USGS
science for a changing world

Alaska Geochemical Database (AGDB)—Geochemical Data for Rock, Sediment, Soil, Mineral, and Concentrate Sample Media

Data Series 637

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

Cover art: Watercolor in pen and ink, by Daniel Granitto, 2011.

Alaska Geochemical Database (AGDB)—Geochemical Data for Rock, Sediment, Soil, Mineral, and Concentrate Sample Media

By Matthew Granitto, Elizabeth A. Bailey, Jeanine M. Schmidt, Nora B. Shew, Bruce M. Gamble, and Keith A. Labay

Data Series 637

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: 2011

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod

To order this and other USGS information products, visit http://store.usgs.gov

Suggested citation:
Granitto, Matthew, Bailey, E.A., Schmidt, J.M., Shew, N.B., Gamble, B.M., and Labay, K.A., 2011, Alaska Geochemical Database (AGDB)—Geochemical data for rock, sediment, soil, mineral, and concentrate sample media: U.S. Geological Survey Data Series 637, 31 p. pamphlet and database, 1 DVD. Available at *http://pubs.usgs.gov/ds/637*.

Contents

Figures

Tables

Conversion Factors

Inch/Pound to SI

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)

SI to Inch/Pound

Multiply	By	To obtain
	Volume	
liter (L)	0.2642	gallon (gal)
	Mass	
gram (g)	0.03527	ounce, avoirdupois (oz)

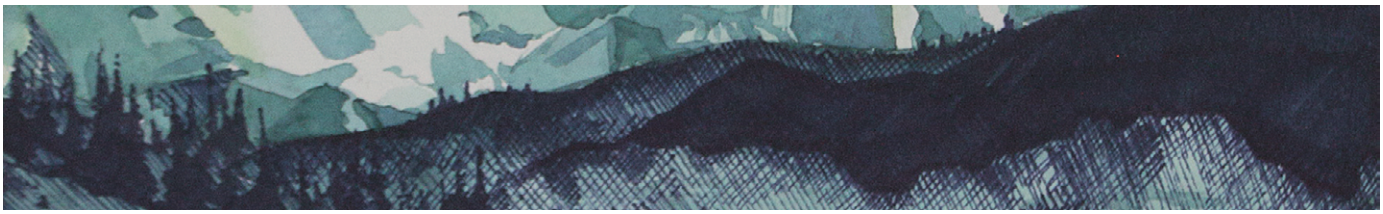
Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F=(1.8×°C)+32$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$°C=(°F-32)/1.8$$

Concentrations of chemical constituents in water are given either in milligrams per liter (mg/L) or micrograms per liter (µg/L).



Alaska Geochemical Database (AGDB)—Geochemical Data for Rock, Sediment, Soil, Mineral, and Concentrate Sample Media

By Matthew Granitto, Elizabeth A. Bailey, Jeanine M. Schmidt, Nora B. Shew, Bruce M. Gamble, and Keith A. Labay

Abstract

The Alaska Geochemical Database (AGDB) was created and designed to compile and integrate geochemical data from Alaska in order to facilitate geologic mapping, petrologic studies, mineral resource assessments, definition of geochemical baseline values and statistics, environmental impact assessments, and studies in medical geology. This Microsoft Access database serves as a data archive in support of present and future Alaskan geologic and geochemical projects, and contains data tables describing historical and new quantitative and qualitative geochemical analyses. The analytical results were determined by 85 laboratory and field analytical methods on 264,095 rock, sediment, soil, mineral, and heavy-mineral concentrate samples. Most samples were collected by U.S. Geological Survey (USGS) personnel and analyzed in USGS laboratories or, under contracts, in commercial analytical laboratories. These data represent analyses of samples collected as part of various USGS programs and projects from 1962 to 2009. In addition, mineralogical data from 18,138 nonmagnetic heavy mineral concentrate samples are included in this database. The AGDB includes historical geochemical data originally archived in the USGS Rock Analysis Storage System (RASS) database, used from the mid-1960s through the late 1980s and the USGS PLUTO database used from the mid-1970s through the mid-1990s. All of these data are currently maintained in the Oracle-based National Geochemical Database (NGDB). Retrievals from the NGDB were used to generate most of the AGDB data set. These data were checked for accuracy regarding sample location, sample media type, and analytical methods used. This arduous process of reviewing, verifying and, where necessary, editing all USGS geochemical data resulted in a significantly improved Alaska geochemical dataset. USGS data that were not previously in the NGDB because the data predate the earliest USGS geochemical databases, or were once excluded for programmatic reasons, are included here in the AGDB and will be added to the NGDB. The AGDB data provided here are the most accurate and complete to date, and should be useful for a wide variety of geochemical studies. The AGDB data provided in the linked database may be updated or changed periodically. The data on the DVD and in the data downloads provided with this report are current as of the date of publication (2011).

Introduction

The U.S. Geological Survey (USGS) began scientific investigations in Alaska in 1889, shortly after Alaska was purchased from the Russian Empire in 1867, but much Alaskan scientific data hasn't always been readily accessible to the public. The USGS participated in the Congressionally-funded, multi-agency Minerals Data Information Rescue in Alaska (MDIRA) Program from 1997 to 2003 in an effort to make its Alaska scientific data digital, correct, user-friendly, and accessible. A large portion of the MDIRA effort is represented by this dataset, which contains geochemical data for Alaska produced by in-house and contract analytical laboratories of the Geologic Division of the USGS from 1962 through 1995 (Bailey and others, 1999). The Alaska Geochemical Database (AGDB) includes analyses of rocks, sediments (collected from streams, lakes, and various sources), soils, minerals, and heavy-mineral concentrates (derived from stream sediments, soils or rocks) compiled in part during the MDIRA process, together with analyses of a variety of geologic materials from 1996 through 2010.

Many of the analyses were the result of mineral resource investigations conducted by the Alaska Mineral Resource Assessment Program (AMRAP) in the mid-1970s through early 1990s; some of the data were produced in support of other USGS programs such as geologic mapping, volcano hazards, development of assessment techniques, and energy resource investigations.

Most of these data were originally entered into the USGS Rock Analysis Storage System (RASS) or PLUTO databases. The RASS database, which contained over 730,000 data records, was used by the Geologic Division from the late 1960s through the late 1980s to archive geochemical data

produced by the Branch of Exploration Services primarily from mineral resource studies. The PLUTO database, which contained over 530,000 data records, was used by the Geologic Division from the late 1960s through 1997 to archive geochemical data from the Branch of Analytical Chemistry and successor branches. More recent geochemical data are presently stored in the commercial Laboratory Information Management System (LIMS) database (1996 to present). Data from these databases have been merged and are maintained in the Oracle-based National Geochemical Database (NGDB) which currently contains nearly 2 million data records. Many of these determinations have been previously published in hardcopy or digital USGS Open-File Reports by the original sample submitters or analysts; however, some have never been published or publically available.

These data were generated over a span of nearly 50 years from 1962 to the present (2010) in USGS laboratories or, under contracts, in commercial analytical laboratories. Upon completion of the sampling and analysis, data for the original RASS and PLUTO databases were keypunched for digital storage. Automated data entry from a LIMS was begun in the late 1980's in the PLUTO database. Over the years, USGS scientists recognized several problems with the databases. Three primary issues were: (1) erroneous or missing sample locality coordinates, (2) sample media were often not adequately identified, and (3) analytical methods were poorly identified (often missing from the early PLUTO database structure). Beginning in 1998, a major review of Alaska samples from these databases was initiated to confirm sample and lab numbers for analyses, correct errors in sample site locations, add sample site locations when missing, and to correctly identify the sample media and analytical protocols for each record. This re-processing consisted of comparing the recorded digital information to the information found in original sample submittal forms, the original analytical reports, published reports, field notebooks and field sheets (sample locality maps), and follow-up discussions with submitters and analysts (when available). As necessary, additional data fields were added to the database structure to more fully describe sample preparation methods and sample media. Geochemical datasets that could have been entered in RASS and PLUTO but were not have been incorporated into this database, resulting in the addition of thousands of sample data records to the NGDB. The RASS and PLUTO analytical data were not checked in great detail, but obvious errors were corrected.

The AGDB is the most current, complete and accurate data compilation for new and historical geochemical analyses of rock, sediment, soil, and concentrate samples. The AGDB also contains mineralogical information from optical examination of the nonmagnetic fractions of heavy-mineral concentrate samples. In addition, geochemical data from reanalyses of sediment samples collected under the Atomic Energy Commission National Uranium Resource Evaluation (NURE) Hydrogeochemical and Stream Sediment Reconnaissance (HSSR) program (henceforth called NURE), and reanalyses (2007–2009) of AMRAP sediment samples from the Alaska Range have been included in the AGDB. Data from other recent projects in Taylor Mountains Quadrangle, southwestern Alaska, Fortymile district, and statewide Surveys and Analysis are also included in the AGDB.

Geographic Setting

The geographic boundaries of the AGDB include all of Alaska as well as Federal and International waters of the Arctic and Pacific Oceans that are reasonable for inclusion. The current AGDB data extents are 71.6000 °N. latitude, 50.1834 °N. latitude, 129.9939 °W. longitude, and 173.1 °E. longitude (fig. 1).

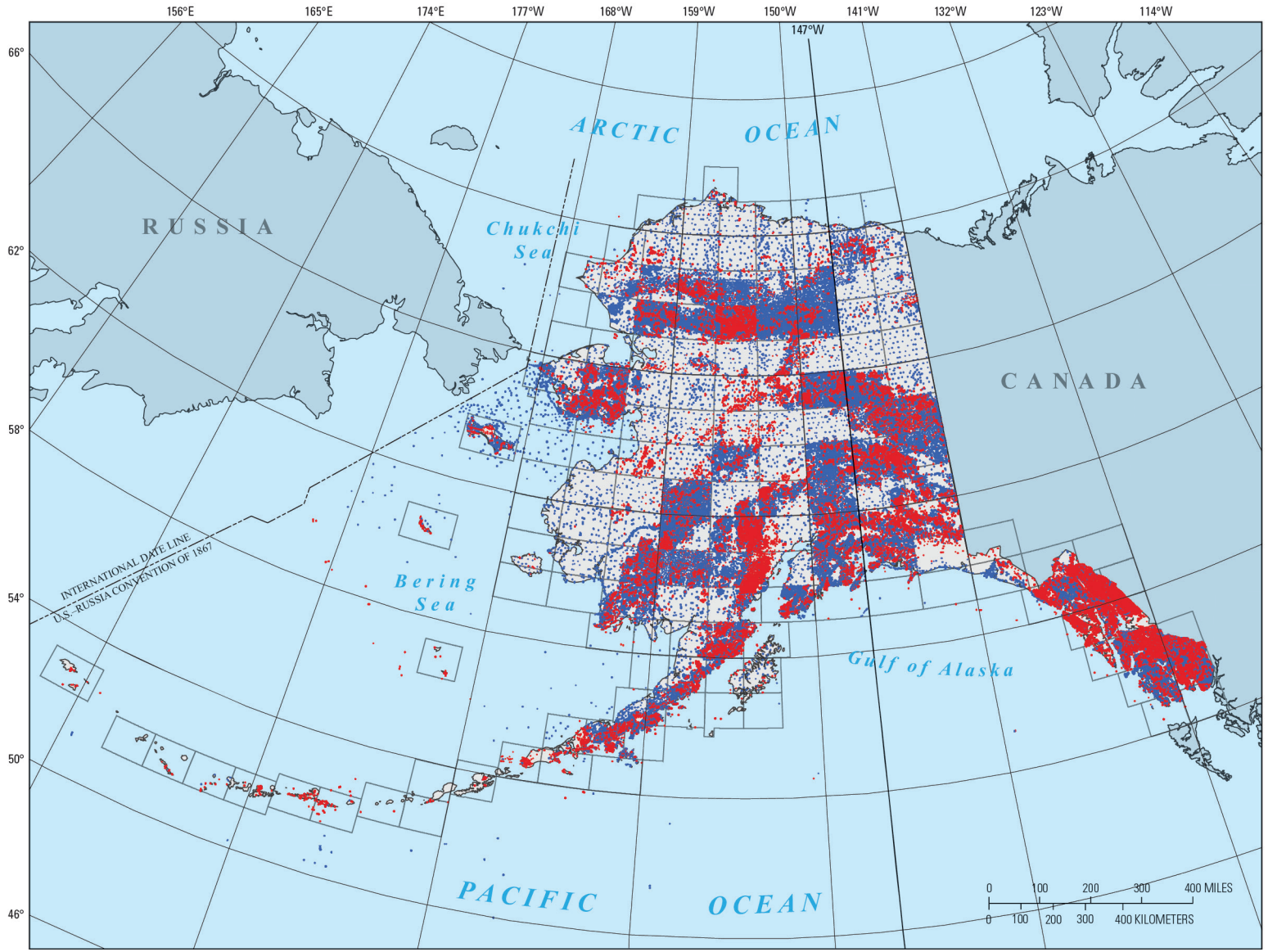
Methods of Study

Sample Media and Collection

Analyses of 108,909 rock, 92,701 sediment, 6,869 soil, 7,357 mineral, and 48,209 heavy-mineral concentrate samples are incorporated into the AGDB. Samples in the AGDB were collected between 1962 and 2009 and prepared according to a variety of USGS standard methods (variously described in Arbogast, 1990, 1996; Miesch, 1976; Taggart, 2002) or by NURE methods (described in Smith, 1997). The database includes analyses of 1,589 stream-sediment samples, originally collected during the USGS Heavy Metals and AMRAP programs (1966–1995), and reanalyzed during 2007–2010 by the Federal Lands in Alaska-Geologic Studies project (Bailey and others, 2010; Gamble and others, 2010); analyses of 3,015 rock, sediment, soil, mineral, and concentrate samples collected during the 2004–2010 Taylor Mountains quadrangle project (Klimasauskas, Miller, Bradley, Bundtzen and others, 2006; Klimasauskas, Miller, Bradley, Karl and others, 2006; Bailey and others, 2007; Klimasauskas, Miller and Bradley, 2007) and data from 6,686 NURE sediment samples that were reanalyzed—4,804 for the National Geochemical Survey project between 1998–2008 (U.S. Geological Survey, 2004) and 1,640 for AMRAP projects. Existing NGDB data from Alaskan water, leachate, and organic samples (including humus and peat) have been excluded from this database but will be compiled, verified, and included in a later version of the AGDB.

Analytical Techniques

Eighty-five different field and laboratory analytical methods were used to produce geochemical data included in the AGDB. The various analytical methods reflect the evolution of analytical chemistry from 1960 to 2009. Appendix 1 provides a complete list of the analytical methods included in the AGDB and descriptive information for each. The **AnalyticMethod** table in the AGDB provides detailed



Base modified from the Alaska Department of Natural Resources. The U.S.–Russia Convention of 1867 line is adapted from the U.S. Dept. of State Treaty Document 101–22, available at *http://www.state.gov/g/oes/ocns/opa/convention/c28187.htm.* Boundary is approximate and for display purposes only.

Figure 1. Geographic area covered by Alaska Geochemical Database. The state of Alaska is in the center of the map. Gray rectangles are the 1:250,000 topographic quadrangle outlines. The red dots are sediment, soil, and heavy-mineral concentrate sample localities. The blue dots are rock and mineral sample localities. The heavy black line at 147°W longitude separates the derivative datasets for sediments and rocks into east and west tables.

information about techniques and the **AnalyticMethodBiblio** table contains citations for the analytical methods.

Quality Assurance/Quality Control

Data on quality assurance and quality control (QA/QC) procedures and methods varied through the period of time associated with AGDB data and are not incorporated into the database. Data from field sample-site duplicates and laboratory analytical replicates (splits of a single sample to check laboratory precisions and accuracy) are included in the database. The USGS and contract laboratories reporting these analyses use a variety of constituent standards (for example, USGS geochemical reference rock standard STM–1, nepheline syenite) and blanks for their internal QA/QC controls (Arbogast, 1990; Taggart, 2002). Data for these reference samples are not included in the database, but a description of commonly used reference materials is available from the USGS at *http://minerals.cr.usgs.gov/geo_chem_stand/*.

Characteristics of the Relational Database

Because of the scope and complexity of geologic materials collected for diverse Alaskan studies, a relational database structure was designed for data storage. The AGDB was constructed in Microsoft Access 2003 and 2010 as a tool to be used for data synthesis and analysis and as an archive of data collected during the studies. The database structure and format are a modification of that used by the National Geochemical Database (NGDB) because more than 85 percent of the data were originally retrieved from the NGDB (Smith and others, 2003). This tabular relational database contains field site and sample observations and laboratory analyses of samples collected at point locations.

Contents

Because of its 3.45 gigabyte size, the AGDB is contained in two linked databases, **AGDB.mdb** and **AGDB_Chem.mdb** (version 2003), or **AGDB.accdb** and **AGDB_Chem.accdb** (version 2010), to accommodate the MS Access limitation of 2 gigabytes for its databases. Data are contained in 39 tables, which are described in table 1. The three primary tables consist of quantitative analytical results, sample data, field site information, geologic data, and mineralogy data. From these three relational datasets, 33 analytical output data tables were created for various sections of data. Analytical method information and analytical method bibliography core-lookup tables provide needed reference for quantitative results, and a reference table of field name definitions was included to assist the user in understanding database field names and contents. In this report, names of tables and queries cited are in boldface; field names of tables and queries are italicized

Structure

Data are contained in three primary tables, **Geol, Chem**, and **Mnrlgy**, and relationships are defined to link these tables (fig. 2). This structure provides for efficient storage of information and for data verification. For example, all valid analytical results must also have corresponding sample information. Relationships between these tables are depicted as lines in figure 2. **Geol** is linked to **Chem** by including a common field (*LAB_ID*) in both tables. Therefore, a chemical value cannot exist without having a corresponding sample in **Geol**. This is a one-to-many relationship; that is, a single sample may have many analytical results (for example, different elements, same element by multiple methods, and so forth). **Geol** is also linked to **Mnrlgy** by including a common field (*LAB_ID*) in both tables in a one-to-one relationship; that is, every sample having mineralogical data also has geospatial and sample media data in **Geol**. Data may be extracted from the AGDB to meet specific user needs by constructing user-defined queries. Example queries are given in Appendix 2.

Relationships between **Geol** and other tables in the AGDB are shown in figure 2. **Geol** contains 264,095 records in 42 fields with information about the sample material collected at each site. Each analyzed sample has a unique *LAB_ID*, as well as a *FIELD_ID* that was provided by the sample collector. *LAB_ID* is a unique identifier assigned to each submitted sample by the analytical laboratory that received the sample. It is a key field that links the sample to its chemical and physical data found in **Chem**. The dates of sample submission and collection are stored in the *DATE_SUBMITTED* and *DATE_COLLECT* fields; however, less than 10 percent of all samples have a collection date recorded. *LATITUDE* and *LONGITUDE* contain the geographic coordinate data, whose precision is set at 0.00001 degree. This precision does not necessarily imply accuracy—in many cases, location information was collected by the sampler to only 3-digit precision by GPS, or recorded as dots of varying sizes on map scales as small as 1:250,000. Associated *SPHEROID* and *DATUM* information are sometimes provided (see Appendix 3, p. 33 for discussion of AGDB spheroid and datum issues). *PRIMARY_CLASS* defines the sample material type, while *SECONDARY_CLASS* and *SPECIFIC_NAME* provide more detailed information about the sample medium. Media type should be carefully noted when assessing data so that data from different sample types are not mistakenly equated. For example, the AGDB contains analyses for copper contained in multiple subsample media types (described in *SPECIFIC_NAME)* derived from one sediment sample site (for example, bulk sediment of various size-fractions, and their panned concentrate fractions of various magnetic susceptibility). Information regarding the collection and preparation of the sample may be found in *METHOD_COLLECTED*, *PREP*, and *MESH_PORE_SIZE*. Most of the entries in *LAB_ID* represent

Table 1. Alaska Geochemical Database (AGDB) tables.

Table name	Type	Description	Primary key field	Fields	Records
Geol[1]	Primary	Spatial, geologic and descriptive attributes for heavy mineral concentrate, mineral, rock, bulk sediment, and soil samples	LAB_ID	42	264,095
Chem[1]	Primary	Table of all chemical data compiled for heavy mineral concentrate, mineral, rock, bulk sediment, and soil samples	CHEM_ID	13	9,666,206
Mnrlgy	Primary	Mineralogy data for non-magnetic heavy mineral concentrate samples	LAB_ID	28	18,138
AnalyticMethod	Primary-lookup	Analytic methods used to obtain chemical and physical data	ANALYTIC_METHOD	4	85
AnalyticMethodBiblio	Primary-lookup	References for analytic methods used to obtain chemical data	ANALYTIC_METHOD_PUB_ID	9	753
HMC_Ag_Mg	Output	Chemical data—silver through magnesium—for heavy mineral concentrate samples	LAB_ID	143	48,096
HMC_Geol	Output	Spatial, geologic and descriptive attributes for heavy mineral concentrate samples	LAB_ID	33	48,096
HMC_Mn_Zr	Output	Chemical data—manganese through zirconium—for heavy mineral concentrate samples	LAB_ID	140	48,096
Mnrl_Ag_Ga	Output	Chemical data—silver through gallium—for mineral samples	LAB_ID	149	7,470
Mnrl_Gd_Pt	Output	Chemical data—silver through magnesium—for sediment samples	LAB_ID	149	7,470
Mnrl_Geol	Output	Spatial, geologic and descriptive attributes for mineral samples	LAB_ID	31	7,470
Mnrl_Rb_Zr	Output	Chemical data—rubidium through zirconium—for mineral samples	LAB_ID	147	7,470
Rock_E_Geol	Output	Spatial, geologic and descriptive attributes for rock samples located east of 147°W	LAB_ID	36	57,791
Rock_E_Majors	Output	Chemical "whole rock" data for rock samples located east of 147°W	LAB_ID	152	57,791
Rock_E_Traces_Ag_Cr	Output	Trace element data—silver through chromium—for rock samples located east of 147°W	LAB_ID	137	57,791
Rock_E_Traces_Cs_Na	Output	Trace element data—cesium through sodium—for rock samples located east of 147°W	LAB_ID	139	57,791
Rock_E_Traces_Nb_Sr	Output	Trace element data—niobium through strontium —for rock samples located east of 147°W	LAB_ID	145	57,791
Rock_E_Traces_Ta_Zr	Output	Trace element data - tantalum through zirconium —for rock samples located east of 147°W	LAB_ID	132	57,791
Rock_W_Geol	Output	Spatial, geologic and descriptive attributes for rock samples located west of 146.99999°W	LAB_ID	36	51,175
Rock_W_Majors	Output	Chemical "whole rock" data for rock samples located west of 146.99999°W	LAB_ID	152	51,175
Rock_W_Traces_Ag_Cr	Output	Trace element data—silver through chromium—for rock samples located west of 146.99999°W	LAB_ID	137	51,175
Rock_W_Traces_Cs_Na	Output	Trace element data—cesium through sodium— for rock samples located west of 146.99999°W	LAB_ID	139	51,175

Table 1. Alaska Geochemical Database (AGDB) tables.—Continued

Table name	Type	Description	Primary key field	Fields	Records
Rock_W_Traces_Nb_Sr	Output	Trace element data - niobium through strontium - for rock samples located west of 146.99999°W	LAB_ID	145	51,175
Rock_W_Traces_Ta_Zr	Output	Trace element data—tantalum through zirconium —for rock samples located west of 146.99999°W	LAB_ID	132	51,175
Sed_E_Ag_Cr	Output	Chemical data—silver through chromium—for bulk sediment samples located east of 147°W	LAB_ID	133	34,712
Sed_E_Cs_Mn	Output	Chemical data—cesium through manganese—for bulk sediment samples located east of 147°W	LAB_ID	132	34,712
Sed_E_Geol	Output	Spatial, geologic and descriptive attributes for bulk sediment samples located east of 147°W	LAB_ID	27	34,712
Sed_E_Mo_Se	Output	Chemical data— molybdenum through selenium —for bulk sediment samples located east of 147°W	LAB_ID	124	34,712
Sed_E_Si_Zr	Output	Chemical data—silicon through zirconium—for bulk sediment samples located east of 147°W	LAB_ID	127	34,712
Sed_W_Ag_Cr	Output	Chemical data—silver through chromium—for bulk sediment samples located west of 146.99999°W	LAB_ID	133	57,982
Sed_W_Cs_Mn	Output	Chemical data—cesium through manganese— for bulk sediment samples located west of 146.99999°W	LAB_ID	132	57,982
Sed_W_Geol	Output	Spatial, geologic and descriptive attributes for bulk sediment samples located west of 146.99999°W	LAB_ID	27	57,982
Sed_W_Mo_Se	Output	Chemical data—molybdenum through selenium —for bulk sediment samples located west of 146.99999°W	LAB_ID	124	57,982
Sed_W_Si_Zr	Output	Chemical data—silicon through zirconium—for bulk sediment samples located west of 146.99999°W	LAB_ID	127	57,982
Soil_Ag_Fe	Output	Chemical data—silver through iron—for soil samples	LAB_ID	153	6,869
Soil_Ga_Pt	Output	Chemical data—gallium through platinum—for soil samples	LAB_ID	150	6,869
Soil_Geol	Output	Spatial, geologic and descriptive attributes for soil samples	LAB_ID	35	6,869
Soil_Rb_Zr	Output	Chemical data—rubidium through zirconium— for soil samples	LAB_ID	142	6,869
FieldNameDictionary	Reference	Field name descriptions for all tables in the database	FIELD_NAME	5	974

[1] Table not included in Excel spreadsheet or ASCII text files.

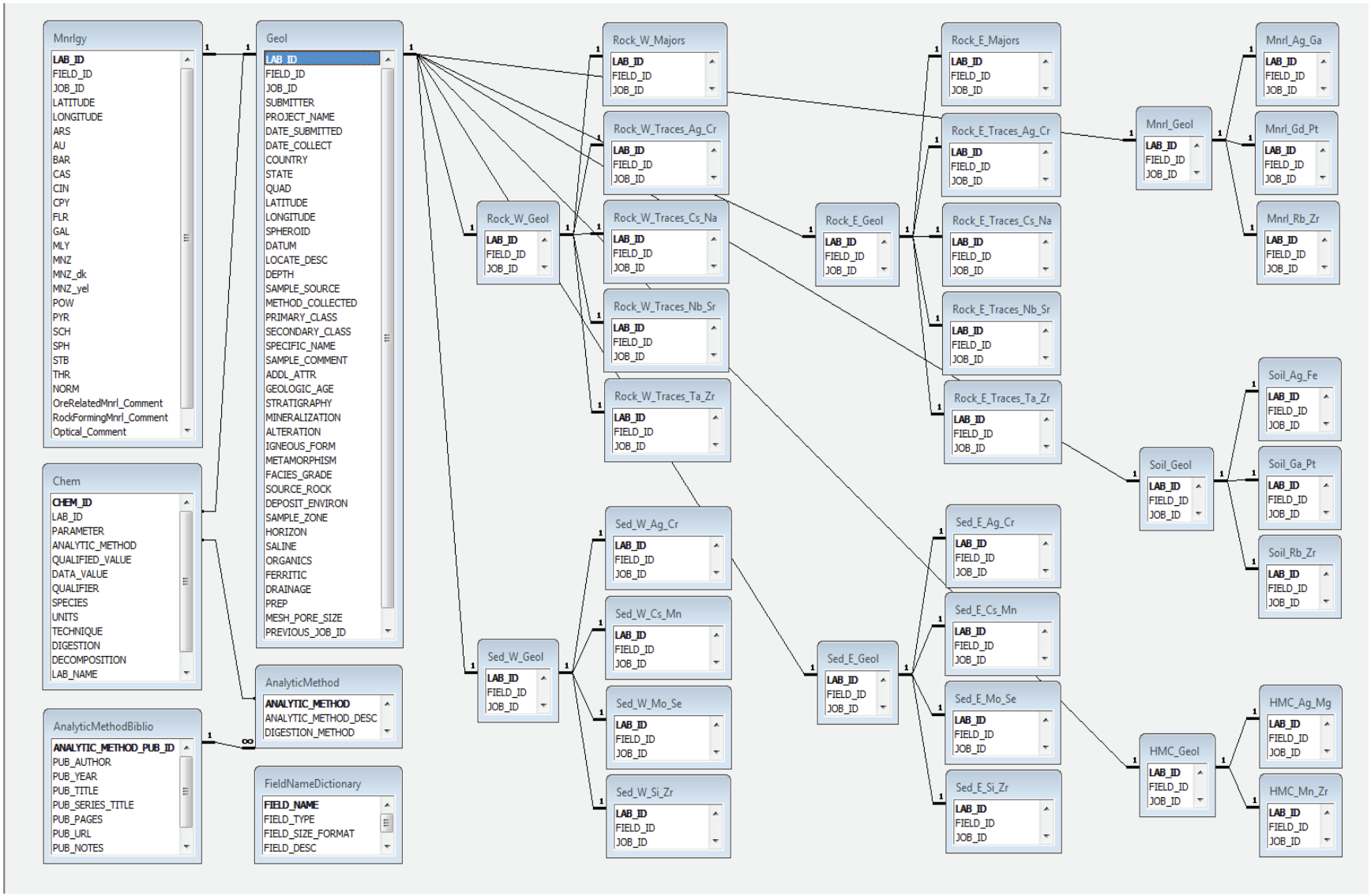


Figure 2. Table relationships in the Alaska Geochemical Database.

samples that were entered in the USGS laboratory information management system and whose data have been archived in the National Geochemical Database (NGDB) (Smith and others, 2003). Thus, the AGDB can be linked to data within the NGDB by using the unique *LAB_ID* field.

The **Chem** table contains 9,666,206 records in 13 fields with laboratory and field analytical measurements, expressed as numeric values. *CHEM_ID* is a unique identifier assigned to each measurement in the table and is a key field of software-assigned integers. Measurements in **Chem** consist of a numeric *DATA_VALUE* and an optional *QUALIFIER*, which is used to describe results such as those that were not detectable or that were estimates based on limits of instrumental detection (for example, "less than" values, such as < 2). *QUALIFIER* entries are "<" or "N," meaning that the element was not detected at concentrations above the lower limit of determination for the method; "L," meaning that the element was detected, but at concentrations below the lower limit of determination for the method; and ">" or "G," meaning that the element was measured at a concentration greater than the upper limit of determination for the method. *QUALIFIED_VALUE* was populated by combining the data in *DATA_VALUE* with its complement in *QUALIFIER*, according to the following conventions: *DATA_VALUE* entries that are accompanied by "<," "N," or "L" entries in *QUALIFIER* are represented in *QUALIFIED_VALUE* as negative numbers (for example, "-2"); and *DATA_VALUE* entries that are accompanied by ">" or "G" entries in *QUALIFIER* are represented in *QUALIFIED_VALUE* as values with 0.00111, 0.01111 or 0.11111 added to them (for example, >0.25 becomes 0.25111, >0.5 becomes 0.51111, and >10 becomes 10.11111).

Measurement characteristics such as units and analytical techniques are identified using a *PARAMETER* code, which is a concatenation of data from the fields *SPECIES*, *UNITS*, *TECHNIQUE*, and *DIGESTION*. There are 879 unique parameters in the AGDB. For example, the parameter "Sb_ppm_AA_P" represents the concentration of antimony, expressed in parts per million, as detected by atomic absorption spectrometry after a partial digestion (dissolution). *PARAMETER* is a 25-character-length field that contains a method summary that can be used as a column name in a data report or spreadsheet. *PARAMETER* entries are used as the field and column names of the output tables and are described in the **FieldNameDictionary** table. Information regarding the method of analysis or measurement used for each entry is found in *ANALYTIC_METHOD*, an abbreviated label in the **Chem** table that links to the **AnalyticMethod** table which is a look-up table that provides additional information on the 85 field and laboratory techniques used for analysis of the various geologic materials. **AnalyticMethod** includes a description of the analytical methods and relevant published references to them and is linked by *ANALYTIC_METHOD_PUB_ID* to bibliographic reference information in **AnalyticMethodBiblio**. Further information regarding methods of preparation for analysis is found in *DECOMPOSITION*. *LAB_NAME* provides information regarding the laboratory or work group responsible for the analysis. Relationships between **Chem** and other tables in the AGDB are shown in figure 2.

To facilitate ease of use when working in Access or when exporting data to Excel, data in **Geol** and **Chem** have been subdivided by sample media and analytical method. Seven "GeolData" output tables have been created from the AGDB. Because of the large number of samples, the rock and sediment datasets have been further subdivided by longitude 147 °W. in order to accommodate the 65,535 row limit of Excel (version 2003). Subdivision of the analytical data created 33 "ChemData" output tables. Each table contains a subset of analytical results for the analysis of a specific sample media subset. For example, **Rock_E_Majors** and **Rock_W_Majors** contain chemical and physical data from rock samples, collected east and west of long 147 °W., respectively, with major elements expressed as oxide concentrations for the "whole rock." Only in these two output tables are major elements presented as oxide concentrations. The **Chem** table contains the original major-element data as received from analytical laboratories. Due to the 256-field limit of Access and the 256-column limit of Excel (version 2003), these datasets have been further subdivided. For example, the trace-element data for rock samples are reported in eight output tables that separate the analyzed elements alphabetically (Ag to Zr) into groups. For sediment, soil, and heavy metal concentrate samples, the major elements are expressed as elemental concentrations. The concentration values in all output tables are in the *QUALIFIED_VALUE* format described previously. Sample location data in *LATITUDE* and *LONGITUDE* are also included in the "ChemData" tables to facilitate ease of use in geographic information system applications. Relationships between the output tables and **Geol** are shown in figure 2. The **FieldNameDictionary** table contains the field name, size, definition, and general data type of the 974 fields used in the AGDB tables as well as the names of tables in which these fields appear.

The **Mnrlgy** table contains 28 fields with optical mineralogical data for 18,138 nonmagnetic heavy-mineral concentrate samples. Fields are included for 18 potentially ore-related minerals, as well as the field *NORM* that indicates that no ore-related minerals were observed in the sample. All other data regarding ore-related minerals are found in *OreRelatedMnrl_Comment*, and data regarding rock-forming minerals are found in *RockFormingMnrl_Comment*. Sample location data are provided in the *LATITUDE* and *LONGITUDE* fields. The relationship between **Mnrlgy** and **Geol** is shown in figure 2. More than 75 percent of these data were generated by Richard B. Tripp from 1975–2009 in support of various USGS mineral resource assessment projects. Mineralogic data were originally recorded in various hardcopy data entry forms and according to different schemes. Mineral abundances were variously reported as: (1) "present" or "not present", (2) "abundant" or "moderate" or "trace" or "absent", (3) as a percentage or as a percentage range, or (4) as counted numbers of mineral particles. The field *Inferred_Comment* contains R. Tripp's comments based on chemical analyses and previous

USGS map publications, and that were written for the Federal Lands in Alaska–Geologic Studies project from 2007–2009 for samples that had been hand ground or consumed during chemical analysis.

Other Data Formats

To serve a wider audience of potential users of the AGDB, all of the Access tables except **Geol** and **Chem** have also been provided as 37 Excel spreadsheets. **Geol** and **Chem** were excluded because all of their data are presented in the 7 GeolData and 26 analytical ChemData output tables. Table 2 lists the spreadsheet files included in the data release. In addition, the Access tables have also been included as tab-delimited ASCII flat files that may be used by various types of applications.

Relational databases can be implemented using a variety of proprietary or nonproprietary software packages. The AGDB is attached to this report in proprietary (Microsoft Office Access 2003 and 2010) and nonproprietary (ASCII tab-delimited) formats. The spreadsheets are presented in Microsoft Office Excel 2003 format.

Database Queries

Queries are the primary mechanism for retrieving information from a database, and consist of requests that are issued to the database in a pre-defined format. Within relational database software packages, queries may be constructed and saved to retrieve data using user-defined criteria. This AGDB publication contains several examples of Access queries that aid the user in viewing and extracting selected datasets. These queries are: (1) select queries that extract data from one or more tables or queries of tables and display the data in tabular form, (2) summary queries that create groups of extracted data from one or more tables or queries of tables and display the data in tabular form, and (3) crosstab queries that summarize data from one or more tables or queries of tables in the form of a spreadsheet. The graphical Query Design Views of these queries are translated into Structured Query Language (SQL) statements that are displayed within the database by clicking on the SQL View of the View box of the Query. Several examples of three query types are presented in Appendix 2.

Data Enhancement, Correction, and Processing

AGDB data retrieved from the NGDB were generated by the analytical laboratories of the USGS and its contractors over a number of decades, beginning in the early 1960s. Upon completion of the chemical analyses, the data were stored in the RASS database (1963 to 1987), the PLUTO database (1979 to 1997), or the commercial LIMS used by USGS analytical laboratories (1996 to present). Beginning in 2002, data from all three sources were combined, reformatted, and standardized into the Oracle-based NGDB. The NGDB is composed of three data tables: **GEO** consisting of sample site, collection, and description information; **CHEM** consisting of sample analysis information; **JOB** consisting of sample submitter information. All three tables were queried to produce most of the AGDB.

To create the AGDB, rock, sediment, soil, and heavy-mineral concentrate sample data were retrieved from the NGDB by using the following criteria: (1) each sample must have a valid and unique laboratory identification number; (2) each sample must have latitude and longitude coordinates that plot within the geographic boundaries of the AGDB (see p. 5); and (3) each sample must be identified as one of these four geologic materials. This dataset was then examined to remove any samples that could be identified as a processed derivative of these media. Rock coatings, partial digestions, leachates, experimental or artificial samples, and some misidentified samples were thereby eliminated. A multi-year effort was made to correct erroneous or incomplete attributes. Several standardized sample descriptive fields were more completely populated using information previously found only in the comment field of PLUTO, the free coding fields of RASS, or the paper sample submittal forms. The geologic material sample-site location and descriptive data were compiled into a single table called **Geol**. Through this process, more than 10,000 samples from Alaska have been added to the NGDB, and tens of thousands more have more accurate locality coordinates as a result of the AGDB data cleanup effort. A detailed explanation of the data enhancement, correction, and processing plan that was used to produce the AGDB is presented in Appendix 3.

Acknowledgments

The authors thank David B. Smith and Carl Abston for their leadership and efforts for the Minerals Data Information Rescue in Alaska (MDIRA) Program. We also thank Dorothy M. Trujillo for her assistance in providing data from the NGDB. Many thanks go to Jill Schneider for making the Alaska Technical Data Unit so accessible, and to Theresa A. Taylor for her aid in the digital scanning and georeferencing of sample locality base maps. Their assistance and encouragement throughout the life of this project is much appreciated. The authors acknowledge Gregory K. Lee and LaDonna M. Choate for their technical review of the manuscript and database; their expertise and assistance were invaluable in improving the work.

Table 2. List of spreadsheets containing Alaska Geochemical Database (AGDB) data.

Spreadsheet name	Information contained in spreadsheet
AnalyticMethod.xls	Analytical methods used to obtain chemical and physical data
AnalyticMethodBiblio.xls	References for analytical methods used to obtain chemical data
FieldNameDictionary.xls	Explanation of field names used for all tables in the database
HMC_Ag_Mg.xls	Chemical data—silver through magnesium—for heavy-mineral concentrate samples
HMC_Geol.xls	Spatial, geologic and descriptive attributes for heavy-mineral concentrate samples
HMC_Mn_Zr.xls	Chemical data—manganese through zirconium—for heavy-mineral concentrate samples
Mnrl_Ag_Ga.xls	Chemical data—silver through gallium—for mineral samples
Mnrl_Gd_Pt.xls	Chemical data—gadolinium through platinum—for mineral samples
Mnrl_Geol.xls	Spatial, geologic and descriptive attributes for mineral samples
Mnrl_Rb_Zr.xls	Chemical data—rubidium through zirconium—for mineral samples
Mnrlgy.xls	Mineralogy data for non-magnetic heavy-mineral concentrate samples
Rock_E_Geol.xls	Spatial, geologic and descriptive attributes for rock samples located east of 147°W
Rock_E_Majors.xls	Chemical "whole rock" data for rock samples located east of 147°W
Rock_E_Traces_Ag_Cr.xls	Trace element data—silver through chromium —for rock samples located east of 147°W
Rock_E_Traces_Cs_Na.xls	Trace element data—cesium through sodium—for rock samples located east of 147°W
Rock_E_Traces_Nb_Sr.xls	Trace element data—niobium through strontium—for rock samples located east of 147°W
Rock_E_Traces_Ta_Zr.xls	Trace element data—tantalum through zirconium—for rock samples located east of 147°W
Rock_W_Geol.xls	Spatial, geologic and descriptive attributes for rock samples located west of 146.99999°W
Rock_W_Majors.xls	Chemical "whole rock" data for rock samples located west of 146.99999°W
Rock_W_Traces_Ag_Cr.xls	Trace element data—silver through chromium—for rock samples located west of 146.99999°W
Rock_W_Traces_Cs_Na.xls	Trace element data—cesium through sodium—for rock samples located west of 146.99999°W
Rock_W_Traces_Nb_Sr.xls	Trace element data—niobium through strontium—for rock samples located west of 146.99999°W
Rock_W_Traces_Ta_Zr.xls	Trace element data—tantalum through zirconium—for rock samples located west of 146.99999°W
Sed_E_Ag_Cr.xls	Chemical data—silver through chromium—for bulk sediment samples located east of 147°W
Sed_E_Cs_Mn.xls	Chemical data—cesium through manganese—for bulk sediment samples located east of 147°W
Sed_E_Geol.xls	Spatial, geologic and descriptive attributes for bulk sediment samples located east of 147°W
Sed_E_Mo_Se.xls	Chemical data—molybdenum through selenium—for bulk sediment samples located east of 147°W
Sed_E_Si_Zr.xls	Chemical data—silicon through zirconium—for bulk sediment samples located east of 147°W
Sed_W_Ag_Cr.xls	Chemical data—silver through chromium—for bulk sediment samples located west of 146.99999°W
Sed_W_Cs_Mn.xls	Chemical data—cesium through manganese—for bulk sediment samples located west of 146.99999°W
Sed_W_Geol.xls	Spatial, geologic and descriptive attributes for bulk sediment samples located west of 146.99999°W
Sed_W_Mo_Se.xls	Chemical data—molybdenum through selenium—for bulk sediment samples located west of 146.99999°W
Sed_W_Si_Zr.xls	Chemical data—silicon through zirconium—for bulk sediment samples located west of 146.99999°W
Soil_Ag_Fe.xls	Chemical data—silver through iron—for soil samples
Soil_Ga_Pt.xls	Chemical data—gallium through platinum—for soil samples
Soil_Geol.xls	Spatial, geologic and descriptive attributes for soil samples
Soil_Rb_Zr.xls	Chemical data—rubidium through zirconium—for soil samples

References

Database References

Within the database, references are cited for analytical methods that were used to determine elemental concentrations in the **Chem** table. Information regarding the method of analysis or measurement used to obtain data is found in **AnalyticMethodBiblio** (table 1). Refer to the *ANALYTIC_METHOD_PUB_ID*—in most cases the USGS Library call number—when researching the analytical method in question.

References Cited

Arbogast, B.F., 1990, Quality assurance manual for the Branch of Geochemistry, U.S. Geological Survey: U.S. Geological Survey Open-File Report 90–668, 184 p.

Arbogast, B.F., ed., 1996, Analytical methods manual for the Mineral Resource Surveys Program, U.S. Geological Survey: U.S. Geological Survey Open-File Report 96–525, 248 p.

Bailey, E.A., Lee, G.K., Mueller, S.H., Wang, Bronwen, Brown, Z.A., and Beischer, G.A., 2007, Major and trace-element data from stream-sediment and rock samples collected in the Taylor Mountains 1:250,000-scale quadrangle, Alaska: U.S. Geological Survey Open-File Report 2007–1196, 7 p. pamphlet and data files, v. 1.1, accessed Jan. 27, 2011, at *http://pubs.usgs.gov/of/2007/1196/*.

Bailey, E.A., Shew, N.E., Labay, K.A., Schmidt, J.M., O'Leary, R.M., and Detra, D. E., 2010, Stream-sediment samples reanalyzed for major, rare-earth and trace elements from ten 1:250,000-scale quadrangles, south-central Alaska, 2007–2008, U.S. Geological Survey Open-File Report 2010–1147, 6 p. pamphlet, v. 1.0, accessed Jan. 28, 2011, at *http://pubs.usgs.gov/of/2010/1147/*.

Bailey, E.A., Smith, D.B., Abston, C.C., Granitto, Matthew, and Burleigh, K.A., 1999, National Geochemical Database, U.S. Geological Survey RASS (Rock Analysis Storage System) geochemical data for Alaska: U.S. Geological Survey Open-File Report 99–433, v. 2.0, accessed on Jan. 26, 2011 at *http://pubs.usgs.gov/of/1999/of99-433/*.

Gamble, B.M., Bailey, E.A., Schmidt, J.M., Shew, N.B., Labay, K.A., ,Granitto, Matthew, O'Leary, R.M., and Detra, D.E., 2010, Stream sediment samples reanalyzed for major, rare earth and trace elements from seven 1:250,000-scale quadrangles, south central Alaska, 2007-2009: U.S. Geological Survey Data Series 532, 4 p. pamphlet, accessed on Jan. 26, 2011 at *http://pubs.usgs.gov/ds/532/*.

Klimasauskas, E.P., Miller, M.L., and Bradley, D.C., 2007, Major- and trace-element concentrations in rock samples collected in 2006 from the Taylor Mountains 1:250,000-scale quadrangle, Alaska: U.S. Geological Survey Open-File Report 2007–1386, 6 p. pamphlet, v. 1.0, accessed Jan. 27, 2011, at *http://pubs.usgs.gov/of/2007/1386/*.

Klimasauskas, E.P., Miller, M.L., Bradley, D.C., Bundtzen, T.K., and Hudson, T.L., 2006, Major- and trace-element concentrations in rock samples collected in 2005 from the Taylor Mountains 1:250,000-scale quadrangle, Alaska: U.S. Geological Survey Open-File Report 2006–1306, 8 p. pamphlet, v. 1.0, accessed January 27, 2011, at *http://pubs.usgs.gov/of/2006/1306/*.

Klimasauskas, E.P., Miller, M.L., Bradley, D.C., Karl, S.M., Baichtal, J.F., and Blodgett, R.B., 2006, Major- and trace-element concentrations in rock samples collected in 2004 from the Taylor Mountains 1:250,000-scale quadrangle, Alaska: U.S. Geological Survey Open-File Report 2006–1070, 8 p. pamphlet, v. 1.1, accessed January 27, 2011, at *http://pubs.usgs.gov/of/2006/1070/*.

Miesch, A.T., 1976, Sampling designs for geochemical surveys—Syllabus for a short course: U.S. Geological Survey Open-File Report 76–772, 128 p.

Neuendorf, K.K.E., Mehl, J.P., Jr., and Jackson, J.A., eds., 2005, Glossary of geology (5th ed.): Alexandria, Va., American Geological Institute, 779 p.

Smith, S.M., 1997, National Geochemical Database reformatted data from the National Uranium Resource Evaluation (NURE) Hydrogeochemical and Stream Sediment Reconnaissance (HSSR) Program: U.S. Geological Survey Open-File Report 97–492, v. 1.4, accessed Jan.27, 2011, at *http://pubs.usgs.gov/of/1997/ofr-97-0492/*.

Smith, S.M., Smith, D.B., and Bailey, E.A., 2003, The U.S. Geological Survey's National Geochemical Database [abs.]; Geological Society of America Abstracts with Program, v. 35, no. 7, p. 366.

Streckeisen, A.L., 1967, Classification and nomenclature of igneous rocks (final report of an inquiry): Neues Jahrbuch für Mineralogie, Abhandlungen, v. 107, no. 2, p. 144–214.

Streckeisen, A.L., 1976, To each plutonic rock its proper name: Earth Science Review, v. 12, p. 1–33.

Taggart, J.E., ed., 2002, Analytical methods for chemical analysis of geologic and other materials, U.S. Geological Survey: U.S. Geological Survey Open-File Report 2002–0223, v. 5.0, accessed Jan. 27, 2011, at *http://pubs.usgs.gov/of/2002/ofr-02-0223/*.

U.S. Geological Survey, 2004, The National Geochemical Survey, database and documentation: U.S. Geological Survey Open-File Report 2004–1001, v. 5.0, accessed January 27, 2011, at *http://pubs.usgs.gov/of/2004/1001/*.

Appendix 1

Analytical Methods

Appendix 1 contains a table of analytical method names and descriptions of the analytical techniques that provided the chemical data of the Alaska Geochemical database.

Table A1–1. Analytical methods.

Analytic Method	Description
AA_CV	mercury by cold-vapor atomic absorption spectrometry after multi-acid digestion and solution.
AA_F_AZ_Fuse	silver, arsenic, bismuth, cadmium, copper, lead, antimony and zinc by flame atomic absorption spectrometry after partial digestion by roast, $K_2S_2O_7$ fusion, HCl-KI, and selective organic extraction with 336-MIBK.
AA_F_AZ_H2O2	silver, arsenic, bismuth, cadmium, copper, lead, antimony and zinc by flame atomic absorption spectrometry after partial digestion with HCl-H_2O_2 and selective organic extraction with 336-MIBK.
AA_F_AZ_HCl	silver, arsenic, bismuth, cadmium, copper, lead, antimony and zinc by flame atomic absorption spectrometry after partial digestion by roast, HCl-KI, and selective organic extraction with 336-MIBK.
AA_F_CX	calcium, magnesium, sodium, potassium and cation exchange capability in soil by flame atomic absorption spectrometry after solution extraction and cation exchange.
AA_F_DTPA	cadmium, cobalt, copper, iron, manganese, nickel, lead and zinc by flame atomic absorption spectrometry after DTPA extraction and cation exchange.
AA_F_Fuse	major and minor elements by flame atomic absorption spectrometry after fusion digestion.
AA_F_Fuse_P	molybdenum and antimony by flame atomic absorption spectrometry after $K_2S_2O_7$ fusion, partial acid digestion, and selective organic extraction with 336-MIBK.
AA_F_H2O	calcium, magnesium, manganese and arsenic in saturation paste of soil by flame atomic absorption spectrometry after solution extraction.
AA_F_HBr	silver, gold and tellurium by flame atomic absorption spectrometry after HBr-Br_2 digestion and selective organic extraction with 336-MIBK.
AA_F_HCl	copper and manganese by flame atomic absorption spectrometry after partial digestion with HCl.
AA_F_HCl_OE	antimony by flame atomic absorption spectrometry after partial digestion with HCl and selective organic extraction with 336-MIBK.
AA_F_HF	major and minor elements by flame atomic absorption spectrometry after multi-acid digestion with HF.
AA_F_HNO3	silver, cadmium, copper, lead and zinc by flame atomic absorption spectrometry after partial digestion with hot HNO_3.
AA_FE	sodium and potassium by flame emission spectrometry (flame photometry) after HF-$HClO_4$ dissolution or $LiBO_2$ fusion.
AA_GF_HBr	gold and tellurium by graphite furnace atomic absorption spectrometry after HBr-Br_2 digestion and selective organic extraction with 336-MIBK.
AA_GF_HF	arsenic, gold, bismuth and tellurium by graphite furnace atomic absorption spectrometry after multi-acid digestion with HF and selective organic extraction with 336-MIBK.
AA_GF_ST	thallium by graphite furnace atomic absorption spectrometry after Na_2O_2 sinter, HCl-HNO_3 dissolution, and selective organic extraction with DIBK.
AA_HG_Acid	selenium by flow injection or continuous flow hydride generation-atomic absorption spectrometry after digestion with HNO_3-HCl-H_2SO_4-$KMnO_4$.
AA_HG_HF	arsenic, antimony, selenium and tellurium by flow injection or continuous flow hydride generation-atomic absorption spectrometry after multi-acid digestion with HF.
AA_HG_ST	arsenic and antimony by flow injection or continuous flow hydride generation-atomic absorption spectrometry after Na_2O_2 sinter digestion.

Table A1–1. Analytical methods.—Continued

Analytic Method	Description
AA_TR	mercury by thermal release and atomic absorption spectrometry after multi-acid digestion (Vaughn-McCarthy method).
AA_TR_P	mercury by thermal release and atomic absorption spectrometry after multi-acid digestion (Vaughn-McCarthy method) and use of a willemite screen.
AES_Acid_P	major and minor elements by inductively coupled plasma-atomic emission spectrometry after unknown partial acid digestion.
AES_AR	major and minor elements by inductively coupled plasma-atomic emission spectrometry after partial digestion with aqua regia.
AES_AZ	silver, arsenic, gold, bismuth, cadmium, copper, molybdenum, lead, antimony and zinc by inductively coupled plasma-atomic emission spectrometry after partial digestion with HCl-H_2O_2.
AES_Fuse	major and minor elements by inductively coupled plasma-atomic emission spectrometry after fusion digestion.
AES_HF	major and minor elements by inductively coupled plasma-atomic emission spectrometry after digestion with HF-HCl-HNO_3-$HClO_4$.
AES_HF_REE	rare earth elements by ion exchange and inductively coupled plasma-atomic emission quantitative spectrometry after HF-HCl-HNO_3-$HClO_4$ digestion.
AES_IE	molybdenum, niobium and tungsten by inductively coupled plasma-atomic emission quantitative spectrometry after HF-HCl-HNO_3-$HClO_4$ digestion and ion exchange separation.
AES_ST	major and minor elements by inductively coupled plasma-atomic emission spectrometry after sinter digestion.
AFS_CV	mercury in aqueous media by flow injection-cold vapor-atomic fluorescence spectrometry.
CB_CHN	carbon, hydrogen and nitrogen by gas chromatography/thermal conductivity (CHN elemental) analyzer after combustion.
CB_IRC	carbon and sulfur by infrared detection after combustion.
CB_TC	total carbon and organic carbon by thermal conductivity detection after combustion.
CB_TT	sulfur by iodometric titration after combustion.
CD	specific conductance by standard method conductivity electrode.
CM_Acid	bromine by colorimetry after acid digestion.
CM_As	arsenic by modified Gutzeit apparatus confined-spot method colorimetry after partial digestion in KOH-HCl and chemical separation.
CM_Cl	chloride by colorimetric spectrophotometry after Na_2CO_3 and ZnO sinter digestion.
CM_cxHM	heavy metal elements by colorimetry after partial extraction in aqueous ammonium citrate solution.
CM_F	fluorine by colorimetric spectrophotometry after acid digestion and chemical separation.
CM_Fuse	major and minor elements by colorimetric spectrophotometry after fusion digestion.
CM_Fuse_P	molybdenum and antimony by colorimetry after partial digestion by KOH fusion (Mo) or KOH fusion-HCl digestion (Sb, rhodamine B).
CM_H2O	sulfate in saturation paste of soil by colorimetric titration after solution extraction.
CM_HF	major and minor elements by colorimetric spectrophotometry after multi-acid digestion with HF.
CM_HNO3	copper, lead and zinc by colorimetry after partial digestion with HNO_3.
CM_PC	uranium by paper chromatography after partial digestion with HNO_3.
CM_ST_P	tungsten by colorimetry after partial digestion with carbonate sinter.
CP	organic carbon, carbonate carbon and totals by computation.
DN	uranium and thorium by delayed neutron activation counting.
EDX	minor elements by energy-dispersive X-ray fluorescence spectrometry.
ES_H2O	boron by semi-quantitative emission spectrography after solution extraction.
ES_Q	major and minor elements by quantitative emission spectrography.
ES_SQ	major and minor elements by semi-quantitative emission spectrography.
FA_AA	gold, silver and platinum group elements by graphite furnace atomic absorption spectrometry after PbO fire assay chemical separation.

Table A1–1. Analytical methods.—Continued

Analytic Method	Description
FA_DC	gold by direct current plasma-atomic emission spectroscopy or atomic absorption spectrophotometry after PbO fire assay chemical separation.
FA_ES	gold and platinum group elements by direct-current arc quantitative emission spectrography after PbO fire assay chemical separation.
FA_MS	platinum group elements by inductively coupled plasma-mass spectrometry after NiS fire assay chemical separation.
FL_HF	beryllium, tin and uranium by fluorometry after multi-acid digestion with HF.
FL_HNO3	selenium by fluorometry after digestion with HNO_3-H_3PO_4.
GRC	uranium by gamma counting.
GV	density, moisture and weight by gravimetry; ash or loss on ignition by weight loss after heating at 900 °C.
GV_Acid	major and minor elements by gravimetry after acid digestion.
GV_CR	major and minor elements by gravimetry for Classical Rock Analysis after unknown digestion method.
GV_Flux	moisture, bound water and total water by heating and weight loss with flux.
GV_Fuse	major and minor elements by gravimetry after fusion digestion.
IC	chloride, fluoride, nitrate, sulfate and phosphate by ion specific chromatography.
ISE_Fuse	chloride, fluoride and iodide by ion specific electrode after fusion digestion.
ISE_H2O	chloride by ion specific electrode after solution extraction.
ISE_HF	chloride by ion specific electrode after multi-acid digestion with HF.
MS_AR	major and minor elements by inductively coupled plasma-mass spectrometry after partial digestion with aqua regia.
MS_HF	major and minor elements by inductively coupled plasma-mass spectrometry after HF-HCl-HNO_3-$HClO_4$ digestion.
MS_ST	major and minor elements by inductively coupled plasma-mass spectrometry after Na_2O_2 sinter digestion.
MS_ST_REE	rare earth elements by inductively coupled plasma-mass spectrometry after Na_2O_2 sinter digestion.
NA	major and minor elements by instrumental neutron activation analysis.
pH	pH by standard method combination pH electrode.
TB_AR	acid-soluble sulfate, sulfur and sulfide by turbidimetry after aqua regia digestion.
TT_Flux	total water by Karl Fischer coulometric titration with flux after combustion.
TT_Fuse	Fe_2O_3 by titration after fusion, decomposition and precipitation.
TT_HCl	carbonate carbon and carbon dioxide (acid soluble carbon) by coulometric titration after $HClO_4$ digestion and extraction.
TT_HF	ferrous oxide by colorimetric or potentiometric titration after HF-H_2SO_4 digestion.
VOL	carbon dioxide or carbonate carbon by evolution after acid decomposition; aka "gasometric" or "manometric".
WDX_Fuse	major and minor elements by wavelength-dispersive X-ray fluorescence spectrometry after $LiBO_2$ fusion digestion.
WDX_Raw	chlorine, iodine and bromine by wavelength-dispersive X-ray fluorescence spectrometry on raw sample.

Appendix 2

Database Query Examples

We provide here several database queries designed to provide data commonly needed by researchers interested in geochemical analyses of geologic materials. These queries are intended as a guide for AGDB users, but the database is amenable to all types of complex queries.

Summary Queries

Summary queries create groups of extracted data from one or more tables or queries of tables and display the data in tabular form. These queries calculate a sum, average, count, or other type of total on records, and then group the information. The query **qsumSampleType** was created to display the various types and subtypes of sample media that were collected and analyzed. This query displays data from the *PRIMARY_CLASS*, *SECONDARY_CLASS*, and *SPECIFIC_NAME* fields of the **Geol** table, while providing the number of samples in each Specific Name subtype (fig. A2–1). These data are sorted by fields in the order listed above. When the query is run, the Datasheet View shows 426 different *PRIMARY_CLASS-SECONDARY_CLASS-SPECIFIC_NAME* combinations:10 for the *PRIMARY_CLASS* category "concentrate," 82 for the *PRIMARY_CLASS* category "mineral," 168 for the *SECONDARY_CLASS* category "igneous rock," 114 for the *SECONDARY_CLASS* category "metamorphic rock," and 61 for the *SECONDARY_CLASS* category "sedimentary rock." *PRIMARY_CLASS* categories "sediment" and "soil" do not have subtypes in *SECONDARY_CLASS* and *SPECIFIC_NAME*.

Another summary query example, **qsumAnalyticMethod,** was created to provide a breakdown of all chemical parameters and their analytical methods that are in the AGDB. This query displays data from *ANALYTIC_METHOD* and *ANALYTIC_METHOD_DESC* in **AnalyticMethod** and *PARAMETER* in **Chem** and provides a count of the number of chemical determinations for each parameter (fig. A2–2). These data are sorted first by *ANALYTIC*_METHOD and then by *PARAMETER*. When the query is run, the Datasheet View shows 994 unique *ANALYTIC_METHOD-PARAMETER* combinations.

Select Queries

Select queries extract data from one or more tables or queries of tables and display the data in tabular form. Two

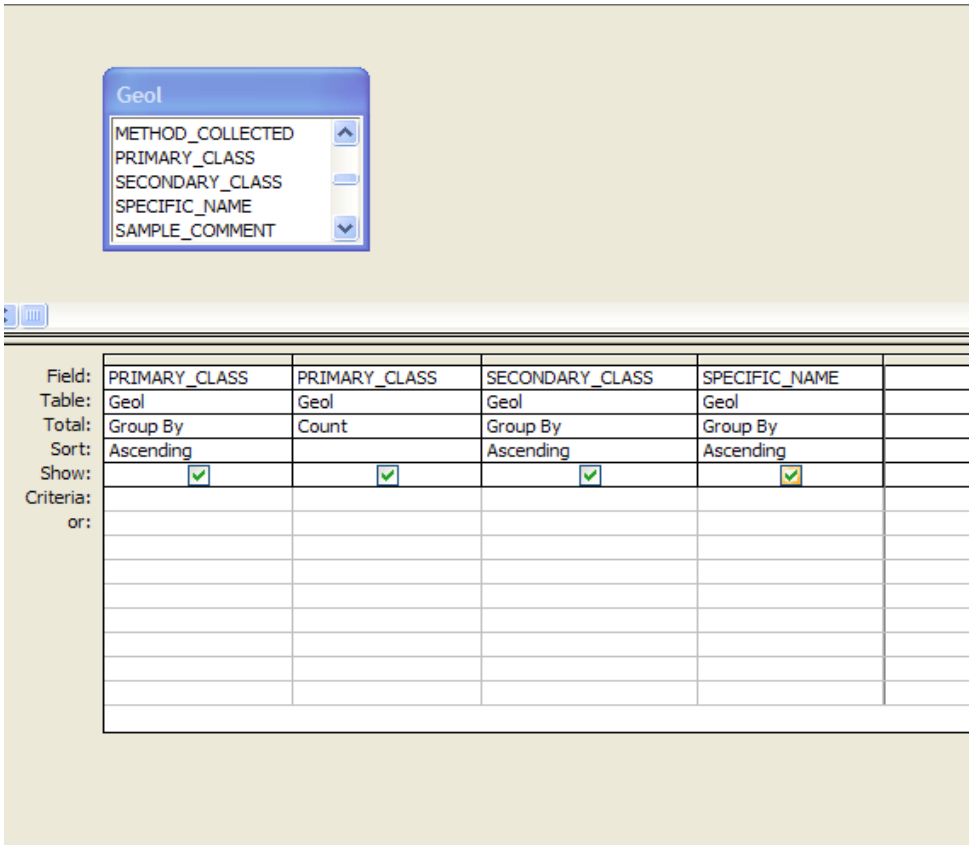


Figure A2–1. Summary query **qsumSampleType** in Query Design View.

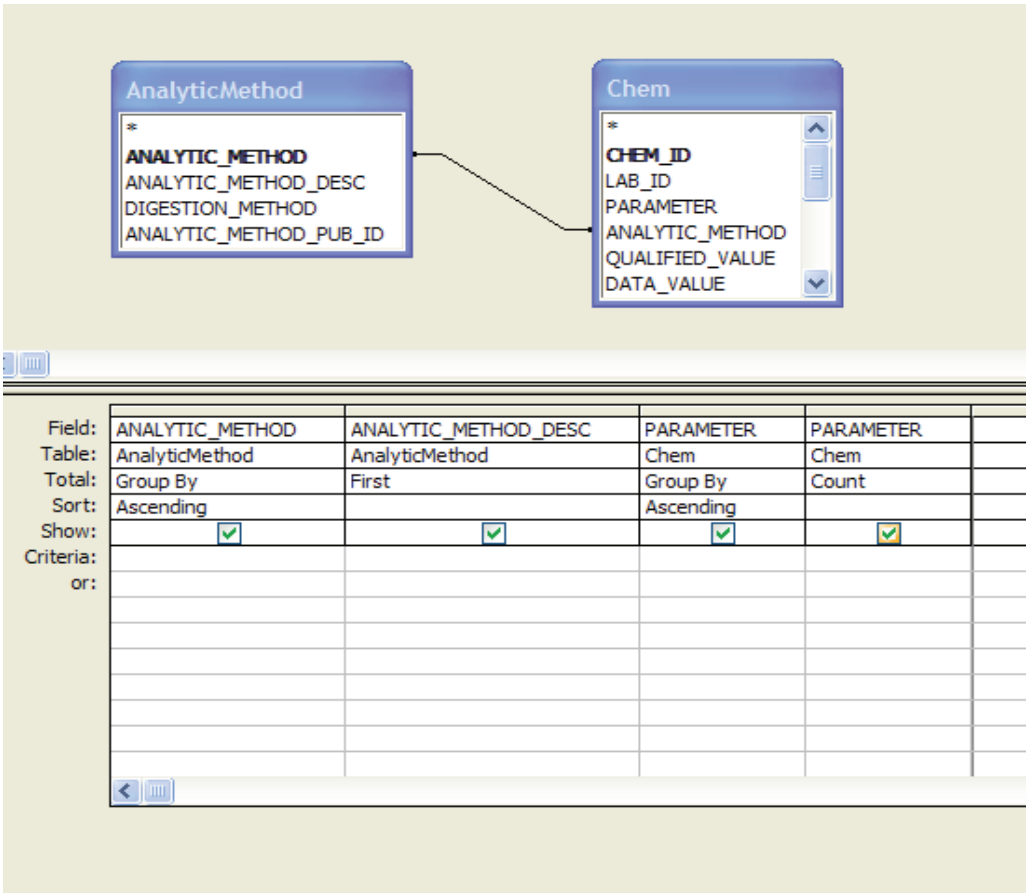


Figure A2–2. Summary query **qsumAnalyticMethod** in Query Design View.

examples of select queries were constructed to select specific data from the nonmagnetic separates (also know as the C3 fraction) of heavy-mineral concentrate samples in the database. First, **qselC3-Chem_Au** was created to show all of the C3 samples that were analyzed for gold (Au), among other elements (fig. A2–3).

This query combines data from two tables: **Geol** (containing *LAB_ID*, *FIELD_ID*, *PROJECT_NAME*, *LATITUDE*, *LONGITUDE*, *DATE_SUBMITTED*, *PRIMARY_CLASS*, and *SPECIFIC_NAME* data) and **Chem** (containing *PARAMETER* and *QUALIFIED_VALUE* data). The tables are linked by the common field *LAB_ID*. A close look at the "Criteria" line of the Query Design View shows the conditions placed on this query: the phrase "C3 fraction*" for *SPECIFIC_NAME* of **Geol** identifies C3 samples; the phrase "Au_ppm_*" for *PARAMETER* of **Chem** identifies gold determinations determined by a variety of analytical methods. The data will be sorted by *FIELD_ID*. When the query is run, the Datasheet View shows 32,459 C3 samples that were analyzed for gold. Query **qselC3-Chem_Au** can be modified as the phrase ">0" is placed as a search condition in the "Criteria" line for *QUALIFIED_VALUE*, and from this we can learn that only 1,833 C3 samples in the AGDB—close to five percent—have Au values greater than the lower limits of detection for the analytical method used to yield those values.

In much the same manner, a second select query, **qselC3-Mnrlgy-Chem_Au,** was created to show all of the C3 samples that were analyzed for gold among other elements (fig. A2–4) and that also have mineralogical determinations.

This query combines data from select query **qselC3-Chem_Au** (containing *LAB_ID*, *FIELD_ID*, *PROJECT_NAME*, *LATITUDE*, *LONGITUDE*, *DATE_SUBMITTED*, *PRIMARY_CLASS*, and *SPECIFIC_NAME* data from **Geol**, and *PARAMETER* and *QUALIFIED_VALUE* data from **Chem**) and from table **Mnrlgy** (containing fields *Au* and *NORM*). The tables are linked by the common field *LAB_ID*. When the query is run, the Datasheet View shows 15,468 C3 samples that were analyzed for gold and also have mineralogical determinations. The query is set up so that gold mineralogical data is referenced but it could be modified so that all mineralogical data is retrieved.

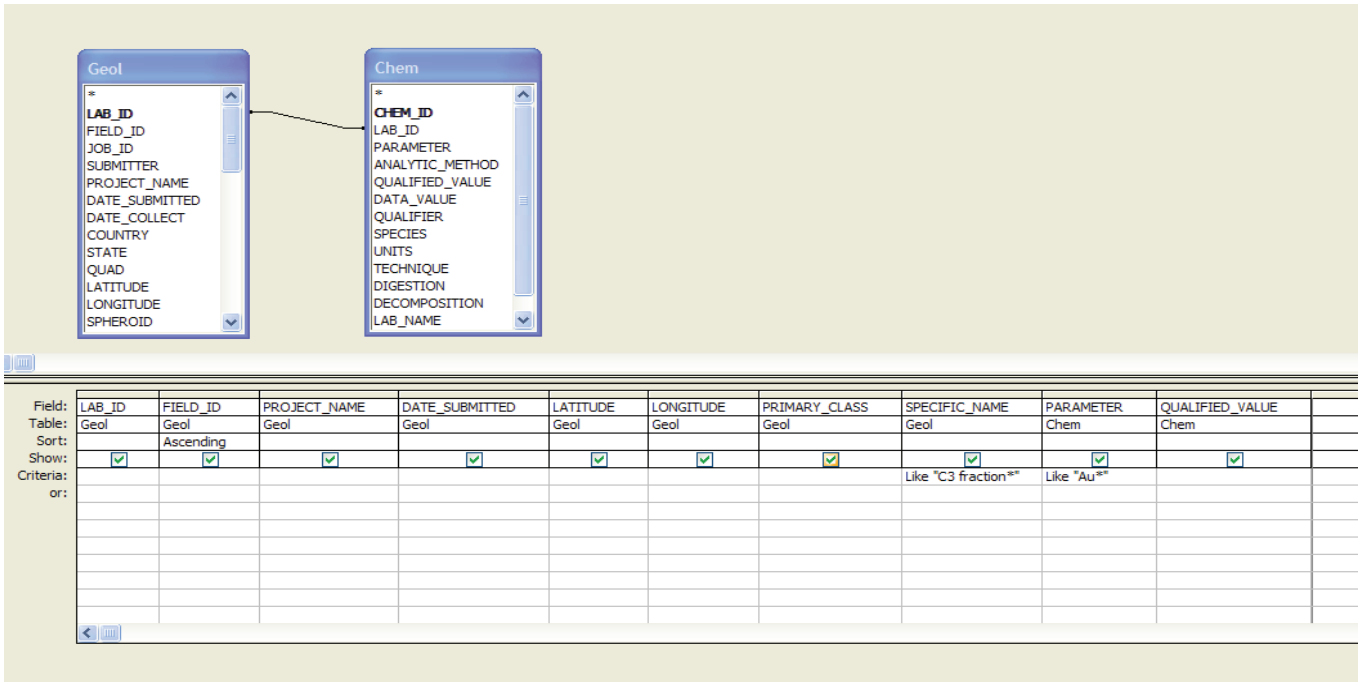


Figure A2–3. Select query **qselC3-Chem_Au** in Query Design View.

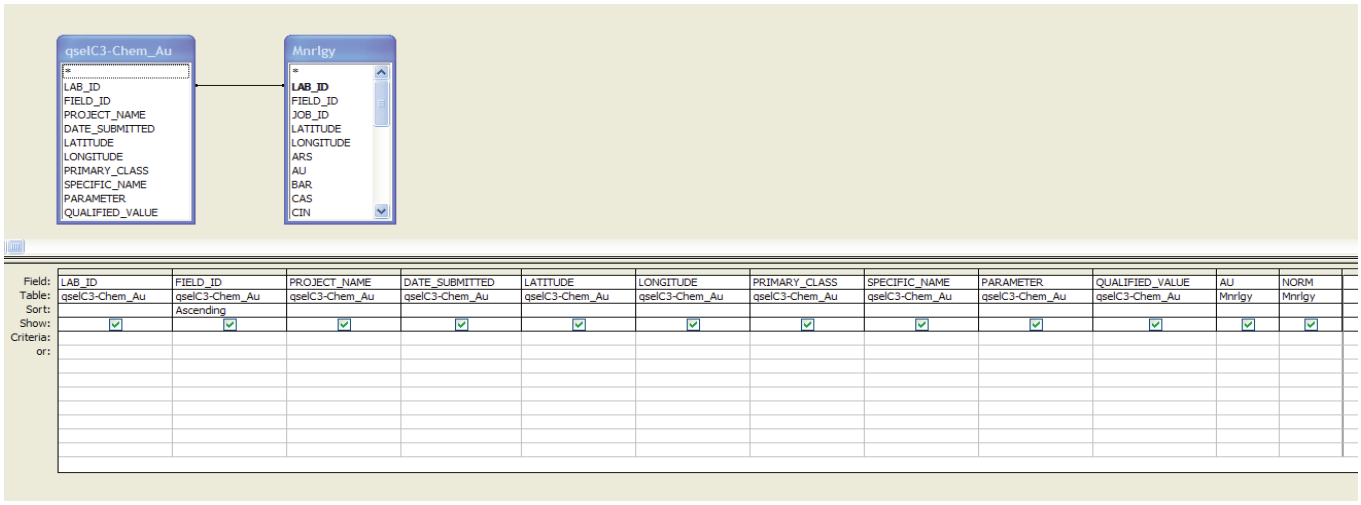


Figure A2–4. Select query **qselC3-Mnrlgy-Chem_Au** in Query Design View.

Crosstab Query

Crosstab queries summarize data from one or more tables or queries of tables in the form of a spreadsheet. The crosstab query **qctabQuad_SampleType** was constructed using the **Geol** table to show the counts of sample types found in 1:250,000-scale quadrangles of the AGDB (fig A2–5).

This crosstab query uses fields *COUNTRY* and *QUAD* as row headings, *PRIMARY_CLASS* is the column heading, and the cells in each column are filled by the *PRIMARY_CLASS* counts of each sample media type. These data are sorted by fields in the order listed above. When the query is run, the Datasheet View shows 156 row entries, and the first four entries show the number and types of samples that were collected from the seafloor beneath marine bodies of water listed in *COUNTRY*, indicating that these samples are outside of quadrangle boundaries but are yet included in the AGDB. Of the 153 1:250,000 scale quadrangles in Alaska, only Kiska has no samples in the AGDB.

Note that the select query **qselC3-Chem_Au** uses the *QUALIFIED_VALUE* field which combines the determined value in *DATA_VALUE* and its qualifier in *QUALIFIER*. Likewise, *PARAMETER* is used rather than *SPECIES* and *UNITS* so that analytical method information could be included with the information regarding determined species and units of reporting for all results. These two fields are critical in the construction of another crosstab query, **qctabIgnRx_Geol_10MajorsChem**, which further aids the user by displaying the data in a flat file or spreadsheet view (fig. A2–6).

This crosstab query was constructed using **Geol** so that *LAB_ID* is the key row heading, and **Chem** so that the unique entries in *PARAMETER* become the column headings, and the cells in each column are filled by the entries in *QUALIFIED_VALUE*. Fields providing sample identification, submittal, location, collection, and description information were also added as row headings to the query from **Geol** so that relevant descriptive information would be available in one Datasheet View. The "Criteria" line of the Query Design View shows the conditions placed on this query: the phrase "igneous" for *SECONDARY_CLASS* of **Geol** identifies igneous rock samples; the phrase ""' Like "Al2*" Or Like "CaO*" Or Like "FeT*" Or Like "Fe2*" Or Like "FeO*" Or Like "K2*" Or Like "LOI*" Or Like "MgO*" Or Like "MnO*" Or Like "Na2*" Or Like "P2*" Or Like "SiO*" Or Like "TiO*"" for *PARAMETER* of **Chem** identifies major-element oxide data. The data retrieved will be sorted by *LAB_ID*. Running **qctabIgnRx_Geol_10MajorsChem** produces a Datasheet View containing all available major-element data, expressed as oxide percentage, as well as all geologic and geographic data for 46,813 igneous rock samples. Like the 33 ChemData tables provided in the database, crosstab queries can be constructed to create unique datasets containing analytical data gathered from a specific sample media, collected within specific geographic locales, and determined by certain analytical methods.

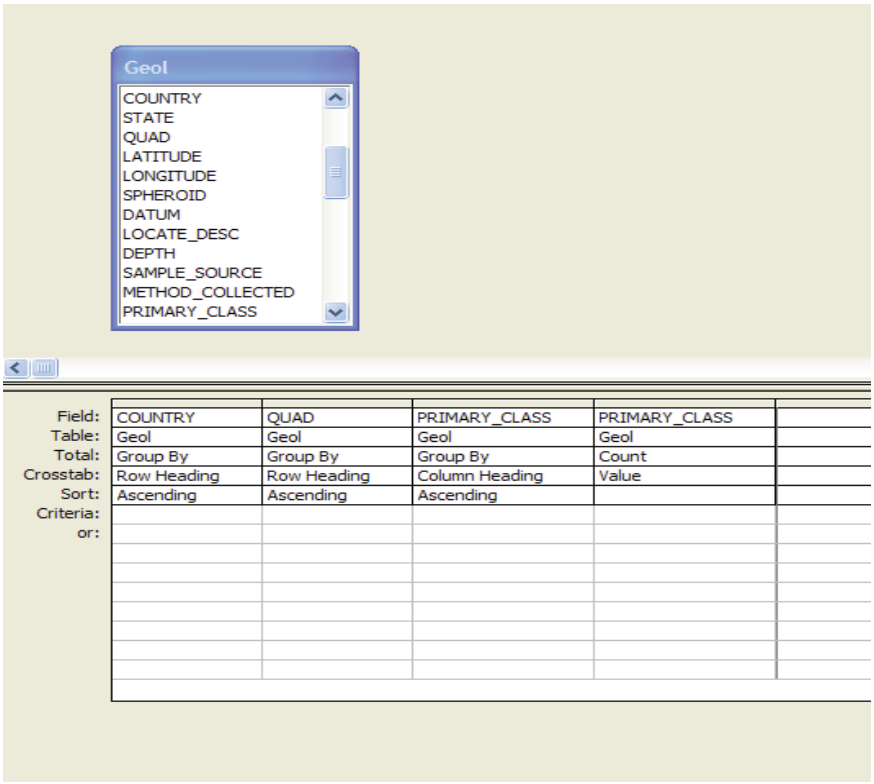


Figure A2–5. Crosstab query **qctabQuad_SampleType** in Query Design View.

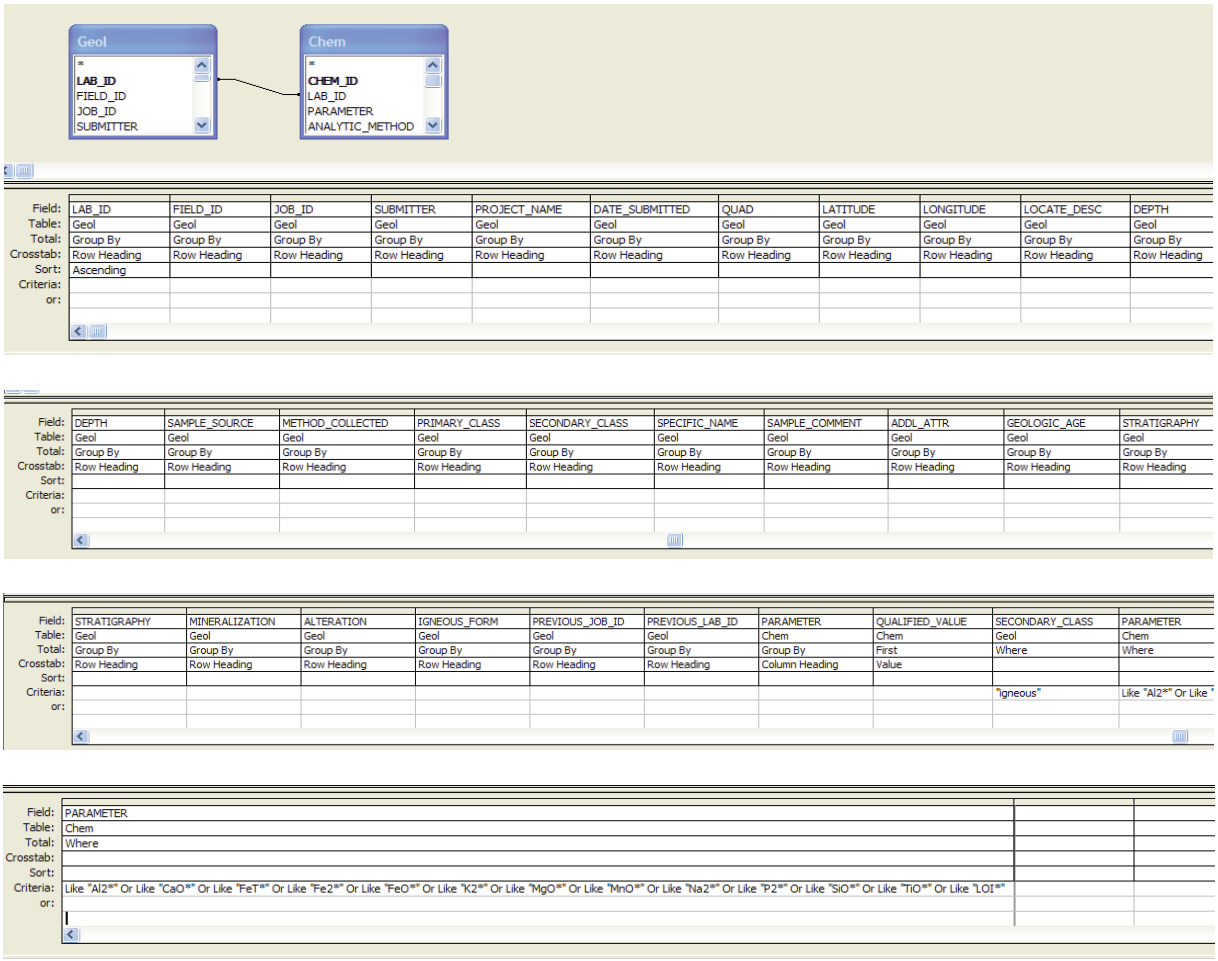


Figure A2–6. Crosstab **query qctabIgnRx_Geol_10MajorsChem** in Query Design View. This query involves many fields and is shown as four panels read left to right and top down.

Appendix 3

Data Enhancement, Correction, and Processing

This appendix describes procedures used to improve the quality of data originally obtained from the National Geochemical Database (NGDB) to create the Alaska Geochemical Database (AGDB) and methods that were used to find, correct, and include additional samples and data. Discussion of these procedures is grouped into the logical attribute sets sample identification, sample submittal, sample locality, sample collection, sample description, sample preparation, chemical analyses, and mineralogical analyses.

Sample Identification

Accurate sample identifiers in geochemical databases link physical samples to their geospatial attributes and chemical determinations. All samples in the NGDB and AGDB have a *FIELD_ID*, a field identifier assigned by the sample collector to the sample submitted for analysis, and a *LAB_ID*, a unique identifier assigned to each submitted sample by the analytical laboratory that received the sample. *LAB_ID* in the NGDB is the unique key field of the **GEO** table, despite the fact that the U.S. Geological Survey (USGS) has had a number of analytical chemistry laboratories, sample control groups, and computer data-processing projects associated with it over the years. *LAB_ID* is also a field in the NGDB **CHEM** table, grouping unique chemical determinations by the sample analyzed. There is potential duplication of *LAB_IDs* for pre-1965 samples because they were submitted for analysis to USGS laboratories operating simultaneously in Washington, DC, Denver, CO, and Menlo Park, CA at a time when there was no perceived need for *LAB_ID* uniqueness because a unified USGS geochemical database had not been developed. Special attention to this fact is required when samples from these older datasets are to be included in the NGDB. Very rarely, *LAB_ID* needed to be corrected (by the NGDB manager) because of archival data entry error. The *LAB_ID* identification stamp was automatically computer generated in the legacy PLUTO database and continues today through the USGS Sample Submittal Form (SSF) that is required for all samples submitted to the NGDB and the Laboratory Information Management System (LIMS). The SSF has undergone a series of revisions to improve its efficiency and ease of use and is currently being redesigned. In addition, a new *LAB_ID* series, YL######, has been created in the NGDB to facilitate the inclusion of relevant datasets that were not issued standard *LAB_IDs* by Sample Control work groups at any work center. There are 843 samples in the AGDB that have these new *LAB_IDs*.

JOB_ID is a laboratory batch identifier assigned by the analytical laboratory that received the samples as a batch (group). In the NGDB, *JOB_ID* is the unique key field of the **JOB** table, linked to the *JOB_ID* fields in the **GEO** and **CHEM** tables. A job may have between one and several hundred samples. Original duplication of *JOB_ID* between the Rock Analysis Storage System (RASS) and PLUTO archival databases has been remedied by the NGDB database manager, enforcing uniqueness in this data field. The following examples illustrate this:

- Previous Branch of Analytical Laboratories (USGS-BAL)-Washington, DC PLUTO *JOBNUMBER "5534"* containing Colorado samples is now called *JOB_ID "W5534"* in the NGDB, in order to avoid confusion with previous Branch of Exploration Research (USGS-BOER)-Denver RASS *JOB_ID "5534"*, now called *JOB_ID "HM5534"* in the NGDB, which contains Alaska samples.

- Previous USGS-BAL-Washington, DC PLUTO *JOBNUMBER "11403"* containing Colorado samples is now called *JOB_ID "X1403"* in the NGDB, in order to avoid confusion with previous USGS-BOER-Denver RASS *JOB_ID "11403"*, which contains Alaska samples.

- Previous USGS-BAL-Denver PLUTO *JOBNUMBER "958"* containing Alaska samples is now called *JOB_ID "Z0958"* in the NGDB in order to avoid confusion with previous USGS-BOER-Denver RASS *JOB_ID "958"* (now called *JOB_ID "HM958"* in the NGDB) containing different Alaska samples.

- Previous USGS-BAL-Denver PLUTO *JOBNUMBER "M287"* containing Colorado samples is not to be confused with previous USGS-BAL-Menlo Park PLUTO *JOBNUMBER "M287"*, now called *JOB_ID "M0287"* in the NGDB, containing Alaska samples.

Rarely, JOB_ID needed to be corrected (by the NGDB manager) because of archival data-entry error. The job number was automatically computer generated in the PLUTO database and continues today through the LIMS for the NGDB database. In addition, a new JOB_ID series, YJ#####, has been created in the NGDB to facilitate the inclusion of relevant datasets that were not issued standard JOB_IDs by Sample Control.

There are 6,686 records of reanalyzed National Uranium Resource Evaluation (NURE) samples in the AGDB. They have been assigned standard *LAB_ID* and *JOB_ID* identifiers and their reanalysis data are stored in the NGDB. Their *FIELD_ID* entries have been corrected so that the NURE *Prime_ID,* the Primary Laboratory Sample Identification Number of the NURE database, is now the sample's *FIELD_ID*, allowing for easy comparison between original analyses in the NURE database and reanalyses of these same NURE samples in the AGDB. This also facilitates the linking of the

NGDB to the NURE database because the NURE *Prime_ID* can be linked to *FIELD_ID* in an Access query.

FIELD_ID has often been modified in the data enhancement process to remedy the truncation of the original data entry. *FIELDNO* in the RASS database was an 8-character field, and since many submitters did not limit the length of their field numbers, many field numbers were truncated by data-entry personnel. At times, this truncation created many apparently identical field numbers in RASS that were in fact different original sample numbers. For example, field numbers "87ABD001S" and "87ABD001C3" truncated to eight characters would be "87ABD001," which is actually the field site number. A bulk stream sediment sample (suffix "S") and a C3 fraction concentrate from a stream sediment sample (suffix "C3") were taken at this site, but truncation obscured this fact. These types of field number errors have now been corrected so that all *FIELD_ID* entries match the field numbers listed in the Request for Analysis (RFA) paperwork that is on file at the NGDB hard copy archives.

Various data-entry formats were used for RASS and PLUTO databases. The AGDB *FIELD_ID* format removes all dashes or spaces between alpha and numeric characters, most slashes (/), and all underscores (_). Spaces between two numeric characters or between two alpha characters have been replaced with a dash (-). Occasionally, a series of field numbers were edited so that they could be sorted effectively. For example, "87ABD1" has been changed to "87ABD001" so that it appears before "87ABD011" or "87ABD101" in an ascending sort. By implementing a standard format and sorting criteria, all samples collected at one field site are grouped together. In cases where only one sample had accurate coordinates, this has provided an effective way of assigning coordinates to samples that had no coordinates or that had been assigned the corner coordinates of the corresponding topographic quadrangle instead of true coordinates. Alternate field numbers that were sometimes provided on RFA coding sheets are retained in the *SAMPLE_COMMENT* field.

Many archived NURE sediment and soil samples have been submitted for reanalysis by the USGS. Although the reanalyses were entered into the RASS, PLUTO, and NGDB databases, there was no standard data entry scheme for these data, and as a result, there was no consistency for the recording of *FIELD_ID*. These samples have been corrected so that *FIELD_ID* contains the *PRIME_ID* from the NURE database, and any alternate IDs have been entered in *SAMPLE_COMMENT*. This correction creates a link between the NGDB and NURE databases on the fields *FIELD_ID* and *PRIME_ID*, respectively. Reanalyzed RASS and PLUTO samples were commonly assigned new *LAB_IDs*, and in many cases the original *LAB_ID* and *FIELD_ID* information may have been incorrect in the submittal. This has been remedied so that the original *LAB_ID* has been entered in *PREVIOUS_LAB_ID*, and the field number has been entered in *FIELD_ID* so that it always contains data generated by the sample submitter.

The field *PREVIOUS_LAB_ID* contains the original *LAB_ID* of a reanalyzed sample that was given a new lab number upon submittal. The *PREVIOUS_LAB_ID* could be entered into the NGDB by way of the SSF, but was not required, which created problems in the NGDB where one sample site may have had multiple sample records, some of which were from the same sample. *PREVIOUS_LAB_IDs* can be generated and populated by creating an Access query where the *FIELD_ID* in one table is linked to *LAB_ID* in a copy of the same table, thereby exposing the original *LAB_ID* and the reanalysis *LAB_ID*.

The field *PREVIOUS_JOB_ID* contains the original batch number (*JOB_ID*) of a reanalyzed sample that was given a new batch number upon submittal. *PREVIOUS_JOB_ID* could be entered into the NGDB via the SSF but was not a required field for data entry. It can be generated and populated by creating an Access query where the *LAB_ID* in one table is linked to *PREVIOUS_LAB_ID* in a copy of the same table, thereby exposing the reanalysis *JOB_ID* and the original *JOB_ID*, which is then entered in *PREVIOUS_JOB_ID*.

Sample Submittal

Sample submittal attributes concern who, when, and sometimes why the sample was submitted for analysis. The *SUBMITTER* field contains the name(s) of the scientist(s) who submitted the sample to the laboratory for analysis. The sample submitter was not necessarily the sample collector. The standard format for *SUBMITTER* has been taken from the USGS Library Catalog format and is as follows: "Schmidt, Jeanine M."; or "Schmidt, Jeanine M., and Karl, Susan M." if there are multiple submitters. In the past, the NGDB has had multiple spellings or formats for one name. This has been corrected so that each person has only one name format in the NGDB. The name of the chemistry project chief or the chemist performing the analyses was often used as the *SUBMITTER*, although they may have had nothing to do with the sample collection or subsequent data interpretation. Because of this, names of secondary sample submitters noted on RFAs have been added to *SUBMITTER* to better reflect the project aims and personnel and to allow for a more efficient linking of *SUBMITTER* with "name" fields in other databases, such as the Library Catalog at *http://library.usgs.gov/*.

The *PROJECT_NAME* field contains the name of the work group or project that allocated funds for the collection and analysis of submitted samples. *PROJECT_NAME* entries can be used to group samples, as well as to indicate the scientific intent behind the samples and their data. In this manner, mineral-resource assessment data can be separated from environmental assessment data. Where data are absent, *PROJECT_NAME* can sometimes be derived from the Project Number (project account number) that is present on the RFA. RFAs commonly had more than one Project Number as the funding account might change during the duration of the project. *PROJECT_NAME* was not a required field in early versions of the SSF. *PROJECT_ACCOUNT* has always been required in the SSF but cannot be easily used to populate *PROJECT_NAME*.

The field *DATE_SUBMITTED* contains the date that the sample was submitted to Sample Control for initial database processing before laboratory sample prep and analysis. *DATE_SUBMITTED* is also useful as it places the samples in time with respect to the analytical methods available when the samples were submitted. Reanalyzed samples may not always have a new *DATE_SUBMITTED* entry, however, because additional analyses could be requested for samples already in the RASS or PLUTO databases without requiring the assignment of a new *LAB_ID* for the samples in question. This can result in situations where a *DATE_SUBMITTED* entry is matched with chemical data derived from an analytical method that was not available until years later. The NGDB currently does not include "date of analysis" information, but it could be gleaned retroactively from RFAs and laboratory reports.

Sample Locality

Because precise sample locality information is critical to any geospatial database, much work was done to obtain more accurate locality data for the AGDB. The *COUNTRY* field contains the name of the Country from which a sample was collected. *COUNTRY* entries do not always match the geographic coordinate entries given for the sample; those samples need to have their locations checked. RASS had no Country field; the *COUNTRY* field in PLUTO consisted of a 2-character code that commonly contained errors and allowed coded entries for oceans and seas. AGDB sample locations located in Federal or International waters of the Arctic and Pacific Oceans and outside of 1:250,000-scale Alaskan quadrangle boundaries have entries such as "Pacific Ocean" or "Bering Sea" in the field *COUNTRY*.

The field *STATE* contains the name of the State from which a sample was collected. *STATE* entries do not always match the geographic coordinate entries given for the same sample; those samples need to have their locations checked. RASS had no State field; the *STATE* field in PLUTO consisted of a 2-character code that commonly contained errors (for example, "AL" entered for Alaska), as well as *Country* codes entered in the wrong field (for example, the "CO" country code erroneously entered in the *State* field, resulting in Colombian records erroneously identified as derived from Colorado). AGDB sample locations located in Federal or International waters of the Arctic and Pacific Oceans and outside of 1:250,000-scale quadrangle boundaries have null entries for *State*. An accurate *STATE* field can be derived via geographic information systems (GIS) applications if geographic coordinates for samples are accurate.

The field *QUAD* contains the name of the 1:250,000-scale quadrangle from which the sample was collected. For some USGS geochemical reconnaissance programs, the 1:250,000 quadrangle name is of primary importance in Alaska, where geotechnical data are routinely sorted and retrieved by quadrangle. As needed, accurate quadrangle information can be derived, using GIS functions, for samples with accurate coordinates.

The *SPHEROID* field contains the name of the reference spheroid or ellipsoid, when recorded, for the latitude and longitude coordinates of the sample site, and the *DATUM* field contains the reference datum for the site coordinates. Spheroid and datum information were not entered in RASS or PLUTO, so most NGDB sample records have no location reference information other than coordinates. Where datum and spheroid information are missing for samples with latitude and longitude values, these older (commonly pre-2000) coordinates were most commonly obtained by digitizing or calculating locations based on USGS topographic maps that are most commonly projected using the NAD27 datum and Clarke 1866 spheroid. More recent samples may have frequently had coordinates measured by geographic positioning systems (GPS). Unless the sample collector specifically changed his GPS readout to be otherwise (in which case he probably thought to provide spheroid and (or) datum information), the default GPS readout is provided as WGS84 spheroid and datum

The field *LATITUDE* contains the latitude coordinate of the sample site, and *LONGITUDE* contains the longitude coordinate of the sample site; both reported in decimal degrees and usually with NAD27 datum and Clarke 1866 spheroid for samples submitted prior to year 2000 (see preceding paragraph). Through the SSF, the submitter entered positive (Northern Hemisphere) and negative (Southern Hemisphere) decimal numbers for latitude and positive (Eastern Hemisphere) and negative (Western Hemisphere) decimal numbers for longitude. Both RASS and PLUTO entered coordinates in the format degrees-minutes-seconds, in varying degrees of precision. The search for accurate coordinates was a major focus of the data cleanup effort that went into the creation of the AGDB. Of the 324,416 Alaska samples researched, 11 percent of these samples gained improved coordinates, while 4 percent remain to be solved. A logical sequence for obtaining missing geographic coordinates was followed, and a discussion of these processes is detailed below.

The first step in obtaining accurate geographic coordinates for records that do not have them was to check the RFAs to identify possible inaccurate data entry. Second, missing coordinates may occur in USGS Open-File Reports (OFRs) where values were published but were never furnished for database entry. Third, some projects have several sample media collected at one site. If one type of media from a site has good coordinates, the same coordinates apply to all other media from this site. Also, projects may have submitted sample splits to separate laboratories for different analyses. Samples were sorted by *FIELD_ID* (see "Sample Identification" section concerning *FIELD_ID*) to group samples by field sites, which often yielded accurate coordinates for *LAB_IDs* that were lacking geographic coordinates.

A different sort of coordinate problem is "corner coordinates," defined as coordinates that plot on the corner , usually the southeast corner, of a 7.5-minute or 15-minute sample-site location topographic base map. In these cases, the actual sample sites are hand-plotted accurately on an original field map, but the map corner coordinates were submitted to Sample

Control for data entry as a kind of shorthand for the location of the samples. This problem is most prevalent in the PLUTO database because the RFA coding sheets requested coordinates for the "SE corner of the 7½′ quadrangle." Samples entered in the PLUTO database may also have been entered in the RASS database and have accurate coordinates there, so sorting samples by *FIELD_ID* will retrieve these matching pairs. Although many corner coordinates have been corrected and updated, there remain samples in the AGDB dataset that have corner coordinates that need to be resolved.

Finally, a search for archived or published sample-site locality maps was conducted in an effort to provide missing geographic coordinates. Many of these maps reside in the geological data archives in the Field Records section of Denver's USGS Library and in the USGS Alaska Technical Data Unit. Digital scanned versions of these maps were then geographically registered, the sample-site locations were identified and digitized, and the resulting coordinates of the sample sites were entered in the *LATITUDE* and *LONGITUDE* fields, with appropriate datum and spheroid information added.

Scientists who submitted samples to be entered in the RASS database had the option of providing locality coordinates as Universal Transverse Mercator (UTM) projection Northings and Eastings, together with the UTM Zone, instead of latitudes and longitudes. UTM coordinates in the NGDB have been converted to geographic coordinates by the use of coordinate translation programs. Inadvertent truncation of the 7-character UTM Northing field to six characters during initial data entry created glaring location errors, but this has since been corrected as well. In the AGDB, all UTM coordinates have been converted to latitudes and longitudes.

Scientists who submitted samples to be entered in the RASS database also had the option of submitting locality information as X–Y coordinates instead of latitude and longitude. In some cases, the X–Y coordinates are referenced to standard systems such as a State Plane projection, but this fact may not be indicated on RFAs or in publications. In other cases, the X and Y values may be coordinates of points located on a local grid superimposed on the sample-site locality map, or they might be the coordinates of points that are measurements on the map from a point of origin (commonly the southeast corner of the map). X–Y coordinates always require the known location of the origin as well as the scale of distance used to create the grid (1in.=1,000 ft, that is) on the sample-site location map. Without a scale or an origin, the sample-site locations would need to be digitized. In the AGDB, all X–Y coordinates have been converted to latitude/longitude.

By international convention, the latitude values of sample-site locations that are in the Northern Hemisphere are positive numbers; negative values identify locations in the Southern Hemisphere. *LATDIR* entries in RASS for samples in this dataset are "N," placing them in the Northern Hemisphere, and latitude entries in PLUTO for AGDB samples are positive numbers. By similar convention, the longitude values of sample-site locations in the Western Hemisphere are negative numbers; positive values identify locations in the Eastern Hemisphere, based on the Greenwich prime meridian line. *LONGDIR* entries in RASS for samples in this dataset are "W," placing them in the Western Hemisphere, and most longitude entries in PLUTO for AGDB samples are negative numbers but there are 117 sample locations from the outer Aleutian Islands that are in the Eastern Hemisphere (west of 180 degrees west longitude) and have positive longitude numbers. Many samples collected in the Western and Northern Hemispheres had erroneous site locations recorded in the NGDB because the hemispheres were incorrectly designated. These errors have been corrected in the NGDB. All samples in the AGDB have positive LATITUDE (Northern Hemisphere) coordinates and all but 117 outer Aleutian samples have negative LONGITUDE (Western Hemisphere) coordinates.

The *DEPTH* field contains information regarding the depth of the sample from the land surface if cored or trenched, accompanied by units of measure. *DEPTH* may refer to a single measurement or a measured interval from which the sample was collected. "Surface" is a valid entry for *DEPTH*. Sample depth was obtained from RFA Sample Comments, from the PLUTO field *DESCRIPT1*, or derived from *FIELD_ID* if the field number contains information that refers to sample depth (for example, DH1A10-20′ indicates that the sample comes from the interval between 10 and 20 feet in drill hole 1A). The current SSF allows the submitter to enter depth value or range, with units, but earlier versions permitted the entry of "Surface (0–25 cm)," "1'–2'," or any other entry desired by the submitter. The depth field *Z_COORD* in RASS was very sparsely populated, did not have units of measure, and could refer to measured distance above land surface or sea level (elevation) as well. The *DEPTH* field in the AGDB is not to be confused with elevation above or below sea level.

The field *LOCATE_DESC* contains descriptive geographic information relating to the location of the sample site. Location descriptions were commonly recorded in the RASS Sample Comment field that was never digitally entered, or in the *DESCRIPT1* field of PLUTO. This field should not contain State or 1:250,000 quadrangle information as there are already fields for that information. Data for *LOCATE_DESC* can be derived from other spatial data layers using GIS layers. It can also be derived from fields in the NURE database when NURE samples are being resubmitted for further analysis.

Sample Collection

Sample collection attributes concern when and how samples were collected, as well as information regarding the environment of the sample source. The field *DATE_COLLECT* contains the date, and time if applicable, when the sample was collected. Collection date information is a relatively new entity in the NGDB and is especially important for samples collected for environmental studies. It is also important in the case of samples submitted for further analysis because reanalyses can take place 30 years after the samples were collected. For reanalysis submittals, the current date was entered in the *DATE_SUBMITTED* field, but *DATE_COLLECT* usually

was not entered, giving the false impression that the samples were recently collected and creating confusion when NGDB reanalyses are compared with the original NURE or AMRAP data. For NURE samples, *DATE_COLLECT* was derived from the *SAMPDAT* field in the NURE database by creating an Access query that linked *Prime_ID* in NURE to *FIELD_ID* in the NGDB. For resubmitted AMRAP samples, *DATE_COLLECT* was derived from the *DATE_SUBMITTED* field of the original sample by creating an Access query where the *PREVIOUS_LAB_ID* in one table was linked to *LAB_ID* in a copy of the same table, thereby exposing the *DATE_SUBMITTED* of the original *LAB_ID*. This date contrasts with the date that the sample was submitted for reanalysis, which is found in the *DATE_SUBMITTED* field ("Sample Submittal" section). In the AGDB *DATE_COLLECT* is in the format mm/dd/yyyy.

The field *SAMPLE_SOURCE* contains information regarding the source or site type as indicated by the sample submitter. Examples of entries for *SAMPLE_SOURCE* are outcrop, stream, mine dump, and prospect pit. It is important to know which samples have been collected at sites directly related to mining, such as underground mines, mine dumps, prospect pits, and mill tailings piles so that their potentially high concentrations of metals can be accurately interpreted. Initially, many stream sediments were coded with a *SAMPLE_SOURCE* of "outcrop" but are actually stream deposits; coding has been changed to reflect this in the AGDB. Much data that once resided in the PLUTO field *DEPOSITN* were migrated to the NGDB's *SAMPLE_SOURCE* in order to describe the environment of deposition at the sample site. Samples from drilled holes (core, cuttings, sludge, and well fluids) were checked for sample depth information that could be used to populate the *DEPTH* field. For soil and plant samples, data that once resided in the PLUTO field *BIOTIC* and *BIOTICSET* were migrated to *SAMPLE_SOURCE* to describe the biotic setting at the sample site. *SAMPLE_SOURCE* is not concerned with the sample preparation processes of segregation, concentration, or extraction that were used to create mineral separates, heavy mineral concentrates, or leachates but is concerned with the source of the parent material from which these sample types were derived.

The *METHOD_COLLECTED* field refers to the field sample collection method—usually single grab, composite, or channel. The field *SMPLCHAR* in PLUTO was used to describe whether the sample was typical or atypical of a larger body or population at the sample site, as well as a description of the collection method. *METHODCOL* in RASS was solely concerned with collection method; therefore, sample typicality is not noted in *METHOD_COLLECTED* of the NGDB.

Sample Description

The enhancement, correction and processing of sample description attributes is complex due to the number of attribute fields and to the significant database design differences between RASS and PLUTO. The *PRIMARY_CLASS* field contains the primary classification of media, and its categories in the AGDB are rock, sediment, soil, mineral, and concentrate. The field *SECONDARY_CLASS* is the secondary classification or subclass of sample media and is used to further describe the basic nature or type of material collected as a sample. The SSF generates *SECONDARY_CLASS* entries for the media rock, organic, and sediment, and possibly for leachate and water. The field *SPECIFIC_NAME* further defines the basic nature or type of material collected as a sample and is an attribute used to modify *PRIMARY_CLASS* and(or) *SECONDARY_CLASS*. The SSF generates *SPECIFIC_NAME* entries for the media mineral, rock, organic, sediment, and miscellaneous, and possibly for leachate and water. A discussion of *PRIMARY_CLASS*, *SECONDARY_CLASS*, and *SPECIFIC_NAME* categories, their attributes, and their interrelationships is provided below.

Rock is defined as an aggregate of one or more minerals, for example, granite, shale, marble; or a body of undifferentiated mineral matter, for example, obsidian; or of solid organic material, for example, coal (Neuendorf and others, 2005). This definition includes ore samples, jasperoid, gossan samples, fault breccia (mylonite or gouge), tektites, and nodules or concretions. In the NGDB, coal is considered an "organic" sample rather than a "rock" sample. This point would certainly be disputed by many geologists and in the AGDB coal samples are classed as sedimentary rocks. The degree of lithification may separate rocks from sediments and is determined by the sample collector or submitter. Saprolite samples are considered rocks. Samples that are derived from rock by processes of mineral separation, particle concentration, or chemical extraction are called minerals, concentrates, or leachates respectively. The *SECONDARY_CLASS* attributes of the class "rock" are "igneous," "metamorphic," "sedimentary," and "unspecified." There can sometimes be difficulty in applying these class distinctions. For example, water-lain tuff may be considered a type of either igneous or sedimentary rock. For the AGDB, the contents of all attribute fields for the rock sample concerned are considered when checking *SECONDARY_CLASS* entries, and consistency is maintained. For instance, where "igneous," "metamorphic," or "sedimentary" rock type could not be discerned the rock is classed as "unspecified."

The entries for *SPECIFIC_NAME* under "igneous" rocks are primarily petrologic terms based on classification according to Streckeisen (1967, 1976). Data from earlier RASS and PLUTO classifications, submitter's comments on RFAs, and other information at hand are considered when entering data in *SPECIFIC_NAME*. If nothing is known of the petrology or chemistry of the sample, its intrusive or extrusive nature is noted. If nothing is known of the form of the igneous body from which the sample was collected, *SPECIFIC_NAME* is considered "unspecified." The SSF provides a significantly abbreviated list of igneous rock names from which to choose. The field *IGNEOUS_FORM* contains information regarding the form or structure of the igneous body from which the sample was collected, whether the rock is considered igneous or metaigneous. The migration of RASS and PLUTO data

into the NGDB has caused the data in *IGNEOUS_FORM* to become less definitive than it was originally. Many igneous structural categories in *ROCKFORM* of PLUTO were not represented in *IGNEOUS_FORM* of RASS, and others were merged to create one more general category. For the AGDB, the data have been returned to their original entry. This field has been further populated using information derived from the *SAMPLE_COMMENT* field that pertains to the igneous structural type found at the site of the collected sample. As *IGNEOUS_FORM* is an attribute in the NGDB for igneous rocks only, it is not possible to define the structures of metaigneous rocks, such as metamorphosed dikes of metadiabase, through the use of this field. This is not the case in the AGDB. Most of these fields of RASS and PLUTO contained code-derived data in a predetermined scheme, meaning that an acceptable attribute option was tied to its code letter or number (for example, "rock" was coded with an "A" in the sample type field of RASS). For igneous rock samples in the AGDB, *ADDL_ATTR* contains code-derived information regarding the igneous texture and structure of the sample (from *TEXTURE* in PLUTO), the sample's quartz-feldspar relationship (from *QUAR_FELD* in PLUTO), as well as other attributes that came from *MODIFIER-IGNEOUS* in RASS. None of this information was transferred from PLUTO or RASS to the NGDB, and was entered in *ADDL_ATTR* of the AGDB. These data also can be used to further define *SPECIFIC_NAME*.

The entries for *SPECIFIC_NAME* under "metamorphic" rocks are derived from rock names from RASS, petrologic composition code data from PLUTO, metamorphic facies code information, and from submitter's comments on RFAs. If nothing is known regarding the petrology, chemistry, or source of the rock from which the sample was collected, *SPECIFIC_NAME* is considered "unspecified." The SSF provides an abbreviated list of metamorphic rock names from which to choose. The field *METAMORPHISM* contains code-derived information regarding the type of metamorphism present at the sample site, with the two entries being "regional" or "contact." Although the data in *METAMORPHISM* was obtained from PLUTO, it is possible to further populate this field by using the "greisen" and "contact metamorphic" entries found in the field *MINERALDP* of RASS, as well as using information derived from *SAMPLE_COMMENT* that pertains to the type of metamorphism present at the sample site. The field *FACIES_GRADE* contains code-derived information regarding the metamorphic facies and grade classification of the rock from which the sample was collected. Although the data in *FACIES_GRADE* were migrated from RASS and PLUTO, it is possible to further populate this field by using information derived from *SAMPLE_COMMENT* that pertains to the type of metamorphic facies or grade present at the sample site. The field *SOURCE_ROCK* describes the type of rock that was metamorphosed to create the metamorphic rock that was sampled. This field only contains the entries "igneous" and "sedimentary" and is very sparsely populated in the NGDB. It could be further populated by using relevant information found in the field *SAMPLE_COMMENT*. It is not certain whether the data in the PLUTO field *METATYPE* were migrated to the NGDB, but if so, the data are most likely found in *ADDL_ATTR*. It is possible to further populate this field using information derived from *SAMPLE_COMMENT* as well as from the metaigneous and metasedimentary entries that were once in *METATYPE*. *SOURCE_ROCK* data are currently transferred to the NGDB by way of the SSF program. For metamorphic rocks, the field *ADDL_ATTR* contains code-derived information regarding the source of the rock (from *METATYPE* in PLUTO), the degree of metasomatism (from *METASOMAT* in PLUTO), the metamorphic texture (from *MTEXTURE* in PLUTO), the mineralogy and chemical composition (from *COMPOSIT* in PLUTO), as well as other attributes that came from *MODIFIER-METAMORPHIC* in RASS. Some of this information was not migrated from PLUTO or RASS to the NGDB and must be entered into *ADDL_ATTR*. The information also can be used to further define *SPECIFIC_NAME*.

The entries for *SPECIFIC_NAME* under "sedimentary" rocks are derived from earlier RASS and PLUTO classifications, submitter's comments on RFAs, and other information at hand. The SSF provides a significantly abbreviated list of sedimentary rock names from which to choose. The field *DEPOSIT_ENVIRON* contains the categories "marine," "continental," and "transitional" that define the environment of deposition for the rock from which the sample was collected; this highly generalized information was entered only in the very sparsely populated *SDEPOSITN* of PLUTO, and is therefore of little use in the NGDB. For sedimentary rocks, *ADDL_ATTR* contains code-derived information regarding the chemistry, mineralogy, and texture of the sedimentary rock from which the sample was obtained. These data were usually migrated from RASS (from *MODIFIER-SEDIMENTARY*) and PLUTO (from *MODIFIER1* and *MODIFIER2*) to the NGDB and also can be used to help further define the entries in *SPECIFIC_NAME*.

Sediment is defined as solid fragmental material that originates from weathering of rocks and is transported or deposited by air, water, or ice, or that accumulates by other natural agents, such as chemical precipitation from solution or secretion by organisms. Sediment forms in layers on the Earth's surface at ordinary temperatures in a loose, unconsolidated form; for example, sand, gravel, silt, mud, till, loess, alluvium (Neuendorf and others, 2005) are all considered sediment. The distinction between sediment and soil can be vague, especially if the soil in question has no true horizon development or the sediment is colluvially derived. Samples derived from sediments by processes of mineral separation, particle concentration, or chemical extraction are called minerals, concentrates, or leachates, respectively. In the NGDB, *SECONDARY_CLASS* attributes of "sediment" are "panned" and "panned only," which refer to concentrated samples. Likewise, *SPECIFIC_NAME* entries for sediments in the NGDB all refer to concentrate samples. "Concentrate" is not yet a *PRIMARY_CLASS* category in the NGDB; but in the AGDB, concentrate samples that have been derived from sediment or

soil have been given a primary classification as concentrates. Concentrate terms are further described below. There are three possible places in the SSF that might generate *SECONDARY_CLASS* attributes for sediment: (1) sample concentration performed by the submitter, (2) sample concentration provided by the analytical laboratory, and (3) *SPECIFICSAMPLETYPE*, which seems the most likely of the three. For sediment samples, *ADDL_ATTR* contains code-derived information regarding the chemistry, mineralogy, texture, and grain size of the sediment that was collected. Some of these data were migrated from RASS (from *DESCRIPT1*) and PLUTO (from *MODIFIER1* and *MODIFIER2*) to the NGDB. Entries from *DEPOSITN* in PLUTO may provide data for *ADDL_ATTR* as well as for *SAMPLE_SOURCE*.

Soil is a term used for the unconsolidated mineral and (or) organic matter on the surface of the Earth that serves as a natural medium for the growth of land plants and has been subjected to and shows the effect of genetic and environmental factors of climate (including water and temperature effects) and macro- and microorganisms, conditioned by relief, acting on parent material over a period of time. Soil differs from the material from which it is derived in many physical, chemical, biological, and morphological properties and characteristics. Engineering geologists refer to soil as all unconsolidated material above bedrock (Neuendorf and others, 2005). These are obviously two very different definitions for soil, showing that there can be a vague distinction between sediment and soil, especially if the soil has no true horizon development or the sediment is colluvially derived. Samples that are derived from soil by processes of mineral separation, particle concentration, or chemical extraction are called minerals, concentrates, or leachates, respectively. There are two possible places in the SSF that might generate *SECONDARY_CLASS* attributes for soil: presubmittal sample concentration (preparation provided by submitter) and postsubmittal sample concentration (preparation provided by analytical laboratory). Another field in the NGDB that describes soil samples is *SAMPLE_ZONE*, which describes the soil zone or horizon from which the sample was collected. Samples from mixed zones cannot be coded accurately in *SAMPLE_ZONE* of the NGDB. The field *HORIZON* describes whether these horizons are well marked or poorly defined. The field *SALINE* describes whether a soil is saline or nonsaline in character. The field *FERRITIC* describes whether a soil is ferritic or nonferritic. The field *ORGANICS* describes whether a soil is organic rich or organically poor to nonorganic in nature. The field *DRAINAGE* describes whether a soil is well drained or poorly drained. These fields do not define the relative degree of these characteristics. It is possible to further populate these fields by using information derived from *SAMPLE_COMMENT*. For soil samples, *ADDL_ATTR* contains code-derived information regarding the chemistry, mineralogy, texture, and grain size of the sediment that was collected. These data were migrated from RASS (from *DESCRIPT1*) to the NGDB.

"Mineral" is defined as a naturally occurring inorganic element or compound having a periodically repeating arrangement of atoms and characteristic chemical composition, resulting in distinctive physical properties. It is also defined as an element or chemical compound that is crystalline and that has formed as a result of geologic processes. Water is not a mineral (although ice is), and crystalline biologic and artificial materials are not minerals (Neuendorf and others, 2005). Ice samples that were submitted for analysis were classified as water in the NGDB. PLUTO had a classification for minerals but RASS did not, and the very few mineral samples in RASS are classed as "other". Mineral samples in the NGDB were usually submitted for analysis after undergoing physical mineral separation processes, though some may be single selected mineral specimens. Multi-mineral separates are classed as concentrates. The NGDB has no *SECONDARY_CLASS* attributes for mineral but the mineral name is entered in *SPECIFIC_NAME*. Entries from *MINERLCOD* in PLUTO consisted of a five-letter code. These entries have been decoded and the mineral names migrated to *SPECIFIC_NAME* in the NGDB. *HOSTMATERL* entries in PLUTO have been migrated to *SOURCE_ROCK* in the NGDB, while *MINERALGRP* entries (for example, sulfides, oxides) were migrated to *ADDL_ATTR*. Mineral group data have been excluded from the AGDB, being of little use when the mineral name is known.

In the NGDB, concentrate is not a category of *PRIMARY_CLASS*, but in the AGDB it has been given a class of its own. Concentrate, as a sample submitted for analysis, is the fraction of solid material remaining after one or a series of physical or mechanical separation or segregation processes have been completed. The source of concentrated material, which may be rock (after crushing), sediment, or soil, is indicated, and the concentration procedure is noted as well. These procedures can involve separation by specific gravity (density), magnetism, and (or) particle size. In the NGDB, separation by particle size (grain size) alone does not constitute a concentrate, but when accompanied by one or more other separation processes, it is classified as such. Samples of mono-mineral separates are entered as "mineral" rather than "concentrate." In the NGDB, only samples with a *PRIMARY_CLASS* entry of "sediment" can be coded as concentrates, but in the AGDB, all sediment, soil, and rock concentrate samples have been coded as concentrates. The *SECONDARY_CLASS* attributes of sediment are "panned" and "panned only," which refer to concentrated samples in the NGDB and in this database. Another similar mineral concentration can be achieved using a Wilfley table. If samples are "panned only," *SPECIFIC_NAME* is a bulk-panned concentrate or heavy-mineral concentrate (called "HMC"). If they are called "panned" in *SECONDARY_CLASS*, it is understood that other separation processes were used as well. Panned sample fractions referred to in the *SPECIFIC_NAME* field are C1 (highly magnetic fraction of heavy-mineral concentrate), C2 (weakly or moderately magnetic fraction of heavy-mineral concentrate), C3 (nonmagnetic fraction of heavy-mineral concentrate), C4 (combination of fractions C2+C3), C5 (combination of fractions C1+C2),

and "lights" (non-heavy-mineral concentrate with magnetite removed).

The separation parameters for these concentrates define the nature of the concentrate type, and are entered in the *PREP* field of the NGDB. A typical entry might be "-35 hmc, bromo @ 2.85 SP, modified Frantz @ 0.25 & 1.75 amp ~ 0.2 & 0.6 amp w 15° fwd slope & 10° side tilt," which means that the sample analyzed was a panned heavy-mineral concentrate, sieved to –35 mesh, with a specific gravity greater than 2.85 after undergoing a density separation by bromoform, split into three magnetic fractions using a "modified Frantz Isodynamic Separator" method with electromagnetic settings of 0.25 and 1.75 amp, which is equivalent to Frantz settings of 0.2 and 0.6 amp with a 15° forward slope and a 10° side tilt. The fractions yielded by this process of separation are ≤0.25 amp fraction (C1), ≥0.25 to ≤1.75 amp fraction (C2), and ≥1.75 amp fraction (C3). The analyzed sample is usually one of these fractions, with the C3 fraction being the most common. Fractions were sometimes combined or their parameters of separation modified upon direction of the sample submitter.

In the NGDB, the category "miscellaneous" represents sample types that are more difficult to classify in the above categories. Some "miscellaneous" samples are precipitates, coatings, synthetic solutions or products, fossils, mill tailings, meteorites, dust, and so forth. Strong arguments can be made to place some of these terms with the sample types listed above, and the NGDB design group is currently investigating and discussing this issue. The *SPECIFIC_NAME* options for miscellaneous samples in the NGDB by way of the SSF log-in process are "peat," "gossan," "limonite," "caliche," and "other," whereby one word is entered to describe this other type of sample. Peat is best classed as a soil type. Gossans are iron-bearing weathered products overlying sulfide deposits and are usually rocks, though they can also be classed as sediment (Neuendorf and others, 2005). In the AGDB, gossan, limonite, and caliche are classified as rock, sediment, or soil, depending on other information that would be used to determine the *PRIMARY_CLASS* choice that is accurate.

The field *SAMPLE_COMMENT* does not exist in the NGDB but has been created for the AGDB. This field contains attributes used to modify *PRIMARY_CLASS*, *SECONDARY_CLASS*, or *SPECIFIC_NAME* and can be used to further describe the sample site, criteria, or any information about the sample that is useful. The contents of *SAMPLE_COMMENT* are not code-derived data but represent comments that have been written or typed on RFAs by sample submitters, data collected from field notebooks and OFRs, or further explanations of the sample that has been added by the database compiler. Potential redundancy between the contents of *SAMPLE_COMMENT* and the code-derived data in other database fields, especially in *ADDL_ATTR* has been removed from the AGDB.

Petrologic and mineralogic information about rock samples can be found in *SAMPLE_COMMENT*. These data can be used to better define the contents of the *SECONDARY_CLASS* and *SPECIFIC_NAME* fields and can assist in the population of other fields in the NGDB. Detailed alteration and mineralization information also is found in *SAMPLE_COMMENT*, and these data may more clearly define the contents of the *ALTERATION* and *MINERALIZATION* fields. *SAMPLE_COMMENT* may contain structure and fabric information about rock samples, as well as comments regarding the relationship of the sample to larger features at the sample site. In the case of igneous rock samples, these data may assist in the population of *IGNEOUS_FORM*. Information regarding the composition of sediment and soil samples also may be found in *SAMPLE_COMMENT*. *SAMPLE_COMMENT* can contain the name of the agency that collected the sample (when not collected by the USGS) or a note that the sample was originally collected under the NURE program. *SAMPLE_COMMENT* may also contain an alternate field number or a non-USGS laboratory identification number that is attached to the sample. In the event that the sample was previously submitted for analysis and received an earlier *LAB_ID* within an earlier *JOB_ID*, this lab number and job number have been transferred to the fields *PREVIOUS_LAB_ID* and *PREVIOUS_JOB_ID*, respectively. *SAMPLE_COMMENT* may contain detailed sample-collection and preparation information. Some of these data should be transferred to the *PREP*, *STABILIZATION*, and *MESH_PORE_SIZE* fields if relevant.

The field *ADDL_ATTR* contains additional attributes used to modify *PRIMARY_CLASS*, *SECONDARY_CLASS*, or *SPECIFIC_NAME* and in the AGDB are derived from sample codes in fields of RASS and PLUTO that do not have current equivalent fields in the NGDB. In the NGDB, *ADDL_ATTR* is a 255-character field that also contains all data from the *SAMPLE_COMMENT* field, which sometimes results in the truncation of entered data. The only code-derived data from RASS that were migrated to *ADDL_ATTR* come from the RASS field *DESCRIPT1*, but only in the cases where *PRIMARY_CLASS* entries are "sediment" ("unconsolidated sediment" in the *SMPLTYPE* field in RASS) but *DESCRIPT1* is not a concentrate ("C," "C1," "C2," or "C3"), or in cases where *PRIMARY_CLASS* entries are "soil." Data entered in the following RASS fields were not included in the NGDB, but have been added to the *ADDL_ATTR* field of the AGDB: *STRUCTURL* refers to the structural setting of the sample ("fracture/joint," "shear or fault"); *MATRIX* refers to the matrix material of the sample, usually a rock ("silica," "Fe/Mn," "carbonate," "clay"); *OXIDATION* refers to the degree of oxidation of a sample, usually a rock ("oxidized," "partially oxidized," "unoxidized"); *ORE/MINAL* refers to the ore mineral group in the sample ("base metals," "precious metals," "mixed base and precious metals," "radioactive," "rare earths"); and *MODIFIER* fields contain data that further define *SPECIFIC_NAME* for rock samples. Code-derived data from PLUTO that were migrated to *ADDL_ATTR* come from fields *MODIFIER1* and *MODIFIER2* if the entry in field *CATEGORY* is "sedimentary rock" or "unconsolidated sediment,"primary and secondary modifier data of sedimentary rock or sediment samples. Code-derived data from PLUTO that were migrated to *ADDL_ATTR* come from *MATERIAL*, if *CATEGORY* is "unconsolidated sediment," which refers to the primary material that makes up

the sediment sample. Other code-derived data from PLUTO that were migrated to *ADDL_ATTR* come from *MINERAL-GRP*. Data entered in the following PLUTO fields were not migrated to the field *ADDL_ATTR*—or to any field—in the NGDB but have been added to the *ADDL_ATTR* field of the AGDB: *TEXTURE* refers to the texture and structure of an igneous rock sample; *QUAR_FELD* refers to the relationship between quartz and feldspar of an igneous rock sample; *METASOMAT* refers to the degree of metasomatism exhibited in a metamorphic rock sample ("clearly," "may be," "no evidence"); *MTEXTURE* refers to the texture exhibited in a metamorphic rock sample ("schistose," "gneissic," "massive," "grain size"); and *COMPOSIT* refers to the mineral composition of a metamorphic rock sample.

ADDL_ATTR in the NGDB is used as the destination for sample comments from the SSF, as well as for the migrated data from the sample comment field *DESCRIPT1* (PLUTO), and from the RASS and PLUTO fields previously listed. Sample site information (for example, adit names, prospect pit numbers, names of and distances from geographic features) that is stored in *ADDL_ATTR* in the NGDB has been transferred to the *LOCATE_DESC* field in the AGDB. As the USGS created the NGDB from the migration of RASS and PLUTO data, some RASS and PLUTO fields were determined to be poorly populated or to contain information that might be considered irrelevant, confusing, or redundant. The AGDB puts all non-code-derived data in *SAMPLE_COMMENT* and all code-derived data from fields with no RASS or PLUTO equivalent field, whether migrated to the NGDB or not, in *ADDL_ATTR*. This would make *ADDL_ATTR* a field containing RASS and PLUTO data only. In creating this AGDB database, the data from some RASS and PLUTO fields excluded from migration have been recovered and reentered into *ADDL_ATTR*, data from other excluded fields have been omitted, and data from a few included fields have been deleted from *ADDL_ATTR* as it has been found to be relatively useless. The contents of the RASS fields *STRUCTURL*, *MATRIX*, *OXIDATION*, *ORE/MINAL*, and *MODIFIER* were not migrated to the *ADDL_ATTR* in the NGDB, but have been added in this database. Although these RASS fields have no equivalent PLUTO fields, the data within them have been deemed critical to the interpretation of information regarding the samples, especially rock samples. Data from the RASS field *DESCRIPT1* were migrated to *SPECIFIC_NAME*. Although this transfer was suitable for rock samples, problems exist for sediment, soil, and organic samples. The only *DESCRIPT1* entries for sediment or soil samples that belong in *SPECIFIC_NAME* are "C1," "C2," "C3," and "concentrate." All others have been transferred from *SPECIFIC_NAME*, or added, to *ADDL_ATTR* except for "stream sediment," which is redundant if "stream" is the *SAMPLE_SOURCE* entry. The contents from the PLUTO fields *TEXTURE*, *QUAR_FELD*, *METASOMAT*, *MTEXTURE*, and *COMPOSIT* were not migrated to the *ADDL_ATTR* in the NGDB but have been added in this database. Though these PLUTO fields have no equivalent RASS fields, the data within them have been deemed critical to the interpretation of information regarding the rock samples with which they were submitted.

A number of data entry points in the SSF create potential problems for migration into the NGDB. The fields "Oxidation" (of rock samples), "Modifier" (of sedimentary rock samples), and "Plant Part" are obvious ones that contain code-derived data. There are other SSF fields whose data may or may not be migrated to ADDL_ATTR in the NGDB: "Sample Comment"; "NURE Resubmitted Sample"; "RASS/PLUTO Resubmitted Sample"; "Reference Sample"; submitter analyzed parameters (pH, alkalinity, conductivity, dissolved oxygen, and temperature); specific sample type data; and numerous fields containing sample preparation information.

Four other fields in the AGDB describe geologic media. *GEOLOGIC_AGE* contains information regarding the geologic age of the sample or the age of the material at the sample collection site. *GEOLOGIC_AGE* contains the names of the geologic time units of Era, Period, and in the case of Tertiary and Quaternary Periods, Epoch. It does not refer to geochronologic age. Data migrated from PLUTO may contain ranges of geologic age (for example, Cambrian–Devonian). This field has had many misspellings and variations of possible ranges of geologic age that need to be standardized. In the case of some sediment or soil samples, *GEOLOGIC_AGE* in the NGDB may refer to the age of the bedrock underlying the sample source or the estimated geologic age of rock outcrops and float samples in the vicinity rather than the age of the sediment or soil samples themselves. The entry of geologic age data in the SSF does not allow for the entry of ranges of geologic time. For samples with accurate coordinates, this field could be populated using a GIS layer from large-scale USGS geologic maps. The field *STRATIGRAPHY* contains stratigraphic unit data, or comments regarding the sample or the sample site. *STRATIGRAPHY* usually contains formal, accepted names of stratigraphic units, igneous bodies, metamorphic zones, and so forth. In these cases, names in the AGDB have been standardized and spelled correctly according to USGS nomenclature. In other cases, stratigraphic comments and details have been migrated to the NGDB, representing information that the submitter felt was important to the interpretation of the geochemical significance of the sample. In the SSF, the submitter is allowed 255 characters to describe stratigraphy.

The field *MINERALIZATION* contains information regarding the mineralization type evident in the sample or at the site where the sample was collected. The scope of *MINERALIZATION* is limited in the NGDB due to lack of population and, in the case of data migrated from PLUTO, lack of concise definition. This field has been further populated using information derived from *SAMPLE_COMMENT* that pertains to the mineralization of the collected sample. If the sample is known to be mineralized but the actual type of mineralization is not noted, the entry "mineralized" has been entered in this field. In the SSF, mineralization is an attribute of rock samples only, but in reality, samples coded as "soil," "sediment," and "miscellaneous" may exhibit types of mineralization as well. The field *ALTERATION* contains data regarding the alteration

type evident in the sample or at the sample site. The scope of the field *ALTERATION* is limited in the NGDB due to lack of population and, in the case of data migrated from PLUTO, lack of concise definition. This field has been further populated using information derived from *SAMPLE_COMMENT* that pertains to the alteration of the collected sample. If the sampled material is known to be altered but the actual type of alteration is not noted, "altered" has been entered in this field. In the SSF, alteration is an attribute of only rock samples, but in reality, samples coded as "soil," "sediment," and "miscellaneous" might exhibit types of alteration as well.

Sample Preparation

Significant sample preparation data enhancement was processed in the creation of the AGDB because the NGDB has been scarcely populated with sample preparation data. The *PREP* field contains sample preparation description or comments. Sample preparation information may be available on the cover sheet of the RFA. It also may be found by linking the sample to a *PROJECT_NAME*, and the project and (or) submitter to an OFR that contains fairly detailed explanation of the sample preparation procedures that were implemented for the various media types collected for that project. Some of this information has been used to populate the field *MESH_PORE_SIZE,* but all the rest populates *PREP*, especially when describing the preparation procedure used for concentrate samples. Also, sample preparation information residing in *ADDL_ATTR* or *SAMPLE_COMMENT* has been moved to *PREP*. *PREP* is poorly populated because there were no preparation fields in RASS or PLUTO, but most sample preparation procedures used by the USGS have changed very little over the years. Generally, USGS samples of earth material were prepared as follows: rock—crushed, pulverized to pass a 100-mesh screen, and mixed; sediment—dried, most organic material removed, sieved to pass an 80-mesh screen; the minus-80 mesh fraction is ground to pass a 100-mesh screen, and mixed; soil—dried, juiced, sieved to pass an 80-mesh screen, the minus-80 mesh fraction is ground to pass a 100-mesh screen, and mixed (however, the analysis of the minus-10 mesh fraction of soil samples is also quite common); heavy mineral concentrate—may vary from submitter to submitter, preparation information taken from RFA and OFR has been entered in *PREP*. In the SSF entry process, there are a number of fields that are concerned with the submitter's preparation procedures and the laboratory's preparation procedures. Some samples are submitted for analysis fully or partially prepared, but the majority of others will require some sort of laboratory preparation prior to analysis. These fields appear to produce data that will be migrated to *ADDL_ATTR* or *PREP* in the NGDB: "Ash/Ashed," "Blend/Blended," "Concentrate/Concentrated," "Grind/Ground," "Sieve/Sieved," "Rock/Sediment/Soil Request for Prep," "Request for Prep," "Completed Treatment (by Submitter)," "Other Completed Treatment (by Submitter)." The field *MESH_PORE_SIZE* contains the mesh size through which a sediment or soil sample was sieved, or pore size through which a water sample was filtered, prior to analysis. Sieve mesh size is of great importance to sediment and soil samples as sieving promotes homogeneity. It is best to know and enter the actual parameters of a sieved fraction (for example, "–10 +80 mesh" rather than "+80" mesh). Sieve size of solid material is expressed using US standard sieve mesh number rather than in millimeters or in micrometers. Filter pore size is expressed in micrometers.

Chemical Analyses

The enhancement, correction, and processing of chemical analyses data involved much research into the analytical methods employed by the Branches of the USGS. A thorough understanding of USGS analytical protocols spanning the past fifty years was crucial in creating the chemical analyses dataset of the AGDB. The unique key field of the NGDB's **CHEM** table is *UNIQ_SEQ_NUM*, software-assigned integers that are created as chemical analyses that are migrated into the NGDB from LIMS or from outside sources. *UNIQ_SEQ_NUM* was not used as the key field of **CHEM** in the AGDB because the table contains many recent analyses that are not currently in the NGDB, but will be migrated to the NGDB at a later date. All analytical records in **CHEM** are accompanied by associated *LAB_IDs*. Samples resubmitted for further analysis have often been assigned new *LAB_ID*s. These samples in the **CHEM** table cannot easily be linked to data from the original samples without linking to the **GEO** table by *LAB_ID*, or by *PREVIOUS_LAB_ID* if it has been fully populated for these resubmittals. Many entries for *JOB_ID* in **CHEM** are unpopulated, but those data also can be obtained by linking **CHEM** to **GEO** using *LAB_ID*. *JOB_ID* in **CHEM** has not been populated in this manner because these chemical analysis records may actually belong to a resubmittal job number that is not in *JOB_ID* of **GEO**.

The measured characteristic *SPECIES* is a chemical attribute of element, ion, or oxide concentration that has an associated data value. Species fields in RASS and PLUTO were very similar, and the NGDB has used PLUTO species names wherever chemical data were present in that database. Very few corrections were required in *SPECIES*, but some species names have been changed in the AGDB to better support sorting. For example, "Carbonate C" and "Organic C" in the NGDB are "C-CO3" and "C-org" in the AGDB. For carbon and sulfur, "Total C" and "Total S" in the NGDB are "C" and "S" in the AGDB because total concentration is assumed for all species unless the sample digestion is known to be partial. In another example, "CV Th" in the NGDB—the coefficient of variance for thorium—is "Th-CV" in the AGDB. The entries of trace elements expressed as oxides have been changed so that *SPECIES* names are elements; the *DATA_VALUE* entries for these traces were converted from oxide concentrations to element concentrations, which helps to reduce the number of possible fields in Access or columns in Excel that are attributes of given species. In the NGDB, samples that have been ashed prior to analysis have been assigned species such

as "ash Mg." The AGDB has changed this attribution so that *SPECIES* would be "Mg," and "after ashing" has been entered in *DECOMPOSITION*. Many entries have been shortened to assist the creation of *PARAMETER*, which is the concatenation of *SPECIES*, *UNITS*, *TECHNIQUE*, and *DIGESTION* (for example, "Loss on Ignition" changed to "LOI"). The fields *NID* (numerical ID) and *DESCRT* (description of the numerical ID) in PLUTO and their RASS equivalents were used to define *SPECIES* and *UNITS* in the NGDB and were migrated to the NGDB as the fields *NID* and *NID_DESC*, and to *STATUS* in part, but these fields were not included in the AGDB.

The field *UNITS* contains the units of concentration or measurement in which the *DATA_VALUE* is expressed in both the NGDB and the AGDB. The units for trace elements expressed as weight percent have been changed so that *UNITS* are parts per million and the *DATA_VALUE* entries multiplied by 10,000 for these elements. The units for most major elements and major-element oxides expressed as parts per million have been changed so that *UNITS* are weight percent; the *DATA_VALUE* entries were divided by 10,000. In the same manner, units of parts per billion have been converted to parts per million and the *DATA_VALUE* entries divided by 1,000. These changes significantly help to reduce the number of possible fields in Access or columns in Excel that are attributes of given species. The entry "percent" in the NGDB has been shortened to "pct" in the AGDB to assist in the creation of *PARAMETER*.

As a result of researching analytical methods used for chemical analysis by the USGS since its inception, the fields *TECHNIQUE*, *DIGESTION*, and *DECOMPOSITION* have undergone significant change. A search for publications that describe these analytical methods has yielded the 753 titles listed in **AnalyticMethodBiblio** of the AGDB, and the 85 methods used to produce the data values in **Chem** are listed in **AnalyticMethod** that details the scope, digestion, decomposition, and detection used. Senior USGS chemists and emeriti were approached with many questions over the past ten years and their answers recorded and archived; this information became a critical part of analytical method description. The association of certain analysts with certain methods, the comparison of dates-of-sample submittal with the general date for the advent of new analytical methods, and the grouping of certain methods with specific laboratories all assisted in accurately defining the analytical methods of detection. Most of the newer data required little editing, but the older data benefited much from the research of analytical methods. Results of the process are, at the least, a good educated guess. This research was used to create the AGDB fields *ANALYTIC_METHOD* and *PARAMETER*. The PLUTO fields *TOA* (code for type of analysis), *DESCCT* (description of the analytical method), and their RASS equivalents were helpful in defining some analytical methods. These data were migrated to *TOA* in the NGDB but were not included in the AGDB.

The field *TECHNIQUE* is the abbreviation of the analytical method used to analyze samples. There are 21 *TECHNIQUE* entries found in **Chem** of the AGDB to accurately reflect the analytical method used. The following AGDB entries have been added to the NGDB: "CB" (combustion), "DN" (delayed neutron counting), "FL" (fluorometry), "GRC" (gamma ray counting), "GV" (gravimetry), "TB" (turbidimetry), and "VOL" (volumetric analysis). Many NGDB entries have been shortened to assist in the creation of *PARAMETER* (for example, "DC-ARC SPEC" changed to "ES" for emission spectrography). The PLUTO fields *LLOA* (combination of the 2-character abbreviation of the laboratory location with the 2-character mnemonic of the analytical method) and *LLOADS* (the description of the *LLOA*) help to define *TECHNIQUE* and *LAB_NAME*. The data in *LLOA* were migrated to the NGDB fields *CENTER* and *PROJECT* but are not included in the AGDB.

The field *DIGESTION* is the abbreviation of the degree of sample digestion—total (T) or partial (P)—required by the *TECHNIQUE* used to analyze the sample for a specific species. Some methods of sample digestion that are called "total" are less "total" than others but still are considered "total" in the NGDB and in this database.

In the NGDB, the field *DECOMPOSITION* contains a brief description of the decomposition method used for a given *TECHNIQUE* in the analysis of the sample and has been expanded in the AGDB also to include comments that further describe this *TECHNIQUE*. There are 109 different types of *DECOMPOSITION* in this database. In many cases, multiple entries that infer the same basic information have been reduced to a general but accurate description that works for all. In two examples, *DECOMPOSITION* contains information that makes the distinction between quantitative emission spectrography and the semiquantitative methods, and between instrumental neutron activation analysis and delayed neutron counting. *DECOMPOSITION* contains key information used in the creation of the entries in *ANALYTIC_METHOD* of **Chem** in the AGDB.

DATA_VALUE and *QUALIFIER* are the fields in the NGDB and the AGDB that contain the numeric results (the analytical values) and, when present, their qualifying modifiers. This data-storage scheme also was used in RASS and PLUTO, so chemical data migration to the NGDB went smoothly, considering the amount of data at hand. Very few values or qualifiers needed to be changed except for the many cases that "0" (zero) had been entered in *DATA_VALUE* of the NGDB, accompanied by the "less than" qualifiers "N" or "L" (Structure section). These values usually represent the lower limits of detection (LLD) for emission spectrography data from PLUTO. Data from RASS were not a problem because the LLD was always included in that database. One of the results of the research of analytical methods used by the USGS is a table of LLDs for nearly all species, specific to the method of analysis used, and related to the time period when the method was employed. This study was used to correct all "zero" values in the AGDB, as well as hundreds of thousands of values in the NGDB. *DATA_VALUE* was also changed in the AGDB if *SPECIES* or *UNITS* was changed in a way that required a conversion of the value. A key drawback of

working with *DATA_VALUE* and *QUALIFIER* together is that two fields are required to create one value, greatly increasing the number of result fields. Also, crosstab queries can only work with one value field at a time, not two. This limitation is overcome by the creation of the derivative *QUALIFIED_VALUE* in the AGDB, which presents chemical values in two different formats (Structure section).

The field *LAB_NAME* was created for the NGDB; although there was some useful laboratory information stored in the PLUTO field *LLOA* (for example, "LAXR" shows the laboratory location "LA" for Lakewood, Colo., and the laboratory mnemonics "XR" for the X-ray spectroscopy laboratory). *LAB_NAME* is more accurately defined in the AGDB to clearly link the chemical data to the analytical laboratory.

Mineralogical Analysis

All mineralogical data in the AGDB were derived from data entry sheets, USGS OFRs, and archival digital spreadsheets; none were obtained from the NGDB. More than 75 percent of the data is attributed to Richard B. Tripp with the remainder attributed to other mineralogists working on other AMRAP projects. The data entries represent numerous abundance schemes, a few of which are: (1) "present" or "not present", (2) "abundant", "moderate", or "trace", (3) percentages from 95 percent to 0.5 percent, (4) percentage ranges, or (5) grain counts. All grain count data have been entered in the field *Ore-relatedMnrl_Comment* as the first entry in that field, with gold data listed first. For some samples there were both grain counts and other types of abundance entries for the same mineral. In cases where there were no "non-grain-count" entries accompanying the grain count entry, the word "present" replaced the grain count data in the mineral field cells. The field *Inferred_Comment* contains Tripp's comments based on chemical analyses and previous USGS map publications, and were written for the Federal Lands in Alaska–Geologic Studies project for samples that had been hand ground or consumed during chemical analysis. The word "inferred" was entered in mineral fields to indicate the probability of a mineral's presence in the sample. The entry "not present" was at times encountered in original data sheets, and these entries were populated as null cells in the AGDB. Null cells in these data sheets equate to "not present" as all ore-related minerals were "looked for," so null cells in the mineral fields of the **Mnrlgy** table indicate that the mineral was not observed in the sample. The data is presented as it was originally recorded and interpreted by the mineralogist. All further interpretation of these data is the responsibility of the data user.

Publishing support provided by:
Denver Publishing Service Center

For more information concerning this publication, contact:
Center Director, USGS Central Mineral
and Environmental Resources Science Center

Box 25046, Mail Stop 973
Denver, CO 80225
(303) 236-1562

Or visit the Central Mineral and Environmental Resources Science Center Web site at:
http://minerals.cr.usgs.gov/