must be tightly controlled if the GIS Services Division's customized GIS applications and toolsets are to be used.

Within the GIS folder, five sub-directories exist: Documents, Layers, Projects, Raster and Vector; which are detailed below:

- The **Documents** sub-directory stores non-georeferenced files, such as NGS monument datasheets, scanned deeds and plats, and digital pictures of buildings. Files stored within this directory are used primarily as hyperlink targets for map features in the GIS. Each file in this directory's sub-folders are indexed on *.html files stored at the root of the Documents sub-directory. The *.html filenames correspond to a GIS feature's unique identifier assigned by the data custodian.
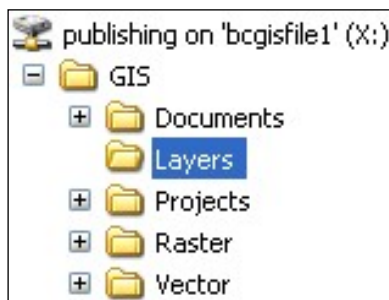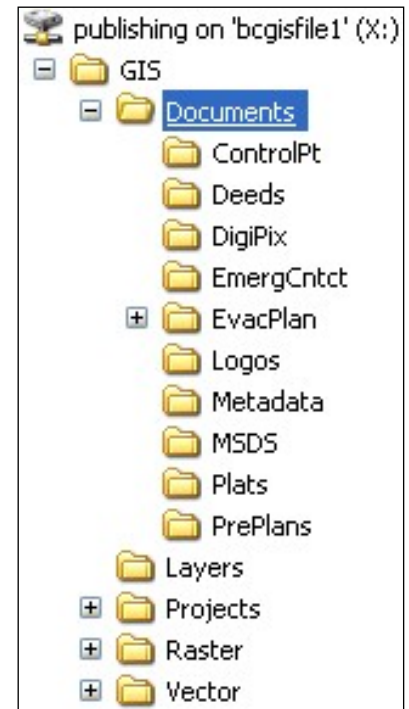
- The **Layers** sub-directory contains *.lyr files for all published vector and raster layers. The files in this directory support the need for a consistent GIS experience by standardizing the symbology and other properties of published data. ArcMap users are strongly encouraged to focus solely on this directory when adding data to their mapping project.

- The **Projects** sub-directory is where standard map projects are stored. ArcMap projects (*.mxd) are stored at the root of this sub-directory. There are two sub-folders within the Projects directory. The Layouts sub-directory contains Cartographic templates (*.mxt) used to ensure uniform map production. The PMFs sub-directory contains the compiled *.pmf files that are viewed by end users within BooneMap.
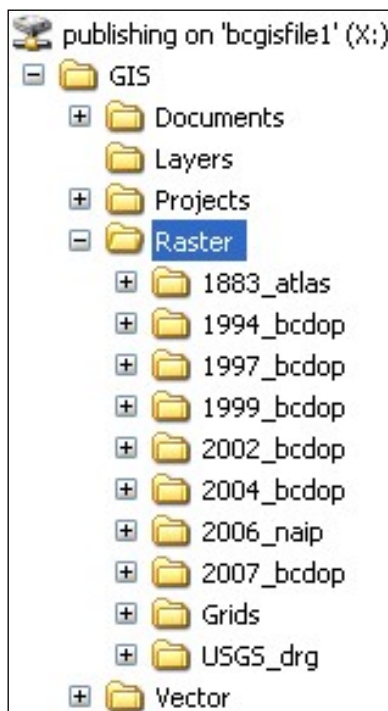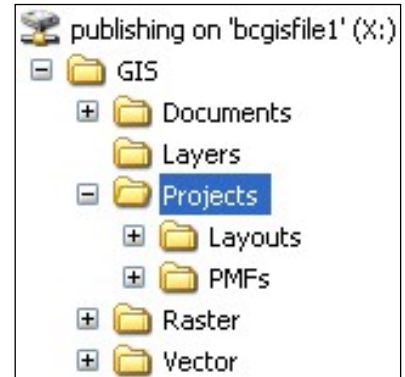
- The **Raster** sub-directory is the location where all georeferenced raster data is stored. Each year of aerial imagery is stored in its own file geodatabase or Mr.SID files inside of sub-folders pertaining to the year the imagery was captured. The 1883_atlas folder contains scanned georeferenced images from an 1883 atlas of Northern Kentucky. Non-imagery raster data sets such as hillshades and DEMs are housed in the Grids folder. USGS digital raster graphics are stored in the USGS_quads folder.
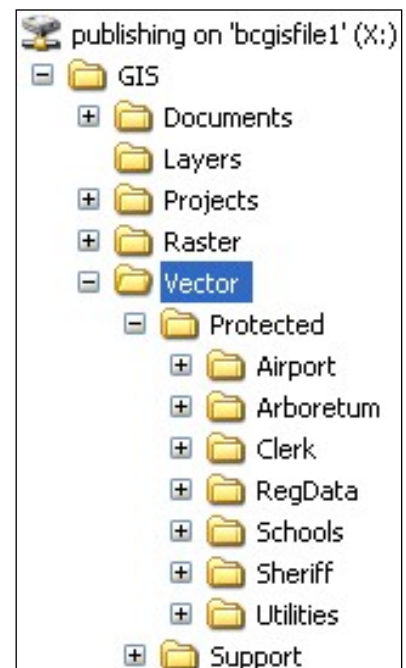
- The **Vector** sub-directory contains georeferenced vector data. This consists mostly of shapefiles and file-based geodatabases that have been published from ArcSDE. There are two sub-folders contained within – Protected and Support. The Support folder stores miscellaneous layers used in special map projects. The Protected folder houses sensitive data layers that have restricted access. Directory level permissions are in place to control access to these sensitive layers.

Data Transformation from Production to Publication

Understanding that the production data schema may not be the same as the schema of the publication data, the logical question is "how does Boone County manage these differences and routinely create published data with ease of repeatability and data integrity". The answer is by using the Data Interoperability extension for ArcGIS Desktop.

This extension was developed specifically to work within ESRI's technology, which conforms to Boone County's standard GIS technical architecture. The toolset is graphically driven eliminating the need to employ high-level programming. The GIS Services Division found the extension to be very easy to learn and use. This extension has proven to be extremely useful in organizing, constructing, and executing routine data transformation procedures that are compiled and stored as models inside of ArcToolbox. The figure below illustrates the relationship between the production and publication geodatabases and the role that the Data Interoperability extension plays.
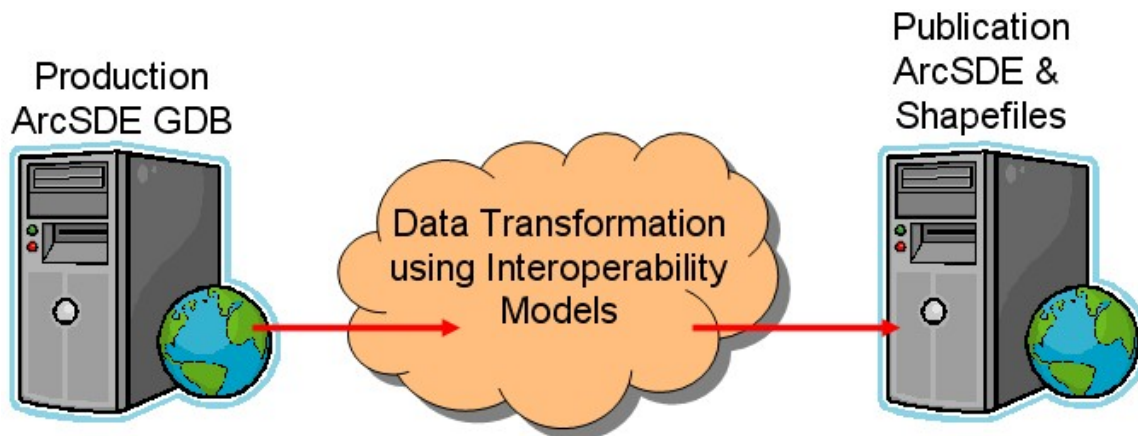

Figure 8: Production Publishing Paradigm

Each production layer has its own publishing translation, which is stored within a comprehensive Data Interoperability model that contains all publishing translations. The Data Manager can enable and disable individual translations to specify which layer's are to be updated. Enabling the contents of the entire Interoperability model will update the entire *Publishing* geodatabase.

Layer Files

The Layers directory on the X drive currently contains over 200 ESRI layer files (*.lyr) that are used to standardize how the published data appears to end users. Boone County's layer files can be considered part of the Publishing environment, although production layers can have layer files associated with them as well. Each published layer has one or more layer files associated with it.

The GIS Services Division pre-configures the layer files with standard symbology (color, pattern, etc.) so that all users will view the content the same. Layer files also contain pre-configured field aliases to assist users in understanding the content. A lot of time has

been spent fine tuning each layer's labeling schemes which are also contained within the layer file configuration. Other characteristics, such as the primary display field, definition queries, transparency factors and map tips are also encompassed by the layer files.

The Layers directory has two sub-folders – one containing layer files for the users detailed by the "Publishing Environment – Connected Environment" section above (i.e. enterprise geodatabase consumers); and another folder containing identically configured layer files (except for the data source) for the users detailed by the "Publishing Environment – Disconnected Environment" section above (i.e. file-based geodatabase consumers).

Departmental or Non-Enterprise Layers

Not all GIS layers are intended for widespread distribution. Some layers are temporary in nature or intended for use by only one person or a smaller group such as a department. In an attempt to standardize how these layers are stored and used, Boone County has issued the following guidelines to GIS users.

Layers that are temporary in nature or are intended only to be used by one person should be not be served up via the X drive. These layers should instead be stored either on the user's local hard drive or on a secured personal folder on the network.

Small groups such as a department within an agency requiring collective access to a layer have a network share established to house such layers. Appropriate permissions are in place to ensure that only those involved in the project have access.