Colombia multielement geochemical data set

Recursos del Caribe, S.A. (www.cbmap.net)
Carl E. Nelson (carlericnelson@gmail.com)

The Colombia Multielement Geochemical Data Set, a compilation of analytical data originally published by the Servicio Geológico Colombiano (SGC-2010) and the Mineral Deposits Research Unit (MDRU-2016), provides access to and displays the published geochemical data for Colombia.

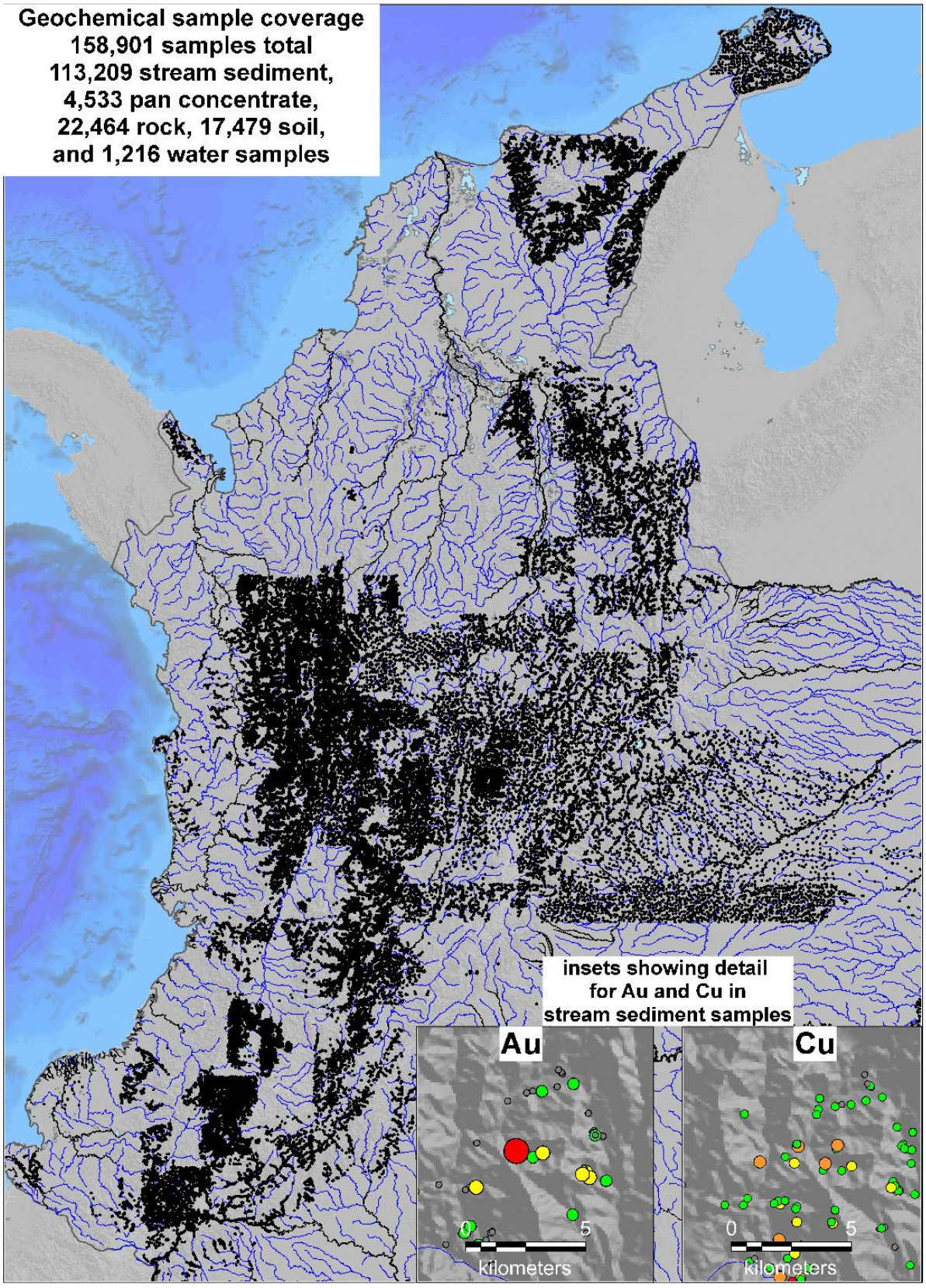
Recursos del Caribe (RdC-2025):

Using the SGC-2010 data set as a starting point (168,656 samples), RdC converted geochemical data (recorded as text) to numeric fields, eliminated 772 duplicate entries, corrected 542 data entry errors, entered 54,553 missing locations, eliminated 65,513 empty records (samples with no data), and eliminated three records that contained electron microprobe analyses of a single mineral. RdC then added the MDRU-2016 data (76,092 samples) and eliminated 4692 empty records.

The RdC-2025 Colombia Multielement Geochemical Data Set provides published geochemical data in an easy-to-read format that can be confidently applied to exploration. The RdC-2025 data set replaces the RdC-2020 data set and is designed to work either on a stand-alone basis or with ColombiaMap, a GIS data set for geology, mines and mineral occurrences, and land status (www.cbmap.net).

Sample types and elements analyzed are listed below; sample locations are displayed on the following page.

Sample Type	Number	Elements analyzed
Pan Concentrate	4,533	Ag, Al, As, Au, B, Ba, Be, Bi, Ca, Cd, Co, Cr, Cu, Fe, Ga, Hg, K, Mg, Mn, Mo, N, Na, Nb, Ni, P, Pb, Pd, Pt, S, Sb, Sc, Se, Sn, Sr, Th, Ti, Tl, U, V, W, Y, Zn, Zr
Rock	22,464	Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ge, Hf, Hg, Ho, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, N, Na, Nb, Ni, P, Pb, Pd, Pr, Pt, Rb, S, Sb, Sc, Se, Si, Sm, Sr, Ta, Tb, Te, Th, Ti, Tl, U, V, W, Y, Yb, Zn
Soil	17,479	Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ge, Hf, Hg, Ho, I, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, N, Na, Nb, Nd, Ni, Os, P, Pb, Pd, Pr, Pt, Rb, Re, Ru, S, Sb, Sc, Se, Si, Sm, Sn, Sr, Ta, Tb, Te, Th, Ti, Tl, Tm, U, V, W, Y, Yb, Zn, Zr
Stream Sediment	113,209	Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Ce, Cl, Co, Cr, Cs, Cu, Dy, Er, Eu, F, Fe, Ga, Gd, Ge, Hf, Hg, Ho, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, N, Na, Nb, Nd, Ni, P, Pb, Pd, Pr, Pt, Rb, Re, Ru, S, Sb, Sc, Se, Si, Sm, Sn, Sr, Ta, Tb, Te, Th, Ti, Tl, U, V, W, Y, Yb, Zn, Zr
Water	1,216	Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ge, Hf, Hg, Ho, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, Na, Nb, Nd, Ni, Pb, Pd, Pr, Pt, Rb, Re, Ru, Sb, Sc, Se, Si, Sm, Sn, Sr, Ta, Tb, Te, Th, Ti, Tl, U, V, W, Y, Yb, Zn, Zr


Geochemical sample coverage
158,901 samples total
113,209 stream sediment,
4,533 pan concentrate,
22,464 rock, 17,479 soil,
and 1,216 water samples
insets showing detail
for Au and Cu in
stream sediment samples
Au
Cu
0
5
kilometers
0
5
kilometers

The Colombia Multielement Geochemical Data Set includes the following subdirectories: RdC-2025, SGC-2009, SGC-2010, and MDRU-2016. The SGC and MDRU subdirectories contain published geochemical data from which the RdC-2025 data set was constructed. Edits, conversions, and corrections made during construction of the RdC-2025 data set are described below and are recorded in an MSAccess database (CO_Geochem_2020.accdb).

Servicio Geológico Colombiano (SGC-2009):

The SGC published a geochemical anomaly map of Colombia (Prieto et al., 2009) that displays anomalous areas for a variety of element combinations. The raw data used to plot the anomalous areas was not published. Recursos del Caribe (RdC) captured the anomalous areas as polygons and plotted the anomalies on a shaded relief base map.

Servicio Geológico Colombiano (SGC-2010):

The SGC released multielement geochemical data for 168,656 samples in 2015, including 72,536 stream sediment samples, representing a compilation of data from 150 surveys conducted through 2010. The SGC-2010 data set includes geochemical data for 103,143 samples; 65,513 records contain sample descriptions but no analytical data. The SGC data (in MS Access format) was available for several years as a free download (no license required) from the SGC website at: http://geoportal.sgc.gov.co/geoportalsgc/catalog/quicklink/basesDatosPublicacion.page. This link is no longer active and the SGC-2010 data set is no longer posted on the website.

Mineral Deposits Research Unit (MDRU-2016):

In 2016, the MDRU, in conjunction with the SGC, published an atlas that displays 56-element geochemical data for 76,092 stream sediment samples. The MDRU-2016 data set of 56 shapefiles, one for each element, was available for several years as a free download on the SGC website.

RdC consolidated the 56 separate MDRU-2016 shapefiles into a single file that contains all of the geochemical data for all of the samples. This format allows the user to review (and process) of all of the MDRU stream sediment data without opening multiple files.

Pricing

The Colombia Multielement Geochemical Data Set eliminates the time and cost involved in making the conversions and corrections that are needed to put the published data sets in a format that can be reliably applied to exploration. Prospective clients are invited to “test drive” the data set before purchase.

Colombia Multielement Geochemical Data Set: US$ 36,000.00

Colombia multielement geochemical data set description
Conversions and corrections made by RdC

Servicio Geológico Colombiano (SGC-2009) map:

The SGC published a geochemical anomaly map of Colombia (Prieto et al., 2009) that displays anomalous areas for a variety of element combinations. The raw data used to plot the anomalous areas was not published. Recursos del Caribe (RdC) captured the anomalous areas as polygons, in MapInfo tables or ArcGIS shapefiles, and plotted the anomalies on a shaded relief base map.

Servicio Geológico Colombiano (SGC-2010) data set:

The SGC released multielement geochemical data for 168,656 samples in 2015, including 72,536 stream sediment samples, representing a compilation of data from 150 surveys conducted through 2010. This data set includes geochemical data for 103,143 samples; 65,513 records are empty (contain sample descriptions but no analytical data). The SGC data (in MS Access format) was available for several years as a free download (no license required) from the SGC website at: http://geoportal.sgc.gov.co/geoportalsgc/catalog/quicklink/basesDatosPublicacion.page. This link is no longer active and the SGC-2010 data set is no longer posted on the website.

Six sample location fields are present in the original SGC-2010 data set containing up to three sample location coordinate pairs. (Colombia uses five different Transverse Mercator zones). RdC selected the appropriate coordinate pair for plotting according to the procedure described below. Unfortunately, many of the SGC-2010 sample locations are obviously out of place. In addition, the SGC-2010 data set provides geochemical data in text rather than in numeric format and, in accordance with local convention, commas in place of decimal points. RdC made no changes, in order to allow review of the original SGC-2010 data set.

Samples were located using the coordinate system recorded in the SGC-2010 data set under Origin_Coord_Original and locations as recorded under the Coord_Este_Original and Coord_Norte_Original fields. Samples with “Magnas Bogota” under Origin_Coord_Original were located using the Bogota Transverse Mercator zone, Magnas-Sirgas datum. They were coded as “BOG-M” in RdC’s added Coordsys_Used field. Samples with “Bogota” under Origin_Coord_Original were located using the Bogota Transverse Mercator zone, Bogota datum. They were and coded as “BOG-B” in the Coordsys_Used field. For samples recorded as “Este-Este,” “Este,” or “3-Este Este,” RdC used the East Transverse Mercator field, Bogota datum and coded them as “E-B.” Samples recorded as “2-Este Centro” and “Este-Centro” were located using the East Central Transverse Mercator Zone, Bogota datum and were coded as “EC-B.” Samples recorded as

"Magnas Oeste" were located using the West Transverse Mercator zone, Magna-Sirgas datum (only 4 samples). Some samples had no Coord_Este_Original or Coord_Norte_Original. Of these, 488 samples recorded as "4-Oeste" were located using the West Transverse Mercator zone, Bogota datum with X, Y locations from Coord_Este_Bogota and Coord_Norte_Bogota. They were coded as "BOG-B."

Samples recorded as "UTM" were located using UTM zone 18 North WGS 84 and were coded as "UTM18-W." Samples (3100) recorded as "UTM zona 18 N" or "UTM zona 18 North, WGS84" were obviously Lat/Long coordinates (not UTM) and so were located using Lat/Long WGS84 and were coded with "LL-W." Samples recorded as "WGS" were located using the Bogota Transverse Mercator Zone, Bogota datum and were coded as "BOG-B." Another 1933 records were built using UTM Zone 18 North, WGS84 datum and were coded as "UTM 18-W."

Samples with no entry for Origin_Coord_Original and with conflicting entries (4124) for Coord_x_Original versus Coord_x_Bogota were located using Coord_x_Bogota in the Bogota Transverse Mercator Zone and were coded as "BOG-B." Samples with no entry for Origin_Coord_Original and with the same (4129) entries for Coord_x_Original versus Coord_x_Bogota were also located using Coord_x_Bogota in the Bogota Transverse Mercator Zone and were also coded as "BOG-B."

Mineral Deposits Research Unit (MDRU-2016) data set:

The SGC published a Geochemical Atlas for Colombia (2016) that provides stream sediment geochemical anomaly maps at a scale of 1:6,000,000. The Atlas can be viewed interactively at: http://srvags.sgc.gov.co/JSViewer/Atlas_geoquimico_2016/. The original MDRU-2016 data set was available for several years for free download (https://miig.sgc.gov.co/Paginas/Resultados.aspx?k=atlas%20geoquimico%202016) and was updated in 2018 with additional samples.

The MDRU-2016 data set structure of 56 separate data files, one for each element, impedes review of the geochemical data for any given sample; a user has to open each of 56 separate files. The MDRU-2016 data set structure also impedes statistical analysis (e.g., correlation coefficients, principal components analysis, factor analysis). To remedy these issues, RdC built one file that contains all of the samples (as rows) and all of the geochemical data (as columns). The steps taken to create this "transposed" data set are as follows.

RdC first compiled a full list of unique samples using the PK_ID_MUES field in the original MDRU-2016 data set. This primary key was used to link MDRU sample locations to data stored in each of the 56 separate MDRU files. Sample locations and geochemical data (with fields for 77 elements) were saved into RdC's revised MDRU-2016 data set in LatLong using the WGS84 datum.

In RdC's revised MDRU-2016 data set, geochemical data excluded by MDRU is coded as -555; cells for which no geochemical data is available are coded as -999, and geochemical data reported as less than the detection limit is set equal to one half the detection limit. MDRU removed geochemical data judged as unreliable eliminating, for instance, half of all values for arsenic, 25% of all values for gold, and 75% of all values for silver. Most of the excluded geochemical values were generated using emission spectrometry. Excluded geochemical data is stored in the original MDRU geodatabase located in the subdirectory MDRU_db_orig.

The MDRU-2016 data set offered by RdC consists of a single file containing multielement geochemical data for 76,092 stream sediment samples, including 15,639 samples from the SGC-2010 data set updated with MDRU data. Each sample contains, in addition to geochemical analyses, an entry for sample type, sample description, observations. The original MDRU data set (MDRU_db_orig) contains additional detail for digestion and analytical technique for each element.

Recursos del Caribe, S.A. (RdC-2025) data set:

Recursos del Caribe S.A. conducted a thorough review of the SGC-2010 and MDRU-2016 data sets and, in 2020, converted SGC-2010 geochemical data from text to numeric format; made 54,553 corrections to incomplete and/or mistaken sample locations (32% of the data set); made 31,538,672 text edits, 542 corrections to incorrectly recorded analytical data; and resolved 772 duplicate data entries.

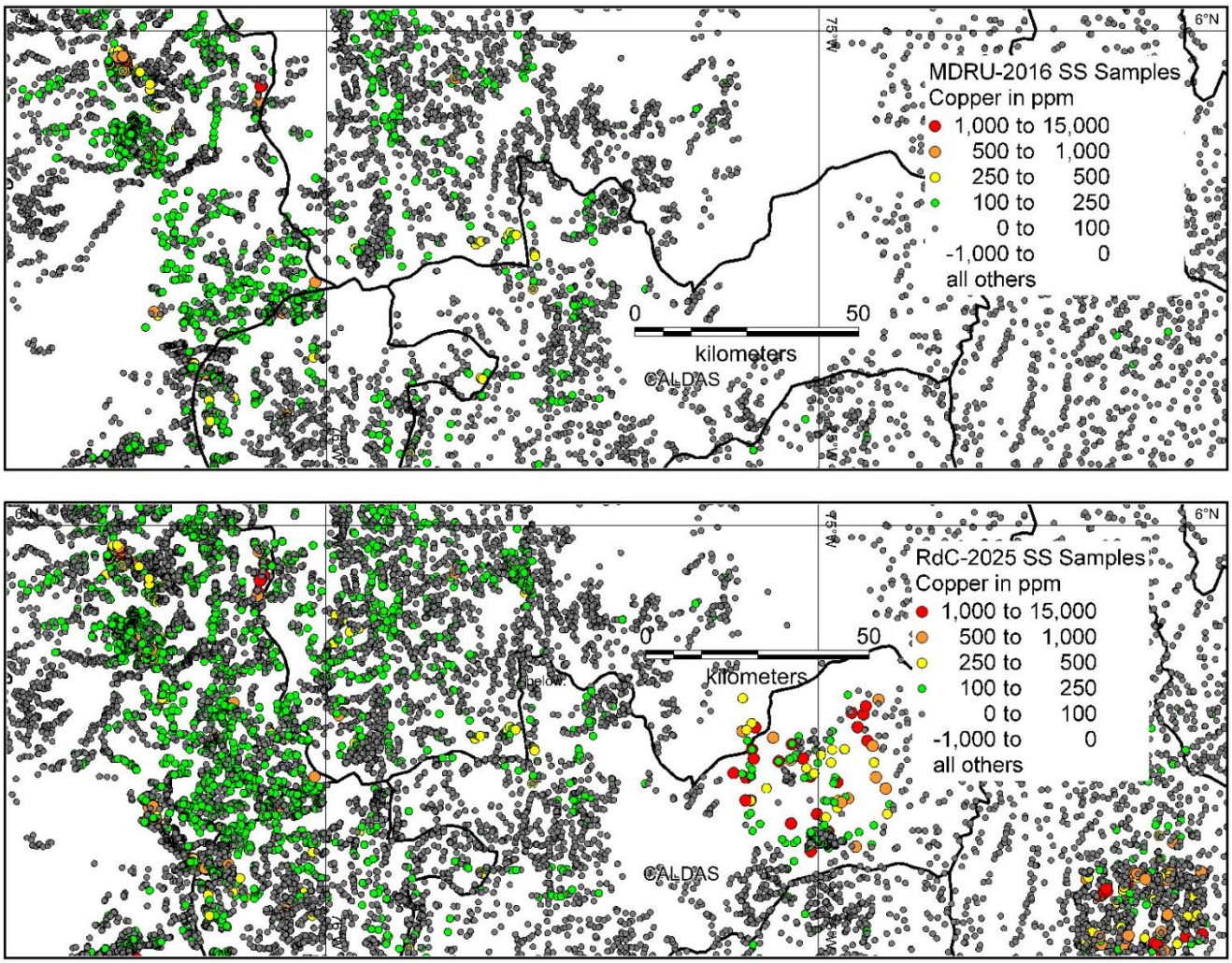
In 2025, RdC combined the MDRU-2016 and SGC-2010 data sets resulting in a single file that contains all of the available geochemical data. Empty records (no geochemical data) from the SGC and MDRU data sets were eliminated along with three samples containing microprobe analyses of individual minerals. A new field indicates the source of the data (SGC-2010 or MDRU-2016). Analytical data from MDRU-2016 supplemented with sample descriptions from SGC-2010 are coded as SGC-2010/MDRU-2016. New corrections, in addition to those made in 2020, were made to analytical data. Gold values were corrected for 656 samples from the SGC data set. Gold values were replaced with -888 for 4498 samples with high emission spec detection limits. These changes improve the reliability of the RdC-2025data set. Analytical values of -999 indicate no analysis was performed; values of -555 indicate an analysis was performed but was excluded by MDRU.

RdC then put the RdC-2025 data set into GIS format; added drainages, digital elevation models, shaded relief, infrastructure, a topographic map index, and departmental boundaries; and, finally, created geodatabases (ArcGIS) and workspaces (MapInfo) to facilitate display of the data.

The RdC-2025data set provides the compiled and corrected geochemical data in a single MS Access database, COGeochem.accdb. Another MS Access

database, COGeochem_2020.accdb, provides a detailed record of conversions, corrections and changes applied to the original SGC and MDRU data.

RdC chose not to eliminate thousands of emission spec analyses from the combined RdC-2025 data set. Many older emission spec values (e.g., copper) are useful for identifying anomalies even if the analyses are imprecise. The advantage of this approach, in addition to including data for 38,000 additional samples, is illustrated below. Significant stream sediment copper anomalies that appear when plotted using RdC-2025 data are lost if plotted using only the MDRU-2016 data.



Appendix: MSAccess tables and queries

CO_Geochem.accdb tables (the RdC-2025 data set)

GEOCHEM_ALL: A table containing the combined SGC-2010 and MDRU-2016 data along with edits, conversions, and corrections incorporated in 2020 and 2025.

MAPINFO_MAPCATALOG: Used by MapInfo to display "mappable" tables in the GEOCHEM_ALL data set. This table is used to plot points from ODBC (Open DataBase Connectivity) queries that are described below.

GEOCHEM_PAN_CONCENTRATE_QUERY: 4,533 pan concentrate samples
GEOCHEM_ROCK_QUERY: 22,464 rock samples
GEOCHEM_SOIL_QUERY: 17,479 soil samples
GEOCHEM_STREAM_SEDIMENT_QUERY: 113,209 stream sediment samples
GEOCHEM_WATER_QUERY: 1216 water samples

CO_Geochem_2020.accdb tables (the RdC-2020 data set)

1_TWO_ASSAYS_PER_ELEMENT_EDITS: two available geochem values (772)
2_FIND_AND_REPLACE_EDITS: commas and spaces in text fields, replaced (93,616)
3_DIVIDE BY 1000 EDITS: Changes to ppb values incorrectly recorded as ppm (978)
4_2017_2025 UPDATES: Replace high values generated by high detection limits (30,521)
5_DELETED_2025: Three samples with electron microprobe analyses of individual minerals.

ASSAYS_PAN_CONCENTRATE: analytical results for 4547 pan concentrate samples.
ASSAYS_ROCK: analytical results for 66,463 rock samples.
ASSAYS_SOIL: analytical results for 21,265 soil samples.
ASSAYS_STREAM_SEDIMENT: analytical results for 72,604 stream sediment samples.
ASSAYS_WATER: analytical results for 3774 water samples.

MAPINFO_MAPCATALOG: A table that MapInfo uses to display "mappable" tables in the COGeochem data set. This table is used to plot points from ODBC (Open DataBase Connectivity) tables that are described below.

MUESTRAS_ORIG_ASSAY: This table, in text format, replicates the 2010 SGC data set, along with minor changes made by the SGC through November 2017. The 2017 changes include 776 new values for As (4), Ag (50), Ti (2), Ta (111), Sr (4), Na (1), Mn (1), Mg (120), Fe (478) and Ca (1).

MUESTRAS_PROCESSED (numeric substitutions and data entry errors): This text table contains a "processed" version of the 'muestras_orig_assay' table. First, null values in the

original SGC assay fields are converted to -999, indicating that no analysis is available. Then, the following changes are made to the SGC analytical data:

1) A total of 772 instances for which a single sample contains two conflicting analytical values (e.g., Ag appears both as <0.2 ppm and as <0.2 ppb) are corrected (see the table: '1_TWO_ASSAYS_PER_ELEMENT_EDITS').
2) A total of 31,538,672 additional text edits, described below, are listed in the table: '2_FIND_AND_REPLACE_EDITS.'
    a) Cell values of -X and <X were treated as below a lower detection limit (ldl) of X and were assigned a cell value equal to 1/2 of the ldl;
    b) Cell values of >Y and GY were treated as greater than an upper detection limit (udl) of Y and are assigned a cell value of Y+1;
    c) Blanks (zero length strings) and entries such as trz, N, and ND were converted to a value of -999 to indicate no data for 30,223,825 cell values.
    d) Spaces in the data (e.g., 1 327) were replaced with commas (e.g., 1,327). Commas, marking the decimal point in accordance with convention in Colombia, were replaced with decimal points for 1,314,847 cell values.

The Find and Replace Edits table includes the following columns: **uid**: unique, auto-numbered ID (primary key); **cellvalue**: analytical data present as text in an assay field in the original SGC data set and in the table 'muestras_analisis2012;' **thefieldname**: the assay field in which the cellvalue was found in 'muestras_analisis2012'; **thetablename**: the name of the table in which the cellvalue was found. For analytical data from the SGC data set, thetablename is always 'muestras_analisis2012'; **valuetype1**: categorizes types of cellvalues. Used in organizing bulk edits to the data; **valuetype2**: categorizes types of cellvalues. Used in organizing bulk edits to the data; **newval**: a new cellvalue in the table 'muestras_processed' that replaces all instances of a particular cellvalue in the table 'muestras_analisis2012;' **updated_status**: indicates whether a cellvalue in a particular field has been changed. If set to 'OK', the cellvalue in the 'muestras_analisis2012' table has been searched for and changed to newval in the 'muestras_processed' table. To make additional changes to the "muestras_processed" table and, at the same time record those changes, first make the change to newval and replace 'OK' in the updated_status field with a zero-length string. Then use the form 'Update Assay Fields' to replace each newly edited cellvalue in the 'muestras_processed' table; **last_updated_to**: which newval was last used in replacement of cellvalue in the indicated field of 'muestras_processed'; **last_updated_date**: the date of last replacement of this cellvalue in the indicated field of 'muestras_processed', using the 'Update Assay Fields' form; **note**: explanation regarding the choice of newval for replacement of the original cellvalue; **temporder**: a field that was used during hand edits of the 'processing_edits_log' table.

3) Corrections, still in text format, in the '3_DIVIDE_BY_1000_EDITS' table include conversions for analytical data (to ppb) for analyses improperly recorded in the SGC data set as ppm. A total of 542 emission spec values were corrected. The SGC-2010 data set contains some (fewer than 1%) suspiciously high emission spec values for Ag, Au, As, B, Ba, Be, Bi, Br, Cd, Co, Cr, Cs, Cu, Ga, I, La, Mn, Mo, Nb, Ni, P, Pb, Pd, Pt, Sb, Sc, Se, Sn, Sr, Ta, Te, Ti, V, W, Y, Zn, and Zr. Rather than discard all emission spec values (a total of 702,320 analyses), the RdC-2020 data set leaves emission spec data for these elements untouched.

4) The table '4_2017_UPDATES' includes a total of 5379 edits made to gold values that are listed in the SGC data set as less than a high detection limit (<20, <10, <5, <2, <1, <0.5 and <0.1 ppm Au). All of these less-than-a-high-detection-limit values were changed to: -888 to avoid spurious gold anomalies that would result if less-than-a-high-detection-limit (dl) values were retained or changed to 1/2 of the dl. Similarly, for 6406 Ag analyses listed as less-than-a-detection-limit of 1, 5, 10 or 20 ppm, cellvalues were also changed to -888. A total of 2114 Cu and 5929 Mo values listed as less-than-a-detection-limit of 5 ppm or higher were changed to -888 along with 18 Pb, 9526 Zn and 319 Ni values listed as less-than-a-detection-limit of 20 ppm or higher. These add up to a total of 25,204 changes made to eliminate spurious anomalous values for Au, Ag, Cu, Pb, Zn and Ni related to high detection limits. Finally, a series of 138 data entry errors for cobalt (all with uniformly high values of 5%) were changed to 5 ppb (detection limit).

MUESTRAS_PROCESSED (sample location errors): The 'muestras_processed' table includes sample location coordinates in decimal latitude and longitude format (WGS 84 datum). These locations incorporate all of the corrections made by RdC to the original SGC sample locations. For instance, latitude longitude coordinates are missing for a total of 54,553 samples (32%) in the original SGC data set. Consequently, latitude longitude coordinates from the SGC data set were not used to locate samples. Instead, new decimal latitude and longitude coordinates (WGS84 datum) were generated using coordinates from the 'coord_este_original' and 'coord_norte_original' fields and the appropriate zone as recorded in the 'origen_coord_original' field. This step provided locations for two-thirds of the samples (114,387).

For the remaining (54,269) samples, the 'coord_original' fields were either empty or were incorrect. Most of these samples were successfully located by using one of remaining Bogota datum coordinate sets. All of these locations were then verified using location information provided in the fields for 'departamento' and 'plancha'. This process revealed 1132 data entry errors, now corrected, in the zone listed in the original SGC data set.

Finally, a total of 1571 sample locations (1% of the data set) plot within quadrangles that do not closely coincide with the quadrangles recorded in the original SGC data set. These locations are plotted, for now, near latitude 0 and longitude 0 until their actual locations can be resolved.

Five tables in the RdC-2020 data set incorporate the changes described above.

CO_Geochem_2020.accdb queries (the RdC-2020 dataset)

ASSAY_QUERY: takes cell values for each element in the 'muestras_processed' table and converts them from text into numeric entries. Data reported in different units is merged in the process. For instance, Cu recorded for some samples as Cu_percent, for others as Cu_ppm and for still others as Cu_ppb is merged into a single column (reported in units of ppb, ppm or pct as indicated).

PAN_CONCENTRATE_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 4547 pan concentrate samples.

ROCK_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 65,598 rock samples.

SOIL_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 21,265 soil samples.

STREAM_SEDIMENT_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 72,536 stream sediment samples.

WATER_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 3774 water samples.

INFORME_LIST: a subset of Assay_Query containing a list of analytical reports.

PROJECT_LIST: a subset of Assay_Query containing a list of projects; listed projects refer to reports that contain the raw analytical data.

MAKE_ASSAYS_PAN_CONCENTRATE_TABLE_QUERY: takes the **PAN_CONCENTRATE_ASSAYS_QUERY** and remakes the **ASSAYS_PAN_CONCENTRATE** table.

MAKE_ASSAYS_ROCK_ TABLE: takes the ROCK_ASSAYS_QUERY and remakes the ASSAYS_ROCK table.

MAKE_ASSAYS_SOIL_ TABLE: takes the SOIL_ASSAYS_QUERY and remakes the ASSAYS_SOIL table.

MAKE_ASSAYS_STREAM_SEDIMENT_ TABLE: takes the STREAM_SEDIMENT_ASSAYS_QUERY and remakes the ASSAYS_ATREAM_SEDIMENT table.

MAKE_ASSAYS_WATER_ TABLE: takes the WATER_ASSAYS_QUERY and remakes the ASSAYS_WATER table.