

In cooperation with the New Mexico Interstate Stream Commission

Usage and Administration Manual for a Geodatabase Compendium of Water-Resources Data—Rio Grande Basin from the Rio Arriba-Sandoval County Line, New Mexico, to Presidio, Texas, 1889–2009

Open-File Report 2010–1331

U.S. Department of the Interior
U.S. Geological Survey

Cover: View of the Rio Grande from Santa Elena Canyon near the confluence of the Rio Grande and Terlingua Creek in Big Bend National Park, Texas, April 2010. The mountain range in the background is the Chisos Mountains. Photograph by Daniel K. Pearson, U.S. Geological Survey.

Usage and Administration Manual for a Geodatabase Compendium of Water- Resources Data—Rio Grande Basin from the Rio Arriba-Sandoval County Line, New Mexico, to Presidio, Texas, 1889–2009

By Thomas E. Burley

In cooperation with the New Mexico Interstate Stream Commission

Open-File Report 2010–1331

**U.S. Department of the Interior
U.S. Geological Survey**

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2010

This and other USGS information products are available at <http://store.usgs.gov/>
U.S. Geological Survey
Box 25286, Denver Federal Center
Denver, CO 80225

To learn about the USGS and its information products visit <http://www.usgs.gov/>
1-888-ASK-USGS

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:

Burley, T.E., 2010, Usage and administration manual for a geodatabase compendium of water-resources data—Rio Grande Basin from the Rio Arriba-Sandoval County line, New Mexico, to Presidio, Texas, 1889–2009: U.S. Geological Survey Open-File Report 2010–1331, 63 p.

Acknowledgments

The author would like to thank Dale Doremus, Peggy Barroll, Cindy Stokes, Beiling Liu, and Russell Isaac of the New Mexico Interstate Stream Commission and the New Mexico Environment Department for guidance on source data and compendium legacy information; Nancy Ragland and Tabitha Kirkland of the Texas Commission on Environmental Quality for assistance with the TCEQ SWQMIS database; and John D. Peine of the U.S. Geological Survey Southern Appalachian Field Branch and the University of Tennessee and Silvia Terziotti of the U.S. Geological Survey North Carolina Water Science Center for their insightful comments which improved the document.

Blank Page

Contents

Abstract	1
1. Introduction	1
1.1. Purpose and Scope	3
1.2. Study Area	3
2. Methods	3
2.1. Data Source Overview	3
2.2. Data Research and Geodatabase Design	4
2.3. Data Sources	4
2.3.1. Updatable Data Sources	4
2.3.2. Static Data Sources	8
2.4. Data Loading from Staged Data Sources	8
2.5. Post-Load Data Quality Control	8
2.6. Managing Data for Continuity	10
2.7. Building the Collecting Agency, Data Source, and Alternate ID Tables	11
3. NMISC Geodatabase Refresh Process Steps	11
3.1. Refresh Updatable Data Sources	11
3.2. Delete the GEO_TBL_SITE Feature Class	11
3.3. Run the Pre-NMISC Data Refresh Script	11
3.4. Run the Make-USGS-Unique-Sites Script	13
3.5. Run the SQL Loader Queries	14
3.6. 12-Step Quality Control Process	14
3.7. Run the Post-NMISC Data Refresh Script	14
3.8. Create and Enable the GEO_TBL_SITE Geodatabase Point Feature Class	15
4. Long-Term Maintenance Considerations	16
5. References Cited	18
Glossary	20
Appendix 1. Updatable Data Source Profiles	23
1.1. U.S. Geological Survey (USGS) National Water Information System (NWIS) Data—Groundwater (GW) Field Water-Level Measurements	23
1.2. U.S. Geological Survey (USGS) National Water Information System (NWIS) Data—Groundwater (GW) Daily Water-Level Measurements	24
1.3. U.S. Geological Survey (USGS) National Water Information System (NWIS) Data—Water-Quality (QW) Instantaneous Field Measurements	25
1.4. U.S. Geological Survey (USGS) National Water Information System (NWIS) Data—Daily Surface-Water (SW) Discharge Values	26
1.5. U.S. Environmental Protection Agency (USEPA) Modern Storage and Retrieval Repository (STORET) Water-Quality Data	27
1.6. International Boundary and Water Commission (IBWC) Daily Surface-Water Discharge Data	28
1.7. International Boundary and Water Commission (IBWC) Water-Quality Data	29
1.8. Texas Water Development Board (TWDB) Groundwater Data	30
1.9. Texas Commission on Environmental Quality (TCEQ) Water-Quality Data	32

Appendix 2.	Static Data Source Profiles	33
2.1.	New Mexico Environment Department (NMED) Water-Quality Data	33
2.2.	Boyle-Parsons Water-Quality Data	34
2.3.	El Paso County Water Improvement District (EPCWID) Water-Quality Data	35
2.4.	S.S. Papadopoulos (SSPA) Surface-Water Data Compilation	36
2.5.	S.S. Papadopoulos (SSPA) Groundwater Data Compilation	38
2.6.	S.S. Papadopoulos (SSPA) Water-Quality Data Compilation	39
2.7.	New Mexico State University (NMSU) Surface-Water Discharge Data Compilation	40
2.8.	U.S. Environmental Protection Agency (USEPA) Legacy Storage and Retrieval Repository (STORET) Water-Quality Data	42
2.9.	Daniel B. Stephens and Associates (DBSA) Water-Quality Data Compilation	43
2.10.	City of Las Cruces (COLC) Groundwater-Level Data	44
2.11.	Sustainability of semi-Arid Hydrology and Riparian Areas (SAHRA) Water- Quality Data Compilation	45
2.12.	L.V. Wilcox and J. Williams Water-Quality Data—Recovered by New Mexico Interstate Stream Commission (NMISC)	46
2.13.	Wilson, Orr, White, and Roybal Data	47
2.14.	Stabler, H. USGS Data (1911)—Recovered by Daniel B. Stephens and Associates	48
2.15.	New Mexico Interstate Stream Commission (NMISC) Data Attributed to the U.S. Geological Survey (USGS)	49
Appendix 3.	Relational Geodatabase Quality Control (QC)	51
Appendix 4.	Geodatabase Table Dictionary	54
Appendix 5.	Example Microsoft (MS) Access Database Queries	59
Appendix 6.	NMISC Geodatabase User Tips	60
6.1.	Database Queries	60
6.2.	User Frequently Asked Questions (FAQ)	61
Appendix 7.	New Mexico Interstate Stream Commission (NMISC) Geodatabase— Geographic Information System (GIS) Functionality Overview	62
Appendix 8.	Additional Resources	63

Figures

1.	Map showing spatial extent of existing data for the Rio Grande Basin, from the Rio Arriba-Sandoval County line, New Mexico, to Presidio, Texas	2
2.	An example of the complex and interrelated nature of the various data sources identified as relevant to the New Mexico Interstate Stream Commission compendium	5
3.	A diagram showing the geodatabase tables and interrelations	6
4.	Screenshot for part of a U.S. Environmental Protection Agency data preprocessing script written by using the Visual Basic for Applications (VBA) programming language	7
5.	Screenshot for part of a geodatabase Structured Query Language loader query for the site table	9
6.	A conceptual representation of the geodatabase refresh process	12
7.	The top level of the data folder directory associated with the geodatabase	13

2-1. Illustration of potential data issues for awareness of users from the Stavros S. Papadopulos surface-water data source	37
2-2. Illustration of potential data issues for the awareness of users from the New Mexico State University surface-water data source	41

Tables

4-1. TBL_SITE table of the geodatabase. The table contains records for surface-water (discharge), groundwater, and water-quality sites	54
4-2. TBL_SAMPLE table of the geodatabase. The table contains data records associated with instantaneous sample-event data	55
4-3. TBL_RESULT table of the geodatabase. The table stores data records for parameter result values	56
4-4. TBL_DAILY_RESULT table of the geodatabase. The table stores data result records associated with daily values	57
4-5. TBL_SOURCE table of the geodatabase. The table contains data sources for the primary data tables of the geodatabase and serves as a domain table from the source identifier (ID) field (source_ID) where it occurs in other tables	58
4-6. TBL_PARAM table of the geodatabase. The table contains all hydrologic parameters associated with result values in the geodatabase and serves as a domain table for the parameter identifier (ID) field (param_ID) where it occurs in other tables	58
4-7. TBL_ALT_ID table of the geodatabase. The table contains site records that have known alternate site identifiers (IDs). An alternate site ID is another known identifier for a physical on-the-ground site	58
4-8. TBL_AGENCY table of the geodatabase. The table contains agency records of agencies associated with the data sources and serves as a domain table for the agency identifier (ID) field (agency_ID) where it occurs in other tables	58

Blank Page

Usage and Administration Manual for a Geodatabase Compendium of Water-Resources Data—Rio Grande Basin from the Rio Arriba-Sandoval County Line, New Mexico, to Presidio, Texas, 1889–2009

By Thomas E. Burley

Abstract

The U.S. Geological Survey, in cooperation with the New Mexico Interstate Stream Commission, developed a geodatabase compendium (hereinafter referred to as the “geodatabase”) of available water-resources data for the reach of the Rio Grande from Rio Arriba-Sandoval County line, New Mexico, to Presidio, Texas. Since 1889, a wealth of water-resources data has been collected in the Rio Grande Basin from Rio Arriba-Sandoval County line, New Mexico, to Presidio, Texas, for a variety of purposes. Collecting agencies, researchers, and organizations have included the U.S. Geological Survey, Bureau of Reclamation, International Boundary and Water Commission, State agencies, irrigation districts, municipal water utilities, universities, and other entities. About 1,750 data records were recently (2010) evaluated to enhance their usability by compiling them into a single geospatial relational database (geodatabase). This report is intended as a user’s manual and administration guide for the geodatabase. All data available, including water quality, water level, and discharge data (both instantaneous and daily) from January 1, 1889, through December 17, 2009, were compiled for the study area. A flexible and efficient geodatabase design was used, enhancing the ability of the geodatabase to handle data from diverse sources and helping to ensure sustainability of the geodatabase with long-term maintenance. Geodatabase tables include daily data values, site locations and information, sample event information, and parameters, as well as data sources and collecting agencies. The end products of this effort are a comprehensive water-resources geodatabase that enables the visualization of primary sampling sites for surface discharges, groundwater elevations, and water-quality and associated data for the study area. In addition, repeatable data processing scripts, Structured Query Language queries for loading prepared data sources, and a detailed process for refreshing all data in the compendium have been developed. The geodatabase functionality allows users to explore spatial characteristics of the data, conduct spatial analyses, and pose questions to the geodatabase in the form of queries. Users can also customize and extend the geodatabase, combine it with

other databases, or use the geodatabase design for other water-resources applications.

1. Introduction

The U.S. Geological Survey (USGS), in cooperation with the New Mexico Interstate Stream Commission, developed a geodatabase compendium (hereinafter referred to as the “geodatabase”) of available water-resources data for the reach of the Rio Grande from Rio Arriba-Sandoval County line, N. Mex., to Presidio, Tex. All readily available water-resources data were identified for the Rio Grande Basin from the Rio Arriba-Sandoval County line, N. Mex., to Presidio, Tex. In addition to compiling readily available surface-water-quality and stream-discharge data, groundwater (GW) data were compiled for the alluvial-fill basins in the study area (fig. 1). The geodatabase provides detailed information regarding the sampling locations (referred to as sites) and associated water-resources data. The geodatabase was designed so that any available surface-water-quality, stream-discharge, or groundwater data can be integrated into a single repository for the Rio Grande, both in the study area described herein as well as outside the study area.

A geodatabase is a spatially enabled database that is capable of handling both spatial and nonspatial data (Zeiler, 1999). A geodatabase provides a framework and an interactive tool to aid in the understanding of spatial and temporal trends in water quality and quantity. A geodatabase not only provides the data management and efficiency capabilities of a relational database but also gives additional capabilities for using and analyzing the spatial and temporal contexts of the data in an integrated fashion. With a geodatabase, geographically referenced data can be manipulated by using a geographic information system to produce maps, make relational database queries, and provide a foundation for various types of spatial analysis. Traditional database methods can also be used to query data using Structured Query Language (SQL). A database query is a question posed to a database whereby any data records meeting the query criteria are returned.

2 Usage and Administration Manual for a Geodatabase Compendium of Water-Resources Data—Rio Grande Basin

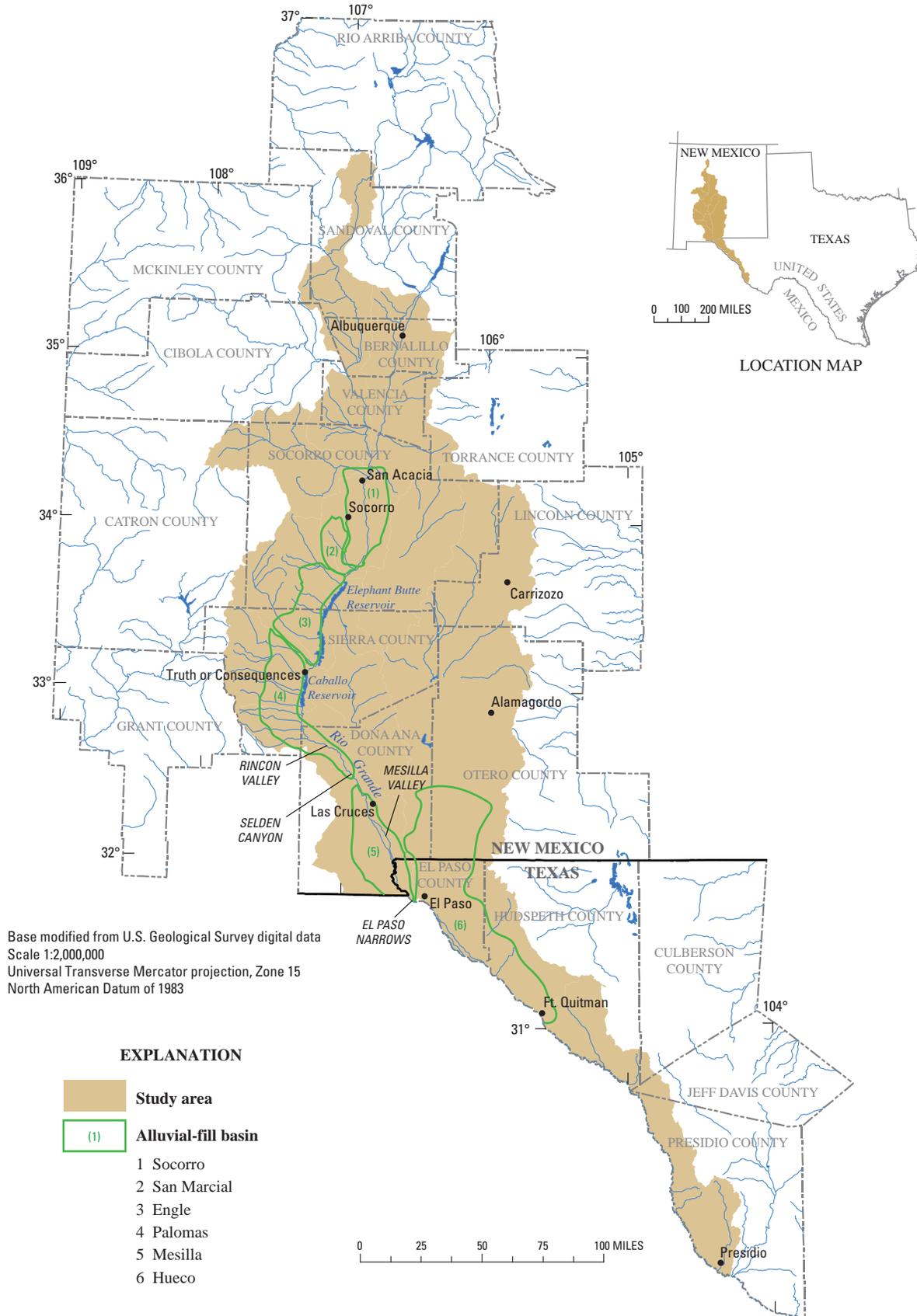


Figure 1. Spatial extent of existing data for the Rio Grande Basin, from the Rio Arriba-Sandoval County line, New Mexico, to Presidio, Texas.

Shah and Maltby (2010) developed a geodatabase of salinity data for the same study area used in the development of this comprehensive geodatabase of water-resources data—the Rio Grande Basin from the Rio Arriba-Sandoval County line, N. Mex., to Presidio, Tex. Shah and Maltby (2010, p. 1) noted “the geodatabase was designed as a tool for water-resource management and includes readily available digital data sources from the USGS, U.S. Environmental Protection Agency (USEPA), New Mexico Interstate Stream Commission (NMISC), Sustainability of semi-Arid Hydrology and Riparian Areas (SAHRA), Paso del Norte Watershed Council, numerous other State and local databases, and selected databases maintained by the University of Arizona and New Mexico State University.” The data compiled by Shah and Maltby (2010) were also compiled for the geodatabase described in this report.

1.1. Purpose and Scope

This report is the usage and administration manual for the geodatabase of water-resources data compiled for the Rio Grande Basin from the Rio Arriba-Sandoval County line, N. Mex., to Presidio, Tex. The June 2010 version of the geodatabase contains water-resources data from January 1, 1889, through December 17, 2009, including water-quality, water-level, and discharge data (both instantaneous and daily). Detailed instructions for using and maintaining these data in the geodatabase are provided; because this geodatabase is intended for ongoing use, database maintenance issues are given special consideration. The appendixes provide additional information on the data sources as well as examples of how to use the geodatabase.

1.2. Study Area

The geodatabase documented in this report was developed for the same study area used for another geodatabase developed by Shah and Maltby (2010), who provide the following description of the study area (p. 3):

The Rio Grande overlies six alluvial-fill basins in the study area—Socorro, San Marcial, Engle, Palomas, Mesilla, and Hueco (Wilkins, 1998). Detailed descriptions of the geologic structure of the alluvial-fill basins are given in Chapin (1971), Hawley (1978), Riecker (1979), Hawley and Kennedy (2004), Hawley and others (2005), and Hutchison (2006). In the study area, the Rio Grande is generally a gaining stream; ground-water inflow contributes to streamflow throughout Rincon and Mesilla Valleys. The Rio Grande is the ultimate point of discharge for the regional flow system which includes both the shallow alluvial aquifer of Rincon and Mesilla Valleys and the deeper regional flow system (Wilson and others, 1981; Bexfield and Anderholm, 1997; Hibbs and others, 2003). However, during

periods of drought, some reaches become losing streams (Wilson and others, 1981; Nickerson, 1995; Anderholm, 2002) when increased groundwater pumping causes drawdown and reversed gradients (Conover, 1954; Frenzel and others, 1992).

Whereas Shah and Maltby (2010, p. 1) focused on salinity-related data in the geodatabase they designed, the geodatabase documented by this report is intended as a compendium of all available water-resources data for the study area that can be used and updated on an ongoing basis. Salinity is among the most critical issues for the study area; Shah and Maltby (2010, p. 1) noted other investigators’ findings that

high concentrations of dissolved solids in this part of the Rio Grande Basin have been noted for almost 100 years (Stabler, 1911). The problems associated with high salinity are of growing concern for water-resource managers as rapid urban growth in cities along the Rio Grande in the United States and Mexico causes increased water demand and changes urban, agricultural, and environmental conditions and water uses. For example, the quality of water in the Rio Grande is becoming increasingly important as more surface water is proposed for diversion from the river for potable and non-potable uses (Langman, 2009). Historically, high concentrations of salinity in this reach have been attributed to (1) reservoir evaporation, which increases the concentration of dissolved solids in the remaining water in storage; (2) displacement of shallow saline groundwater during irrigation, which subsequently contributes to irrigation return flow; (3) erosion and dissolution of natural mineral deposits containing high concentrations of dissolved solids; and (4) inflow of deep saline or geothermal groundwater (Pillsbury, 1981; Allison and others, 1990; Moore and Anderholm, 2002; Philips and others, 2003).

2. Methods

2.1. Data Source Overview

The NMISC provided the USGS Texas Water Science Center (TXWSC) a copy of all data associated with a previous compilation of water-resources data done in the early 2000s. This previous compendium included data from sources that are not accessible online or were compiled in previous data compilations (hereinafter referred to as static data sources), or were from sources that could be accessed online (hereinafter referred to as updatable data sources) (Dale Doremus, New Mexico Interstate Stream Commission, written commun., 2009). Twenty-four data sources representing irrigation districts, municipal water utilities, universities, Federal, State,

and local agencies, and other entities were included in the geodatabase of water-resources data; these sources were identified with input from the NMISC and New Mexico Environment Department (NMED) staff. For the geodatabase described in this report, datasets available online were downloaded to ensure completeness and to help prevent propagation of data integrity problems from previous compilations and static files. An annotated list of the data sources and associated details is provided in the appendixes of this manual.

2.2. Data Research and Geodatabase Design

Commonalities across data sources and individual data files were carefully researched. The NMISC provided more than 1,750 individual data files in a variety of formats, such as delimited text files, spreadsheets, and Microsoft (MS) Access databases. Data files were carefully examined to determine redundancy with other files, as well as the condition of the data in terms of formatting and any other characteristics of concern that would need to be addressed. Data files were flagged as not usable if their contents were structured in such a way that rendering them into a usable format would require substantial effort that could not be automated using a programming language.

Original data from the entity that collected the data were obtained whenever possible. One of the biggest challenges was determining the relations among the various sources of data (referred to hereinafter as data originators or collecting agencies) and among those that compiled data from various sources (referred to hereinafter as data compilations or data sources) (fig. 2). Determining the relations among the various sources of data was further complicated by the large number of data files. In many cases, a data source (such as an agency's database) contained original data and data from other sources. An example of such a source is the Environmental Protection Agency Storage and Retrieval Repository (STORET) database (U.S. Environmental Protection Agency, 2010), a Web-accessible, federally supported database that aggregates data from state environmental management agencies. Relations among the various sources were determined and used to guide identification of files that contained (to the best of the author's knowledge) the most comprehensive and accurate set of data for each source.

A flexible, streamlined geodatabase was designed by using the information gained from source files to ensure that all necessary data fields were included (fig. 3). The USGS National Water Information System (NWIS) database (U.S. Geological Survey, 2010) design was used as a conceptual starting point by building from the idea that one site could have multiple sampling events, and each sampling event could have multiple samples, each associated with multiple data values. This conceptual design was further developed and revised based on characteristics of the final 24 data sources included in the geodatabase.

Differentiating between collecting agency and data source is necessary for managing data from 24 sources. Record-level

metadata documents where each data record originates. For example, previous compilations of Rio Grande data, such as the database developed by Stavros S. (S.S.) Papadopoulos (SSPA) (Dale Doremus, New Mexico Interstate Stream Commission, written commun., 2009) or EPA STORET (U.S. Environmental Protection Agency, 2010), contained data aggregated from many different collecting agencies. Using data from these sources as an example, data taken from either of these two databases would have a source of SSPA or EPA STORET. The number of collecting agencies contained in each data source, however, could number 10 or more; the collecting agency for a record from the SSPA database might be the Bureau of Reclamation (BR) whereas a collecting agency for a record from EPA STORET might be the National Park Service (NPS). Differentiating between collecting agency and data source becomes even more complex when site records are considered. Anomalous data were reconciled to the extent practicable. The same site often occurred in multiple data sources, but for the purpose of the geodatabase, only one record for a site is used in the site table (TBL_SITE). Often, a site loaded from one source had multiple associated samples as well as multiple collecting agencies and multiple sources. As a result, the collecting agency and data source fields were included on all relevant tables including the site table, sample table (TBL_SAMPLE), and daily result table (TBL_DAILY_RESULT) (fig. 3).

During the design process, consideration was given to the issue of long-term maintenance and the ability to update the geodatabase with new data. Not all data sources contained the same level of detail with supporting data tables (for example, not all groundwater records contained well construction information), so a relatively simple approach was used encompassing most of the available data. A simpler design (with relatively few data and lookup tables and uncomplicated relations among data) facilitates the update and maintenance process.

2.3. Data Sources

The various data sources were inspected to determine whether they could be downloaded from the Internet and for data integrity evaluation. Data sources were handled differently depending on whether they were updatable (currently [2010] being updated) or static (no longer being updated).

2.3.1. Updatable Data Sources

Public data that were readily available were downloaded. Some updatable data sources such as USGS NWIS and EPA STORET provide data distributed by either states or individual counties. When possible, the larger spatial extent was used to ensure a comprehensive download of data. The respective data source site tables were then imported into Environmental Systems Research Institute (ESRI) ArcGIS software to create point feature locations in a map (Environmental Systems Research Institute, 2010). The point feature locations were compared to the area of interest. The study area (fig. 1) was

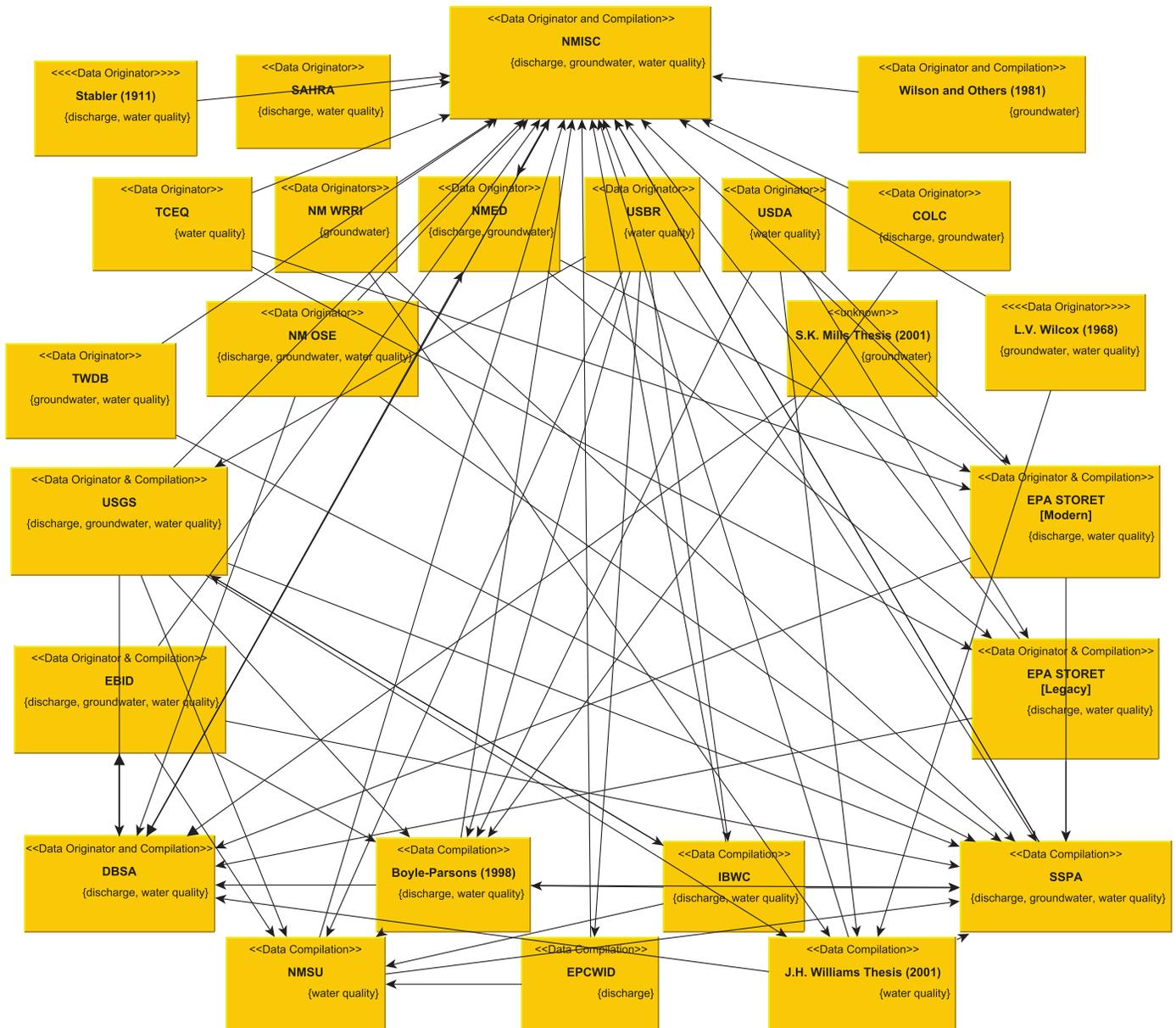


Figure 2. An example of the complex and interrelated nature of the various data sources identified as relevant to the New Mexico Interstate Stream Commission compendium. Each box represents a source, and on each source are noted the types of data contained in the source and if any other sources are contained in that source (for example, a previous data compilation). (See glossary for abbreviations as well as the appendixes for more information on the specific data sources.)

delineated as an ESRI shapefile and used as a filter to select sites within the study area. A half-kilometer buffer was used around the study area boundary so that sites on the perimeter could be evaluated to make sure their coordinates were correct.

Preprocessing scripts (single-use programs written in a given programming language) were written to automate data formatting on an ongoing (repeatable) basis. The use of such scripts increases consistency, reduces human error, and improves quality control by facilitating data integrity

and consistent formatting. Scripts were written by using the Visual Basic for Applications (VBA) programming language (fig. 4). VBA was used primarily because most data sources were “staged” (preprocessed) in MS Access databases, which are able to run VBA scripts by using the VBA Editor built into MS Access. The site table, sample table, result table, and in some cases parameter table (TBL_PARAM) format was used to stage the data for loading. All scripts were written and maintained as text files (.txt) and copied into MS Access

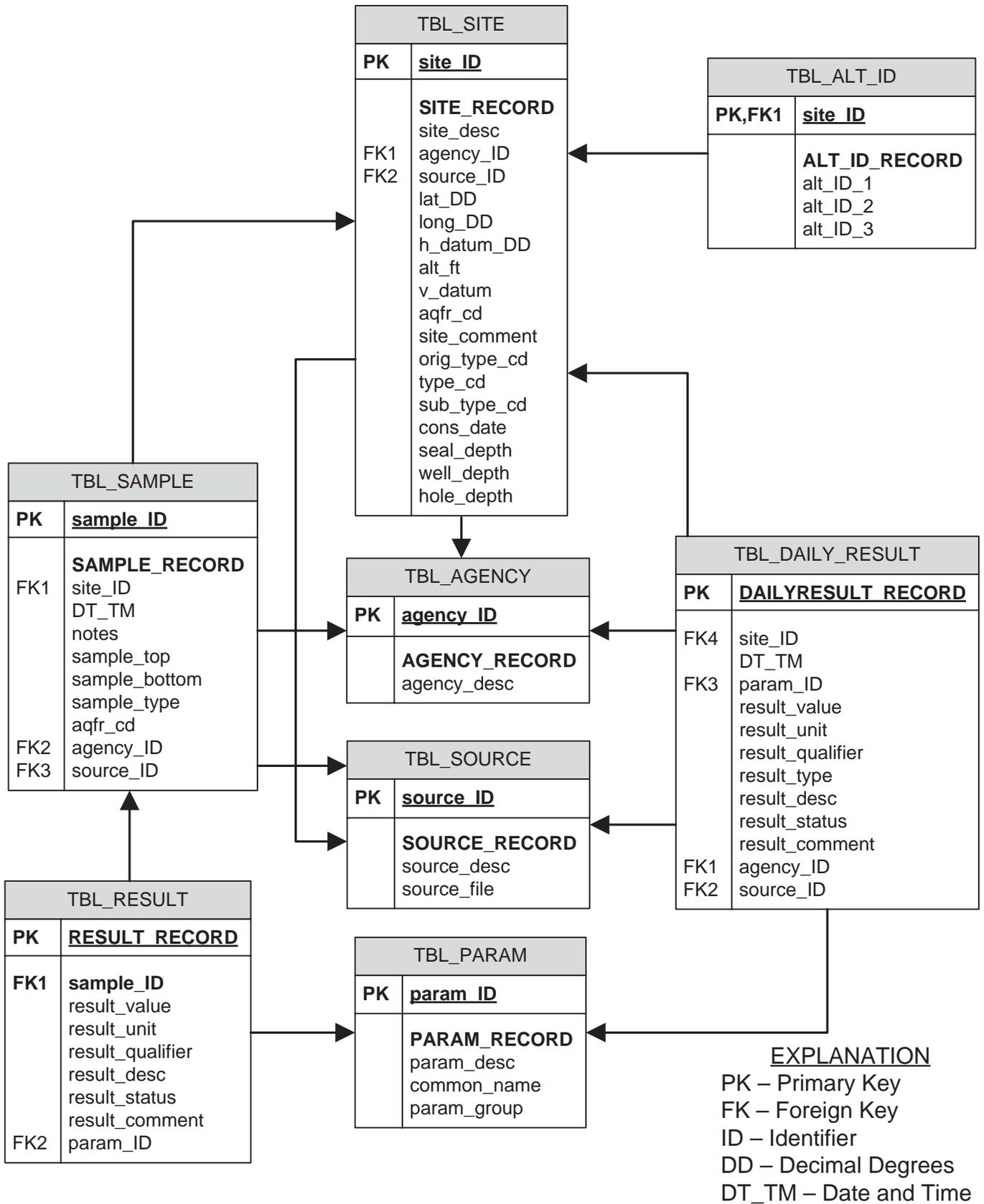


Figure 3. A diagram showing the geodatabase tables and interrelations. Tables 4–1 through 4–8 in appendix 4 provide information on the table field definitions and field data types.

```

51 ' Microsoft Office 12.0 Object Library
52 ' Microsoft Scripting Runtime
53 ' #####
54
55 Public Sub Load_EPA_Modern_Data
56 On Error GoTo ErrHandler
57 Dim basepath As String
58 basepath = ""
59 Dim Msg As String
60 Mst = ""
61
62 Msg = "Enter the location of the EPA Modern STORET data files." & vbCrLf & _
63     "EXAMPLE: D:\NMISC\EPA\Modern\"
64 basepath = InputBox(Msg)
65
66 If Right(basepath, 1) <> "\" Then
67     basepath = basepath & "\"
68 End If
69
70 Dim fso As New FileSystemObject
71 'Enter folder to search here, currently passed in via basepath variable
72 Dim folder As Scripting.folder
73 Set folder = fso.GetFolder(basepath)
74 'Defines the directory of files from the basepath
75 Dim file As Scripting.file
76
77 'For evaluating the file to make sure it's the right type
78 Dim filename As String
79
80 Dim pRS As New ADO.DB.Recordset
81 ' reads the lines in the text file
82 Dim strline As String
83 'String for holding SQL statements
84 Dim pSQL As String
85
86 ' Recordset for working with the sample table when creating the result table
87 Dim sRS As New ADO.DB.Recordset
88
89 ' array for the field names
90 Dim pFieldNameArray() As String

```

Figure 4. Screenshot for part of a U.S. Environmental Protection Agency data preprocessing script written by using the Visual Basic for Applications (VBA) programming language. Script comments can be seen as the lines having a leading apostrophe which is the VBA programming language convention for comments.

VBA editor for each data source when needed; this way, the scripts are maintained in one place outside of the MS Access databases.

To the extent practicable, the scripts were written to accommodate the data characteristics of the source files and to accommodate changes to the source files that might occur over time. An example of something that might change over time is

a computer file path. Extensive comments were used throughout all scripts to document the functionality and the code itself. Comments in the script can be identified by text lines that have a leading apostrophe (fig. 4). Comments are a different color when viewed in the VBA Editor, which helps them stand out. Other text editors also have the ability to display text files of a known programming language type (in this

case, VBA) in a color-coded format. A script file header was also included that describes key characteristics of each script, usage notes, an example of how to run the script by using the VBA Editor, the script purpose, the data source it was designed for, major requirements and dependencies for the script to run correctly (assuming no changes in data source file formats), and VBA Editor language library references.

2.3.2. Static Data Sources

Most static sources of data were manually prepared and staged in MS Access databases so they could be loaded into the geodatabase. The only exception was the EPA Legacy STORET data, which were downloaded and prepared by using a preprocessing script. Because Legacy STORET data were no longer updated after December 31, 1998 (U.S. Environmental Protection Agency, 2009), these data are a static source. If they were not already in Access, the static data files were imported into an MS Access database. Once the data were in an MS Access database, the original data tables were appended with the necessary fields (such as a “sample ID” field) to establish the relations with other tables, or new tables of data were derived from the original source and formatted for loading into the geodatabase. The site table, sample event table, data result table, and, in some cases, parameter list table were created for each static data source. Further details are included in the appendixes regarding the individual data sources, descriptions of the data sources, whether the data were considered “static” or “updatable,” descriptions of preprocessing steps (for example, a sample table derived from the original [raw] results table for staging purposes), the names and locations of text file VBA scripts, processing steps unique to each data source, and potential “user beware” data issues discovered during the data-screening process.

2.4. Data Loading from Staged Data Sources

Structured Query Language (SQL), a database computer language for managing data in relational databases, was used to write SQL statements for loading and retrieving data from the geodatabase (referred to hereinafter as “SQL loader queries”) (Lorents and Morgan, 1998). The SQL loader queries were used once all data from the various MS Access source staging databases were processed and staged (fig. 5). Use of the SQL loader queries provides a repeatable and consistent process that reduces the potential for error associated with hand-entering data into the MS Access tables. Five of the eight geodatabase tables (fig. 3) have their own SQL loader query. Loading of information pertaining to the collecting agency (TBL_AGENCY), data source (TBL_SOURCE), and the alternate ID (TBL_ALT_ID) is done using a VBA script discussed in section 3. Any data records flagged as containing daily data were loaded into the MS Access table called “TBL_DAILY_RESULT”; sample event-related result

records were loaded into TBL_RESULT. SQL language formatted comments were included to help distinguish the individual queries composing the set of SQL loader queries. It should be noted that MS Access does not handle SQL comments, and the labels shown in green (fig. 5) need to be removed before each query is copied into the SQL view query window.

Similar to the VBA preprocessing scripts, the SQL loader queries are maintained as separate text files. The SQL loader queries provide information on the specific files from which each source is being queried, the specific tables in each MS Access database, and the specific fields for each data source. The SQL loader queries also cross-reference information between the source data table fields and table fields in the geodatabase. For example, the site table SQL loader query for loading surface-water (SW) discharge sites of the International Boundary and Water Commission (IBWC) uses the “IBWC_ID” field located in a table called “Sites” in the IBWC SW MS Access source staging database called “IBWC_SW.mdb” at the local file location of “\NMISC\IBWC\IBWC_SW\” (fig. 5). The SQL loader queries incorporate lookup domain tables (when available) for referencing code definitions. The source data file paths shown in figure 5 can be modified as needed for the individual user’s computer; any change in an MS Access source staging database file location, the database file name, the database table name(s), or any database table field name(s) will require that all SQL loader queries be updated accordingly.

2.5. Post-Load Data Quality Control

Data loading is an iterative process that requires trial and error to determine how each individual source query statement in the SQL loader queries works by itself and in conjunction with the 24 sources of data included in the geodatabase. Adjustments were made in the SQL loader queries and in the data source tables themselves until duplicate sites, duplicate result records, and orphaned records were resolved. Additional detail related to major adjustments (for example, a table used to filter out sites from one source to avoid duplication with the same sites being loaded from another source) is outlined in the appendixes.

Various methods were used to perform quality-control checks of the data after running the SQL loader queries. All data tables in the geodatabase are assessed for orphan and duplicate records. Duplicate records and missing field values were initially identified by visual inspection. The development of a 12-step data quality-control (QC) process was necessary to identify duplicate table records as well as unmatched or orphaned records within tables. An orphan record occurs when a table has a record that lacks an associated (linked) record in a related table. An example of an orphan record is a sample record table with no associated result records in the result table. These sometimes occur and may necessitate adjustments to data loaders or preprocessing scripts. Data integrity

```

1  /*USGS Sites*/
2  SELECT site_ID, site_desc, agency_ID, source_ID, lat
3  type_cd, sub_type_cd, cons_date, seal_depth, well_de
4  FROM [D:\NMISC\USGS\USGS_MASTER_SITE.mdb].[USGS_DIST
5  WHERE site_ID IN (SELECT DISTINCT site_id FROM [D:\N
6  site_ID IN (SELECT DISTINCT site_id FROM [D:\NMISC\U
7  site_ID IN (SELECT DISTINCT site_id FROM [D:\NMISC\U
8  site_ID IN (SELECT DISTINCT site_id FROM [D:\NMISC\U
9  UNION
10 /*EPA Legacy STORET*/
11 SELECT s.Station AS site_ID, s.Station_Na AS site_c
12 s.Longitude AS long_DD, "" AS h_datum_DD, "" AS alt_
13 "SW" AS type_cd, s.Station_Ty AS sub_type_cd, "" AS
14 FROM [D:\NMISC\EPA\NMISC_EPA_Legacy_Data.mdb].NMISC_
15 WHERE s.Agency NOT IN (SELECT * FROM [D:\NMISC\EPA\N
16 AND s.Station NOT IN (SELECT * FROM [D:\NMISC\EPA\NM
17 AND s.Station IN (SELECT DISTINCT station FROM [D:\N
18 UNION
19 /*EPA Modern STORET*/
20 SELECT s.site_id AS site_ID, s.site_desc AS site_desc
21 s.long_dd AS long_DD, s.h_datum_dd AS h_datum_DD, ""
22 "SW" AS type_cd, s.type_cd AS sub_type_cd, "" AS cor
23 FROM [D:\NMISC\EPA\NMISC_EPA_Modern_Data.mdb].NMISC_
24 WHERE s.agency_ID NOT IN (SELECT * FROM [D:\NMISC\EF
25 AND s.site_id IN (SELECT DISTINCT site_id FROM [D:\N
26 UNION
27 /*IBWC SW*/
28 SELECT s.IBWC_ID AS site_ID, s.STATION_NM AS site_desc
29 "" AS h_datum_DD, "" AS alt_ft, "" AS v_datum, "" AS
30 "SW" AS type_cd, s.EPA_TYPE_1 AS sub_type_cd, "" AS
31 FROM [D:\NMISC\IBWC\IBWC_SW\IBWC_SW.mdb].Sites AS s
32 UNION

```

Figure 5. Screenshot for part of a geodatabase Structured Query Language loader query for the site table.

issues may also occur when the source database does not have sound data integrity. For example, a relationship between two data tables based on a table field in a source database that has a null, or empty, value can result in null values when the data are loaded into the project schema. Instances of incorrect null values were resolved in the source data if present. The QC process uses relational database integrity concepts, such as permanent linkages between related tables used to retrieve related data across tables (Lorents and Morgan, 1998), which provide an objective means to verify data compiled from so many different sources. The semi-automated approach facilitated record-by-record examination of the final database which included 5,614 unique sites, 1,050,698 water quality instantaneous result value records, and 1,822,599 daily value result records. A 12-step QC process is outlined and the actual SQL queries are listed in appendix 3.

Duplicate records identified after the initial database was compiled were removed to the extent possible; the resolution of duplicate sites and duplicate data is an ongoing maintenance issue for such a large-scale data compilation. In particular, numerous duplicate sites based on the Site ID were discovered. Many data sources, such as preexisting data compilations, assigned sites new site IDs without keeping the original site ID. For example, instances were discovered where a site, attributed to a particular source, did not have a site ID following the naming convention of the source. So, two or more sources might have contained the same physical on-the-ground site as a record yet had different site IDs. When available, “alternate” site IDs were written to the TBL_ALT_ID table, which serves as an important tool for data assessment and maintenance. The alternate ID table provides, where known and available, any alternate site IDs for a site.

All result fields in source data files need to be scrutinized to identify duplicate values; all fields associated with a sample event record in the sample table and an associated result table record in the result table were used to determine whether or not the combination of an event and result was unique. The NMISC compendium schema reflects a semi-normalized design with, for example, a sample event table that contains information pertaining to the samples (for example, sample depth, sample notes, sample date, and sample time), and then a result table with fields containing information pertaining to each parameter result record. The goal of database normalization is to reduce duplication of data in a database (Morris, 2005). For example, if a sample event date was included as a field with each parameter result value in a table, the sample event date would be duplicated across all parameter result records in the data table. During the QC process, the sample table and result table were combined to assess for duplicate data values by searching through all fields related to each result value. In some cases it is possible that a sample depth or sample note field might differentiate two sample records or even two result table records that otherwise appear similar. Likewise a result comment field might differentiate between two result records that otherwise appear the same.

Results and sample event data were loaded as they existed in the MS Access databases containing the 24 data sources, including keys (identifiers that locate specific rows in a table [Lorents and Morgan, 1998]) that cross-walked (that is, could be directly imported and used) in the geodatabase to differentiate between records on the basis of values of fields in the tables for each sample event and result combination. No attempts were made to interpret the differences between similar records. For example, no attempt was made to interpret the meaning of two result comments with slightly different wording. Such an example might be a result comment of “CALM CONDITIONS” while another record having all values the same except for a result comment of “CALM CONDITIONS, CLOUDY.” The two are technically different, and determining which one might be correct would require interpretation; given the size of the database this was not feasible.

2.6. Managing Data for Continuity

Because some of the records provided by the NMISC contained additional fields or data values that were added to expound upon the data contained in the records from the various sources, a process had to be developed to manage for data continuity. The additional data provided by the NMISC staff included fields to enhance usability of the compiled data, or data to better characterize compiled records such as site types or water-quality constituent names. It was important to properly synchronize these data with the data from the compiled sources, using automated processing steps. An approach was devised to use a separate reference MS Access database for maintaining reference copies of the geodatabase tables containing custom data. This separate MS Access reference database named “DATA_REFRESH.mdb” is designed to

work with a VBA script that is run after all data sources have been recompiled and all data QC checks in appendix 3 have been run. This script is contained in the text file called Post_NMISC_Data_Refresh.txt and is located in the final source data folder tree location described in the refresh process steps in section 3.

The post-NMISC data refresh script builds the collecting agency, data source, and alternate ID tables from the geodatabase data tables, and then populates custom-defined data such as source descriptions and source file names from the reference copy of the table maintained in the “DATA_REFRESH.mdb” database. The reference copy of a table contains data defined by the NMISC and is where administrators of the geodatabase would refine user-defined fields or update existing fields and their data values in the reference tables. This script also populates data fields in the site table and parameter table including site types, site subtypes, parameter common names, and parameter groups.

The post-NMISC data refresh script automates the geodatabase refresh process by updating the geodatabase tables using data from the associated reference tables on the basis of the table primary keys (a unique identifier meaning that no two rows in the table can have the same value in the column(s) identified as the primary key [Lorents and Morgan, 1998]). The updated table is then exported back to the DATA_REFRESH.mdb database for use with future data refreshes. The reference tables in the reference database can then be updated manually by the user for any records that have a null value for a user-defined field. If a null value is found in a reference table after a data refresh, it is likely that the refreshed database contains new data not previously present. In such a case the new data will not have an associated record in the reference table and thus will have nothing to update against. But, if the user then updates the field of interest in the reference table and runs the script again, the particular geodatabase table field will be updated, assuming no other appreciable changes.

The post-NMISC data refresh script handles updates for the site table, parameter table, agency table, and source table. The site, parameter, agency, and source tables all contain custom-defined data that cannot be compiled directly from the MS Access source staging databases. Table names in the DATA_REFRESH.mdb database are nearly the same as their counterpart table in the geodatabase, except that the word “_REFERENCE” is appended to the end of each table name. For example, the table TBL_SITE is named TBL_SITE_REFERENCE in the DATA_REFRESH.mdb Access database.

The alternate ID table (TBL_ALT_ID) is maintained as a stand-alone table in the DATA_REFRESH.mdb table with the name TBL_ALT_ID_REFERENCE. An alternate site ID is any unique identifier for a site other than the primary site ID defined in the source data. If new alternate IDs are discovered, the sites (if not already present) and their associated alternate IDs need to be added to the TBL_ALT_ID_REFERENCE table along with the associated primary site ID. This can be done by manually keying site records into the TBL_ALT_ID_REFERENCE table or by using automated SQL methods.

More details about the post-NMISC data refresh script can be found in the script file usage header as well as in the extensive code comments included in the script.

2.7. Building the Collecting Agency, Data Source, and Alternate ID Tables

The collecting agency (TBL_AGENCY), data source (TBL_SOURCE), and Alternate ID (TBL_ALT_ID) tables are built by using the script described in section 2.6. These three tables, along with the parameter table (TBL_PARAM), serve as domain tables for the data tables where their respective primary key fields exist as foreign key fields (index keys used to link tables [Lorents and Morgan, 1998]). The collecting agency ID codes and source ID codes are compiled into their respective domain tables from all existing data tables after the data tables have been loaded and reviewed for quality. The collecting agency descriptions, data source descriptions, and data source file paths were manually entered into the reference tables. The post-NMISC data refresh script documents the names of any collecting agencies that are directly loaded from their MS Access source staging database.

The alternate ID table was compiled by using data fields defined as alternate IDs from the source data. For example, an SSPA database (see appendix 2) contained alternate site IDs and no other data. The NMISC staff also provided a list of site IDs with known alternate IDs for a number of sites. SQL queries were used to combine the alternate ID data from these sources into one table. A one-to-one relationship exists between the alternate ID table and the primary site table (TBL_SITE) to help facilitate data queries and inspection of data with the alternate ID information.

3. NMISC Geodatabase Refresh Process Steps

This section discusses the steps to refresh (update) the geodatabase (fig. 6). It is important to make a backup copy of any original (unmodified) data file before modifications are made to any of the source files, VBA scripts, or SQL loader queries. This also applies to the geodatabase. Backing up a working copy of the original data and any associated files allows for a recovery option if a problem arises while modifying the data or any programming code (scripts and SQL loader queries).

3.1. Refresh Updatable Data Sources

All updatable data sources need to be refreshed by following the source-specific instructions outlined in appendix 1. It is up to the user to determine on the basis of project or agency needs when an updatable data source should be

refreshed; the frequency of change for each data source is variable. Data for each updatable source will need to be downloaded and processed with the associated VBA script in their respective MS Access source staging database file. All VBA scripts are run in their respective MS Access source staging database file except for the scripts associated with the Texas Water Development Board (TWDB) data or IBWC water-quality data. The locations of the MS Access source staging databases can be determined by the “source_file” field in the source table (TBL_SOURCE). Figure 7 shows the top level of the data folder directory provided to NMISC for the data sources included in the geodatabase. Computer hard drive letters are dependent on the computer on which the data are stored, but file paths will remain the same regardless of the computer drive letter (as long as the locations of data files in the data folder directory (fig. 7) provided to the NMISC in June 2010 remain unchanged).

The data preprocessing scripts were written to allow the user to specify file paths that may change over time. Keeping all data files in the same location over time can help reduce confusion if different users carry out the data refresh process. The previous raw data files can be deleted or archived to another location, and the new raw data files can be placed in the same location. All text files containing preprocessing scripts can be found at: \NMISC\Scripts\VB\ based on the “Scripts” folder (fig. 7). It is important that the user read the VBA script usage header comments to understand script function, in addition to read the script code comments to understand the script components.

3.2. Delete the GEO_TBL_SITE Feature Class

The GEO_TBL_SITE feature class needs to be deleted by using ESRI ArcCatalog before proceeding with a database refresh. To delete the feature class, open ESRI ArcCatalog, select the GEO_TBL_SITE feature class in the NMISC geodatabase, right click, and select “delete.” This step will automatically delete any ESRI Geodatabase relationship classes associated with the feature.

3.3. Run the Pre-NMISC Data Refresh Script

The script contained in the text file named Pre_NMISC_Data_Refresh.txt needs to be run by using the VBA editor according to the script usage header instructions in the geodatabase. This script is used for processing steps needed before a refresh of the geodatabase is completed. It is important that the user read the VBA script usage header and script code comments and understand what the script is doing. This script will drop all table constraints including table relationships and primary keys. The script then drops all primary data tables including the GEO_TBL_SITE geodatabase feature class copy of the site table. Finally, the script recreates all primary data tables by using the SQL code described in the Create_NMISC_Schema.txt file. All text files containing processing

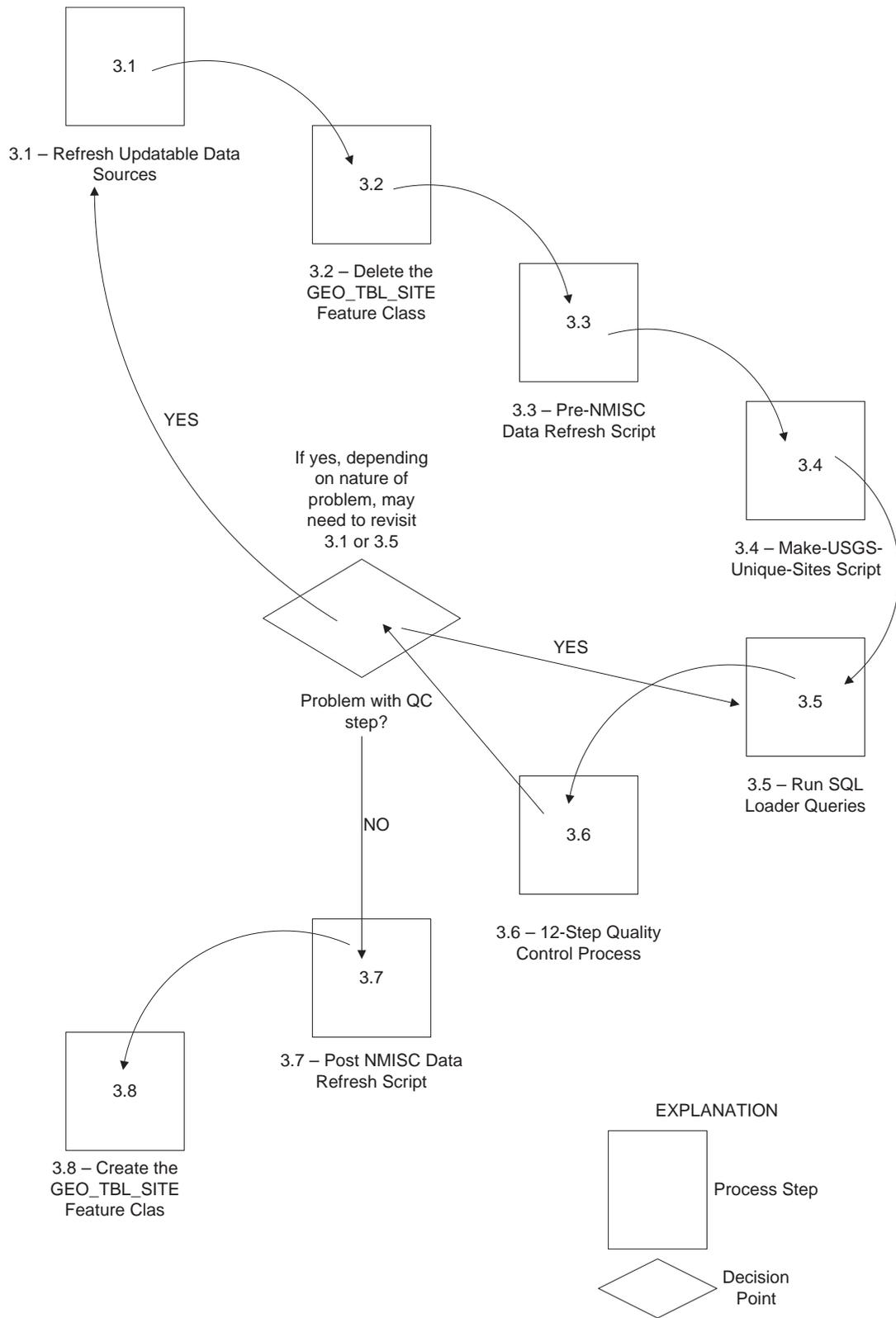


Figure 6. A conceptual representation of the geodatabase refresh process.

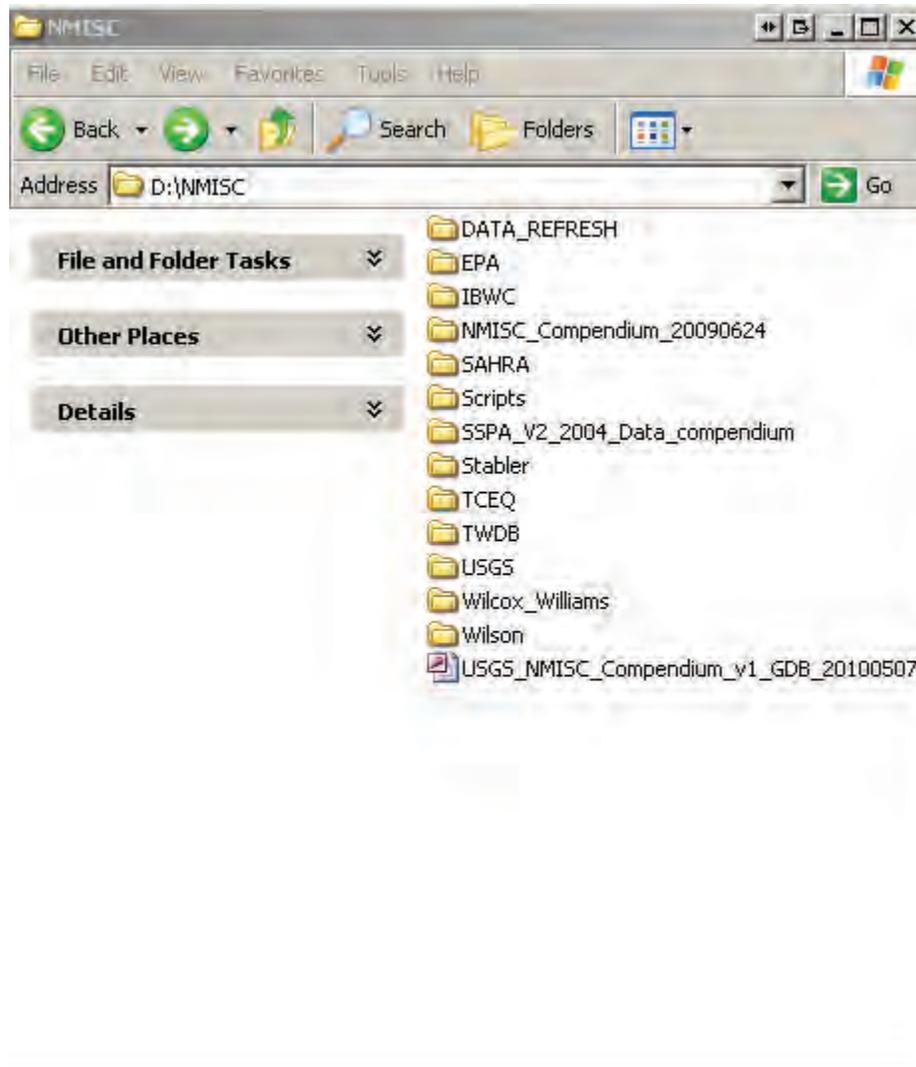


Figure 7. The top level of the data folder directory associated with the geodatabase.

scripts can be found at: \NMISC\Scripts\VB\ in the project data directory structure (fig. 7).

3.4. Run the Make-USGS-Unique-Sites Script

The script contained in the file “ResolveDuplicateSitesFromMultipleNWISWebSources.txt” needs to be run according to the script usage header instructions. The purpose of this script is to resolve duplication of sites that can occur when retrieving data from multiple NWIS Web interfaces (for example, NWIS QW Daily, QW Field, SW, GW Daily, and GW Field). Some sites are returned from more than one NWIS Web interface because multiple data types may be collected at that site. This script resolves instances of duplicate site records on the basis of NWIS site IDs and is dependent on the SQL loader query for USGS sites (found in the “Union_USGS_Sites.txt” file). A table created

to maintain the unique set of NWIS site IDs is contained in the MS Access database “USGS_MASTER_SITE.mdb” at \NMISC\USGS\ in the NMISC geodatabase structure (fig. 7). The site SQL loader query is then used to check for USGS sites already in the table contained in the database. A source ID code of “USGS” is used only for site records obtained from USGS NWIS and will not be found in any other primary data table. The source codes for any data table records obtained from USGS NWIS will characterize the respective USGS source (for example, USGS_QWField).

All text files containing scripts can be found at \NMISC\Scripts\VB\; all SQL code and SQL loader queries can be found at \NMISC\Scripts\SQL\ in the data directory structure (fig. 7). The SQL loader query contained in the “Union_USGS_Sites.txt” need to be saved as a query named “Query_USGS_Sites” in the NMISC geodatabase before running the script.

3.5. Run the SQL Loader Queries

The SQL loader queries each need to be copied into an NMISC geodatabase SQL-view query window. There are five SQL loader queries that will need to be copied: one for the parameter table, one for the daily value results table, one for the results table, one for the samples table, and one for the sites table. SQL-format comments were inserted between each source SQL query statement in the maintaining text files to facilitate differentiation among the individual queries that make up the SQL loader queries. MS Access cannot handle SQL comments, and the labels shown in green in figure 5 need to be removed before each SQL loader query is copied into the SQL view query window. An easy way to do this is to just copy everything into the SQL-view query window and to scroll down the window, deleting the comments denoted by “/*” and “*/” tags.

Specifying the source-data file paths in the SQL loader queries as shown in figure 5 with a computer hard drive letter was unavoidable because MS Access SQL requires a complete file path; however, the file paths can be modified in the SQL loader queries located in the text files found at \NMISC\Scripts\SQL\ in the NMISC data file structure (fig. 7). Any change in the location of an MS Access source staging database file, database file name, database table name(s), or any database table field name(s) will require that all table SQL loader queries be updated accordingly. The SQL loader query needs to be saved in the NMISC geodatabase by using the following names:

- Sites SQL loader query (UnionSites_NMISC.txt) is saved as “Query_Site_Sources”
- Samples SQL loader query (UnionSamples_NMISC.txt) is saved as “Query_Sample_Sources”
- Results SQL loader query (UnionResults_NMISC.txt) is saved as “Query_Result_Sources”
- Parameters SQL loader query (UnionParameters_NMISC.txt) is saved as “Query_Param_Sources”
- Daily Results SQL loader query (UnionDailyValueResult_NMISC.txt) is saved as “Query_Daily_Result_Sources”

To append the data returned by using the SQL loader queries into the appropriate data tables, the table queries need to be called by separate SQL statements. The SQL append statements (located in the \NMISC\Scripts\SQL\ directory) need to be saved in the geodatabase with the following names:

- Sites append query (TBL_SITE_Append_SQL.txt) is saved as “Load_Sites”
- Samples append query (TBL_SAMPLE_Append_SQL.txt) is saved as “Load_Samples”
- Results append query (TBL_RESULT_Append_SQL.txt) is saved as “Load_Results”

- Parameters append query (TBL_PARAM_Append_SQL.txt) is saved as “Load_Param_Table”
- Daily Results append query (TBL_DAILY_RESULT_Append_SQL.txt) is saved as “Load_Daily_Values/Water_Levels_Results”

To load the data into the geodatabase tables from the MS Access source staging databases, double-click on each append query and the associated SQL loader query will be called. Because of the magnitude of data records being pulled together from 24 files, the append queries may run for several minutes or longer depending on the computer and its specifications. During this time, the opened NMISC geodatabase will be relatively unresponsive but the normal MS Access query status bar will typically be visible near the bottom right side of the screen. As a result, a data refresh of the NMISC geodatabase is best done on a computer that can be left alone and not used while each append query is running so that computer central processing unit (CPU) resources and memory can be allocated to MS Access. All text files containing scripts can be found at: \NMISC\Scripts\VB\ in the data directory structure (fig. 7).

3.6. 12-Step Quality Control Process

The 12-step QC check process needs to be conducted after all data tables have been compiled. Appendix 3 outlines the SQL-based QC routines as well as the exact SQL statements, and section 2.5 discusses the approach and reasoning behind the methods used. If duplicates or orphaned records are discovered, it is up to the user to determine the circumstances specific to the duplicate and/or orphaned records. Determining the cause of duplicate data is an iterative process. There is no single systematic way to specify how duplicate data should be handled. It is highly likely that adjustments will need to be made to the associated table SQL loader query. It is also possible that the issue may lie with the actual source data and that the respective source preprocessing scripts will need to be modified accordingly. A backup copy of an original unmodified data file needs to be made before any major modification is done to a source file or any of the script or SQL maintaining text files. Backing up a working copy of the original data allows for a recovery option if a problem arises while modifying the data to account for duplicates or other issues.

3.7. Run the Post-NMISC Data Refresh Script

The script contained in the text file named Post_NMISC_Data_Refresh.txt needs to be run according to the script usage header instructions by using the NMISC geodatabase VBA editor. The purpose of the script is to carry out necessary steps after doing a complete refresh of the NMISC geodatabase.

The specific functions of this script include the following:

- Builds the agency (TBL_AGENCY) table and source (TBL_SOURCE) table by querying the agency IDs and

source IDs from all data tables, and then updates the additional respective table fields by using SQL update queries against the “Reference” tables in the DATA_REFRESH.mdb MS Access database.

- Updates the parameter table against the parameter reference table in the DATA_REFRESH.mdb file for the common names and parameter group fields.
- Updates the site table against the site reference table in the DATA_REFRESH.mdb file for the site type codes and subtype codes.
- Builds the alternate ID table (TBL_ALT_ID) by creating the table and then inserting the alternate ID data records from the reference copy of the table maintained in the DATA_REFRESH.mdb MS Access database.
- Creates the database constraints including table primary keys as well as table relationships.

The reference table approach allows for user-defined fields to be manually updated by the user in the DATA_REFRESH.mdb file tables. The updated geodatabase data tables handled by this script (all except for the TBL_ALT_ID table) are exported back to the DATA_REFRESH.mdb database to over-write the previous respective reference table copy when the script is run. The tables in the reference database can then be updated manually by the user for any records that have a null value for a user-defined field. One example might be a new agency table (TBL_AGENCY) record that does not have an agency description for the associated agency ID. Such a scenario likely indicates that the refreshed database contains new data that were not present in the previous configuration of the geodatabase. In such a case the new data will not have an associated record in the reference table and thus will have nothing to update against for any associated user-defined fields. If the user then updates the field of interest in the reference table (in this case, a description for the new agency), however, and runs the script again, the particular geodatabase table field then needs to be updated, assuming no other significant changes. In the case of the alternate ID table, known sites that have one or more alternate IDs will need to be entered manually or by using SQL into the TBL_ALT_ID_REFERENCE table.

The user needs to inspect the agency table (TBL_AGENCY), source table (TBL_SOURCE), and parameter table (TBL_PARAMETER) after running the script to see if any new agencies, parameters, or sources have been added. If new agencies, parameters, or sources have been added, the user can manually enter a value in the associated DATA_REFRESH.mdb table (named accordingly but with “_REFERENCE” appended to the table name). The script then needs to be rerun and all tables rechecked.

The post-NMISC data refresh script handles such updates for the site table, parameter table, agency table, source table, and alternate ID table, all of which contain custom-defined data that cannot be compiled directly from the MS Access

source staging databases. Currently (2010), the user-defined fields in the reference tables located in the DATA_REFRESH.mdb file that are used to populate the associated named field in the primary data tables of the NMISC geodatabase are as follows:

- TBL_SITE_REFERENCE: type_cd, sub_type_cd, alt_ID
- TBL_PARAM_REFERENCE: param_group, common_name
- TBL_AGENCY_REFERENCE: agency_desc
- TBL_SOURCE_REFERENCE: source_desc, source_file
- TBL_ALT_ID: site_ID, alt_ID_1, alt_ID_2

More details about the script and its mechanics can be found in the script file-usage header as well as in the extensive code comments throughout the script. All VBA/VBscript text files and SQL loader query files can be found at: \NMISC\Scripts\VB\ based on the “Scripts” folder (fig. 7).

3.8. Create and Enable the GEO_TBL_SITE Geodatabase Point Feature Class

The ESRI ArcGIS ArcCatalog application is used to turn the TBL_SITE table into a point feature class viewable in ArcMap. The steps to do this outlined below are based on ArcGIS version 9.3.1 (Environmental Systems Research Institute, 2010):

1. Open ArcCatalog
2. Browse in the ArcCatalog Table of Contents window to the NMISC geodatabase file. The Table of Contents window is the Windows Explorer-style window on the left-hand side that enables navigation to data files.
3. Expand the file so that all tables in the geodatabase can be seen.
4. Select the TBL_SITE table and right-click to display menu options. Select the option called “Register with Geodatabase.”
5. Right-click again on the TBL_SITE table and select “Create Feature Class.” This step will open up a second side menu labeled “From XY Table” that the user needs to select.
6. The “Create Feature Class From XY Table” window should open. In it make sure that the X Field is set to the long_DD field, and that the Y Field is set to the lat_DD field. Then click on “Coordinate System of Input Coordinates” which will open the Spatial Reference Properties window. Click the button labeled “Select...” which will open a third window that enables browsing to

a coordinate system. A geographic datum is needed for locations noted by latitude and longitude; depending on the application and intent, a projected coordinate system could also be used. Be sure to check any information available on the horizontal geographic datum of the sites if there is concern regarding the level of accuracy.

7. Proceed to select the appropriate coordinate system and then click “OK” in the Spatial Reference Properties window.
8. Make sure that the Output location for the feature class is pointing to the NMISC compendium geodatabase. **IMPORTANT:** Change the name of the feature class to be created to “GEO_TBL_SITE” when specifying the name to save the new feature class.
9. Click “OK” on the “Create Feature Class From XY Table” window.
10. Right-click on the NMISC geodatabase file in the ArcCatalog Table of Contents window and select “Refresh.” The GEO_TBL_SITE point feature class should now be visible.
11. Reconfigure all relationships associated with the site-table feature class. The relationships need to mirror the relationship classes set up in the MS Access database (see fig. 3.) associated with TBL_SITE.

NOTE: The GEO_TBL_SITE table can only be used with all other geodatabase data tables in the ESRI ArcGIS environment. The TBL_SITE table can only be used with all other geodatabase data tables in the MS Access environment. This limited use is because the configured table relationships in MS Access are set to work with the nonspatial TBL_SITE table, and configured relationship classes in ESRI ArcGIS will only work with the spatially enabled GEO_TBL_SITE table.

4. Long-Term Maintenance Considerations

This section provides an overview of some things to consider for the long-term maintenance of the geodatabase. The stewardship of such a diverse compilation of data requires considerable forethought and planning with regards to potential new data sources as well as the 24 individual sources included in the June 2010 version of the geodatabase. In addition, changes in updatable source-data formats are as yet unforeseeable. Documentation and metadata are critical for capturing the evolution of any complex database with regards to data format changes, field methodology changes, and technology changes (Burley and Peine, 2009). The following list describes some user actions that could result in lost or corrupted data, difficulty in carrying out the process steps as outlined herein,

or difficulty using the SQL loader queries and preprocessing scripts:

- Never manually key data into the geodatabase (including manually updating using SQL); always use the geodatabase refresh process described in section 3.3. Manually keying data into NMISC geodatabase tables will result in data loss if a geodatabase refresh is subsequently done. The source for the records of interest will guide determination of the data being “static” or “updatable” as described in section 2.3. If the data are an “updatable” data source as defined in section 2.3’, the alternate approach of entering the updated data into the appropriate field of the reference copy of the table in the DATA_REFRESH.mdb file can be used (see sections 2.6 and 3.7). The post-NMISC data refresh script will then need to be modified to accommodate that particular table and data field if that field is not already handled in the script.
- Ad hoc entering of data into the compiled tables will make tracking of data in the defined refresh process steps difficult. Instead, stage the data (if the data constitute an entirely new source not represented in the 24 sources included in the June 2010 version of the geodatabase) in a formatted MS Access source staging database file and load the data following the steps outlined in section 3.3. Properly loading new data will require creating an additional site, sample, result, parameter, and daily result (if applicable) SQL loader query statement for the new source. These new SQL loader query statements would be added to each of the appropriate existing SQL loader queries. If the data are part of an existing source, cleaning, formatting, and appending the new data to the existing site, sample, result, and parameter tables of the appropriate MS Access source staging database will facilitate data integration.
- Documentation is a critical part of managing data for long-term use, and detailed documentation will be necessary to manage the geodatabase over time; incomplete or inaccurate documentation erodes the utility of a complex database (Burley and Peine, 2009). Upkeep of this user’s manual as well as maintaining data administration log files will allow others to determine what changes have been made over time relative to the original compilation and methods. For example, keeping a simple text file log that documents each instance a section 3 process step is implemented and any issues that arise, changes made to accommodate unexpected issues, and any other items of interest for each time the geodatabase is refreshed will create a documentation log necessary for and valuable to maintenance over time. Even seemingly trivial modifications such as the changing of a field name in an NMISC geodatabase table can result in unforeseeable consequences.

Documenting every change to the geodatabase or to SQL scripts used to process data (such as reasons the name of a given field was changed, related changes made to process steps, preprocessing scripts, other SQL scripts, or other changes), along with the name of the person who made the changes, will allow any user to track the evolution of the geodatabase. In addition, keeping the processing script usage file headers and code comments current for any changes made will allow others to determine the reasoning behind changes and document why changes were necessary.

- Maintaining backup copies of all original source data files, original SQL loader query maintaining text files, original preprocessing VBA script maintaining text files, as well as the most recent “production ready” version of the NMISC geodatabase is important for long-term maintenance of the geodatabase. Should problems arise with attempted modifications, attempted data refreshes, or both, the backup copies can be used to start over. Backups can also serve as an “out” in case data get irrevocably and unintentionally corrupted. The original files can be maintained as a reference copy that can be used to make clean copies of a particular file(s) in a scenario that requires the database to be rebuilt from the original source files.
- Unforeseeable changes in one of the updatable data sources (appendix 1) might cause problems with preprocessing or data loading. Such changes might include changes in the format of the source data (for example, table or text file structure), changes in the names of key data fields, or changes in the data themselves (for example, parameter codes). Changes of this nature will likely result in errors during the data refresh process. If changes have occurred, such changes may present themselves at any point in the data refresh process. A change in parameter codes, for example, might not become apparent until possibly the 12-step QC process (appendix 3). Other changes such as the change of a key data field name, however, would likely become apparent during the use of data preprocessing script because an error message will appear when the script is running. The course of action for addressing a change in an updatable source will vary case by case and will depend on the type of change that has occurred.
- Changing any of the geodatabase table field names or table names will require updating all SQL loader queries, any QC process steps that use the field that has been changed, the pre-NMISC data refresh and post-NMISC data refresh scripts (sections 3.3 and 3.7, respectively), and the user manual documentation including the data dictionary (tables 4-1 through 4-8) and database diagram (fig. 2).
- The actual file locations within the source data structure are important because the SQL loader queries and scripts depend on consistent file locations. Processing scripts were written to be flexible where practicable. Changing the locations of the MS Access source staging databases will require updating the file paths in the SQL loader queries and in some of the processing scripts (the processing script usage headers can provide specifics in such scenarios). Modifying any USGS source-data location will require updating the “make-USGS-unique-sites” script and the USGS sites SQL loader query. Also, some updatable data sources have source staging databases hard coded in the post-NMISC data refresh script because of the script executing SQL queries containing hard-coded paths. Any of these instances are noted in the post-NMISC data refresh script usage header.
- It is important to use the June 2010 version of the static-source data files (for example, the S.S. Papadopoulos data compilation), because these files have been preprocessed and staged for loading. Draft copies provided to the NMED or the NMISC prior to June 2010 should not be used. Previous draft copies might contain site duplicates among multiple sources or other data integrity problems that were later resolved in the June 2010 version.
- The lifespan of the data formats used for the source staging databases, as well as the overall geodatabase, needs to be considered (Morris, 2005). Change over time in available data file formats and software will require migration to new data file formats, and adaptation of data programming code and likely geodatabase refresh-process steps. It often can be easier to plan for and execute a data file or format migration before it becomes mandatory to migrate. If a process step, file format dependent on a particular software package, or data format is no longer working or compatible with another aspect of the process, data corruption might occur, or user downtime might result.
- Establishment of a data policy and stewardship plan for the geodatabase is critical for maintaining such a complex assemblage of data over time (Burley and Peine, 2009). Identifying appropriate qualified staff members as responsible parties for maintaining the geodatabase, and cross-training additional staff, will help maintain a proficient knowledge of the geodatabase despite inevitable staff turnover and organizational changes. In addition, documenting how the geodatabase is used and devising a mechanism for citing the geodatabase when it is used in environmental studies could help acknowledge the added value provided by such a complex data compilation effort.

5. References Cited

- Allison, G.B., Cook, P.G., Barnett, S.R., Walker, G.R., Jolly, I.D., and Hughes, M.W., 1990, Land clearance and river salinization in the western Murray Basin, Australia: *Journal of Hydrology*, v. 119, p. 1–20.
- Anderholm, S.K., 2002, Water-quality assessment of the Rio Grande Valley, Colorado, New Mexico, and Texas—Surface-water quality, shallow ground-water quality, and factors affecting water quality in the Rincon Valley, south central New Mexico, 1994–95: U.S. Geological Survey Water-Resources Investigations Report 02–4188, 117 p.
- Bexfield, L.M., and Anderholm, S.K., 1997, Water-quality assessment of the Rio Grande Valley, Colorado, New Mexico, and Texas—Ground-water quality in the Rio Grande flood plain, Cochiti Lake, New Mexico, to El Paso, Texas, 1995: U.S. Geological Survey Water-Resources Investigations Report 96–4249, 93 p.
- Boyle Engineering Corporation and Parsons Engineering Science, 1998, Boyle-Parsons Data Compilation.
- Burley, T.E., and Peine, J.D., 2009, National Biological Information Infrastructure Southern Appalachian Information Node (NBII-SAIN) Data Management Toolkit: U.S. Geological Survey Open-File Report 2009–1170, 96 p.
- Chapin, C.E., 1971, The Rio Grande Rift, part 1—Modifications and additions, *in* James, H.L., ed., *Guidebook of the San Luis Basin, Colorado: New Mexico Geological Society, Twenty-Second Field Conference*, p. 191–201.
- Conover, C.S., 1954, Ground-water conditions in the Rincon and Mesilla Valleys and adjacent areas in New Mexico: U.S. Geological Survey Water-Supply Paper 1230, 200 p.
- Environmental Systems Research Institute, 2010, ArcGIS—A complete integrated system: accessed January 2011 at <http://www.esri.com/software/arcgis/index.html>.
- Frenzel, P.F., Kaehler, C.A., and Anderholm, S.K., 1992, Geohydrology and simulation of ground-water flow in the Mesilla Basin, Doña Ana County, New Mexico, and El Paso County, Texas: U.S. Geological Survey Professional Paper 1407–C, 105 p.
- Hawley, J.W., 1978, *Guidebook to the Rio Grande Rift in New Mexico and Colorado: New Mexico Bureau of Mines and Mineral Resources Circular 163*, 241 p.
- Hawley, J.W., and Kennedy, J.F., 2004, Creation of a digital hydrogeologic framework model of the Mesilla Basin and southern Jornada del Muerto Basin: Las Cruces, N. Mex., New Mexico Water Resources Research Institute of New Mexico State University, Technical Completion Report 332, 105 p.
- Hawley, J.W., Kennedy, J.F., Ortiz, Marquita, and Carrasco, Sean, 2005, Digital hydrogeologic framework model of the Rincon Valley and adjacent areas of Doña Ana, Sierra and Luna Counties, New Mexico: Las Cruces, N. Mex., New Mexico Water Resources Research Institute of New Mexico State University, addendum to Technical Completion Report 332, accessed October 22, 2009, at <http://wrrri.nmsu.edu/publish/techrpt/tr332/cdrom/addendum.pdf>.
- Hibbs, Barry, Phillips, Fred, Hogan, James, Eastoe, Chris, Hawley, John, Granados, Alfred, and Hutchison, Bill, 2003, Hydrogeologic and isotopic study of the groundwater resources of the Hueco Bolson aquifer, El Paso, Texas/Juarez, Mexico area: *Hydrological Science and Technology*, v. 19, no. 1–4, p. 109–119.
- Hutchison, W.R., 2006, Groundwater management in El Paso, Texas: University of Texas at El Paso, Ph.D. dissertation, 329 p.
- Langman, J.B., 2009, Traveltime of the Rio Grande in the Middle Rio Grande Basin, water years 2003–05: U.S. Geological Survey Scientific Investigations Report 2007–5293, 34 p.
- Lorents, A.C., and Morgan, J.N., 1998, Database systems—Concepts, management, and applications: Fort Worth, Tex., Dryden Press, 650 p.
- Mills, S.K., 2001, Quantifying salinization of the Rio Grande using environmental tracers: Socorro, N. Mex., New Mexico Institute of Mining and Technology, Master's thesis, 397 p.
- Moore, S.J., and Anderholm, S.K., 2002, Spatial and temporal variations in streamflow, dissolved solids, nutrients, and suspended sediment in the Rio Grande Valley study unit, Colorado, New Mexico, and Texas, 1993–95: U.S. Geological Survey Water-Resources Investigations Report 02–4224, 52 p.
- Morris, P.J., 2005, Relational database design and implementation for biodiversity informatics: Phyloinformatics, *Journal for Taxonomists*, v. 7, p. 1–63, accessed August 2010 at <http://systbio.org/files/phyloinformatics/7.pdf>.
- Nickerson, E.L., 1995, Selected hydrologic data for the Mesilla ground-water basin, 1987 through 1992 water years, Doña Ana County, New Mexico, and El Paso County, Texas: U.S. Geological Survey Open-File Report 95–111, 123 p.
- Phillips, F.M., Hogan, J.F., Mills, S.K., and Hendrickx, J.M.H., 2003, Environmental tracers applied to quantifying causes of salinity in arid-region rivers—Preliminary results from the Rio Grande, southwestern United States of America, *in* Alsharhan, A.S., and Wood, W.W., eds., *Water resources perspectives—Evaluation, management, and policy: Developments in Water Science*, v. 50, p. 327–334.

- Pillsbury, A.F., 1981, The salinity of rivers: *Scientific American*, v. 245, no. 1, p. 54–65.
- Riecker, R.R., ed., 1979, *Rio Grande Rift—Tectonics and magnetism*: American Geophysical Union, 438 p.
- Shah, S.D., and Maltby, D.R., II, 2010, Design and compilation of a geodatabase of existing salinity information for the Rio Grande Basin, from the Rio Arriba-Sandoval County line, New Mexico, to Presidio, Texas, 2010: U.S. Geological Survey Data Series 499, 24 p.
- Stabler, Herman, 1911, Some stream waters of the western United States, *with chapters on Sediment carried by the Rio Grande and the industrial application of water analyses*: U.S. Geological Survey Water-Supply Paper 274, 188 p.
- U.S. Environmental Protection Agency (USEPA), 2009, STORET database access: accessed August 2010 at <http://www.epa.gov/storet/dbtop.html>.
- U.S. Environmental Protection Agency (USEPA), 2010, Storage and retrieval data warehouse, accessed August 2010 at <http://www.epa.gov/storet/>.
- U.S. Geological Survey (USGS), 2010, National Water Information System (NWIS)—USGS water data for the Nation: accessed August 2010 at <http://waterdata.usgs.gov/nwis>.
- Wilcox, L.V., 1968, Discharge and salt burden of the Rio Grande above Fort Quitman, Texas, and salt-balance conditions on the Rio Grande project: U.S. Salinity Laboratory Research Report 113, 263 p.
- Wilkins, D.W., 1998, Summary of the southwest alluvial basins regional aquifer-system analysis in parts of Colorado, New Mexico, and Texas: U.S. Geological Survey Professional Paper 1407–A, 49 p.
- Williams, J.H., 2001, Salt balance in the Rio Grande project from San Marcial, New Mexico to Fort Quitman, Texas: Las Cruces, New Mexico State University, Master's thesis, 80 p.
- Wilson, C.A., White, R.R., Orr, B.R., and Roybal, G.R., 1981, Water resources of the Rincon and Mesilla Valleys and adjacent areas, New Mexico: New Mexico State Engineer Technical Report 43, 514 p.
- Zeiler, Michael, 1999, *Modeling our world—The ESRI guide to geodatabase design*: Redlands, Calif., ESRI Press, 200 p.

Glossary

The report glossary is intended to clarify acronyms and terms used throughout the report and particularly in figure 2.

COLC	City of Las Cruces
CPU	central processing unit
DBSA	Daniel B. Stephens and Associates, Inc.
EBID	Elephant Butte Irrigation District
EPA	Environmental Protection Agency
EPCWID	El Paso County Water Improvement District
ESRI	Environmental Systems Research Institute
GIS	geographic information system
GW	groundwater
IBWC	International Boundary and Water Commission
ID	identifier
LRG	Lower Rio Grande
MS	Microsoft
NMED	New Mexico Environment Department
NMISC	New Mexico Interstate Stream Commission
NMSU	New Mexico State University
NM OSE	New Mexico Office of the State Engineer
NM WRI	New Mexico Water Resources Research Institute
NPS	National Park Service
NWIS	National Water Information System
Primary Key	A unique identifier meaning that no two rows in the table can have the same value in the column(s) identified as the primary key
QC	quality control
Query	A question posed to a database by using Structured Query Language
QW	water quality
SAHRA	Sustainability of semi-Arid Hydrology and Riparian Areas
SQL	Structured Query Language
SSPA	Stavros S. Papadopoulos and Associates, Inc.
STORET	Storage and Retrieval Repository
SW	surface water
SWQMIS	Surface Water Quality Monitoring Information System
TCEQ	Texas Commission on Environmental Quality
TWDB	Texas Water Development Board
TXWSC	Texas Water Science Center
BR	Bureau of Reclamation
USDA	United States Department of Agriculture
USEPA	United States Environmental Protection Agency
USGS	United States Geological Survey
VBA	Visual Basic for Applications
VBScript	Visual Basic Script

Appendixes 1–8

Blank Page

Appendix 1. Updatable Data Source Profiles

Appendix 1.1. U.S. Geological Survey (USGS) National Water Information System (NWIS) Data—Groundwater (GW) Field Water-Level Measurements

- **Source Description:** USGS instantaneous groundwater-level data
- **Data Range:** July 7, 1900 through December 17, 2009
- **Raw Data Source:** NWIS Web: <http://nwis.waterdata.usgs.gov/usa/nwis/gwlevels>
- **Associated Data Preprocessing Script:** located in text file titled NWIS_GW_FieldData_ExtractParse_v2.txt
File location: \NMISC\Scripts\VB\
- **MS Access Staging Database Files:** NMISC_NWIS_GW_Field_Data.mdb
File location: \NMISC\USGS\NWIS_GWField\
Also, USGS_MASTER_SITE.mdb
File location: \NMISC\USGS\
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - The USGS NWIS master parameter list table available from <http://nwis.waterdata.usgs.gov/nwis/pmcodes/> needs to be downloaded and reloaded into the staging database file and renamed “NWIS_MASTER_PARAM_CODES” for each update. Downloading the parameter list each time a refresh is done ensures the parameter list is up to date.
 - The NWIS GW field sites table, which contains the site IDs and from which the site information is loaded through the SQL loader queries, needs to be a copy of the downloadable file (subset to the area of interest—see section 2.3) from the NWIS Web site (<http://waterdata.usgs.gov/nwis/si>) with all site-description information fields selected.
- **Additional Preparation Notes:**
 - It is important that the user read the VBA script usage comments to understand what the script is doing, in addition to reading through the script code comments to understand the script components. Required preparation steps are described in the script.
 - USGS NWIS publicly available sites were obtained from the NWIS Web site download interface for all of Texas and all of New Mexico. All of Texas then all of New Mexico were downloaded because of the restrictions of the NWIS Web download interface. The study area was used to select relevant sites by using the methods described in section 2.3.
 - Section 3.4 has additional notes regarding USGS NWIS Web data type sources.
 - The SQL loader queries document the specific tables used from this database.

Appendix 1.2. U.S. Geological Survey (USGS) National Water Information System (NWIS) Data—Groundwater (GW) Daily Water-Level Measurements

- **Source Description:** USGS daily-value groundwater-level data
- **Data Range:** October 1, 1984 through September 30, 2008
- **Raw Data Source:** NWIS Web: http://waterdata.usgs.gov/nwis/dv/?referred_module=gw
- **Associated Data Preprocessing Script:** located in text file titled NWIS_GW_DailyData_ExtractAndParse.txt
File location: \NMISC\Scripts\VB\
- **MS Access Staging Database Files:** NMISC_NWIS_GW_Daily_Data.mdb
File location: \NMISC\USGS\NWIS_GWDaily\
Also, USGS_MASTER_SITE.mdb
File location: \NMISC\USGS\
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - The USGS NWIS master parameter list table available from <http://nwis.waterdata.usgs.gov/nwis/pmcodes/> needs to be downloaded and reloaded into the staging database as a table and renamed “NWIS_MASTER_PARAM_CODES” for each update. Downloading the parameter list each time a refresh is done ensures the parameter list is up to date.
 - The NWIS GW daily sites table, which contains the site IDs and from which the site information is loaded by the SQL loader queries, needs to be a copy of the downloadable file (subset to the area of interest – see section 2.3) from the NWIS Web site <http://waterdata.usgs.gov/nwis/si> with all site-description information fields selected.
- **Additional Preparation Notes:**
 - It is important that the user read the VBA script usage header comments to understand what the script is doing, in addition to reading through the script code comments to understand the script components. Included in the script header are notes necessary for preparation.
 - USGS NWIS publicly available sites were obtained from the NWIS Web site download interface for all of Texas and all of New Mexico. The approved NMISC study area was used to subset out sites of relevance by using the methods described in section 2.3.
 - Section 3.4 covers additional notes regarding USGS NWIS Web data type sources.
 - The SQL loader queries document the specific tables used from this database.

Appendix 1.3. U.S. Geological Survey (USGS) National Water Information System (NWIS) Data—Water-Quality (QW) Instantaneous Field Measurements

- **Source Description:** USGS instantaneous water-quality data
- **Data Range:** March 15, 1905 through December 4, 2009
- **Raw Data Source:** NWIS Web: <http://nwis.waterdata.usgs.gov/usa/nwis/qwdata>
- **Associated Data Preprocessing Script:** located in text file titled NWIS_QW_FieldData_ExtractAndParse.txt
File location: \NMISC\Scripts\VB\
- **MS Access Staging Database Files:** NWIS_QW_Field_Data.mdb
File location: \NMISC\USGS\NWIS_QWField\
Also, USGS_MASTER_SITE.mdb
File location: \NMISC\USGS\
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - The USGS NWIS master parameter list table available from <http://nwis.waterdata.usgs.gov/nwis/pmcodes/> needs to be downloaded and reloaded into the staging database as a table and renamed “NWIS_MASTER_PARAM_CODES” for each update. Downloading the parameter list each time a refresh is done ensures the parameter list is up to date.
 - The NWIS QW field sites table, which contains the site IDs and from which the site information is loaded by the SQL loader queries, needs to be a copy of the downloadable file (subset to the area of interest—see section 2.3) from the NWIS Web site <http://waterdata.usgs.gov/nwis/si> with all site-description information fields selected.
- **Additional Preparation Notes:**
 - It is important that the user read the VBA script usage header comments to understand what the script is doing, in addition to reading through the script code comments to understand the script components. Included in the script header are notes necessary for preparation.
 - USGS NWIS publicly available sites were obtained from the NWIS Web site download interface for all of Texas and all of New Mexico. The approved NMISC study area was used to subset out sites of relevance by using the methods described in section 2.3.
 - Section 3.4 covers additional notes regarding USGS NWIS Web data type sources.
 - The SQL loader queries document the specific tables used from this database.

Appendix 1.4. U.S. Geological Survey (USGS) National Water Information System (NWIS) Data—Daily Surface-Water (SW) Discharge Values

- **Source Description:** USGS daily surface-water values
- **Data Range:** January 1, 1899 through October 12, 2009
- **Raw Data Source:** NWIS Web: http://waterdata.usgs.gov/nwis/dvstat/?referred_module=sw
- **Associated Data Preprocessing Script:** located in text file titled NWIS_SW_Discharge_DailyData_ExtractAndParse.txt
File location: \NMISC\Scripts\VB\
- **MS Access Staging Database Files:** NWIS_SW_SITES_DATA.mdb
File location: \NMISC\USGS\NWIS_SWDaily\
Also, USGS_MASTER_SITE.mdb
File location: \NMISC\USGS\
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - The USGS NWIS master parameter list table available from <http://nwis.waterdata.usgs.gov/nwis/pmcodes/> needs to be downloaded and reloaded into the staging database file and renamed “NWIS_MASTER_PARAM_CODES” for each update. Downloading the parameter list each time a refresh is done ensures the parameter list is up to date.
 - The NWIS SW sites table, which contains the site IDs and from which the site information is loaded by the SQL loader queries, needs to be a copy of the downloadable file (subset to the area of interest—see section 2.3) from the NWIS Web Site Information site <http://waterdata.usgs.gov/nwis/si> with all site-description information fields selected.
- **Additional Preparation Notes:**
 - It is important that the user read the VBA script usage header comments to understand what the script is doing, in addition to reading through the script code comments to understand the script components. Included in the script header are notes necessary for preparation.
 - USGS NWIS publicly available sites were obtained from the NWIS Web site download interface for all of Texas and all of New Mexico. The approved NMISC study area was used to subset out sites of relevance by using the methods described in section 2.3.
 - Section 3.4 covers additional notes regarding USGS NWIS Web data type sources that pertain to NWIS GW Field Water-Level Data.
 - The SQL loader queries document the specific tables used from this database.

Appendix 1.5. U.S. Environmental Protection Agency (USEPA) Modern Storage and Retrieval Repository (STORET) Water-Quality Data

- **Source Description:** Water-quality data downloaded from USEPA Modern STORET
- **Data Range:** October 11, 1937 through November 9, 2007
- **Raw Data Source:** USEPA STORET Web site: http://www.epa.gov/storet/dw_home.html
- **Associated Data Preprocessing Script:** located in text file named Load_EPA_Modern_Data_v2.txt
File location: \NMISC\Scripts\VB\
- **MS Access Staging Database File:** NMISC_EPA_Modern_Data.mdb
File location: \NMISC\EPA
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - MASTER_STORET_PARAMS table is a master parameter list of USEPA Modern STORET parameters used to reference parameter descriptions.
 - AGENCIES_NOT_TO_LOAD table contains USEPA Modern STORET agency codes. It is designed to resolve duplicate data occurring in USEPA Legacy STORET and also to filter out data that are obtained directly from some of the agencies represented by the codes in the table.
- **Additional Preparation Notes:**
 - It is important that the user read the VBA script usage header comments to understand what the script is doing, in addition to reading through the script code comments to understand the script components. Included in the script header are notes necessary for preparation.
 - Sample sites were obtained from the USEPA Modern STORET database for all of Texas and New Mexico. The approved NMISC study area was used to make a subset of sites of relevance by using the methods described in section 2.3.
 - The SQL loader queries document the specific tables used from this database.

Appendix 1.6. International Boundary and Water Commission (IBWC) Daily Surface-Water Discharge Data

- **Source Description:** Daily mean-discharge data from IBWC
- **Data Range:** May 10, 1889 through November 30, 2009
- **Raw Data Source:** IBWC Web site: http://www.ibwc.gov/Water_Data/histflo1.htm
- **Associated Data Preprocessing Script:** located in text file named Clean_And_Load_IBWC_SW_Data.txt
File location: \NMISC\Scripts\VB\
- **MS Access Staging Database File:** IBWC_SW.mdb
File location: \NMISC\IBWC\IBWC_SW
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - The IBWC names the discharge site data files on the basis of an internal “binary ID” that was inserted into this table for reference.
 - The sites loaded into the NMISC geodatabase are documented in the code of the above preprocessing script. The text file “binary names” are used to determine which site the downloaded file represents.
- **Additional Preparation Notes:**
 - It is important that the user read the VBA script usage header comments to understand what the script is doing, in addition to reading through the script code comments to understand the script components. Included in the script header are notes necessary for preparation.
 - Discharge values were converted from cubic meters per second to cubic feet per second by using the preprocessing script.
 - The SQL loader queries document the specific tables used from this database.

Appendix 1.7. International Boundary and Water Commission (IBWC) Water-Quality Data

- **Source Description:** Water-quality data from IBWC
- **Data Range:** January 11, 1996 through April 15, 2009
- **Raw Data Source:** IBWC Web site: <http://www.ibwc.gov/CRP/monstats.htm>
- **Associated Data Preprocessing Script:** located in text file named Clean_And_Load_IBWC_QW_Data.txt
File location: \NMISC\Scripts\VB\
- **MS Access Staging Database File:** IBWC_QW.mdb
File location: \NMISC\IBWC\IBWC_QW
- **Script Runtime Database:** the database used to run the script from: IBWC_QW Loader.mdb
File location: \NMISC\IBWC\IBWC_QW
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - The Sites table in the staging database file contains site IDs and geographic coordinates for the sites represented in the data.
 - It is important that the user read the VBA script usage header comments to understand what the script is doing, in addition to reading through the script code comments to understand the script components. Included in the script header are notes necessary for preparation.
 - The SQL loader queries document the specific tables used from this database.

Appendix 1.8. Texas Water Development Board (TWDB) Groundwater Data

- **Source Description:** Groundwater elevations and groundwater-quality data from TWDB
- **Data Range:** January 1, 1900 through September 15, 2007
- **Raw Data Source:** TWDB Web site: *http://www.twdb.state.tx.us/publications/reports/GroundWaterReports/GWDatabaseReports/GWdatabaserpt.asp*
- **Associated Data Preprocessing Scripts:** located in text file named *Format_TWDB_Data.txt* and *Update_TWDB_Sample_Result_IDs.txt*
Location: \NMISC\Scripts\VB\
- **TWDB-specific SQL queries for preprocessing:** located in text file named *TWDB_SourceDatabase_SubsetQueries.txt* and *TWDB_SourceDatabase_UPDATEqueries.txt*
Location: \NMISC\Scripts\VB\
- **MS Access Staging Database File:** *TWDB_NMISC_Data.mdb*
File location: \NMISC\TWDB
- **Script Runtime Database:** the database used to run the script from: *TWDB_Loader.mdb*
File location: \NMISC\TWDB
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - The *NMISC_TWDB_Sites* table contains the TWDB sites of interest that were filtered by using the methods described in section 2.3 (see note below under Additional Preparation Notes).
 - The well-type table was exported from the original TWDB *GWDB.mdb* database and updated with an “UNKNOWN” well type to ensure data integrity between the well type lookup table and the *well_type* field in the *NMISC_TWDB_Sites* table. An update of the *NMISC_TWDB_Sites* table in the *well_type* field for records that had null values was required.
 - The *entity_codes* table was exported from the original TWDB *GWDB.mdb* database and updated the table with an “UNKNOWN” entity type to ensure data integrity between the *entity_codes* lookup table and the *collecting_agency* field in the *NMISC_TWDB_Sites* table. This step required an update of the *NMISC_TWDB_Sites* table in the *collecting_agency* field for records that had null values. This table is also used in the post-NMISC data refresh script.
 - The *stret_code* table was exported from the original *GWDB.mdb* database.
 - The *reporting_agency* table was exported from the original TWDB *GWDB.mdb* database and updated the table with an “UNKNOWN” agency type to ensure data integrity between the *reporting_agency* table and the *NMISC_TWDB_Sites* table. The original name of the *reporting_agency* table was *rept_agency* in the *GWDB.mdb* file. The *reporting_agency* table is not used in the SQL loader queries but was manually referenced to determine the name of the agency associated with the code in the NMISC geodatabase *TBL_AGENCY* table. The *reporting_agency* table is also used in the post-NMISC data refresh script.
- **TWDB Preparation Process Steps:**
 - Delete the preexisting tables called *NMISC_TWDB_Samples* and *NMISC_TWDB_Results*.
 - Run the script contained in *Format_TWDB_Data.txt* in the loader Access database.
 - Run the script contained in *Update_TWDB_Sample_Result_IDs.txt*
 - Manually run the two update queries contained in *TWDB_SourceDatabase_UPDATEqueries.txt* in the TWDB Access staging database.
 - OPTIONAL: delete the *NMISC_TWDB_Sites* table, export the *master_site* table, and enable as a feature class in ArcGIS to subset the TWDB sites to the NMISC study area as described in section 2.3.

- Manually run the subset samples query that is located in TWDB_SourceDatabase_SubsetQueries.txt in the TWDB Access staging database.
- Manually run the subset results query that is located in TWDB_SourceDatabase_SubsetQueries.txt in the TWDB Access staging database.
- **Additional Preparation Notes:**
 - It is important that the user read the VBA script usage header comments to understand what the script is doing in addition to reading through the script code comments to understand the script components. Included in the script header are notes necessary for preparation.
 - The TWDB sites were filtered by using the methods described in section 2.3 with the area of interest. The master_site table that is created from the preprocessing script needs to be used if you want to refresh the NMISC-specific TWDB sites. Load these sites back in as a table named NMISC_TWDB_Sites.
 - The SQL loader queries document the specific tables used from this database.

Appendix 1.9. Texas Commission on Environmental Quality (TCEQ) Water-Quality Data

- **Source Description:** Water-quality data from TCEQ
- **Data Range:** September 18, 1968 through August 24, 2009
- **Raw Data Source:** Data obtained through direct access to TCEQ SWQMIS system. The TCEQ SWQMIS system requires username and password authentication. TCEQ staff should be contacted if access is needed: <https://www8tst.tceq.state.tx.us/SwqmisWeb/>
- **Associated Data Preprocessing Script:** located in text file named
File location: \NMISC\Scripts\VB\
- **MS Access Staging Database File:** NMISC_TCEQ_Data.mdb
File location: \NMISC\TCEQ
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - The SubmittingEntityCodes table is used to reference the agency descriptions in order to populate the NMISC geodatabase TBL_AGENCY table; population of the TBL_AGENCY table occurs by using the post-NMISC data refresh script. This file was obtained from the above TCEQ direct access site.
 - The TCEQ_Parameter_Codes table is the parameter table for TCEQ data. It is available at this Web location: ftp://ftp.tceq.state.tx.us/pub/WaterResourceManagement/WaterQuality/DataCollection/CleanRivers/public/sw_parm_format.txt
 - Downloading the parameter list each time a refresh is done ensures the parameter list is up to date.
- **Additional Preparation Notes:**
 - It is important that the user read the VBA script usage header comments to understand what the script is doing in addition to reading through the script code comments to understand the script components. Included in the script header are notes necessary for preparation.
 - The SQL loader queries document the specific tables used from this database.
 - TCEQ sample sites were filtered by using the methods described in section 2.3 with the area of interest. Data for only the counties in the NMISC study area were downloaded from the above TCEQ direct access site.

Appendix 2. Static Data Source Profiles

Appendix 2.1. New Mexico Environment Department (NMED) Water-Quality Data

- **Source Description:** Water-quality data provided by the NMISC
- **Data Range:** October 15, 1982 through December 8, 2008
- **Raw Data Source:** Data files obtained from NMISC during project
- **MS Access Staging Database File:** NMED_LRG_Monitoring_1208.mdb and DP_DATA.mdb
File locations: \\NMISC\NMISC_Compedium_20090624\Compedium_1208\SurfaceWater\Sources\NMED\
\\NMISC\NMISC_Compedium_20090624\Compedium_1208\Groundwater\Sources\NMED\Discharge Permit Data\
- **Data Tables Manually Created From Original Source Tables:**
These tables were derived from the original data tables by using SQL in NMED_LRG_Monitoring_1208.mdb:
 - SW_GW_Sites: compiled to contain all sites
 - SW_GW_Sample: derived sample table for all data
 - SW_GW_Results: derived result table for all data
 - SW_GW_Parameters: derived parameter table for all data
 These tables were derived from the original data tables by using SQL in DP_DATA.mdb:
 - sample: derived sample table for all data
 These tables were modified in DP_DATA.mdb
 - DP_Data_Final: sample ID field added to table.
- **Additional Preparation Notes:**
 - The SQL loader queries document the specific tables used from these databases.

Appendix 2.2. Boyle-Parsons Water-Quality Data

- **Source Description:** Water-quality data provided by the NMISC contained in the version of the NMISC compendium prior to the geodatabase described in this report.
- **Data Range:** January 15, 1971 through September 15, 1984
- **Raw Data Source:** Data files obtained from NMISC during project
- **MS Access Staging Database File:** lrg_surfacewatermonitoring.mdb
 File location: \NMISC\NMISC_Compndium_20090624\Compendium_1208\SurfaceWater\
 Data Tables Manually Created From Original Source Tables:
 These tables were derived from the original data tables by using SQL in lrg_surfacewatermonitoring.mdb:
 - Load_These_BoyleParsons_Sites: table derived for Boyle-Parsons sites of interest.
 - BoyleParsons_Samples: table derived for sample records of Boyle-Parsons data.
 - BoyleParsons_Results_Data: table derived for results data associated with Boyle-Parsons.
 - BoyleParsons_Parameters: table derived for the parameters associated with the Boyle-Parsons data.
- **Additional Preparation Notes:**
 - NMISC staff provided criteria for identifying data of interest that Boyle-Parsons data not occurring between January 1, 1934 and December 31, 1963 should be used.
 - The SQL loader queries document the specific tables used from this database.
- **Possible Data Issues for Awareness of Users:**
 - It is known that the previous NMISC compendium databases (file names lrg_surfacewatermonitoring.mdb and lrg_groundwatermonitoring.mdb) have issues with any time data as communicated by NMISC (Dale Doremus, New Mexico Interstate Stream Commission, written commun., 2009). It is suspected that some records were set to a time of 12:00 p.m. to indicate some sort of daily value or average, or 12:00 p.m. was used as a placeholder for data that did not have time data. The data records were not flagged if such an edit had occurred; consequently the time component of a date/time field is not usable.

Appendix 2.3. El Paso County Water Improvement District (EPCWID) Water-Quality Data

- **Source Description:** Water-quality data provided by the NMISC
- **Data Range:** March 25, 2003 through August 31, 2005
- **Raw Data Source:** Data files obtained from NMISC during project
- **MS Access Staging Database File:** EPCWID_data.mdb
File location: \NMISC\NMISC_Compndium_20090624\Compndium_1208\SurfaceWater\Sources\EPCWID\
- **Original Data File:** epcwid_sw_2003_2005.xls
File location: \NMISC\NMISC_Compndium_20090624\Compndium_1208\SurfaceWater\Sources\EPCWID\
- **Data Tables Manually Created From Original Source Tables:**

The following process steps were used to create the Access database format:

- Imported the Excel worksheet called DailyFlow into an MS Access database
- Cleaned data by using SQL and created a new table called Cleaned_EPCWID_Flow
- **Additional Preparation Notes:**
 - The sites associated with these data are drawn from NMISC geodatabase source files other than those associated with the EPCWID data.
 - The SQL loader queries document the specific tables used from this database.

Appendix 2.4. S.S. Papadopulos (SSPA) Surface-Water Data Compilation

- **Source Description:** Surface-water discharge data provided by the NMISC.
- **Data Range:** January 1, 1889 through December 31, 2002
- **Raw Data Source:** Data files obtained from NMISC during project.
- **MS Access Staging Database File:** v2SurfaceWater.mdb
File location: \NMISC\SSPA_V2_2004_Data_compendium\SurfaceWater\
- **Data Tables Manually Created From Original Source Tables by using SQL:**
 - The NMISC_Sites table was created to serve as a limiting list of site IDs to query from the SSPA SW database by using the relationships in the associated SQL loader query.
 - The DONOTLOADFROMSSPA_SW table resolves duplicates for sites and associated data loaded from the SSPA SW database. This table contains sites that occur in both the NMSU and SSPA SW database that are instead being loaded from the NMSU database. The sites in the two databases were compared by using approximately 25 percent of the sites occurring in both databases; the NMSU data had longer periods of record for all.
- **Possible Data Issues for Awareness of Users:**
 - A possibility of rounded discharge values was discovered in comparing the SSPA SW database and what may be original source data files. One example that may indicate data rounding is for site ID 1C27.5 (also known as Bonita Lateral). This Excel file located in the data provided by the NMISC was checked:

Compendium_1208\SurfaceWater\Sources\BOYLE_PARSONS\Boyle cd\Boyle\FLOW\RESERVS\CANAL\BONITAF.xls per the NMSU FinalReport.pdf (page 18 of the actual PDF file). For the date of 2/22/1938 at the aforementioned site, the Excel file shows a discharge result value of 13.4. The value for the same site on the same date in the SSPA SW database shows 13. The next day 2/23/1938 has a value in the Excel file of 10.6, but the SSPA SW database shows a value of 11 (fig. 2–1). It was beyond the extent practicable to try and reconcile any similar Excel data files given the format of the data in addition to the complexity of determining which data in the SSPA SW database were possibly rounded and which data were correct. Judgment or interpretation of which value is correct was avoided.
- **Additional Preparation Notes:**
 - The SQL loader queries document the specific tables used from this database.

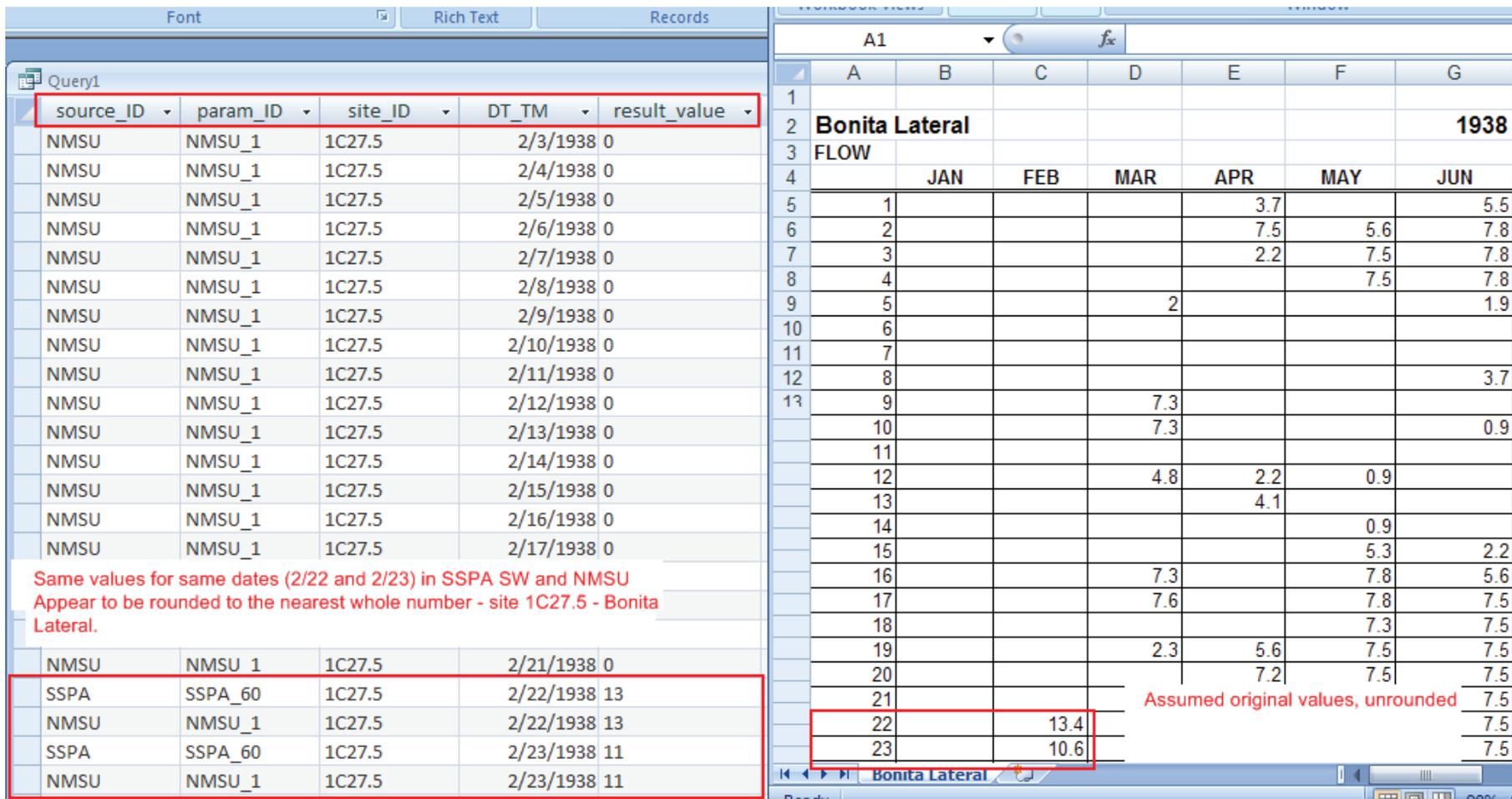


Figure 2-1. Illustration of potential data issues for awareness of users from the Stavros S. Papadopoulos surface-water data source.

Appendix 2.5. S.S. Papadopulos (SSPA) Groundwater Data Compilation

- **Source Description:** Groundwater data provided by the NMISC.
- **Data Range:** August 9, 1931 through June 15, 2003
- **Raw Data Source:** Data files obtained from NMISC during project.
- **MS Access Staging Database File:** v2GroundwaterElevation.mdb
File location: \NMISC\SSPA_V2_2004_Data_compendium\WaterElevation\
- **Data Tables Manually Created From Original Source Tables:**

The following tables were derived from the original data tables by using SQL:

- The DONOTLOADFROMSSPA_GW table was created to resolve site record duplicates between the SSPA GW database and the TWDB data, as well as between the SSPA GW database and the SSPA QW database. Creation of the DONOTLOADFROMSSPA_GW table was accomplished by the inner join on this table in the associated sites SQL loader query.
- The MES_XX table was created to serve as documentation of SSPA database sites of interest that had to be loaded in a different manner because of their data source codes. The primary SSPA GW database SQL loader query uses the table data source codes, and the sites in this table have a data source code attributed to USGS NWIS. The primary loader does not load sites that are attributed to NWIS, but the NMISC staff specified that these sites in particular largely have their water-level data from EBID and the Bureau of Reclamation.
- The MES_XX_SITES_LOCATION table was created to serve as the site records to be loaded from SSPA GW that could not be loaded from any other source. This is a subset of the sites contained in the MES_XX table.
- The MES_XX_SITES_LOCATION_NOLOAD table was created to allow for the MES_XX sites data to be properly loaded from the SSPA GW database. The sites in this table are being loaded from other sources, so the point was to not load the sites but just the data.
- **Additional Preparation Notes:**
 - The SQL loader queries document the specific tables used from this database.

Appendix 2.6. S.S. Papadopoulos (SSPA) Water-Quality Data Compilation

- **Source Description:** Water-quality data provided by the NMISC
- **Data Range:** February 11, 1931 through August 4, 1987
- **Raw Data Source:** Data files obtained from NMISC during project.
- **MS Access Staging Database File:** v3WaterQuality.mdb
File location: \NMISC\SSPA_V2_2004_Data_compendium\WaterQuality\
- **Data Tables Manually Created From Original Source Tables:**
The following table was derived from the original data tables by using SQL:
TBL_Sample table created to serve as a sample table for the data set.
- **Additional Preparation Notes:**
 - A sample_ID field was inserted into the TBL WQ Parameter results table for association with the TBL_Sample table.
 - The NMISC specified that only the data associated with these source codes should be loaded from the SSPA QW database: 4, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21.
 - The SQL loader queries document the specific tables used from this database.

Appendix 2.7. New Mexico State University (NMSU) Surface-Water Discharge Data Compilation

- **Source Description:** Surface-water data provided by the NMISC
- **Data Range:** January 1, 1908 through September 30, 2003
- **Raw Data Source:** Data files obtained from NMISC during project.
- **MS Access Staging Database File:** nmsuqupdate.mdb
File location: \NMISC\NMISC_Compndium_20090624\Compendium_1208\SurfaceWater\Sources\NMSU\
- **Data Tables Manually Created From Original Source Tables:**
The following tables were derived from the original data tables by using SQL:
 - NMSU_DailySites_Compiled_Uncleaned table was created by appending the data from every table that had “Daily” as the last part of the table name in the above database.
 - NMSU_Sites_CLEANED table was created to serve as the site table for this data source.
 - NMSU_DailySites_Cleaned_LoadALL table is the final cleaned data file from which the daily discharge values are queried.
- **Additional Preparation Notes:**
 - The nmsuqupdate.mdb database contains no coordinate data for sites. The latitude/longitude coordinates for 17 sites in the NMSU_Sites_CLEANED table were obtained from the lrg_surfacewatermonitoring.mdb file located at: \NMISC\NMISC_Compndium_20090624\Compendium_1208\SurfaceWater\
 - The remaining sites have null latitude and longitude values in the NMISC geodatabase because the coordinate data were not available.
 - The SQL loader queries document the specific tables used from this database.
 - The NMSU Final Report PDF located at \NMISC\NMISC_Compndium_20090624\Compendium_1208\SurfaceWater\Sources\NMSU was used to assess sites as being potentially duplicative of sites from other sources (for example, IBWC, USGS). Sites identified as duplicates for the latter sources were removed after data were inspected.
 - The NMSU_Result field in the NMSU_DailyData_Cleaned_LoadALL table is the primary data field.
- **NMSU Preparation Process Steps:**
 - The script contained in the file called NMSU_Append_Query.txt located at \NMISC\Scripts\VB\ was used to query all tables that had “Daily” as the last part of the file name.
 - The agency and source fields were manually populated in the compiled table NMSU_DailySites_Cleaned_LoadALL
- **Possible Data Issues for Awareness of Users:**

A possibility of erroneous, zero-result values as placeholders for null daily values where data were not actually collected may be an issue with the NMSU database. A null value (no data) and a value of zero are two completely different things and are not interchangeable. One example of a possible null value is for site ID 1C27.5, Bonita Lateral. This Excel file located in the data provided by the NMISC was checked: Compendium_1208\SurfaceWater\Sources\BOYLE_PARSONS\Boyle cd\Boyle\FLOW\RESERVS\CANAL\BONITAF.xls per the NMSU FinalReport.pdf (page 18 of the actual PDF file). For the date of 2/26/1938 at the aforementioned site, the Excel file shows a null value with empty cells for 2/26/1938 and 2/27/1938, but the Access database shows zeros for both those dates (fig. 2–2). It was beyond the extent practicable to reconcile any similar Excel data files given the format of the data, in addition to the complexity of determining which data in the SSPA SW database were possibly rounded and which data were correct. Judgment or interpretation of which value is the correct value was avoided.

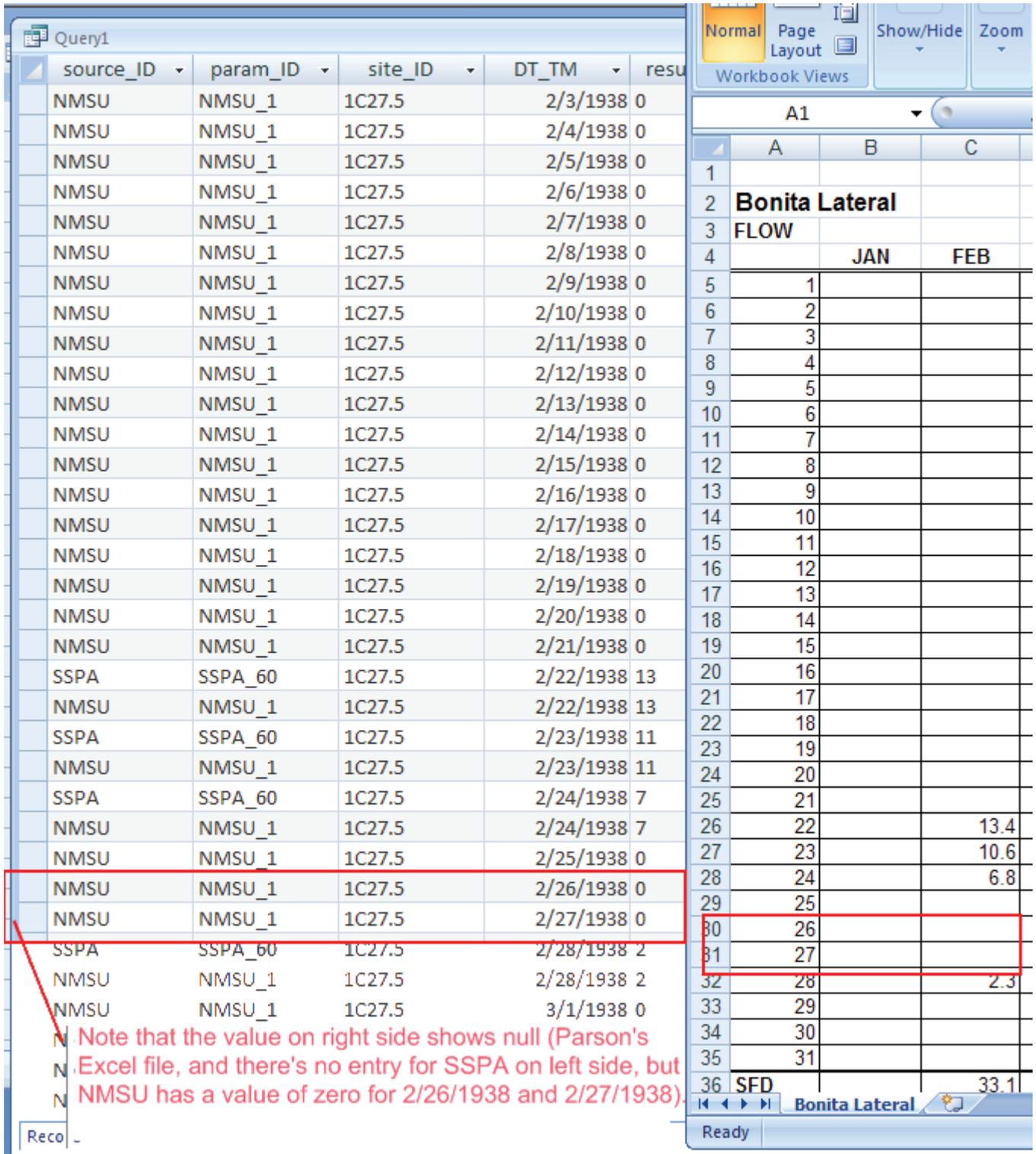


Figure 2-2. Illustration of potential data issues for the awareness of users from the New Mexico State University surface-water data source.

Appendix 2.8. U.S. Environmental Protection Agency (USEPA) Legacy Storage and Retrieval Repository (STORET) Water-Quality Data

- **Source Description:** Water-quality data obtained from the Legacy USEPA STORET system
- **Data Range:** October 11, 1937 through May 9, 1996
- **Raw Data Source:** Downloaded from Web site: *http://www.epa.gov/storet/dbtop.html*
- **MS Access Staging Database File:** NMISC_EPA_Legacy_Data.mdb
File location: \NMISC\EPA\
- **Associated Data Preprocessing Scripts:** Load_EPA_Legacy_Data.txt and Update_EPA_Legacy_Sample_Result_IDs.txt
Location: \NMISC\Scripts\VB\
- **Additional Data Tables of Relevance Contained in the Staging Database File:**
 - The DONOTLOADFROMLEGACY_STORET table filters the site records loaded to resolve site duplication between USEPA Legacy STORET and USEPA Modern STORET.
 - The AGENCIES_NOT_TO_LOAD table filters the data on the basis of the collecting agency associated with records. Agencies contained in this table are not loaded.
 - The EPA_Legacy_Organizations table contains information on the agency codes contained in the EPA Legacy data.
- **Additional Preparation Notes:**
 - It is important that the user read the preprocessing VBA script usage header comments to better understand what the script is doing in addition to reading through the script code comments to understand the script components. Included in the script header are notes necessary for preparation.
 - The SQL loader queries document the specific tables used from this database.
 - Sample sites were obtained from the USEPA Legacy STORET downloaded data for all of Texas and all of New Mexico. The approved NMISC study area was used to subset out sites of relevance by using the methods described in section 2.3.
 - The Update_EPA_Legacy_Sample_Result_IDs.txt script should be run only if the USEPA Legacy data are being completely redone. In such a case, it would be run after first running the Load_EPA_Legacy_Data.txt script.

Appendix 2.9. Daniel B. Stephens and Associates (DBSA) Water-Quality Data Compilation

- **Source Description:** Water-quality data provided by the NMISC
- **Data Range:** August 3, 2004 through January 27, 2005
- **Raw Data Source:** Data files obtained from NMISC during this study.
- **MS Access Staging Database File:** DBSAWaterQuality.mdb
File location: \NMISC\NMISC_Compendum_20090624\Compendum_1208\SurfaceWater\Sources\DBSA\
- **Data Tables and Queries Manually Created From Original Source Tables:**
These tables were derived from the original data tables through SQL in the database:
 - TBL_Sample was created to serve as a data source sample table.
 - QRY_Sample was created to properly load results data.
- **Additional Preparation Notes:**
 - The NMISC specified that only data associated with source code 32 should be loaded from the source Access database.
 - This source contains the Suzanne Mills thesis data (Dale Doremus, New Mexico Interstate Stream Commission, written commun., 2009).
 - The SQL loader queries document the specific tables used from this database.

Appendix 2.10. City of Las Cruces (COLC) Groundwater-Level Data

- **Source Description:** Water-level data provided by the NMISC
- **Data Range:** March 10, 1972 through February 21, 1978
- **Raw Data Source:** Data files obtained from NMISC during this study.
- **MS Access Staging Database File:** COLC.mdb
File location: \NMISC\NMISC_Compndium_20090624\Compendium_1208\Groundwater\Sources\COLC\
- **Original Data File:** CityofLasCrucesWaterLevels.xls
File location: \NMISC\NMISC_Compndium_20090624\Compendium_1208\Groundwater\Sources\COLC\
- **Data Tables Manually Created From Original Source Tables:**
These tables were derived from the original data tables through SQL in the database:
 - COLC_Sites table serves as the site data table.
 - COLC_Sample table serves as the data sample table.
 - COLC_Data_Result table serves as the result data table.
- **Additional Preparation Notes:**
 - The SQL loader queries document the specific tables used from this database.

Appendix 2.11. Sustainability of semi-Arid Hydrology and Riparian Areas (SAHRA) Water-Quality Data Compilation

- **Source Description:** Water-quality data obtained from the SAHRA FTP project Web site.
- **Data Range:** January 28, 2000 through August 10, 2006
- **Raw Data Source:** SAHRA FTP site: <ftp://ftp.hwr.arizona.edu/pub/incoming/jhogan/>
- **MS Access Staging Database File:** RioGrandeWQ_ODM_working_v3_NoMisVal.mdb
File location: \NMISC\SAHRA\
- **Data Tables Manually Created From Original Source Tables:**

The following table was derived from the original data tables by using SQL:

 - Sample2 table was created as an improved version of the original sample table.
- **Additional Preparation Notes:**
 - The SQL loader queries document the specific tables used from this database.

Appendix 2.12. L.V. Wilcox and J. Williams Water-Quality Data—Recovered by New Mexico Interstate Stream Commission (NMISC)

- **Source Description:** Water-quality data provided and recovered into a usable format by the NMISC
- **Data Range:** January 15, 1934 through December 15, 1963
- **Raw Data Source:** Data files obtained from NMISC during this study.
- **MS Access Staging Database File:** Wilcox_Williams.mdb
File location: \NMISC\Wilcox_Williams\
 - **Original Data File:** WilcoxData_williams_thesis_corrected.xls
File location: \NMISC\Wilcox_Williams\
 - **Data Tables Manually Created From Original Source Tables:**
These tables were derived from the original data tables by using SQL:
 - SITES table serves as the site record table.
 - SAMPLES table serves as the samples record table.
 - RESULT table serves as the result data table.
 - PARAM table serves as the parameter table.
 - **Additional Preparation Notes:**
 - The SQL loader queries document the specific tables used from this database.

Appendix 2.13. Wilson, Orr, White, and Roybal Data

- **Source Description:** Water-quality data provided and recovered into a usable format by the NMISC
- **Data Range:** April 22, 1936 through September 18, 1977
- **Raw Data Source:** Data files obtained from NMISC during this study.
- **MS Access Staging Database File:** WilsonData.mdb
File location: \NMISC\Wilson\
 - **Original Data File:** Wilson_and_others_WQdata.xls
File location: \NMISC\Wilson\
 - **Data Tables Manually Created From Original Source Tables:**
These tables were derived from the original data tables by using SQL:
 - WilsonWells table serves as the site record table.
 - Wilson_Sample table serves as the sample record table.
 - Wilson_WQ_Results table serves as the result data table.
 - Wilson_USGS_Param table serves as the parameter table.
 - DONOTLOADFROMWILSON table serves to resolve duplicate site records between this dataset and SSPA datasets.
- **Additional Preparation Notes:**
 - The SQL loader queries document the specific tables used from this database.

Appendix 2.14. Stabler, H. USGS Data (1911)—Recovered by Daniel B. Stephens and Associates

- **Source Description:** Water-quality data provided and recovered into a usable format by the NMISC
- **Data Range:** January 8, 1905 through April 30, 1907
- **Raw Data Source:** Data files obtained from NMISC during this study.
- **MS Access Staging Database File:** Stabler_Data.mdb
File location: \NMISC\Stabler\
 - **Original Data File:** NMED DBSA Stabler Deliverable_20100423.xls
File location: \NMISC\Stabler\
 - **Data Tables Manually Created From Original Source Tables:**
These tables were derived from the original data tables by using SQL:
 - Stabler_Sites table serves as the site record table.
 - Stabler_Sample table serves as the sample record table.
 - Stabler_Data table serves as the results data table.
 - Stabler_Params table serves as the parameter table.
 - **Additional Preparation Notes:**
 - The SQL loader queries document the specific tables used from this database.

Appendix 2.15. New Mexico Interstate Stream Commission (NMISC) Data Attributed to the U.S. Geological Survey (USGS)

- **Source Description:** Data identified by the NMISC as nonpublic USGS data. The original source of data is unknown.
- **Data Range:** March 20, 1926 through December 31, 2007
- **Raw Data Source:** Data files obtained from the NMISC as in the previous version of the NMISC compendium.
- **MS Access Staging Database File:** NMISC_LRG_USGS_DATAOFINTEREST.mdb
File location: \NMISC\NMISC_Compndium_20090624\Compndium_1208\
NMISC_LRG_Data_Attributed_To_USGS\
- **Data Tables Manually Created From Original Source Tables:**
 - All data in the aforementioned Access database were extracted from the original NMISC compendium data-base files lrg_surfacewatermonitoring.mdb and lrg_groundwatermonitoring.mdb located at: \NMISC\NMISC_Compndium_20090624\Compndium_1208\SurfaceWater\ and \NMISC\NMISC_Compndium_20090624\Compndium_1208\Groundwater\
- **Additional Preparation Notes:**
 - The SQL loader queries document the specific tables used from this database.
 - All of these data were classified for the agency as “attributed to USGS in NMISC LRG Compendium 12/08”, and the source was classified as NMISC LRG Compendium 12/08. Classifying these data with an agency ID of USGS would not be correct because of where the data were obtained from and because of uncertainty given how the data were acquired and ended up in the LRG compendium database. It is difficult for certain to determine if data that did not match data acquired from USGS NWIS Web are truly USGS or not.
- **Possible Data Issues for Awareness of Users:**
 - This data source requires a note of caution because it likely duplicates data obtained directly from USGS NWIS, but the extent of duplication is unknown. It was included at the request of the NMISC as it was reported to contain some nonpublic (not accessible from NWIS Web) USGS New Mexico data, but the exact records, sites, and number of records are unknown. It is not known how the data were originally obtained by the NMISC. In addition, the nonpublic data of interest may have been provisional (not approved) values and thus may not have been correct.
 - Many of the sites associated with these data do not follow the USGS site ID name convention (8-digit and 15-digit numeric IDs). All alternate IDs that were found associated with these sites were used, but a considerable number had site IDs of different formats that could not be determined as to how those IDs were created.
 - It is known that the LRG compendium databases (lrg_surfacewatermonitoring.mdb and lrg_groundwatermonitoring.mdb) have issues with any time data as communicated by NMISC (Dale Doremus, New Mexico Interstate Stream Commission, written commun., 2009). It is suspected that some records were set to a time of 12:00 p.m. to indicate some sort of daily value or average, or 12:00 p.m. was used as a placeholder for data that did not have time data. The data records were not flagged if such an edit had occurred; consequently, the time component of the date/time field is not usable.
 - Differentiation between instantaneous sample results and daily values was not consistently made, so those data records could be any one of the possible values for that field used in the LRG compendium databases (daily mean, daily mean estimated, estimated, instantaneous, monthly total, regular, regular estimated, or replicate). 98,700 data records of 162,972 have a result description of “unknown” which means they could be daily values or instantaneous values; consequently, all data were considered instantaneous given concern over the accuracy of the data that did have a description.
 - Some parameter codes deviated from the USGS parameter code convention (five numeric digits).

• **Data Preparation Process Steps:**

- Extracted all data attributed to USGS from the lrg_surfacewatermonitoring.mdb and lrg_groundwatermonitoring.mdb files.
- Combined the sites from both databases into one master site table.
- Extracted all sites and data (both instantaneous and daily) obtained from USGS NWIS and combined into similar site and result data tables.
- Took any alternate ID that fit the USGS 8-digit and 15-digit site-naming convention from the StationAssignments and SiteAssignments tables in the lrg_surfacewatermonitoring.mdb and lrg_groundwatermonitoring.mdb, respectively, and replaced the primary site ID of those data (if not in the USGS naming convention) with those alternate IDs. The purpose of this step to switch primary IDs with alternate IDs was to use any information indicating a site and associated data as USGS to compare with the actual USGS NWIS data downloaded from NWIS Web for duplicate resolution.
- Ran a matching query based on the site ID, date (because any time data from the previous LRG compendium databases are not reliable), parameter ID code, and result value to determine which match between the two data sets. These are data to filter out because these data are presumed to be duplicate if they matched based on these four data fields. Those data records were not loaded from the LRG compendium data attributed to USGS. The rule of thumb was that any data obtained from the USGS NWIS Web site supersedes any other records attributed to USGS from other sources.
- Compiled the data that did not match the USGS NWIS Web data and organized them into the site, sample, result, and parameter table format.

Appendix 3. Relational Geodatabase Quality Control (QC)

- **Instantaneous Result Data Duplicate Assessment**

Combining the sample table (TBL_SAMPLE) and result table (TBL_RESULT) into one “flat” denormalized table allows for the data to be fully assessed record-by-record based on a site, its sample information, and the associated supporting data fields. It is important to include all table fields because even a result comment can differentiate between two records that otherwise appear similar. If the source data were provided including data having such characteristics, selecting records in an effort to resolve data that “could” be duplicative was beyond the scope of what could be objectively done. This query combines the two tables (for assessment/quality control purposes ONLY—the database cannot be used with this combined table):

```
SELECT s.site_ID, s.sample_ID, s.DT_TM, s.sample_top, s.sample_bottom, s.sample_type, s.aqfr_cd, s.notes,
s.agency_ID, s.source_ID, t.param_ID, t.result_value, t.result_unit, t.result_qualifier, t.result_desc, t.result_status,
t.result_comment INTO SAMPLE_RESULT
FROM TBL_SAMPLE AS s INNER JOIN TBL_RESULT AS t ON s.sample_ID=t.sample_ID;
```

The following SQL code can then be used to assess for duplicates using the compiled SAMPLE_RESULT table created from the above query. All fields except for the sample field, agency_ID field, and source_ID fields are displayed in the result, but the sample field, agency_ID field, and source_ID field are not included in the analysis because they are relational fields that were created as part of the compilation process, and as such are not part of the original data:

```
SELECT s.[site_ID], s.[sample_ID], s.[DT_TM], s.[sample_top], s.[sample_bottom], s.[aqfr_cd], s.[result_value],
s.[result_qualifier], s.[param_ID], s.[sample_type], s.[notes], s.[agency_ID], s.[source_ID], s.[result_unit],
s.[result_desc], s.[result_status], s.[result_comment] FROM SAMPLE_RESULT AS s WHERE (((s.[site_ID])
In (SELECT [site_ID] FROM SAMPLE_RESULT As Tmp GROUP BY [site_ID],[DT_TM],[sample_
top],[sample_bottom],[aqfr_cd],[result_value],[result_qualifier],[result_desc],[param_ID],[sample_type],[result_
status],[result_comment] HAVING Count(*)>1 And [DT_TM] = s.[DT_TM] And [sample_top] = s.[sample_top]
And [sample_bottom] = s.[sample_bottom] And [aqfr_cd] = s.[aqfr_cd] And [result_value] = s.[result_value] And
[result_qualifier] = s.[result_qualifier] And [result_desc] = s.[result_desc] And [result_status] = s.[result_status]
And [result_comment] = s.[result_comment] And [param_ID] = s.[param_ID] And [sample_type] = s.[sample_
type])))
ORDER BY s.[site_ID], s.[DT_TM], s.[sample_top], s.[sample_bottom], s.[aqfr_cd], s.[result_value], s.[result_
qualifier], s.[param_ID], s.[sample_type];
```

The June 2010 geodatabase version contained only two records that were returned as duplicates. These records are associated with NWIS site 08353000 with samples collected on 10/29/1969. The data existed in this manner in the original downloaded data text file. The reason they are differentiated is because one had a time of 00:00 and the other had a time of null in the original file, and the automated preprocessing script for NWIS QW Field data handles such records as separate events because it is not known whether or not they truly are two distinct events – technically they are different because one had a time of 00:00 and one had a time of null.

- **Daily Result Data Duplicate Assessment**

This query allows the Daily Result Value (TBL_DAILY_RESULT) table to be assessed for records that occur more than once on the basis of all data fields.

```
SELECT TBL_DAILY_RESULT.site_ID, TBL_DAILY_RESULT.DT_TM, TBL_DAILY_RESULT.param_ID,
TBL_DAILY_RESULT.result_value, TBL_DAILY_RESULT.result_unit, TBL_DAILY_RESULT.result_qualifier,
TBL_DAILY_RESULT.result_type, TBL_DAILY_RESULT.result_desc, TBL_DAILY_RESULT.result_status,
TBL_DAILY_RESULT.result_comment, TBL_DAILY_RESULT.agency_ID, TBL_DAILY_RESULT.source_ID
FROM TBL_DAILY_RESULT
WHERE (((TBL_DAILY_RESULT.site_ID) In (SELECT [site_ID] FROM [TBL_DAILY_RESULT] As Tmp
GROUP BY [site_ID],[DT_TM],[param_ID],[result_value],[result_unit],[result_qualifier],[result_type],[result_
desc],[result_status],[result_comment] HAVING Count(*)>1 And [DT_TM] = [TBL_DAILY_RESULT].[DT_
TM] And [param_ID] = [TBL_DAILY_RESULT].[param_ID] And [result_value] = [TBL_DAILY_RESULT].
[result_value] And [result_unit] = [TBL_DAILY_RESULT].[result_unit] And [result_qualifier] = [TBL_DAILY_
RESULT].[result_qualifier] And [result_type] = [TBL_DAILY_RESULT].[result_type] And [result_desc] = [TBL_
DAILY_RESULT].[result_desc] And [result_status] = [TBL_DAILY_RESULT].[result_status] And [result_com-
ment] = [TBL_DAILY_RESULT].[result_comment])))
ORDER BY TBL_DAILY_RESULT.site_ID, TBL_DAILY_RESULT.DT_TM, TBL_DAILY_RESULT.param_ID,
TBL_DAILY_RESULT.result_value, TBL_DAILY_RESULT.result_unit, TBL_DAILY_RESULT.result_qualifier,
TBL_DAILY_RESULT.result_type, TBL_DAILY_RESULT.result_desc, TBL_DAILY_RESULT.result_status,
TBL_DAILY_RESULT.result_comment;
```

- **Daily Result Data Duplicate Result Assessment**

The following query examines for more than one record per site for any given date. The query may return valid data if various types of daily values are available for a site on a date, so any returned records need to be carefully inspected.

```
SELECT s.site_ID, s.agency_ID, t.DT_TM, t.param_ID, Count(t.result_value) AS CountOfresult_value
FROM TBL_SITE AS s INNER JOIN TBL_DAILY_RESULT AS t
ON s.site_ID = t.site_ID
GROUP BY s.site_ID, s.agency_ID, t.DT_TM, t.param_ID
HAVING (((Count(t.result_value))>1))
ORDER BY Count(t.result_value) DESC;
```

- **Duplicate Site Assessment**

This query assesses the site table (TBL_SITE) for duplicate site IDs:

```
SELECT s.[site_ID], s.[lat_DD], s.[long_DD], s.[agency_ID], s.[source_ID], s.[site_desc]
FROM TBL_SITE AS s
WHERE (((s.[site_ID]) In (SELECT [site_ID] FROM [TBL_SITE] As Tmp GROUP BY [site_ID] HAVING
Count(*)>1 )))
ORDER BY s.[site_ID];
```

- **Relational Database Integrity—Orphaned Record Assessment**

An orphaned record is an occurrence of a record in a data table that does not have a record in a related table. One example might be a sample record in a sample table that does not have a site record in the site table. The below queries look at the data tables relative to their related tables to identify cases of orphaned records.

- **Identify Site Table Records Having No Sample Table Records**

```
SELECT s.site_ID, s.site_desc, s.agency_ID, s.source_ID, s.lat_DD, s.long_DD
FROM TBL_SITE AS s LEFT JOIN TBL_SAMPLE AS t
ON s.site_ID = t.site_ID
WHERE t.site_ID Is Null
AND s.site_ID NOT IN (SELECT DISTINCT site_ID FROM TBL_DAILY_RESULT)
```

- **Identify Site Table Records Having No Daily Result Table Records**

```
SELECT TBL_SITE.site_ID, TBL_SITE.agency_ID, TBL_SITE.source_ID
FROM TBL_SITE LEFT JOIN TBL_DAILY_RESULT ON TBL_SITE.site_ID=TBL_DAILY_RESULT.site_ID
WHERE (((TBL_DAILY_RESULT.site_ID) Is Null)) AND TBL_SITE.site_ID NOT IN (SELECT DISTINCT
site_ID FROM TBL_SAMPLE);
```

- **Identify Sample Table Records Having No Result Table Records**

```
SELECT TBL_SAMPLE.site_ID, TBL_SAMPLE.sample_ID, TBL_SAMPLE.agency_ID, TBL_SAMPLE.
source_ID, TBL_SAMPLE.DT_TM
FROM TBL_SAMPLE LEFT JOIN TBL_RESULT ON TBL_SAMPLE.sample_ID = TBL_RESULT.sample_ID
WHERE (((TBL_RESULT.sample_ID) Is Null));
```

- **Identify Sample Table Records Having No Site Table Records**

```
SELECT DISTINCT TBL_SAMPLE.site_ID, TBL_SAMPLE.agency_ID, TBL_SAMPLE.source_ID
FROM TBL_SAMPLE LEFT JOIN TBL_SITE ON TBL_SAMPLE.site_ID=TBL_SITE.site_ID
WHERE (((TBL_SITE.site_ID) Is Null));
```

- **Identify Result Table Records Having No Parameter Table Records**

```
SELECT s.sample_ID, s.param_ID
FROM TBL_RESULT as s LEFT JOIN TBL_PARAM as t
ON s.[param_ID] = t.[param_ID]
WHERE t.param_ID Is Null;
```

- **Identify Result Table Records Having No Sample Table Records**

```
SELECT TBL_RESULT.sample_ID, TBL_RESULT.param_ID, TBL_RESULT.result_value
FROM TBL_RESULT LEFT JOIN TBL_SAMPLE ON TBL_RESULT.sample_ID = TBL_SAMPLE.sample_ID
WHERE (((TBL_SAMPLE.sample_ID) Is Null));
```

- **Identify Daily Result Table Records Having No Site Table Records**

```
SELECT DISTINCT TBL_DAILY_RESULT.site_ID, TBL_DAILY_RESULT.source_ID  
FROM TBL_DAILY_RESULT LEFT JOIN TBL_SITE ON TBL_DAILY_RESULT.site_ID=TBL_SITE.site_ID  
WHERE (((TBL_SITE.site_ID) Is Null));
```

- **Identify Parameter Table Records Having No Associated Result and Daily Result Table Records**

```
SELECT p.param_ID, p.param_desc  
FROM TBL_PARAM AS p LEFT JOIN TBL_RESULT AS r ON p.param_ID=r.param_ID  
WHERE r.param_ID Is Null  
AND p.param_ID NOT IN (SELECT DISTINCT param_ID FROM TBL_DAILY_RESULT)
```

Appendix 4. Geodatabase Table Dictionary

Table 4–1. TBL_SITE table of the geodatabase. The table contains records for surface-water (discharge), groundwater, and water-quality sites.

[ID, identifier; NMISC, New Mexico Interstate Stream Commission]

Table column	Data type	Field size	Field description
SITE_RECORD	Counter	N/A	Auto-incrementing integer for records in the table.
site_ID	Text	75	PRIMARY KEY; the unique site ID associated with the site record.
site_desc	Text	255	The descriptive name for the site, or a description of the site location.
agency_ID	Text	75	FOREIGN KEY; the agency associated with the spatial location data only. This data field could represent the primary steward of the site, or simply the agency that was associated with the data collected at the site in the data source file.
source_ID	Text	75	FOREIGN KEY; the source of the site record spatial location data only. This data field could be an agency, an online-accessible database, or a previous data compilation. The source IDs in this table may not return collected sample or daily result data in a query because they are not necessarily the same.
lat_DD	Double		The latitude coordinate for the site.
long_DD	Double		The longitude coordinate for the site.
h_datum_DD	Text	50	The horizontal geographic datum of the latitude and longitude coordinates.
alt_ft	Text	50	The altitude of the site in feet above sea level datum.
v_datum	Text	50	The vertical datum of the site.
aqfr_cd	Text	50	An aquifer code associated with the site. This field is used for daily value water-level sites; otherwise, the aqfr_cd field on the TBL_SAMPLE table can be used (in cases of water-quality data collected in a well at various levels).
site_comment	Text	255	Notes, comments, general text field for information related to the site.
orig_type_cd	Text	50	The site “type” code (for example, discharge, drain, and so forth) from the original data source.
type_cd	Text	50	The type code for the site assigned from this NMISC data compilation.
sub_type_cd	Text	50	The subtype code for the site (typically will be more specific than the general type code).
cons_date	Text	50	The known construction date for the site.
seal_depth	Text	50	Feet below the land surface; the depth to the bottom of the seal (related to the construction of the site).
well_depth	Text	50	Feet below the land surface; the depth of the finished well, this is the greatest depth to which the well can be sounded, or the reported depth at which the well was finished.
hole_depth	Text	50	Feet below the land surface; the depth to which the hole was drilled.

Table 4-2. TBL_SAMPLE table of the geodatabase. The table contains data records associated with instantaneous sample-event data.

[ID, identifier]

Table column	Data type	Field size	Field description
SAMPLE_RECORD	Counter	N/A	Auto-incrementing integer for records in the table.
site_ID	Text	75	FOREIGN KEY; the unique site ID for the site associated with the sample-event record.
sample_ID	Text	150	PRIMARY KEY; the unique ID for the sample event record.
DT_TM	DATETIME		The date and time of the sample event record.
notes	Text	255	Notes, comments, and so forth related to the sample event record.
sample_top	Text	100	Feet below the land surface; the topmost depth of the sample (applies to well/subsurface sites that have screened intervals or similar construction; this data field field in such an example case denotes the depth in feet to the top part of a screen in a well where a sample was collected). This field is also used in cases of stream/surface-water samples to represent sample depth.
sample_bottom	Text	100	Feet below the land surface; the bottommost depth of the sample (applies to well/ subsurface sites that have screened intervals or similar construction. The sample_bot- tom data field in such an example case denotes the depth in feet to the bottom part of a screen in a well where a sample was collected).
sample_type	Text	255	A general field for information related to the type of sample collected and represented by this sample record (for example, could be routine, duplicate types, and so forth).
aqfr_cd	Text	50	An aquifer code associated with the sample event collected from a particular site. This field is applicable for data that are not daily values or daily water levels.
agency_ID	Text	75	FOREIGN KEY; the agency associated with the sample event record in the electronic data. This data field could represent the data collecting agency, or the data distributing agency (in some cases it is not known which the agency represents), or possibly both.
source_ID	Text	75	FOREIGN KEY; the source of the actual sample event record. This data field could refer to an agency, an online accessible database, a previous data compilation, or collection of files.

Table 4–3. TBL_RESULT table of the geodatabase. The table stores data records for parameter result values.

[ID, identifier]

Table column	Data type	Field size	Field description
RESULT_RECORD	Counter	N/A	PRIMARY KEY an auto-incrementing integer for records in the table.
sample_ID	Text	150	FOREIGN KEY; the unique ID of the parent sample event record in the sample table.
result_value	Double		The actual measured value for the associated parameter of the result record.
result_unit	Text	255	The measurement units of the result record value.
result_qualifier	Text	255	A general field for any result value qualifiers (for example, a symbol denoting a detection limit); this is sometimes also known as a value remark.
result_desc	Text	255	A general field for any related descriptive text associated with the result.
result_status	Text	255	A general field for information related to the overall status of the result (for example, approved, and so forth).
result_comment	Text	255	A general field for notes related to the result value.
param_ID	Text	150	FOREIGN KEY; unique parameter ID for the measurement represented by the result value.

Table 4-4. TBL_DAILY_RESULT table of the geodatabase. The table stores data result records associated with daily values.

[ID, identifier]

Table column	Data type	Field size	Field description
DAILYRESULT_RECORD	Counter	N/A	PRIMARY KEY; auto-incrementing integer for records in the table.
site_ID	Text	75	FOREIGN KEY; the unique ID for the site with which the daily result value is associated.
DT_TM	DATETIME		The date associated with the daily result value record.
param_ID	Text	150	FOREIGN KEY; the unique ID of the parameter associated with the daily result value record.
result_value	Double		The actual daily value for the associated parameter of the result record.
result_unit	Text	255	The measurement units of the daily result value record.
result_qualifier	Text	255	A general field for any daily result value qualifiers (for example, a symbol denoting a detection limit); this is sometimes also known as a value remark.
result_type	Text	255	A general field for the type of daily value.
result_desc	Text	255	A general field for any related descriptive text associated with the daily result value record.
result_status	Text	255	A general field for information related to the overall status of the daily result value (for example, approved, and so forth).
result_comment	Text	255	A general text field for notes, comments, and so forth associated with the daily result value record.
agency_ID	Text	75	FOREIGN KEY; the agency associated with the daily result value in the electronic data. This could represent the data collecting agency, or the data distributing agency (in some cases it is not known which the agency represents), or possibly both.
source_ID	Text	75	FOREIGN KEY; the source of the actual daily result value record. The value in this data field could refer to an agency, an online-accessible database, a previous data compilation, or collection of files.

Table 4–5. TBL_SOURCE table of the geodatabase. The table contains data sources for the primary data tables of the geodatabase and serves as a domain table from the source identifier (ID) field (source_ID) where it occurs in other tables.

Table column	Data type	Field size	Field description
SOURCE_RECORD	Counter	N/A	Auto-incrementing integer for records in the table.
source_ID	Text	75	PRIMARY KEY; the unique ID for the data source.
source_desc	Text	255	A description of the source.
source_file	Text	255	The actual file name of the source electronic data loaded into the NMISC data compilation.

Table 4–6. TBL_PARAM table of the geodatabase. The table contains all hydrologic parameters associated with result values in the geodatabase and serves as a domain table for the parameter identifier (ID) field (param_ID) where it occurs in other tables.

Table column	Data type	Field size	Field description
PARAM_RECORD	Counter	N/A	Auto-incrementing integer for records in the table.
param_ID	Text	255	PRIMARY KEY; the unique parameter ID.
param_desc	Text	255	A description of the parameter.
common_name	Text	255	A common name for the parameter.
param_group	Text	255	A group identifier for the parameter.

Table 4–7. TBL_ALT_ID table of the geodatabase. The table contains site records that have known alternate site identifiers (IDs). An alternate site ID is another known identifier for a physical on-the-ground site.

Table column	Data type	Field size	Field description
ALT_ID_RECORD	Counter	N/A	Auto-incrementing integer for records in the table.
site_ID	Text	75	PRIMARY KEY and FOREIGN KEY; the unique parameter ID.
alt_ID_1	Text	75	An alternate site ID for the associated primary site ID. This data field can be used to group sites by name, location, or other generalized information.
alt_ID_2	Text	75	An alternate site ID for the associated primary site ID. This data field can be used to group sites by name, location, or other generalized information.
alt_ID_3	Text	75	An alternate site ID for the associated primary site ID. This data field can be used to group sites by name, location, or other generalized information.

Table 4–8. TBL_AGENCY table of the geodatabase. The table contains agency records of agencies associated with the data sources and serves as a domain table for the agency identifier (ID) field (agency_ID) where it occurs in other tables.

Table column	Data type	Field size	Field description
AGENCY_RECORD	Counter	N/A	Auto-incrementing integer for records in the table.
agency_ID	Text	75	PRIMARY KEY; the unique ID for the agency.
agency_desc	Text	255	A description of the agency.

Appendix 5. Example Microsoft (MS) Access Database Queries

Below are example queries showing how data might be queried from the NMISC geodatabase. The SQL code below needs to be copied and pasted into the MS Access Query Design SQL-view window (do not add tables to the query). The [Enter Parameter to Search] criteria will prompt the user with a popup window to enter specific criteria (for example, a parameter common name such as specific conductance) that will then be used in the query. It should be noted that the SQL wildcard character symbol, an asterisk "*", can be used before and after a parameter name in the below queries to capture any parameter common name similar to it. This wildcard works only for parameter names and makes the query more flexible and generalized. For example, when running the first query below, enter *specific conductance* and the query will return any record that has the exact term as the parameter common name.

- **Query Instantaneous Results (TBL_RESULT) by Parameter Common Name**

```
SELECT TBL_SITE.site_ID, TBL_SAMPLE.DT_TM, TBL_PARAM.common_name, TBL_RESULT.result_qualifier, TBL_RESULT.result_value, TBL_RESULT.result_unit, TBL_AGENCY.agency_ID, TBL_AGENCY.agency_desc, TBL_SOURCE.source_ID, TBL_SOURCE.source_desc
FROM TBL_SOURCE INNER JOIN (TBL_AGENCY INNER JOIN (TBL_PARAM INNER JOIN ((TBL_SITE INNER JOIN TBL_SAMPLE ON TBL_SITE.site_ID = TBL_SAMPLE.site_ID) INNER JOIN TBL_RESULT ON TBL_SAMPLE.sample_ID = TBL_RESULT.sample_ID) ON TBL_PARAM.param_ID = TBL_RESULT.param_ID) ON (TBL_AGENCY.agency_ID = TBL_SITE.agency_ID) AND (TBL_AGENCY.agency_ID = TBL_SAMPLE.agency_ID)) ON (TBL_SOURCE.source_ID = TBL_SITE.source_ID) AND (TBL_SOURCE.source_ID = TBL_SAMPLE.source_ID)
WHERE TBL_PARAM.common_name LIKE [Enter Parameter Common Name to Search];
```

- **Query Instantaneous Results (TBL_RESULT) by Site ID**

```
SELECT TBL_SITE.site_ID, TBL_SAMPLE.DT_TM, TBL_PARAM.common_name, TBL_RESULT.result_qualifier, TBL_RESULT.result_value, TBL_RESULT.result_unit, TBL_SAMPLE.agency_ID, TBL_SAMPLE.source_ID
FROM (TBL_SITE INNER JOIN TBL_SAMPLE ON TBL_SITE.site_ID = TBL_SAMPLE.site_ID) INNER JOIN (TBL_PARAM INNER JOIN TBL_RESULT ON TBL_PARAM.param_ID = TBL_RESULT.param_ID) ON TBL_SAMPLE.sample_ID = TBL_RESULT.sample_ID
WHERE TBL_SITE.site_ID = [Enter Site ID to Search]
```

- **Query Instantaneous Results (TBL_RESULT) by Site ID and Parameter Common Name**

```
SELECT TBL_SITE.site_ID, TBL_SITE.site_desc, TBL_SITE.type_cd, TBL_SAMPLE.DT_TM, TBL_PARAM.common_name, TBL_RESULT.result_value, TBL_RESULT.result_qualifier, TBL_RESULT.result_unit, TBL_SAMPLE.agency_ID, TBL_SAMPLE.source_ID
FROM TBL_PARAM INNER JOIN ((TBL_SITE INNER JOIN TBL_SAMPLE ON TBL_SITE.site_ID = TBL_SAMPLE.site_ID) INNER JOIN TBL_RESULT ON TBL_SAMPLE.sample_ID = TBL_RESULT.sample_ID) ON TBL_PARAM.param_ID = TBL_RESULT.param_ID
WHERE TBL_SITE.site_ID = [Enter Site ID to Search] AND TBL_PARAM.common_name LIKE [Enter Parameter Common Name to Search];
```

- **Query Instantaneous Results (TBL_RESULT) by Parameter Common Name and Minimum Detection Value**

```
SELECT TBL_SITE.site_ID, TBL_SAMPLE.DT_TM, TBL_RESULT.result_value, TBL_RESULT.result_unit, TBL_RESULT.result_qualifier, TBL_SAMPLE.agency_ID, TBL_SAMPLE.source_ID
FROM TBL_PARAM INNER JOIN ((TBL_SITE INNER JOIN TBL_SAMPLE ON TBL_SITE.site_ID = TBL_SAMPLE.site_ID) INNER JOIN TBL_RESULT ON TBL_SAMPLE.sample_ID = TBL_RESULT.sample_ID) ON TBL_PARAM.param_ID = TBL_RESULT.param_ID
WHERE TBL_PARAM.common_name LIKE [Enter Parameter Common Name to Search] AND TBL_RESULT.result_value > [Enter Minimum Result Value];
```

- **Query Daily Value Results (TBL_DAILY_RESULT) by Site ID**

```
SELECT TBL_SITE.site_ID, TBL_PARAM.common_name, TBL_DAILY_RESULT.param_ID, TBL_DAILY_RESULT.result_value, TBL_DAILY_RESULT.result_unit, TBL_DAILY_RESULT.result_type, TBL_DAILY_RESULT.result_desc, TBL_DAILY_RESULT.agency_ID, TBL_DAILY_RESULT.source_ID
FROM TBL_SITE INNER JOIN (TBL_PARAM INNER JOIN TBL_DAILY_RESULT ON TBL_PARAM.param_ID = TBL_DAILY_RESULT.param_ID) ON TBL_SITE.site_ID = TBL_DAILY_RESULT.site_ID
WHERE TBL_SITE.site_ID = [Enter Site ID To Search For];
```

Appendix 6. NMISC Geodatabase User Tips

This section covers important items of interest to users of the NMISC compendium. These tips will help guide the use of the geodatabase to answer questions based on the geodatabase design, relational database principles, and the characteristics of the data associated with the variety of sources.

Appendix 6.1. Database Queries

- The NMISC geodatabase likely is not considered duplicate-free in terms of duplicate sites or duplicate result data values. There is an onus on the user to further scrub data given the complex issue of alternate site IDs (see section 2.5 and further down in this section).
- The report geodatabase diagram (fig. 3) can be used to guide database queries designed either in the MS Access Query Design Window or when writing SQL code. The relationships shown in the diagram determine how the various tables and fields relate to each other.
- TBL_PARAM, TBL_AGENCY, and TBL_SOURCE serve as domain tables for the other primary data tables. The purpose of a domain table is to contain all possible values for a particular field whenever that field is found on a primary data table. As such, the primary key fields of a respective table serve as foreign key fields on the other primary data tables. A foreign key domain field on any particular table (for example, the agency_ID field in the sample table TBL_SAMPLE) may not necessarily have all acceptable values in that data table. In other words, the TBL_AGENCY domain table may have agency codes that do not appear in the sample table but that do appear in the site table TBL_SITE.

- One workaround to determine all values of a foreign key domain table field that occur in any particular table is to use the SQL keyword DISTINCT in a query on the foreign key domain field of interest. For example, this query will return all agency ID codes that occur in the sample table along with the agency ID code description from the TBL_AGENCY domain table:

```
SELECT DISTINCT s.agency_ID, t.agency_desc FROM TBL_SAMPLE AS s INNER JOIN TBL_AGENCY AS t ON s.agency_ID = t.agency_ID
```

The returned results can then be used to identify agency(s) of interest found in a particular table. The same thing can be done for all other primary data tables that contain a foreign key domain field of interest.

- The GEO_TBL_SITE table can only be used with all other data tables in the ESRI ArcGIS environment. The TBL_SITE table can only be used with all other geodatabase tables in the MS Access environment. This selective use is because the configured table relationships in MS Access will only work with the nonspatial TBL_SITE table, and the configured relationship classes in ESRI ArcGIS will only work with the spatially enabled GEO_TBL_SITE table.
- The “USGS” source ID code only applies to USGS site records in the site table and not any USGS sample table records or daily result value records, which is further explained in section 3.4. The source ID code of “USGS” is used only for site records obtained from USGS NWIS and will not be found in any other primary data table. The source codes for any data table records obtained from USGS NWIS will characterize the respective USGS data source (for example, USGS_QWField is the source ID code for USGS water quality sample data).
- Alternate site IDs are stored in the TBL_ALT_ID table. This table is designed with a one-to-one relationship with the primary site table, and contains multiple alternate ID fields. This table and the table fields are defined in the data dictionary and the purpose of this table with duplicate site resolution is discussed in section 2.5.
- The data source called “NMISC data attributed to USGS” requires a note of caution. This data set likely duplicates data obtained directly from USGS NWIS, but the extent of duplication is unknown. It was included at the request of the NMISC as it was reported to contain some nonpublic (not accessible from NWIS Web) USGS New Mexico data, but the exact records, sites, and number of records are unknown. It is not known how the data were originally obtained by the NMISC. In addition, the nonpublic data of interest may have been provisional (not approved) values and thus may not have been correct. See appendix 2 for more information regarding the specifics of this source.

Appendix 6.2. User Frequently Asked Questions (FAQ)

This section contains possible usage FAQs from a user's perspective.

- **I have extracted data from the NMISC geodatabase for my study area. Are the data ready for analysis or is there anything else I should be aware of regarding duplicate data or other?**
- There is an onus on any user to examine the data they extract from the NMISC geodatabase for both data duplicates as well as site duplicates. When so many data sources in addition to preexisting compilations are combined, it is nearly impossible to resolve all site duplicates among sources because of issues such as alternate site IDs and missing information from previous data compilations. Section 2.5 discusses the methods that were used for duplicate site and duplicate data resolution. The TBL_ALT_ID table needs to be used in conjunction with the primary data table to further scrub data based on the issue of alternate site IDs for the same physical on-the-ground sites among many of the geodatabase sources.
- **Why are values in the source_ID field not the same in different tables?**
- Database design necessitates normalization of data to reduce redundancy. The source_ID field is a table field that is associated with a domain table called TBL_SOURCE. TBL_SOURCE contains the possible values that source_ID might have. It is not required that every value in TBL_SOURCE occur in every table where the source_ID field is found. The geodatabase diagram along with the data dictionary (appendix 4) show which table fields are associated with a domain table and will be labeled with "Foreign Key" or "FK" accordingly. Appendix 4 also discusses this issue and an example method for determining the values of such fields in the primary data tables.
- **What is the purpose of the Alternate ID table and the Alternate ID fields?**
- Known alternate IDs for a given site are common unique identifiers given to a site by one or more agencies that collect data at that location. These can be thought of as alias names for a physical location. For example, a site in USGS NWIS may have an eight-digit unique site ID, and then the same site may have data collection done by the IBWC under a different site ID with a completely different format. The collection of site IDs allows the data user to identify all data from various sources at that location. For the purpose of database design, selection of a "primary" site ID is needed and then is used as the primary unique identifier and, for the site table, the table Primary Key. In some cases, sources do not document the alternate IDs for locations associated with data. In such cases, known alternate ID information can be used to further clean compiled data for possible duplicates based on the primary site ID, alternate site ID, and other information such as the user's on-the-ground knowledge of the site(s) of interest.

Appendix 7. New Mexico Interstate Stream Commission (NMISC) Geodatabase—Geographic Information System (GIS) Functionality Overview

The NMISC compendium is spatially enabled as a geodatabase that allows the data to be viewed and queried by using ESRI ArcGIS Software. ArcGIS relationship classes have been set up in the June 2010 version of the geodatabase that mirror the MS Access database relationships. The relationship classes enable relational database-style selections across tables so that the spatial aspect of the geodatabase and GEO_TBL_SITE feature class can be used to answer questions and determine management priorities. Following are some general instructions for using the geospatial features of the geodatabase in ArcGIS version 9.3:

- Open a new ArcMap .MXD file or an existing .MXD file associated with the study area of interest.
- Add the GEO_TBL_SITE point feature class, the NMISC study area (included in the June version of the geodatabase), and any other GIS data of interest.
- Add the TBL_SAMPLE, TBL_RESULT, TBL_DAILY_RESULT, TBL_PARAM, TBL_AGENCY, and TBL_SOURCE geodatabase tables to the ArcMap document. They will not be visible in the main ArcMap Table of Contents window on the left side of the screen but can be seen by clicking on the “Source” tab at the bottom of the Table of Contents window.
- Select one or more sites of interest by clicking on the site record(s) in the GEO_TBL_SITE table, or by spatially selecting sites relative to other GIS data in the ArcMap document (for example, all sites within a particular NMISC study area subbasin). Individual sites can also be selected by clicking directly on their point feature in ArcMap by using the Selection tool.
- Open the GEO_TBL_SITE attribute table (if not already open). Click Options→Related Tables to propagate the selection to related fields in a related table of interest. Any tables related by relationship classes will be displayed. For example, one or more sites can be selected, and then the selection can be propagated to the sample table. This will open the Sample table, and the display of the Sample table can be limited to just the selected records by clicking on the “Selected” button near the bottom middle of the sample table attribute window.
- The selected samples can then be propagated to the Results table in a similar manner by clicking Options→Related Tables and selecting TBL_RESULT.
- SQL can also be used to select records from any table (for example, samples for a particular site that occur between a specified date range). The Select By Attributes functionality is accessed by opening a table attribute view window and then selecting Options Select By Attributes. There a SQL statement can be written and executed. The records of interest will be selected, and then the selection can be propagated to a related table as described above.
- It is also possible to start from the bottom-up by querying the Daily Result or Result table for a particular parameter code and any results above a certain detection limit by using the SQL Select By Attributes tool. The selected results records can then be propagated to the Sample table to determine when those result values of interest occurred by the Date/Time field (DT_TM) as well as the site ID associated with those selected sample records. The selected samples can then be propagated as described up to the site table (GEO_TBL_SITE) and will be displayed as selected features in ArcMap.

NOTE: The GEO_TBL_SITE table can only be used with all other geodatabase tables in the ESRI ArcGIS environment. The TBL_SITE table can only be used with all other geodatabase tables in the MS Access environment. This limitation exists because the MS Access table relationships are only configured to work with the nonspatial TBL_SITE table, and geodatabase relationship classes are only configured to work in ESRI ArcGIS with the spatially enabled GEO_TBL_SITE table.

Appendix 8. Additional Resources

Numerous free and publicly available resources related to concepts and tools used to create the June 2010 version of the geodatabase are available on the Internet. Examples of resource topics include databases, SQL, and ESRI ArcGIS. No claim is made by the USGS regarding the accuracy of the material found on the associated Web sites. Any opinions reflected on the listed Web sites are solely those of the respective Web site authors. Inclusion here does not constitute endorsement by the USGS. All example Web sites were accessed June 2010.

- Microsoft Office Access query tips: <http://www.fmsinc.com/tpapers/queries/>
- A general reference for database SQL (though Access does not implement all of the commands found at this site): <http://www.1keydata.com/sql/sql-syntax.html>
- SQL quick-reference sheet: http://www.cryer.co.uk/brian/sql/sql_crib_sheet.htm
- Microsoft SQL reference: <http://office.microsoft.com/en-us/access/CH062526881033.aspx>
- ArcGIS 9.3 SQL reference: http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=257&pid=247&topicname=SQL_reference
- ArcGIS 9.3 information on building SQL expressions: http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName>About_building_an_SQL_expression
- ArcGIS 9.3 information on creating and maintaining geodatabase relationship classes: http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=2582&pid=2576&topicname=An_overview_of_creating_and_maintaining_relationship_classes
- Comparing relates, joins, and relationship classes in ArcGIS 9.3: http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Deciding_between_relationship_classes,_joins,_and_relates

