A GeoData Management Practicum

Gar Clarke – GIT Coordinator, New Mexico Office of the State Engineer

Statistical modeling of water resources usually demands an evaluation of historical trends and the establishment of current conditions to predict future impacts. Linking diverse data on wells, water levels, surface runoff, hydrogeology, topography, precipitation, evapotranspiration, soils, vegetation, and land use with the 3-D analytical requirements for surface and groundwater models can produce a chaotic soup of data.

To protect their integrity, these data must be managed with appropriate organizational and quality assurance procedures. The identification, harvesting, assessment, acceptance, integration, and documentation of information demands systematic procedures to validate the data. Most of the published information on data development focuses on large "enterprise" systems that have huge complexities not realized in a small organization. For the "small shop," with an environment of deadlines and billable hours, the following approaches are offered to help establish a geospatial data framework.

Project Planning

Data Requirements: What data are needed to solve a problem? For spatial data, the important steps are defining the features, establishing accuracy thresholds, and identifying the content. A simple matrix can describe a task or subtask, list the data layers required to accomplish the task, show scale/accuracy thresholds, define the contents or attributes for each layer, and list possible sources for these data. This exercise helps to prioritize data needs and define what is either ideal or marginally acceptable, and creates a "data shopping list" where type, content, and accuracy are articulated.

This process also helps define attributes or fields that may categorize or stratify a feature. Attempt to categorize attributes as core, mandatory, or optional. Core attributes (such as a Hydrologic Unit Code for a basin) are essential to keeping records unique, providing a means to link to other data tables, and/or noting source or accuracy. Mandatory attributes, such as depth of well and static water level, are essential to the project's success and include analytical requirements. Optional attributes may be temporary flags or comments, or may direct the user to a picture or a document.

Expansiveness: Many projects outgrow their original parameters. Consider what practical steps can be taken to ensure that the database design, hardware, software, and analytical capacity can grow without having to retool, rethink, or redo. For example, when designing a groundwater model, bear in mind what expansion of data standards and design would allow its applicability to another region.

Consider what practical steps can be taken to ensure that the database design, hardware, software, and analytical capacity can grow without having to retool, rethink, or redo.

Data Acquisition

Data Sourcing: Many hydrology projects combine data from local, state, federal, and even private sources, some of which may be redundant. Use the "data shopping list" with a tiered approach by accessing and downloading federal, state, and local data, in that order. In each of these arenas, first access the "data clearinghouse" sources, which contain data from multiple sources that can be downloaded free of charge. The federal equivalent is the Geospatial Onestop, and many states have similar services (see sidebar). Depending on the size of the study area and your data requirements, individual federal agencies (such as for soils data), state agencies (for bridge locations), and local governances (for parcels) may be accessed for specific data.

A systematic approach to data acquisition can minimize costs. When acquiring data, be sure to document the who, what, and why. This is a form of internal accounting, but is not to be confused with metadata. A simple spreadsheet suffices to record the person who acquired the data, contacts, date acquired, storage location, and miscellaneous comments. This spreadsheet can be used to document alterations such as conversion or migration processes that were applied to each dataset during the course of a project.

Units and Standards

Spatial Standards: The world of mapping has a slew of specifications incorporated into federal and international standards. In adopting geodata standards for a project, an essential first stop is the Federal Geographic Data Committee (FGDC), whose standards must be used for many projects funded with federal dollars. The FGDC represents a marriage of federal agencies that have incorporated local, state, and professional interests to provide "layer" standards for major feature datasets, such as hydrology, cadastral, and transportation. Incorporating project specifications into the federal standards can prevent compliance problems later.

Globe Parameters: Most projects are defined by a geographic jurisdictional area. Federal agencies may use longitude/latitude, state projects may use UTM (Universe Transverse Mercator), and local governances often use State Plane Coordinate Systems. Establish a projection and a spatial unit measurement standard early on to guard against future problems. Recognize that each projection has both a unit of measurement and a "datum," such as the North American Datum established in 1983 (NAD83), which replaced one established in 1927 (NAD27). Some software will convert projections dynamically.

Migratory Converts: Migrating or converting geospatial data can inadvertently turn good data into bad data. The essential key is to standardize coordinate projections, units of measurement, attribute needs, data warehousing, naming conventions, and procedures. For example, define and document systematic procedures to convert CAD drawings into geographic data, compress imagery, import spreadsheets to append to a geodatabase, and validate data accuracy and content.

Organization and Documentation

Data Warehousing: Data organization can be facilitated by creating the following primary directories:

- *Development* – where all the conversion, migration, integration, and alterations will occur before data layers are considered "analysis-ready." This directory may have subdirectories that categorize data by source (federal, state, local), thematic layer (hydrology, transportation, boundaries), or region (state, regional, county). A structured work area directory for temporary files or interim products may also be necessary.
- *Data Home* – for storing data once they have been projected, consolidated, verified, and prodded. These "base data layers" need restricted-write access to protect the integrity of the data. Separating the digital imagery from vector or feature datasets facilitates performance and access.
- *Production Area* – for analysis of results and work products. In many cases, this space will duplicate the structure of *Data Home*, adding a work area for temporary or interim files. The Production Area must be backed up daily; in some cases files must be saved out separately.
- *Administration* – for the business side of a project. This is where correspondence, reports, standards material, digital pictures, sample materials, and other supporting documentation belong.

Document: In the geospatial world, three main areas are important to document: feature layer metadata, processing logs, and procedure. All federally sponsored geospatial projects require compliance with FGDC standards for metadata, but any documentation is preferable to none, regardless of funding. Documentation can be as simple as creating a text file with the same prefix as the feature layer file name, listing what was done, who did it and when, how well it was done, and if it will be repeated. Only the analyst who produced the data has the knowledge to provide these metadata. This will simplify replication in the future as memory dims or staff changes.

These approaches are not a comprehensive listing of geodata practices nor may they be appropriate for all users. In fact, some of the suggestions will grate on a few professional purists. But the suggestions are based on experiences that small shops may find useful. The websites above may clarify some of the terms and suggestions noted.

Contact Gar Clarke at george.clarke@state.nm.us.

Data Sources

National Atlas

www.nationalatlas.gov/atlasftp.html

Federal Agencies

FEMA Flood Data
www.fema.gov/fhm/mh_main.shtm

Geospatial Onestop
www.geo-one-stop.gov

NOAA NGS Geodetic Control Data
www.ngs.noaa.gov

USACE National Inventory of Dams
crunch.tec.army.mil/nid/webpages/nid.cfm

USDA Aerial Photography
www.apfo.usda.gov

USDA Data Gateway
datagateway.nrcs.usda.gov/NextPage.asp

USDA NRCS Soilmart
soildatamart.nrcs.usda.gov

USGS Earth Resources Observation and Science
edcsns17.cr.usgs.gov/EarthExplorer/

USGS Mapping Products
mcmcweb.er.usgs.gov/status/

USGS National Water Information System
waterdata.usgs.gov/nwis/

State Clearinghouses

Arizona: *www.land.state.az.us/alris/*
New Mexico: *rgis.unm.edu*
Texas: *www.tnris.state.tx.us*
Utah: *agrc.utah.gov/agrc_sgid/sgidintro.html*

Standards

Federal Geographic Data Committee
www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/index_html/?searchterm=nssda

International Organization for Standardization
www.iso.org/iso/en/ISOOnline.frontpage

National Institute of Standards and Technology/Federal Information Processing Standards
www.itl.nist.gov

Open Geospatial Consortium
www.opengeospatial.org