

ASSESSMENT REPORT GEOCHEMISTRY (GARDED)
DIGITAL RELEASE 2
May 1, 2024

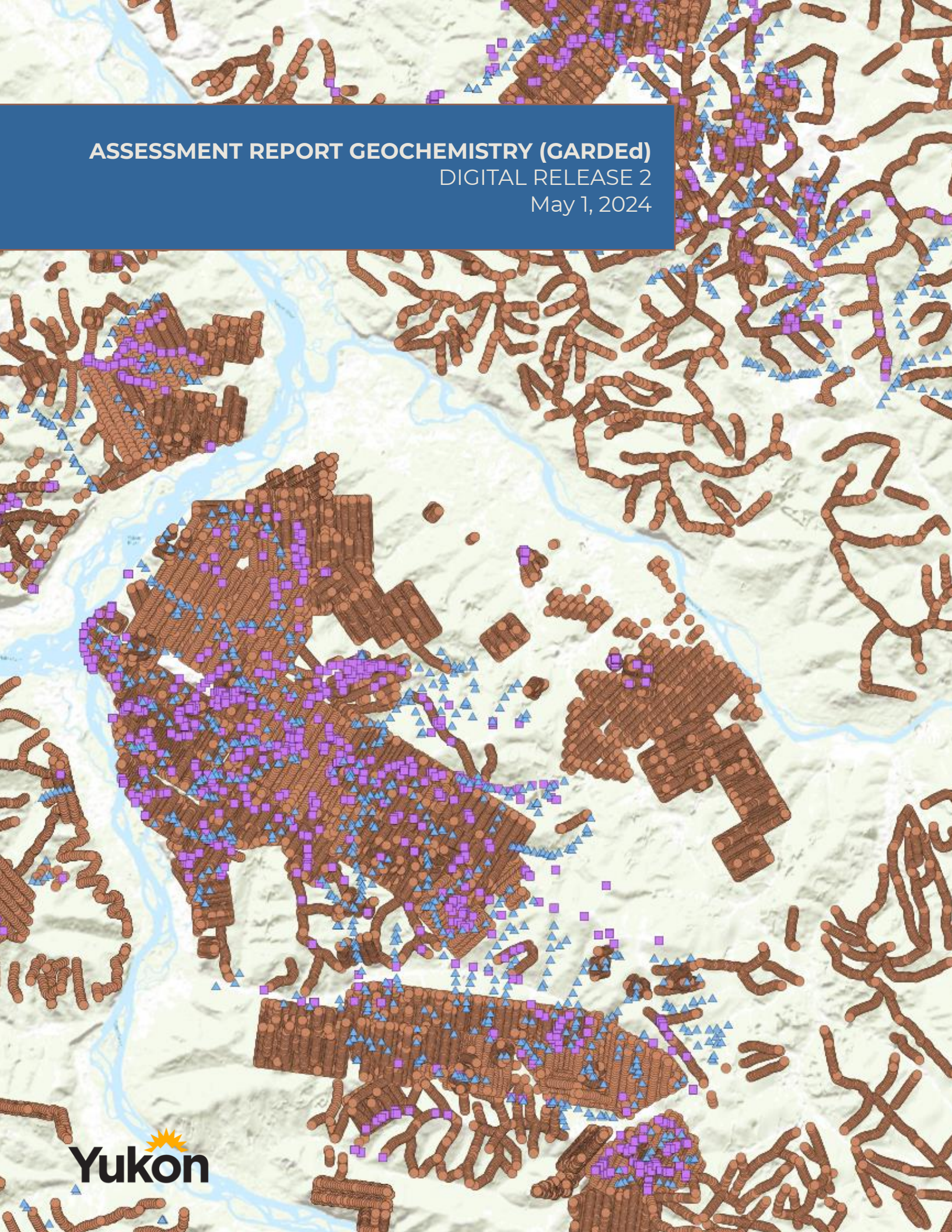


Table of Contents

1. Introduction	3
2. Database description	5
2.1. Data extraction	5
2.2. Geopackage instance of the GARDEd.....	5
2.3. Flat file version of data in the GARDE database.....	7
3. Accessing data in the geopackage instance of the GARDEd.....	8
3.1. DBever connection with geopackage database	8
3.2. ArcGIS Pro connection with GARDEd	9
3.2.1. Connecting ESRI geodatabase project of the GARDEd	9
3.2.2. ArcGIS Pro connection with geopackage instance.....	10
3.3. Data Preparation and Analysis Workflow in ArcGIS Pro, using an example of the ESRI geodatabase	10
3.3.1. Relate geochemical data to assessment reports	10
3.3.2. Data constraints and symbology.....	12
4. References	16
5. Appendix 1 - Field description of the geopackage	17
6. Appendix 2 - Description of views in the geopackage instance	20
6.1. 'long_table' view description	20
6.2. 'wide_table' view description	20
7. Appendix 3 - Sources of potential errors.....	21
8. Appendix 4 - field descriptions of ESRI geodatabase and CSV flat files.....	23

1. Introduction

This manual outlines how to access data stored in the Geochemical Assessment Report Data Extracted database (GARDEd), described in open report (Klyukin 2024). The GARDEd includes geochemical data from 408 assessment reports (ARs), describing 354,296 surficial samples. Data are mostly from assessment reports in the Dawson Range submitted since 2003 though a few are from elsewhere in the territory (Figure 1). Samples described in the database belong to one of four categories: rock, soil, stream sediments (referred to in GARDEd as silt), and vegetation. Most of the samples are analyzed for between 32 and 52 elements. While there are numerous analytical techniques used for sample characterization, most of the sample material was analyzed using ICP-MS (inductively coupled plasma-mass spectrometry) or ICP-AAS/AES (inductively coupled plasma-atomic absorption/emission spectroscopy). Samples were mostly prepared by weak acid digestion, with samples assayed for PGE and precious metals prepared by fire assay. Most of the data were extracted from assessment reports originally submitted in electronic form as opposed to scanned copies. The data have been extracted using a program developed by YGS. This program can be provided on a request, by contacting YGS-Minerals@yukon.ca or Yury.Klyukin@yukon.ca. In cases where users encounter inconsistencies between the data in GARDEd and the assessment report, we kindly request that any errors be reported to the same email addresses.

This manual describes the database (Section 2) behind the data in the accompanying release, includes an outline of the data model scheme used in the geopackage instance, and a guide on accessing, viewing, and extracting the data (Section 3). The data release associated with this manual is presented as three separate instances: the geopackage (gpkg) database; and flat file versions, consisting of comma separated value (CSV) and ArcGIS compatible featureclass database files. The geopackage instance contains the most complete version of the records and users familiar with tools needed to operate SQLite databases (of which geopackage is a geospatial derivative) are encouraged to connect GARDEd with the tool of their choice. Users not accustomed to databases are advised to use the flat file version of the database or use this manual to view and generate custom CSV files from the geopackage instance.

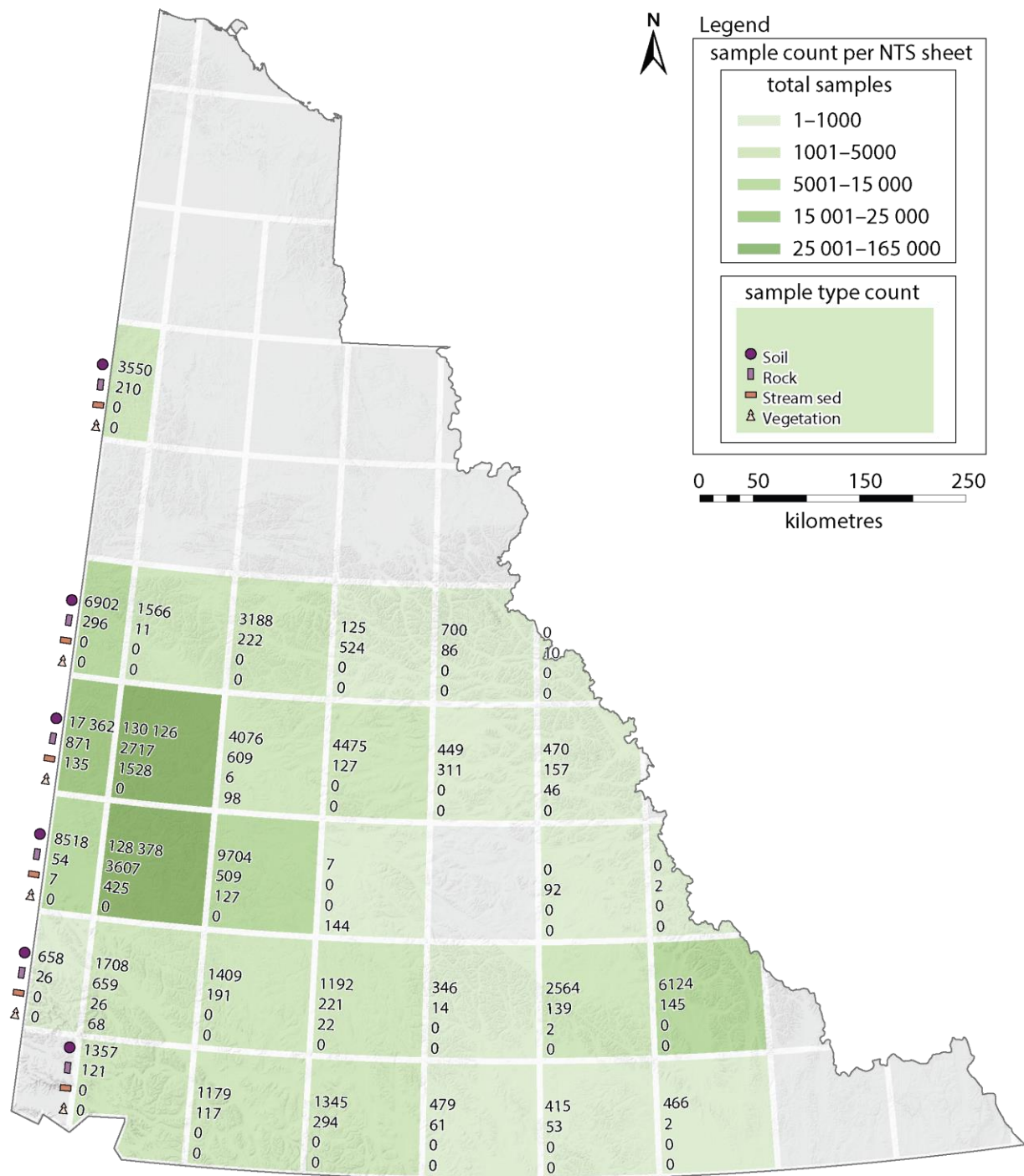


Figure 1 Spatial distribution of extracted samples data for each 250k NTS sheet, included in the present release. The colour gradation applies to total count of samples, labels within each NTS map sheet indicate the count for each sample type: soil, rock, stream sediment and vegetation.

2. Database description

The GARDE database consists of three instances: a geopackage, ESRI geodatabase and CSV files. The geopackage instance provides more comprehensive spatial and descriptive data and stores data as close to those reported in assessment reports as possible, while the CSV instance focuses on basic descriptive and analytical details and is more accessible.

2.1. Data extraction

The data in the GARDEd were extracted using a data extraction tool, which is currently being developed by the Yukon Geological Survey. Currently the tool does not work with scanned reports, and only works with tables stored in PDF, XLS/XLSX and CSV files. This tool was developed in Python and operates in a Jupyter notebook. In its current state, the tool requires a background in Python programming. Upon extraction of the data, the tool enforces some quality assurance and quality control measures, for example verifying that sample locations match with the assessment report footprint, or that sample type and date of collection are reported. Access to it can be provided on an individual basis.

2.2. Geopackage instance of the GARDEd

The geopackage instance of the GARDEd follows the simplified data model of Norris and Fortin (2019; Figure 2) matching that of the British Columbia Geological Survey assessment sourced surface sediment geochemical database. Similar to work of Norris and Fortin (2019), the geopackage instance of the GARDEd consists of three tables with the prefix code_, which describe laboratories, methods, and sample preparation techniques, and tables with the prefix data_ that store data related to specific aspects of the assessment reports. These include the format data as presented in assessment reports (data_source_info), metadata from assay certificates (data_cert), the analytical data from certificates (data_analyte), and a central table that stores spatial and descriptive data of the samples (data_sample). Figure 2 presents the entity-relationship diagram illustrating the linkage between tables within the geopackage instance of the database.

The tables shown in Figure 2 store data as close to the source as possible. The geopackage instance also contains dynamically generated tables (views), derived from data_ and code_ tables. The views functionality is described in Appendix 2, but because the views were used to generate ESRI geodatabase and CSV files, the following should be stated here about the views:

- The 'long_table' recalculates the concentration / limit of detection to metal form, if it is reported in oxides/carbonates, and changes the units from those originally used in the report to those specified in Table 1.
- The 'wide_table' view presents data from 'long_table' view pivoted on analytes and does not include values for minimum limit of detection. The abundance for each

element in this case is stored in a separate column, with units specified in the column header. Coordinate values in this view are rounded to 6 decimal places.

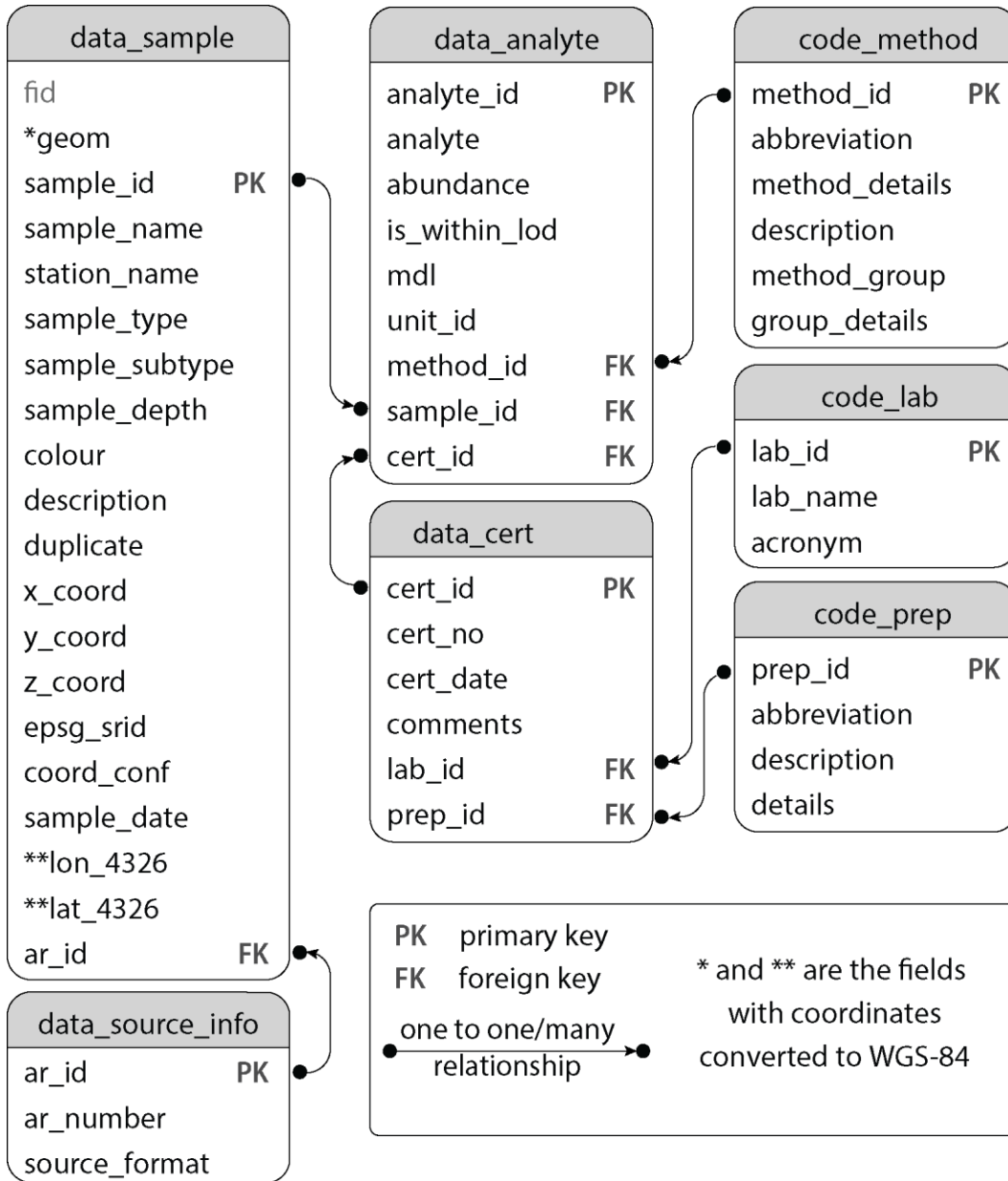


Figure 2 Entity-relationship diagram for the GARDEd (modified after Norris & Fortin, 2019). Modifications include added geom and lon_4326 lat_4326 fields for 'data_sample' table (marked by * and **) and the field 'is_within_lod' added to 'data_analyte' table. The geom field contains the longitude and latitude for sample location converted to WGS84 (EPSG 4326) formatted as a geospatial field, according to geopackage specification. Primary Key (PK) is a unique identifier for each record in a table. Foreign Key (FK) is a column in a table that refers to the primary key of another table. Descriptions for each field in the tables and types of data stored are in Appendix 1.

2.3.Flat file version of data in the GARDE database

For broader integration into standard GIS and analytical software, a denormalized flat file version of the database is provided in CSV file format and ESRI geodatabase. To do this, the data_ and code_ tables are combined for each sample type (Figure 2), using 'wide_table' view described in Section 2.2 and 6.1. CSV files included in flat file version exported such that each file contain only one type of the sample type: rock, soil, silt, and vegetation. If the same element in the sample was analyzed by multiple analytical techniques, additional measurements are stacked in adjacent rows, duplicating descriptive/spatial data. Compared to geopackage instance, the flat file version only includes basic sample and analytical details, and some data has been modified or removed to fit into a single table instead of a relational database. For instance, concentration units are standardized in the flat file version of the data to those in Table 1. These modifications mean that the data in the flat file version is not exactly the same as stored in data_ tables in the geopackage instance.

In the flat file version, the analytical technique is assigned a generalized description, identifying the type of analysis performed, using the value from the 'method_details' field (Appendix 1,

Table 7

Table 7

Table 7). For the measurements in assessment reports where major elements are reported in oxides form, the flat file version contains abundance and analytes recalculated to elemental form. Units of concentration are standardized to those shown in Table 1.

Table 1 concentration units used in dynamically generated tables in geopackage instance, as well as in CSV instance. CSV instance also have elements reported in oxide form, recalculated to the elemental form.

Unit	Elements
%	Al, Ca, Fe, K, Mg, Mn, Na, P, S, Si, Ti TC (<i>total carbon</i>), TS (<i>total sulfur</i>), LOI (<i>loss on ignition</i>)
ppb	Au
ppm	Ag, As, B, Ba, Be, Bi, Br, Cd, Ce, Cl, Co, Cr, Cs, Cu, Dy, Er, Eu, Ga, Gd, Ge, Hf, Hg, Ho, In, Ir, La, Li, Lu, Mo, Nb, Nd, Ni, Pb, Pd, Pr, Pt, Rb, Re, Rh, Sb, Sc, Se, Sm, Sn, Sr, Ta, Tb, Te, Th, Tl, Tm, U, V, W, Y, Yb, Zn, Zr

3. Accessing data in the geopackage instance of the GARDEd

The geopackage instance of the GARDEd stores data as similar to the source as possible, including sample description and original coordinates. While the geopackage instance of the GARDEd can be connected directly to GIS software, users may experience performance issues while working with a relational database from GIS software. To address this issue, we encourage the use of the denormalized flat file version of the database in either CSV or ESRI geodatabase (.gdb) format, or suggest users review the geopackage instance and extract a subset of the data from it using saved views.

Use of a database administration tool is beneficial for working with the geopackage instance, as it allows for efficient management and data analysis, as well as defining parameters for extracting a subset of data. For example, a user could export samples from a specific area, work with a limited list of chemical elements or only access soil samples for which assessment reports explicitly specified C horizon.

This section outlines the steps necessary to access the data using a database administration tool (DBeaver) and GIS software (ArcGIS Pro).

3.1.DBeaver connection with geopackage database

To review and access the most complete instance of the data stored in the GARDEd database, users are advised to use the open-source database administration software DBeaver. It can be downloaded and installed following the guide from official website (<https://dbeaver.io>). Once DBeaver is installed, a user can connect to the GARDEd database and access the data in it by following these steps:

1. Open DBeaver and click on the 'New Connection' button.
2. Select 'SQLite' from the list of available databases and click 'Next'.
3. Browse to the location of the GARDEd gpkg file and select it.
4. Click 'Test Connection' to ensure that the database is accessible and click 'Finish' to connect the database to DBeaver.
5. It will add the GARDEd to the 'Database Navigator' pane on the left side of the screen.
6. There are two options to access the data:
 - a. in the 'Tables' section each table from Figure 2 can be accessed individually, and values there can be filtered using headers of the tables, or
 - b. The 'Views' section has two dynamically created tables that concatenate most of the data for each sample in long and wide format – named 'long_table' or 'wide_table' correspondingly. 'long_table' view corresponds to the table where chemical data is presented with one column containing all the abundance values and other columns listing the elements, abundance units and other metadata regarding the analysis. 'wide_table' view present data the same way CSV instance present it – the abundance for each element stored in a separate column.

- c. Both of these dynamically generated tables ('long_table' and 'wide_table' views) have concentration values and units modified compared with those originally stored in 'data_analyte' table, such that units of concentration and elements match those specified in the Table 1.
- 7. Both 'long_table' and 'wide_table' views by default show coordinates as two separate fields for latitude/longitude: lat_4326 and lon_4326, where 4326 stands for CRS EPSG code for WGS 84 datum. However, these views can be modified to include the field 'geom' from 'data_sample' table. The 'geom' field show the sample location on the map preview in the value window of the DBeaver.
 - a. If the map preview is not available, then right click on any cell within 'geom' column (in data_sample table for example), and there follow to 'View/Format -> set 'geometry' format -> Geometry.
- 8. Map preview for 'wide_table' and 'long_table' views can be enabled by removing '--' characters from the line '--data_sample.geom' in script of the 'long_table' and from the line '--geom' in script for the 'wide_table'. Those scripts can be accessed by navigating to the tabs 'Properties -> Definition' in in each table.
- 9. Selecting and filtering of the data in views/tables in DBeaver is possible by clicking on a triangle next to the table header, opened in Data preview and defining filtering parameters.
- 10. Once data has been selected and filtered for export, user can use the dialog 'Export data' at the bottom of the screen to define the format of the output. Export of the data to CSV files allows the data to be easily picked up in a GIS software of the choice.

3.2.ArcGIS Pro connection with GARDEd

3.2.1. Connecting ESRI geodatabase project of the GARDEd

GARDEd release includes the ESRI geodatabase project, saved in the folder GARDEd_ESRI, prepared using ArcGIS Pro v 3.1.0 and CSV flat file version of the GARDEd. This section describes how to add to the project the Assessment Report Geochemistry (GARDEd).lyrx file, but the same instructions applicable to the older version of the layer file GARDEd.lyr:

1. Open ArcGIS Pro and create new project.
2. In the 'Catalog' pane on the right side of the screen, add the folder GARDEd_ESRI by right clicking on the Folders, 'Add folder connection' and navigate to the saved folder.
3. Browse to the location where the Assessment Report Geochemistry (GARDEd).lyrx file is saved.
4. Drag and drop the file into the 'Contents' pane of the ArcGIS Pro project.
5. The Assessment Report Geochemistry (GARDEd).lyrx file should now be loaded and visible in the 'Contents' pane.

You have now successfully loaded the GARDEd_ESRI project and run the GARDEd.lyr and Assessment Report Geochemistry (GARDEd).lyrx files in ArcGIS Pro.

3.2.2. ArcGIS Pro connection with geopackage instance

We recommend to either generate custom CSV files from geopackage instance using steps described in the Section 3.1 or take advantage of a project described in previous section (3.2.1). However, the geopackage instance can also be accessed from ArcGIS Pro, following these steps:

1. Open ArcGIS Pro and create new project.
2. In the 'Catalog' pane, click on the 'Project' tab, and right click on the 'databases' section, choose 'Add Database'. Find and select the GARDEd.gpkg file and connect it to the project.
3. Once connected, find GARDEd.gpkg in the databases of the 'Catalog' pane, and expand connected database.
4. Add 'data_sample' to the project by selecting 'Add to Project'.
5. To add geochemical data, add to the project the table 'data_analyte'.
6. To increase performance, each single element (analyte) should be added separately, through joining 'data_sample' with the subset of the 'data_analyte' table, created by exporting filtered data:
 - a. Right click on the 'data_analyte' table, select Data -> Export Table, defining the name of the output table and applying filter to the Analyte field being equal to the element of interest.
7. 'data_sample' layer added to the project in the Step 4 can be joined with newly created table using 'backup_sample_id' with 'sample_id' in filtered instance of the 'data_analyte' table. Resulting joined table can be used for data reviewal, conditional color coding of the XY points etc.

3.3.Data Preparation and Analysis Workflow in ArcGIS Pro, using an example of the ESRI geodatabase

3.3.1. Relate geochemical data to assessment reports

Once the GARDEd is loaded following the steps in section 3.2.1, it is useful to relate it to the Assessment Report Footprint feature class. This relate allows access to information about, and links to, the underlying assessment report from which the data was extracted. To add this feature class, we recommend doing the following :

1. **Add the Feature Classes to the Map:**

- If the feature classes are not already added to your map, add both the 'Assessment Report Geochemistry' and 'Assessment Report Footprints' feature classes to the map by navigating to their locations in the Catalog pane, right-clicking on them, and selecting 'Add to Current Map'.
- Download the Assessment Report Footprints feature class here:

https://map-data.service.yukon.ca/GeoYukon/Geological/Assessment_Report_Footprints/

2. **Enable Relationships View:**

- Right click on 'Assessment Report Geochemistry' in the contents pane

- Expand 'joins and relates'
- Select 'Add Relate '.

3. Define the Relationship:

- In the 'Add Relationship' dialog, select the 'Assessment Report Footprints' layer as the Related Table.
- Choose the 'REPORT_NUMBER' field as the Relationship Field.
- Click 'OK' to create the relationship.

Add Relate ? x

Layer Name or Table View
Assessment Report Geochemistry (GARDEd) [v] [Folder icon]

Input Relate Field
REPORT_NUMBER [v] [Gear icon]

Relate Table
Assessment Report Footprints [v] [Folder icon]

Output Relate Field
REPORT_NUMBER [v] [Gear icon]

Relate Name
AssessmentReports

Cardinality
One to many [v]

Figure 3: Add Relate settings

4. Review and Test the Relationship:

- Click on a map feature in the 'Assessment Report Geochemistry' layer.
- Related records will be highlighted in the 'Assessment Report Footprints' layer based on the matching 'REPORT_NUMBER' field.

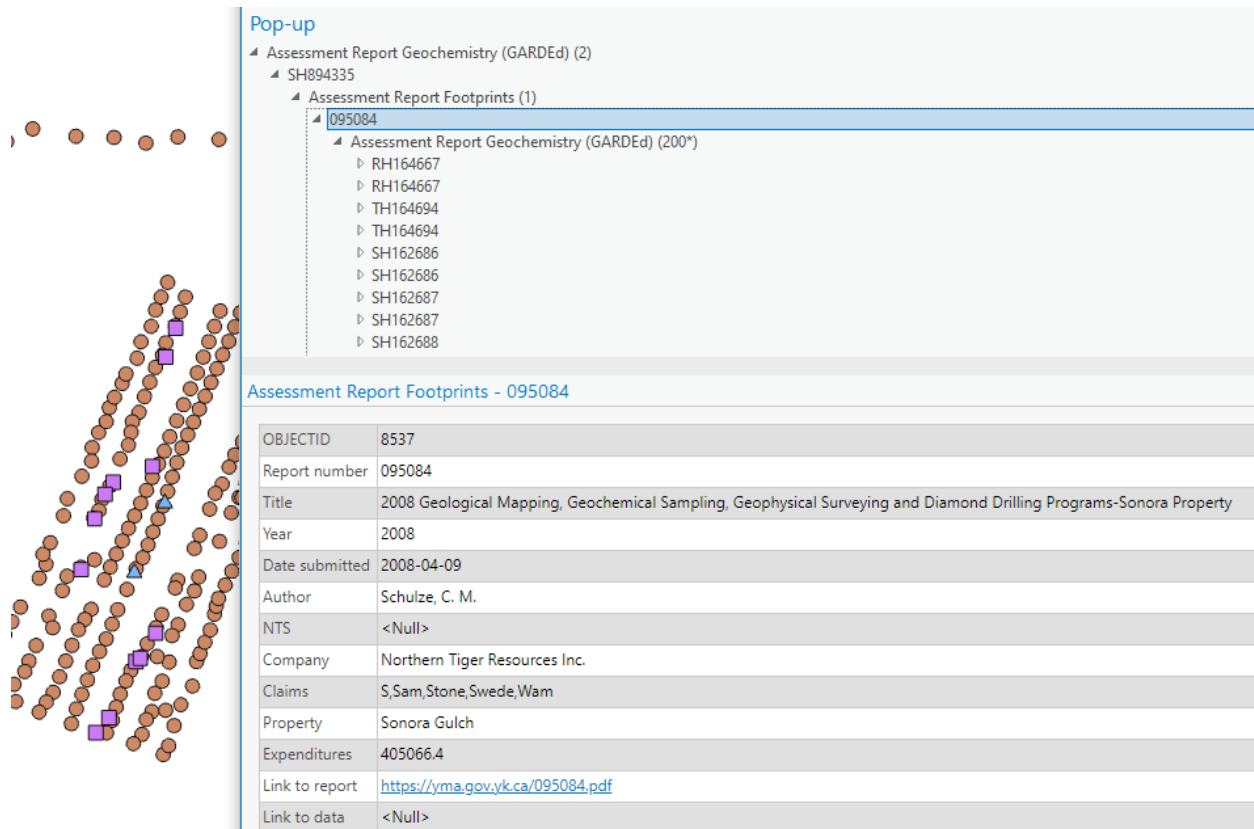


Figure 4: All of the data from the Assessment Report Geochemistry layer and the Assessment Report Footprints layers are now available in the pop-up when a data point is selected.

3.3.2. Data constraints and symbology

Given the database's substantial size, there might be a performance degradation, especially when engaging in computationally intensive tasks like custom symbology creation, label addition, analysis, or geoprocessing within ArcGIS Pro.

To mitigate these issues, we propose the following workflow:

1. Initial Data Exploration:

- Begin by reviewing the data without applying any symbology or modifications.

2. Data Constraint:

- Utilize a definition query to restrict the dataset based on specific attribute values, such as sample type.
- Spatially confine the data by defining an Area of Interest (AOI) and exporting a subset of data manageable by a given system.

3. Data Manipulation and Analysis:

- Proceed with necessary operations, such as symbolization based on geochemistry, gridding, or labeling.

Example

Below is an illustration on how to implement this workflow step by step in ArcGIS Pro to focus on soil samples, delineate an Area of Interest (AOI), and identify potential gold anomalies:

1. Initial Data Exploration:

- Load the dataset into ArcGIS Pro as instructed in the previous section (Figure 5)

2. Data Constraint:

- Use a definition query to restrict the dataset to only include soil samples (Figure 6)
 1. Open the layer Properties.
 2. Select 'Definition Query.'
 3. Create a 'New Definition Query.'
 4. Input the condition: SAMPLE_TYPE is equal to SOIL (SAMPLE_TYPE = SOIL).
- Note: Despite this filtering, the dataset may still be quite large, with over 200,000 soil samples.

3. Refining the Dataset with an AOI:

- Let's further refine the dataset to focus on soil samples east of the Yukon River and White River confluence (Figure 7):
 1. Select the target samples.
 2. Right click on the feature class in the contents pane.
 3. Select Data > Export Features to create new dataset with only the selected features.

4. Data Manipulation and Analysis:

- Customize the visualization of your data according to your needs. For example, let's symbolize the data to highlight areas with higher concentrations of gold-in-soil:
 1. Utilize graduated color and size formatting to emphasize regions with elevated gold concentrations (Figure 8).

Following these steps, allows effectively confine the dataset, delineate an AOI, and symbolize the results to identify potential gold anomalies within ArcGIS Pro.

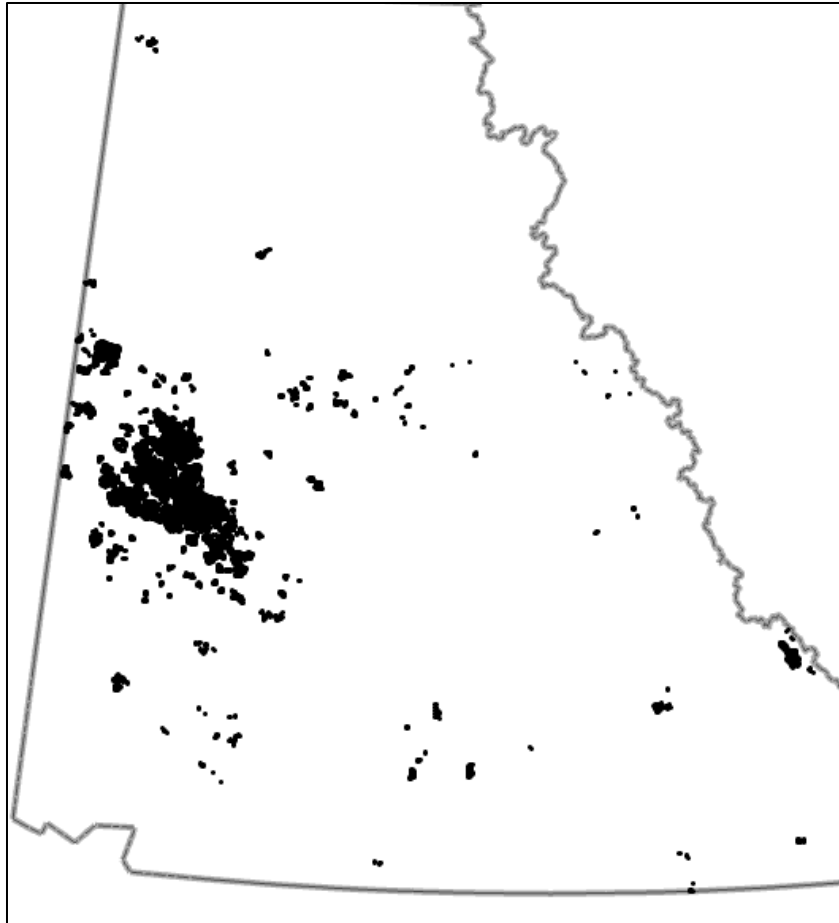


Figure 5: Unsymbolized Assessment Report Geochemistry dataset loaded into a new workspace in ArcGIS PRO

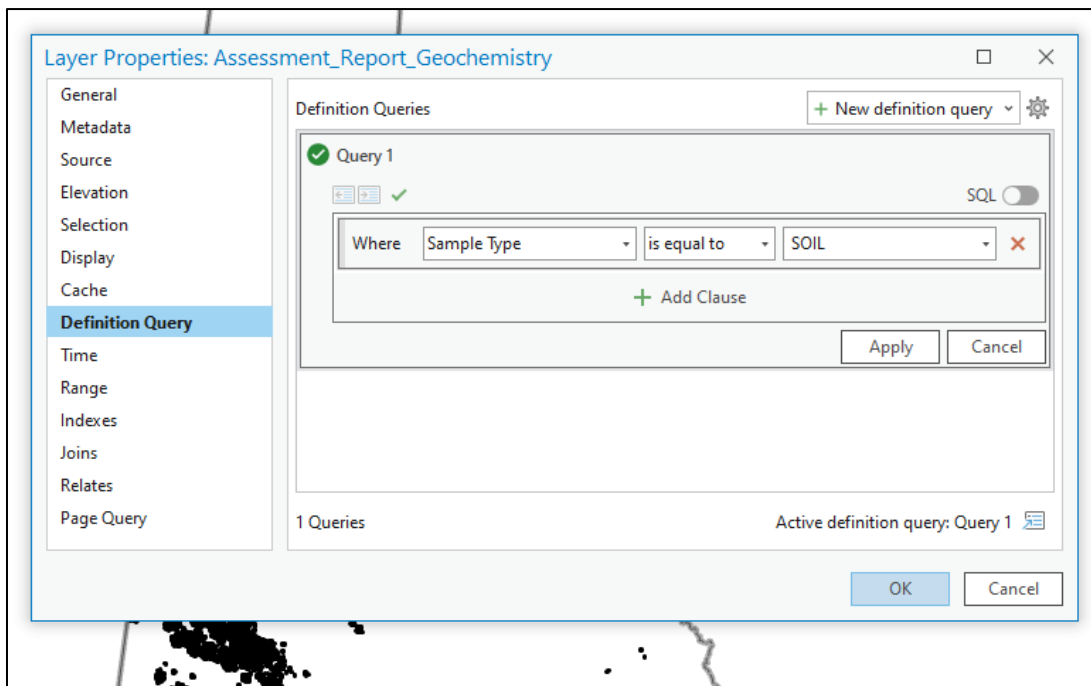


Figure 6: Constraining the dataset to Soil Samples using a definition query

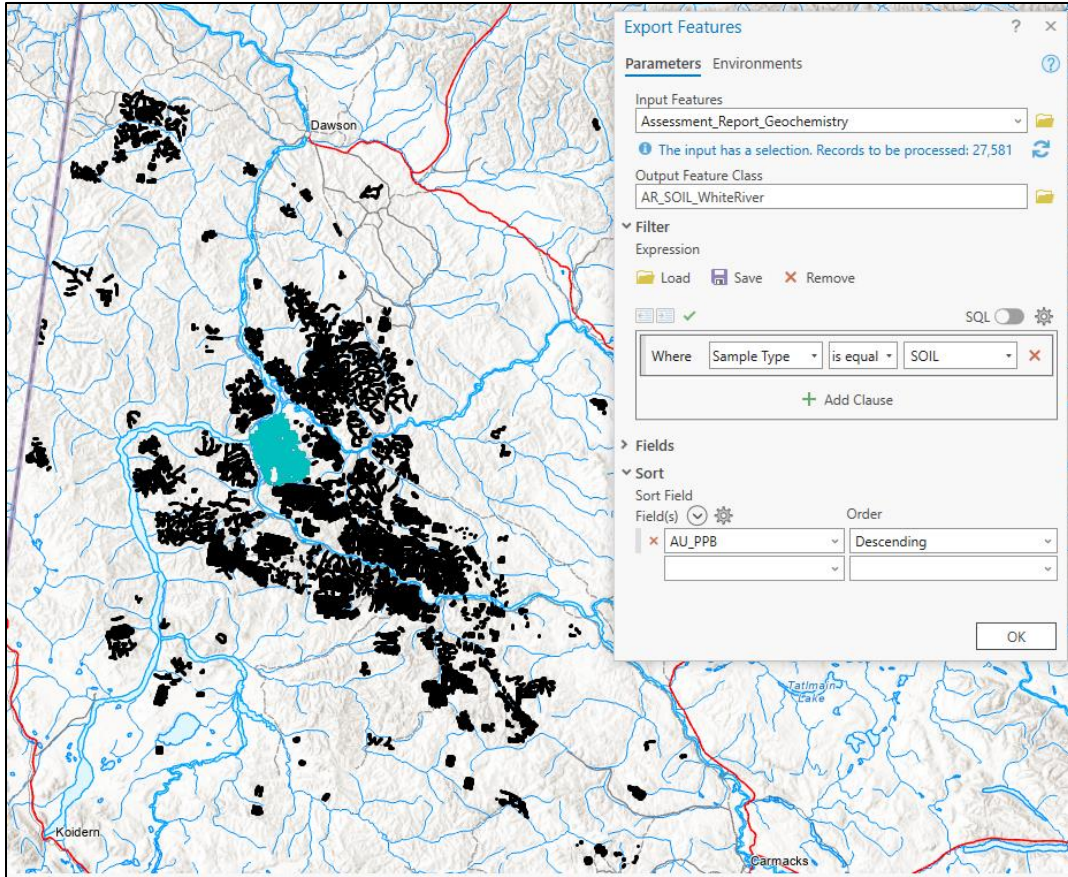


Figure 7: Constraining the data with an AOI and exporting to a new feature class

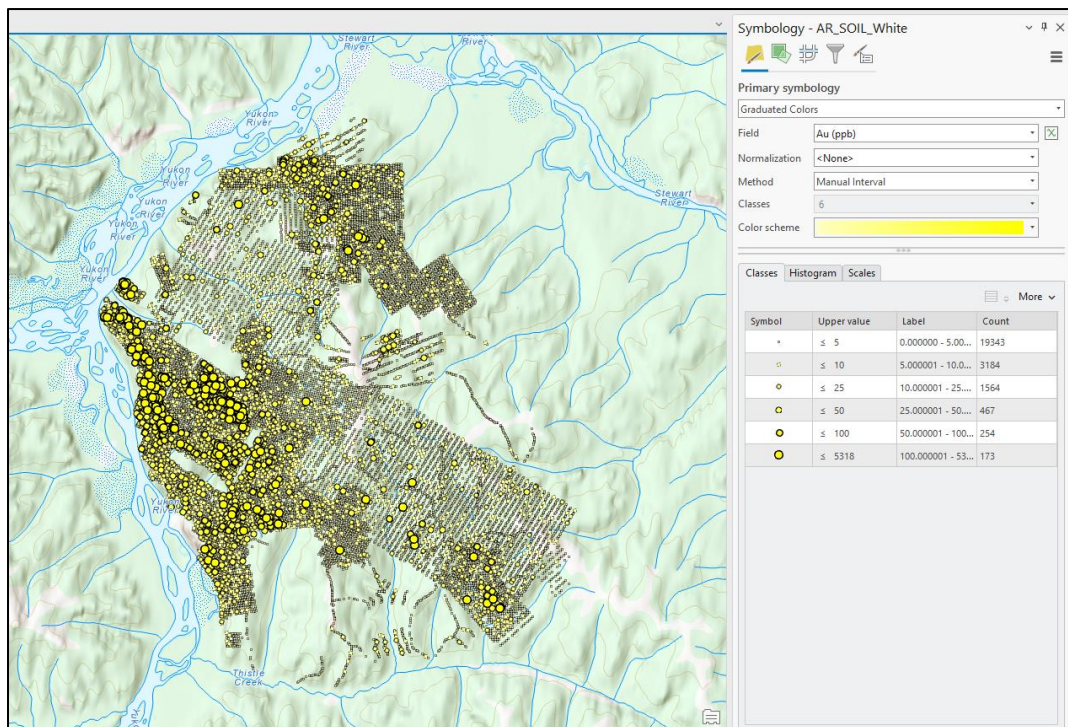


Figure 8: Au-in-Soil symbolized with graduated symbols and a yellow colour gradient.

4. References

Klyukin, Y., 2024. Surficial geochemical data extracted from assessment reports: Development and initial release of the database. In: Yukon Exploration and Geology Technical Papers 2023, L.H. Weston and Purple Rock Inc. (eds.), Yukon Geological Survey, p. 51–56.

Norris, J., & Fortin, G. (2019). Assessment report-sourced surface sediment geochemical database: Development and initial data release from the Interior Plateau. British Columbia Geological Survey GeoFile, 4(10).

5. Appendix 1- Field description of the geopackage

This appendix contains descriptions of the seven tables found in the geopackage instance of GARDEd. These tables, modified from Norris & Fortin (2019), describe the data types and brief description for each field. Fields in italics have been added to tables 2 and 3 and are not found in Norris and Fortin (2019).

Field Name	Type	Description
sample_id	Integer	Unique integer assigned to each sample.
sample_name	Text	Original sample name, as-is from the report.
station_name	Text	When applicable, alternate sample location name from the report.
sample_type	Text	Media sampled (e.g. soil, silt, rock, vegetation), as reported.
sample_subtype	Text	Additional details on sampled media (e.g. A Horizon, B Horizon, etc.).
sample_depth	Text	Sample depth (in cm), as-is from the source report.
colour	Text	Colour of the sampled material, as reported.
description	Text	Description of the sample material or location. This column contains notes from QA/QC.
duplicate	Text	Sample name of duplicate sample from the same location.
x_coord	Real	X coordinate of sample location, reported in original units as-is from the report.
y_coord	Real	Y coordinate reported as-is from the original report.
z_coord	Text	Elevation (in meters) of sample location if reported.
epsg_srid	Integer	European Petroleum Survey Group (EPSG) standard spatial reference system identifier (SRID) of original coordinates.
coord_conf	Text	Confidence in coordinate locations: 'h' for high confidence (samples where tabular coordinates were in the report, usually from GPS), 'm' for moderate confidence (sample coordinates digitized from georeferenced, originally digital maps), 'l' for low confidence (used for scanned sources, whether original coordinates were in tabular or map format).
sample_date	Text	Date the sample was collected in the field (YYYY-MM-DD), if reported.
ar_id	Integer	Links to data_source_info for assessment report data attributes
<i>geometry</i>	<i>Text</i>	<i>Stores WGS-84 (EPSG code 4326) coordinates, recalculated from those of x_coord, y_coord using epsg_srid, that are compliant with compliant with the geopackage specifications.</i>

Table 2 data_sample, is the central table for the geopackage instance of the GARDEd. It contains both spatial data and metadata for the samples. In cases where values were manually adjusted during data extraction to correct for errors, the value in the destination field would have the adjusted value, while the value in the 'Description' field would contain a note describing the changes compared to the assessment report. For example, if the 'x_coord' value in the report was originally recorded as 947000

but should have been 497000, the 'x_coord' field now displays the corrected value (497000), while the value in the 'Description' field would have the following note added: 'NOTE: x_coord corrected from reported 947000 to 497000.'

Field name	Type	Description
analyte_id	Integer	Unique integer id for each analytical determination.
Analyte	Text	Element e.g., Au, As.
abundance	Real	Reported abundance in original concentration units. Negative sign indicates abundance below detection limit results. Insufficient sample size and NULL values are excluded from the database.
is_within_lod	Text	Boolean variable to indicate if abundance is within range between lower and upper detection limits (1) or outside of it (0).
mdl	Text	Lower (minimum) detection limit.
unit_id	Text	Describe unit of measurement as reported in AR. For cases, where unit is not specified, it is indicated by a prefix 'infer_' to distinguish inferred units from those specified explicitly.
method_id	Integer	Links to code_method for details on the analytical method.
sample_id	Integer	Links to data_sample for sample location and metadata.
cert_id	Integer	Links to data_cert for details on the assay certificate the analysis is sourced from.

Table 3 data_analyte: this table contains analytical details and detection limits for all samples.

Field name	Type	Description
ar_id	Integer	Unique integer id for each assessment report
ar_number	Text	Assessment report number, used in YGS to track report.
source_format	Text	Format of the source report: 'digital' for report where digital data (spreadsheets, databases, etc.) were available; 'georef-ocr' for reports where sample locations were digitized from georeferenced images and results ocr'd from pdf; and 'ocr' for reports where results and locations were extracted from scanned version of reports by optical character recognition.

Table 4 data_source_info assessment report information, describing way how data was collected for each assessment report.

Field name	Type	Description
cert_id	Integer	Unique integer id for each certificate.
cert_no	Text	Lab-assigned assay certificate number.
cert_date	Date	Certificate date as reported by the lab (YYYY-MM-DD).
comments	Text	Additional notes on the certificate.
lab_id	Integer	Integer id for lab of origin links to code_lab table.
prep_id	Integer	Integer id linking to the prep_method table.

Table 5 data_cert stores assay certificates metadata.

Field name	Type	Description
lab_id	Integer	Unique integer id for each lab.
lab_name	Text	Name of the lab.
acronym	Text	Commonly used acronym for the lab.

Table 6 code_lab laboratory metadata.

Field name	Type	Description
method_id	Integer	Unique integer id for each method.
abbreviation	Text	Lab abbreviation for the method/package used.
description	Text	Description of the analytical package.
method_details	Text	Additional details on method.
method_group	Text	Grouping (coded) assigned to simplify data products.
group_details	Text	Grouping of methods.

Table 7 code_method metadata on analytical technique used by the laboratory.

Field name	Type	Description
prep_id	Integer	Unique integer id for each prep method.
abbreviation	Text	Method preparation code from the lab.
description	Text	Basic method description.
details	Text	Detailed method description.

Table 8 code_prep details on sample preparation methods.

6. Appendix 2- Description of views in the geopackage instance

6.1. 'long_table' view description

The code of the 'long_table' view contains SQLite scripts that perform a variety of calculations and transformations on data related to different analytes and units and reporting them in long format. This view used as an intermediate step before taking the data and pivoting it on the elements such that each element would receive a single column in 'wide_table'. The code joins data_ and code_ tables to retrieve additional information such as lab names, method groups, and certification details.

The calculations performed by a script include conversions from measurements and detection limits of those reported in oxides and carbonates to the pure metal form, as well as scaling concentration and detection limits for metals from original units to standardized units, mentioned in the Table 1. For example, there are calculations to convert Al₂O₃ abundance and minimum detection limit values reported in ppm to the Al concentration / mdl in % and replaces originally existing units with standardized.

6.2. 'wide_table' view description

The code of the 'wide_table' view creates a dynamic table by selecting specific columns from the 'long_table' and using the MAX function with conditional statements to pivot the data based on the values of 'elem_unit' – a field generated in 'long_table' view and populated by analyte and its unit separated by underscore character. MAX function used to select the maximum value for abundance in the case where the same element in the same sample of a given assessment report was analyzed by the same technique more than once. In cases where the same element in the sample of a given assessment report was analyzed by different techniques, grouping the data by values in ar_number', 'sample_name', and 'group_details' allows to stack additional measurements in adjacent rows, duplicating descriptive/spatial data. This view rounds the abundance to 5 decimal places, and also rounds latitude and longitude WGS-84-converted coordinates to 6 decimal places.

7. Appendix 3- Sources of potential errors

The GARDEd may have potential errors associated with inaccurate reporting of original data in assessment reports or introduced during data extraction and reporting. These errors can be categorized into two types: typos in existing data and omitting/including wrong data in the report. We encourage to report encountered errors between the data in GARDEd and the assessment report to YGS by contacting YGS-Minerals@yukon.ca or Yury.Klyukin@yukon.ca. Here are the specific errors and how they were addressed:

- Some reports contain data for samples that are not described in the report itself. In such cases, the existing data is captured in GARDEd and attributed to the corresponding assessment report, even if the report describes the results of a different project.
- Some coordinates for samples had obvious typos (swapped or missing digits, digits neighbouring on the keyboard entered instead of desired one). In such cases, the coordinates were corrected by inferring the type of error and/or approximating the sample location based on neighboring points. These cases are flagged in the description field of the 'data_sample' table, starting with 'NOTE:' in uppercase, followed by a description of the discovered error and the original coordinate from the report.
- Some reports may have potential errors related to elevation. In cases where the report authors included elevation in feet but stated it was in meters and discrepancies were obvious, such cases were corrected and noted in the description field of the 'data_sample' table starting with 'NOTE:' as in the previous case.
- The UTM zone for the samples in the report may be incorrect or missing. To address this, the samples were plotted on the map overlaying the polygon for the corresponding report, extracted from the Assessment Report Footprint database. For cases where a change in UTM zone allowed the samples to match the assessment report footprint, the UTM zone was reassigned to the samples and saved in the corresponding field ('epsg_srid' of the 'data_sample' table).
- Sample names stated in the assessment report's tables describing spatial and descriptive parameters did not match the sample names in the assay certificates.
 - Lists of unmatched samples were provided for manual review during extraction. Samples with missing chemical data were plotted on the map together with the assessment report footprint to evaluate potential data loss and relevance of these samples to the report.
 - In cases where a change in naming nomenclature was apparent (e.g., spatial table follows the pattern 'SID-00##', but assay certificate reports the sample results following the pattern 'SID00##'), the naming convention from the assay certificates was applied to the table with spatial and descriptive parameters.
 - Samples that had data only about spatial or chemical aspects were ignored and not saved in GARDEd.

- Some reports did not include original assay certificates, and chemical data was extracted from the tables provided by the report's authors. This may result in the following challenges:
 - The analytical techniques used by the laboratory are not included and will be indicated as 'unspecified' in GARDEd.
 - Limits of detection are omitted. In such cases, if the authors of the report indicated cases for below the detection limit as a negative value or using the character '<', the limit of detection (LOD) is inferred from that value for a given element in a given report and assigned to all samples of the same type. If the below limit of detection is reported as half of the LOD or zero, these values are captured in GARDEd as is.
 - Units for analyte abundance are not included (% , ppm). If the report provided maps with concentrations of certain elements and specified abundance units, those units were compared against the chemical data, and if the range is matching, then assigned to the respective element. If units were not stated anywhere in the report, the units were guessed by the person who extracted the data. Such cases are indicated by the 'Infer_' prefix in before the unit, saved in 'unit_id' field in 'data_analyte' table in the geopackage instance. Because original units not passed to other instances (CSV, ESRI geodatabase), this may introduce further errors in flat file versions, as they contain recalculated abundance values.
- In most cases, the coordinates of the samples were extracted from tabulated data, in some cases, sample labels and positions were recovered through georeferencing. These reports are indicated by the 'georef-ocr' value in 'data_source_info' and by the 'coordinate confidence' value 'm' in the flat file instance (Figure 2 or Appendix 1 Table 4). While the sample's location was reviewed and compared against the assessment report footprint, and sample names must match those for samples in the assay certificates, georeferencing may introduce errors, either in the sample names or in coordinates. For example, the sample's order may not match that on the original map, or georeferenced map was done with error in one of the axes.
- Some reports included in GARDEd release were tested for feasibility of the optical character recognition (OCR), using Tesseract (<https://github.com/tesseract-ocr/tesseract>). Numerical format was enforced on the fields in the tables that are expected to bear such type of data (coordinates, abundance). Results of the OCR were manually reviewed and corrected, however, by there might contain potential errors like wrong digit recognized or misplaced(absent) decimal point.

8. Appendix 4- field descriptions of ESRI geodatabase and CSV flat files

FIELD	Alias	Type	Attribute Description
AR_NUMBER	Report Number	Text	Assessment report number
SAMPLE_NAME	Sample Name	Text	Original sample name, as-is from the report.
SAMPLE_TYPE	Sample Type	Text	Media sampled (e.g. soil, silt, rock, vegetation), as reported.
LONGITUDE_DD	Longitude	Double	Longitude extracted from the geometry field.
LATITUDE_DD	Latitude	Double	Latitude extracted from the geometry field.
COORDINATE_CONFIDENCE	Coordinate Confidence	Text	Confidence in coordinate locations: 'h' for high confidence (sample coordinates were reported in tabular format, usually from GPS), 'm' for moderate confidence (sample coordinates digitized from georeferenced, originally digital maps), 'l' for low confidence (used for scanned sources, whether original coordinates were in tabular or map format).
ANALYTICAL_METHOD_FAMILY	Analytical Method Family	Text	Type of analytical method(s) used during assay process.
AU_PPB	Au (ppb)	Float	Gold content obtained through assay testing, expressed in parts per billion (ppb).
AL_PCT	Al (%)	Float	Aluminum content obtained through assay testing, expressed in percent.
CA_PCT	Ca (%)	Float	Calcium content obtained through assay testing, expressed in percent.
FE_PCT	Fe (%)	Float	Iron content obtained through assay testing, expressed in percent.
K_PCT	K (%)	Float	Potassium content obtained through assay testing, expressed in percent.
MG_PCT	Mg (%)	Float	Magnesium content obtained through assay testing, expressed in percent.
MN_PCT	Mn (%)	Float	Manganese content obtained through assay testing, expressed in percent.
NA_PCT	Na (%)	Float	Sodium content obtained through assay testing, expressed in percent.
P_PCT	P (%)	Float	Phosphorus content obtained through assay testing, expressed in percent.
S_PCT	S (%)	Float	Sulfur content obtained through assay testing, expressed in percent.
SI_PCT	Si (%)	Float	Silicon content obtained through assay testing, expressed in percent.
TI_PCT	Ti (%)	Float	Titanium content obtained through assay testing, expressed in percent.
AG_PPM	Ag (ppm)	Float	Silver content obtained through assay testing, expressed in parts per million (ppm).
AS_PPM	As (ppm)	Float	Arsenic content obtained through assay testing, expressed in parts per million (ppm).

FIELD	Alias	Type	Attribute Description
B_PPM	B (ppm)	Float	Boron content obtained through assay testing, expressed in parts per million (ppm).
BA_PPM	Ba (ppm)	Float	Barium content obtained through assay testing, expressed in parts per million (ppm).
BE_PPM	Be (ppm)	Float	Beryllium content obtained through assay testing, expressed in parts per million (ppm).
BI_PPM	Bi (ppm)	Float	Bismuth content obtained through assay testing, expressed in parts per million (ppm).
BR_PPM	Br (ppm)	Float	Bromine content obtained through assay testing, expressed in parts per million (ppm).
CD_PPM	Cd (ppm)	Float	Cadmium content obtained through assay testing, expressed in parts per million (ppm).
CE_PPM	Ce (ppm)	Float	Cerium content obtained through assay testing, expressed in parts per million (ppm).
CL_PPM	Cl (ppm)	Float	Chlorine content obtained through assay testing, expressed in parts per million (ppm).
CO_PPM	Co (ppm)	Float	Cobalt content obtained through assay testing, expressed in parts per million (ppm).
CR_PPM	Cr (ppm)	Float	Chromium content obtained through assay testing, expressed in parts per million (ppm).
CS_PPM	Cs (ppm)	Float	Cesium content obtained through assay testing, expressed in parts per million (ppm).
CU_PPM	Cu (ppm)	Float	Copper content obtained through assay testing, expressed in parts per million (ppm).
DY_PPM	Dy (ppm)	Float	Dysprosium content obtained through assay testing, expressed in parts per million (ppm).
ER_PPM	Er (ppm)	Float	Erbium content obtained through assay testing, expressed in parts per million (ppm).
EU_PPM	Eu (ppm)	Float	Europium content obtained through assay testing, expressed in parts per million (ppm).
GA_PPM	Ga (ppm)	Float	Gallium content obtained through assay testing, expressed in parts per million (ppm).
GD_PPM	Gd (ppm)	Float	Gadolinium content obtained through assay testing, expressed in parts per million (ppm).
GE_PPM	Ge (ppm)	Float	Germanium content obtained through assay testing, expressed in parts per million (ppm).
HF_PPM	Hf (ppm)	Float	Hafnium content obtained through assay testing, expressed in parts per million (ppm).
HG_PPM	Hg (ppm)	Float	Mercury content obtained through assay testing, expressed in parts per million (ppm).
HO_PPM	Ho (ppm)	Float	Holmium content obtained through assay testing, expressed in parts per million (ppm).
IN_PPM	In (ppm)	Float	Indium content obtained through assay testing, expressed in parts per million (ppm).
IR_PPM	Ir (ppm)	Float	Iridium content obtained through assay testing, expressed in parts per million (ppm).

FIELD	Alias	Type	Attribute Description
LA_PPM	La (ppm)	Float	Lanthanum content obtained through assay testing, expressed in parts per million (ppm).
LI_PPM	Li (ppm)	Float	Lithium content obtained through assay testing, expressed in parts per million (ppm).
LU_PPM	Lu (ppm)	Float	Lutetium content obtained through assay testing, expressed in parts per million (ppm).
MO_PPM	Mo (ppm)	Float	Molybdenum content obtained through assay testing, expressed in parts per million (ppm).
NB_PPM	Nb (ppm)	Float	Niobium content obtained through assay testing, expressed in parts per million (ppm).
ND_PPM	Nd (ppm)	Float	Neodymium content obtained through assay testing, expressed in parts per million (ppm).
NI_PPM	Ni (ppm)	Float	Nickel content obtained through assay testing, expressed in parts per million (ppm).
PB_PPM	Pb (ppm)	Float	Lead content obtained through assay testing, expressed in parts per million (ppm).
PD_PPM	Pd (ppm)	Float	Palladium content obtained through assay testing, expressed in parts per million (ppm).
PR_PPM	Pr (ppm)	Float	Praseodymium content obtained through assay testing, expressed in parts per million (ppm).
PT_PPM	Pt (ppm)	Float	Platinum content obtained through assay testing, expressed in parts per million (ppm).
RB_PPM	Rb (ppm)	Float	Rubidium content obtained through assay testing, expressed in parts per million (ppm).
RE_PPM	Re (ppm)	Float	Rhenium content obtained through assay testing, expressed in parts per million (ppm).
RH_PPM	Rh (ppm)	Float	Rhodium content obtained through assay testing, expressed in parts per million (ppm).
SB_PPM	Sb (ppm)	Float	Antimony content obtained through assay testing, expressed in parts per million (ppm).
SC_PPM	Sc (ppm)	Float	Scandium content obtained through assay testing, expressed in parts per million (ppm).
SE_PPM	Se (ppm)	Float	Selenium content obtained through assay testing, expressed in parts per million (ppm).
SM_PPM	Sm (ppm)	Float	Samarium content obtained through assay testing, expressed in parts per million (ppm).
SN_PPM	Sn (ppm)	Float	Tin content obtained through assay testing, expressed in parts per million (ppm).
SR_PPM	Sr (ppm)	Float	Strontium content obtained through assay testing, expressed in parts per million (ppm).
TA_PPM	Ta (ppm)	Float	Tantalum content obtained through assay testing, expressed in parts per million (ppm).
TB_PPM	Tb (ppm)	Float	Terbium content obtained through assay testing, expressed in parts per million (ppm).
TE_PPM	Te (ppm)	Float	Tellurium content obtained through assay testing, expressed in parts per million (ppm).

FIELD	Alias	Type	Attribute Description
TH_PPM	Th (ppm)	Float	Thorium content obtained through assay testing, expressed in parts per million (ppm).
TL_PPM	Tl (ppm)	Float	Thallium content obtained through assay testing, expressed in parts per million (ppm).
TM_PPM	Tm (ppm)	Float	Thulium content obtained through assay testing, expressed in parts per million (ppm).
U_PPM	U (ppm)	Float	Uranium content obtained through assay testing, expressed in parts per million (ppm).
V_PPM	V (ppm)	Float	Vanadium content obtained through assay testing, expressed in parts per million (ppm).
W_PPM	W (ppm)	Float	Tungsten content obtained through assay testing, expressed in parts per million (ppm).
Y_PPM	Y (ppm)	Float	Yttrium content obtained through assay testing, expressed in parts per million (ppm).
YB_PPM	Yb (ppm)	Float	Ytterbium content obtained through assay testing, expressed in parts per million (ppm).
ZN_PPM	Zn (ppm)	Float	Zinc content obtained through assay testing, expressed in parts per million (ppm).
ZR_PPM	Zr (ppm)	Float	Zirconium content obtained through assay testing, expressed in parts per million (ppm).
LOI_PCT	Loss On Ignition (%)	Float	Loss on ignition, expressed as a percentage of the initial sample weight.
TS_PCT	Total Sulphur (%)	Float	Total Sulphur, expressed as a percentage of the initial sample weight.
TC_PCT	Total Carbon (%)	Float	Total Carbon, expressed as a percentage of the initial sample weight.

9. Appendix 5 – changes compared to the release 1, of Feb 5, 2024

After the confidentiality period elapsed and data extraction for older reports recommenced, the latest release encompasses data from 354,296 surficial geochemical samples, an increase from the previously described 307,311 samples. This data was extracted from 408 assessment reports, reflecting a significant rise compared to the 311 reports released in February.