



# Building a Water-Resources Geodatabase for the Rio Grande Basin – San Acacia, New Mexico to Fort Quitman, Texas

ESRI International User Conference – San Diego, CA  
Session: Regional Water Information Systems  
07/13/2010

# Overview – Need and Significance

- Data collected for middle Rio Grande Basin for variety of purposes over several decades by numerous agencies, researchers, and organizations
- Previous attempts to compile these data have not been complete or done in a way that could be repeated
- A well-organized and usable database is needed to identify management priorities and facilitate hydrologic studies

# Hydrologic Geodatabase Overview

- Data organized into a comprehensive spatially-enabled relational database (geodatabase)
- Stores spatial and tabular data for area of interest from multiple sources
- Data can be queried and/or viewed spatially to assist in identifying data gaps, conduct spatial analysis, and support decision-making

# Rio Grande Basin – San Acacia, New Mexico to Fort Quitman, Texas

- Surface-water catchments encompass approximate extents of underlying aquifers of interest
- Surface-water catchments used to geographically filter physical sites and associated data from sources with larger spatial delivery extents



# Data Sources and Types

- Hard-copy reports, previous data compilations, and various collecting entities
- Updated web-accessible raw data sources such as USGS NWIS, EPA STORET and state environmental agencies, many current through day of download
- Surface-water discharges, ground-water elevations, and water-quality data (daily and instantaneous)



# Compilation Methods and Tools

- Data pre-processing scripts stage raw data for loading (e.g., separate source files with site/sample/result/parameter tables)
- Loaders query data from staged source files
  - Facilitate maintenance and updates
  - Document how table fields were mapped from source files to final compiled geodatabase
- Methods scalable for a variety of hydrologic studies (data types, data size)
  - 24 distinct data sources
  - Of those, 10 are pre-existing compilations or raw data aggregation efforts such as EPA STORET which contain data from one or more collecting entities

# Data Pre-Processing Scripts

- Import and format data from raw downloaded files into one “staged” relational database per source
- Scripts help clean data to ensure data integrity
  - Remove extraneous non-printable characters
  - Remove extra spaces
  - Date formatting (MM/DD/YYYY)
  - Separate multiple data types from one field (e.g., result values with comments) into separate fields
- Repeatabile script programs (VBScript, VBA) with code comments as documentation help ensure data consistency and reduce human error



# Data Compilation Loaders

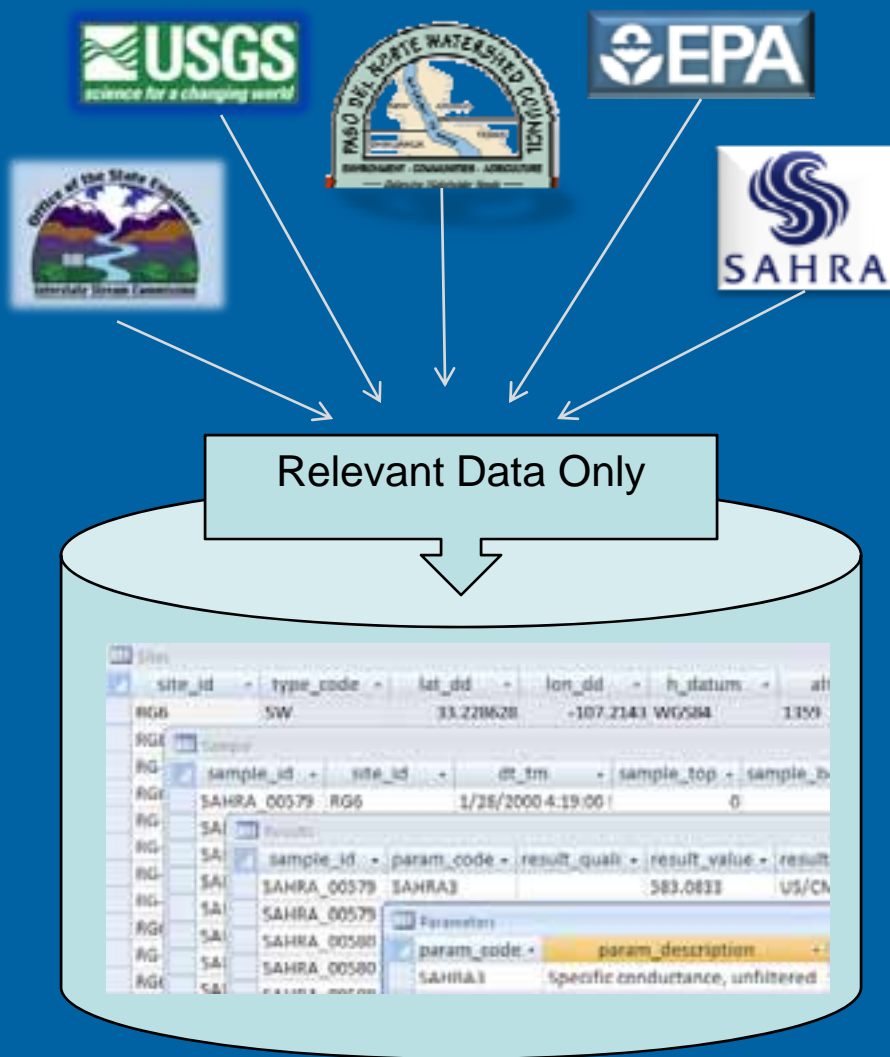
- Load data into compilation database from source “staged” databases
- Cross-walks source file table fields with compilation table fields
- Functions as process documentation; customizable for each source
- Repeatable and consistent – helps reduce human error

# Data Management Challenges: The Devil is in the Details

- Source compilation relational databases – database design and data integrity issues
- Site identification (Site ID's): different ID's & names for the same physical location on the ground
- Duplicate data
- Data recovery efforts and methods used
- Null values vs. zero; result value rounding
- Little or no metadata in most cases



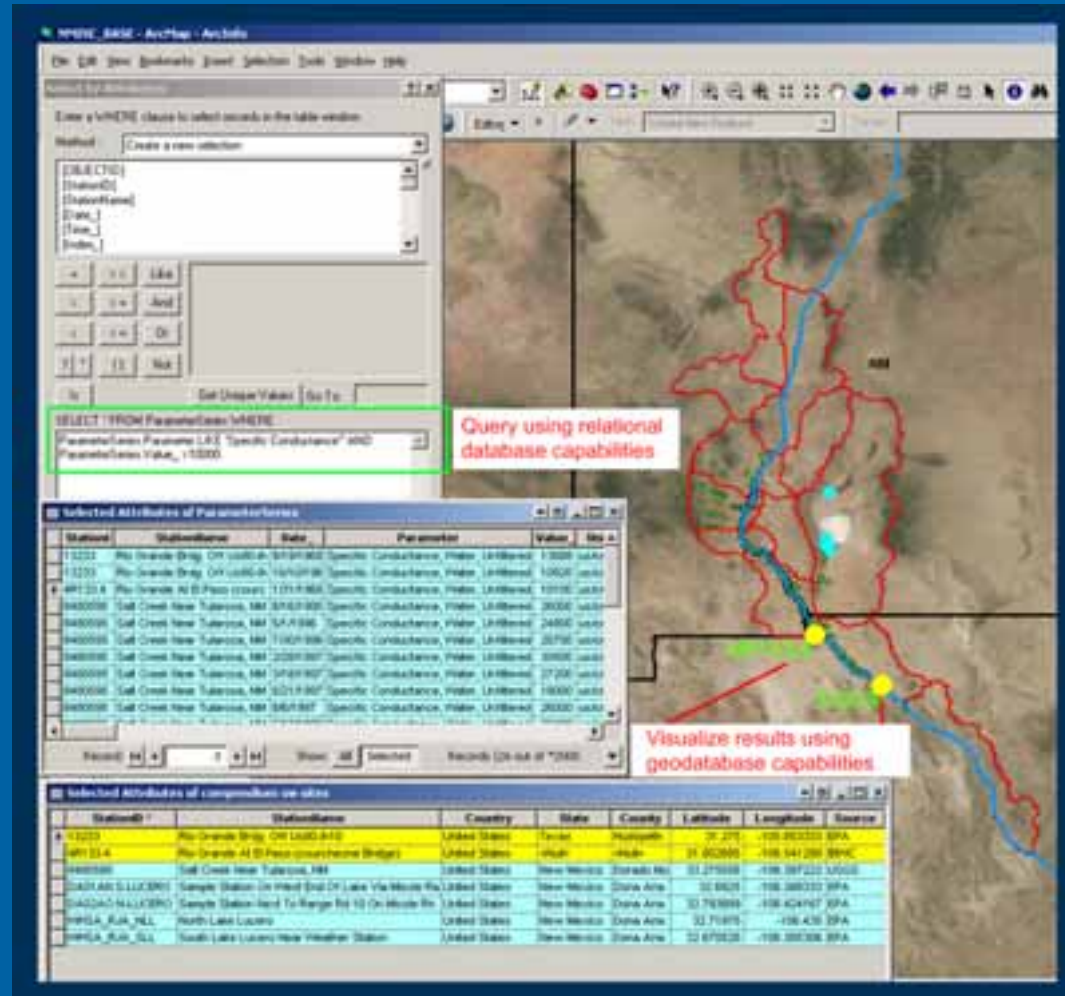
# Direct Benefits to Researchers



- Multiple agencies' data managed in one location
- Record-level metadata to document collecting agencies and data sources (not always the same)
- One comprehensive relational geodatabase as opposed to scattered records stored as flat files in spreadsheets

# Direct Benefits to Researchers (cont.)

- Data can be limited to just relevant parameters/sites using traditional database queries via Structured Query Language (SQL) and/or spatial selections in a GIS
- Data enhanced for map production and spatial and temporal analysis



# Summary

- Usability of source data greatly enhanced
- Facilitates identification of data gaps – spatial, temporal, and thematic
- Multitude of data stored in a well-documented format that can be readily updated
- Sound data management helps support sound science – quality information can only be derived from quality data

# Questions?

Thomas E. Burley

Texas Water Science Center

U.S. Geological Survey

[teburley@usgs.gov](mailto:teburley@usgs.gov)