Regional Geochemical Survey (RGS) data update and release using the newly developed RGS database

Tian Han, and Alexei S. Rukhlov

Ministry of Energy, Mines and Petroleum Resources

BRITISH COLUMBIA

British Columbia Geological Survey GeoFile 2017-11

Ministry of Energy, Mines and Petroleum Resources
Mines and Mineral Resources Division
British Columbia Geological Survey

Regional Geochemical Survey (RGS) data update and release using the newly developed RGS database

Tian Han
Alexei S. Rukhlov

Regional Geochemical Survey (RGS) data update and release using the newly developed RGS database

Tian Han[1a], and Alexei S. Rukhlov[1]

[1]British Columbia Geological Survey, Ministry of Energy, Mines and Petroleum Resources, Victoria, B.C., V8W 9N3
[a]corresponding author: tian.han@gov.bc.ca

Abstract
This GeoFile is a release of the most current and complete province-wide geochemical data set collected under the Regional Geochemical Survey (RGS) program. This data set was generated by integrating RGS data previously published by the British Geological Survey and Geoscience BC. It was compiled from 111 original sources with 64,828 samples and about 5 million determinations analyzed using 18 methods in 18 laboratories. For the ease of use and consistency with previously published data, the data set is in a flat tabular format. Behind the scenes is the newly developed RGS database that operates through a semi-automated process for data compilation, quality control, upload, and product generation. This RGS database provides a solution for consistent, standardized, and long-term data storage and for timely data update and regular data release. The RGS data set is just one of many potential data products that can be generated from the RGS database.

Keywords: Stream-sediment geochemistry, Regional Geochemical Survey (RGS), geochemical database, data product, logical data model, quality control, data update, mineral exploration

1. Introduction

The pioneering work of Vernadsky (1924) and Fersman (1939) established the foundations of applying geochemistry, including stream-sediment, soil, water, and vegetation surveys to mineral exploration. In North America, stream-sediment geochemistry was first applied by Lovering et al. (1950), who investigated the downstream dispersion of copper, lead, and zinc from the San Manuel deposit in Arizona (Meyer et al., 1979). Since then, the geochemistry of stream and lake sediments, along with other media (e.g., moss-mat sediment, till, and vegetation), has become an increasingly effective tool for mineral exploration, bedrock mapping, and environmental monitoring.

Since 1976, the Geological Survey of Canada, the British Columbia Geological Survey (BCGS), and Geoscience BC have managed regional geochemical survey (RGS) programs in British Columbia. The analytical determinations from these programs provide baseline information on the dispersion of elements in drainage systems of the province. The surveys, which sampled stream, lake and moss-mat sediment and water, were originally part of the Uranium Reconnaissance Program (URP) and later, the National Geochemical Reconnaissance (NGR; Darnley et al., 1975; Carter et al., 1978; Friske and Hornbrook, 1991). Province-wide RGS compilations were released by Lett (2005), Lett (2011), and Rukhlov and Naziri

(2015). The latest of these compilations comprises data from 64,828 stream-, lake-, and moss-sediment and water samples, including up to 63 determinations for sediment samples and up to 78 for water samples. Jackaman (2017) released results from archived RGS samples that were re-analyzed for 56 minor and trace elements using inductively coupled plasma-mass spectrometry (ICP-MS) following aqua regia digestion. RGS sampling extends across about 80% of the province at a density of one sample per 7 to 13 km^2 (Fig. 1).

Results from the RGS are subject to management operations such as updating, supplementing, integrating, and validating, and therefore should be treated as a living dataset. We built a new RGS database as a long-term repository to facilitate these operations with the following processes: 1) developing a logical data model based on the intrinsic relationships between basic data entities and essential attributes of the RGS data; 2) building a MS Access relational database using this model; 3) populating the database with data compiled by Rukhlov and Naziri (2015); 4) integrating data from re-analyzed archived samples (Jackaman, 2017) into the database; and 5) automating in-bound and out-bound data flow for loading the database and generating data products. We thus established a streamlined process and an operational database system to manage, maintain, update, and distribute province-wide RGS data routinely and consistently. Herein we: provide summary
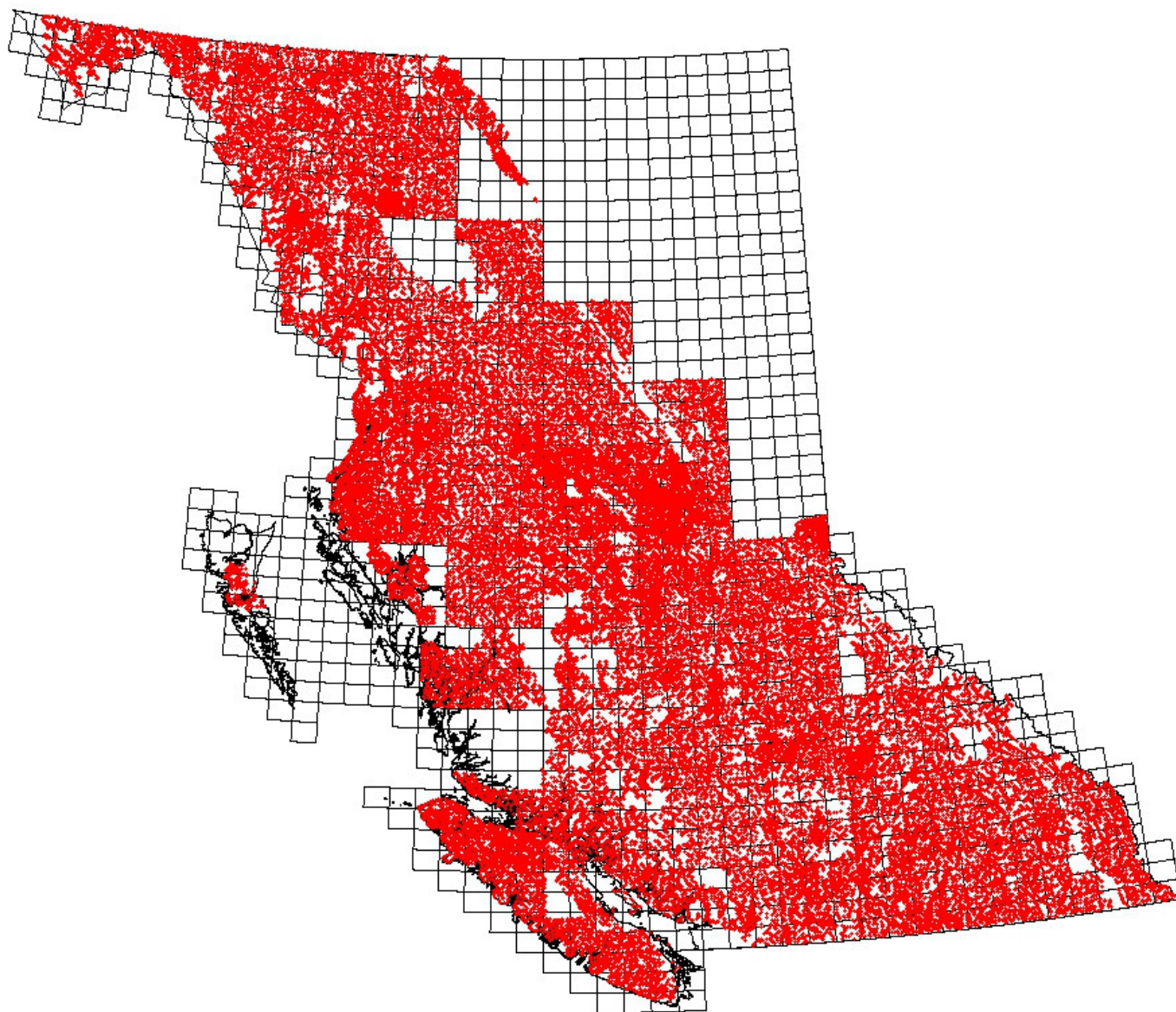
Fig. 1. RGS sample distribution.

statistics derived from the released RGS data set; elaborate on the components, format, and metadata of the data set; detail the development of the database; and describe the operation of the database and dataflow in data management.

2. Summary statistics of the data set

The RGS data set in this GeoFile was generated through the newly developed RGS database, which is populated with data extracted from Rukhlov and Naziri (2015) and Jackaman (2017). Compiled from 111 publications, the data set includes 5,105,684 analytical determinations using different methods on 64,828 samples collected between 1976 and 2013, including stream sediment and water, lake sediment and water, moss-trapped sediment and water and spring sediment and water (Fig. 2).

3. Data product

To maintain consistency with previously published RGS data (in particular, Rukhlov and Naziri (2015) and to make the data easier to use, we present the current data set as three MS Excel files, each of which contains multiple spreadsheets. "RGS2017_data.xlsx" contains RGS data (Table 1), "RGS2017_header.xlsx" contains data column header descriptions (Table 2), and "RGS2017_info.xlsx" contains additional information (Table 3).

4. Database development

Geoscience data live a lifecycle of 6 stages, including plan, acquire, process, analyze, preserve, and publication/share, as defined by Faundeen et al (2013). Except for the first 2, these stages are commonly instantiated in a database environment by the data custodial personnel or organization. It is particularly the case if the geoscience data want to be long-lived and
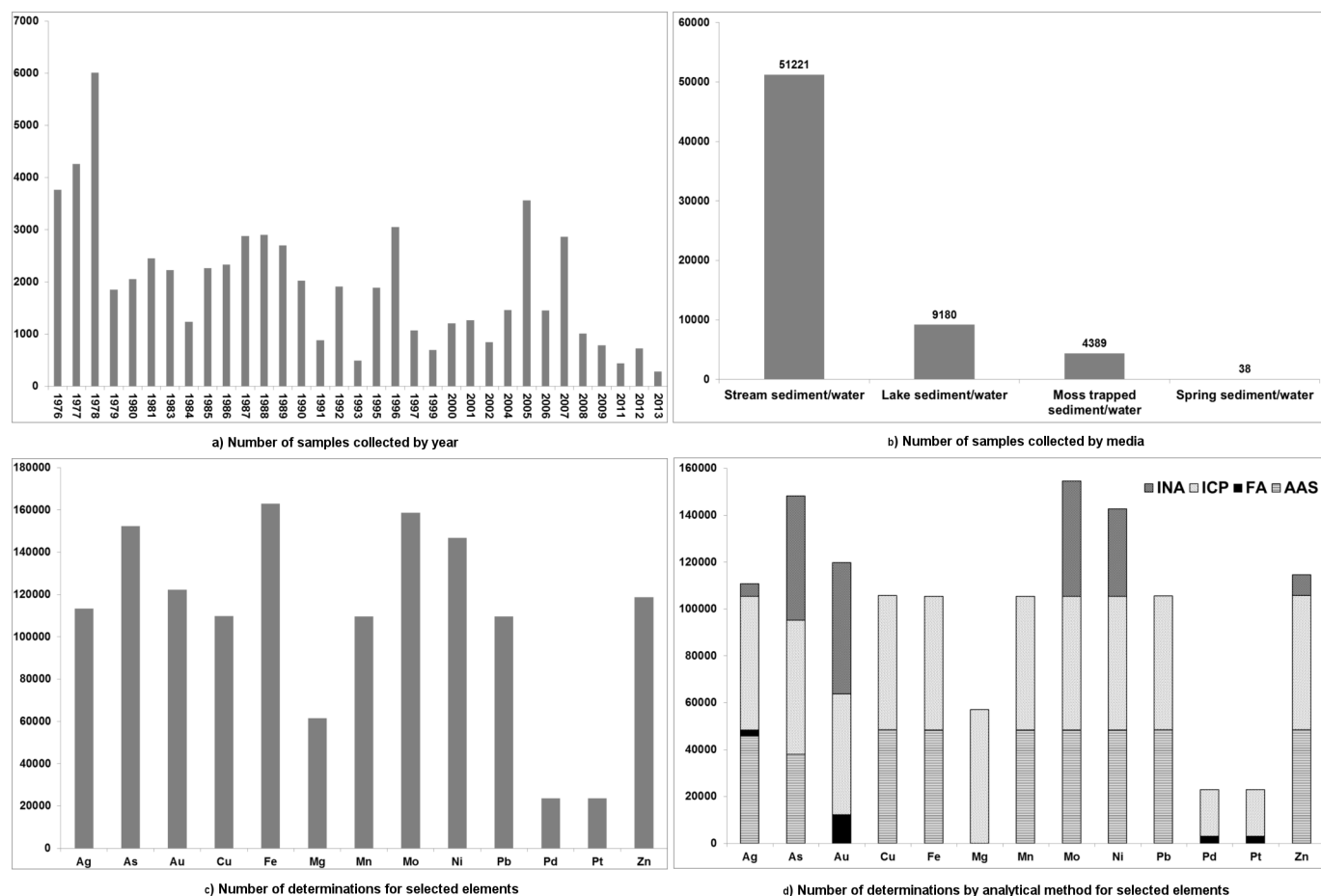
Fig. 2. Summary statistics.

accessible and consumed by many users. For this reason, database plays a central role during the lifespan of many types of geoscience data. Given the evolving nature of the RGS, we developed a formal relational database to house all RGS data to support data management activities through its lifespan. Database development commonly follows a cycle including requirement analysis, logical design, implementation, and database population (Connolly and Begg, 1999). We followed this cycle closely to develop the RGS database. We started by gathering feedback from the mineral exploration industry, examining the intrinsic relationships in the RGS data, and studying the original database developed by Lett (2005 and 2011) and Rukhlov and Naziri (2015), which led to the RGS logical data model (Fig. 3). This model consists of 9 entities: data_lake; data_moss; data_stream; data_sample; data_publish; data_analyte; code_lab; code_method; and code_unit.

Of the four types of RGS samples (lake sediment and water; stream sediment and water; moss-mat trapped sediment and water; and spring sediment and water), each (with the exception of spring sediment and water) carries unique attributes. For example, the lake-sediment and water samples have the attributes of lake area, lake perimeter, lake depth. To avoid using redundant attributes, we used three entities: data_lake, data_moss, and data_stream to house these survey-type

specific attributes, while we used the data_sample entity for common attributes, including sample location and geological information. The data_publish entity is for modeling and storing the publication-related information. The data_analyte entity is for analyte modeling and storing analytical determination-related information. The above 6 entities are considered as data entities. The code_lab; code_method; and code_unit are three metadata entities that are used to model and store such information as analytical method and laboratory. Keeping data and metadata together is important, because geochemical data with missing or incomplete metadata (for instance, an Au concentration without knowing if the corresponding sample was partially or completely dissolved) may cause problems when using these data for mineral identification.

With minimum redundancy, this model can address relationships between sample, publication, analyte, analytical method, and laboratory in cases where: 1) samples were analyzed using different methods at different laboratories; 2) samples were re-analyzed; 3) samples were re-analyzed with different minimum detection limits; and 4) results were released in multiple publications. With this model, we can unambiguously retrieve each determination and its associated information (sample, method, lab, and publication) using Sequential Query Language (SQL). Applying this logical

Table 1. Spreadsheets in 'RGS2017_data.xlsx'

Name	Description	Additional information
Sediment_and_water_master	Providing locations, sample medium, field observations, and the complete analytical data for sediment samples by different methods and routine RGS water samples (pH, conductivity, fluoride by specific ion electrode, uranium by laser-induced fluorescence, and sulphate by turbidimetry in untreated water) - 64,828 samples.	Same as the "Sediment_and_routine_water_master" table in Rukhlov and Naziri (2015) with the addition from Jackaman (2017)
Water_geochemistry	Water geochemistry by inductively coupled plasma mass spectrometry (ICP-MS) and atomic emission spectrometry (ICP-ES), ion chromatography (chloride, fluoride, nitrate, and sulphate), titration (total alkalinity), and reference electrode (total dissolved solids) - 4332 samples.	Identical to the "Water_geochemisrty" table in Rukhlov and Naziri (2015)
Sediment_selective_extraction	Re-analysis of archived RGS sediment samples by non-standard partial extraction using sodium acetate leaching with ICPM-ES finish - 1997 samples.	Identical to the "Sediment_selective_extraction" table in Rukhlov and Naziri (2015)
Laboratories	List of original report designations with information about the format used for values below detection limits for different analytical methods and laboratories that performed the preparation and analysis of the samples - 111 records.	Same as the "Laboratories" table in Rukhlov and Naziri (2015) with the addition from Jackaman (2017)
Geology_at_sample_site	Geology underlying sample site from the provincial digital geology (after Cui et al., 2013) – 64,828 samples.	Identical to the "Water_geochemisrty" table in Rukhlov and Naziri (2015)

Table 2. Spreadsheets in 'RGS2017_header.xlsx'

Name	Description	Additional information
Sediment_and_water_master	Definition of column headers for the "Sediment_and_water_master" spreadsheet in "RGS2017_data.xlsx"	Same as Table 5 in Rukhlov and Naziri (2015) with the addition from Jackaman (2017)
Water_geochemistry	Definition of column headers for the "Water_geochemistry" spreadsheet in "RGS2017_data.xlsx"	Identical to Table 5 in Rukhlov and Naziri (2015)
Sediment_selective_extraction	Definition of column headers for the "Sediment_Selective_extraction" spreadsheet in "RGS2017_data.xlsx"	Identical to Table 5 in Rukhlov and Naziri (2015)
Laboratories	Definition of column headers for the "Sediment_and_water_master" spreadsheet in "RGS2017_data.xlsx"	Same as Table 5 in Rukhlov and Naziri (2015) with the addition from Jackaman (2017)
Geology_at_sample_site	Definition of column headers for the "Sediment_and_water_master" spreadsheet in "RGS2017_data.xlsx"	Identical to Table 5 in Rukhlov and Naziri (2015)

Table 3. Spreadsheets in 'RGS2017_info.xlsx'

Name	Description	Additional information
Sediment_method	Analytes by different methods determined from RGS sediment samples	Identical to Table 1 in Rukhlov and Naziri (2015)
Water_geochemistry	Analytes by different methods determined from RGS water samples	Identical to Table 2 in Rukhlov and Naziri (2015)
Summary	Regional Geochemical Survey summary	Identical to Table 3 in Rukhlov and Naziri (2015)
Data_source	Summary of RGS data sources	Similar to Table 4 in Rukhlov and Naziri, 2015) with the addition from Jackaman (2017)
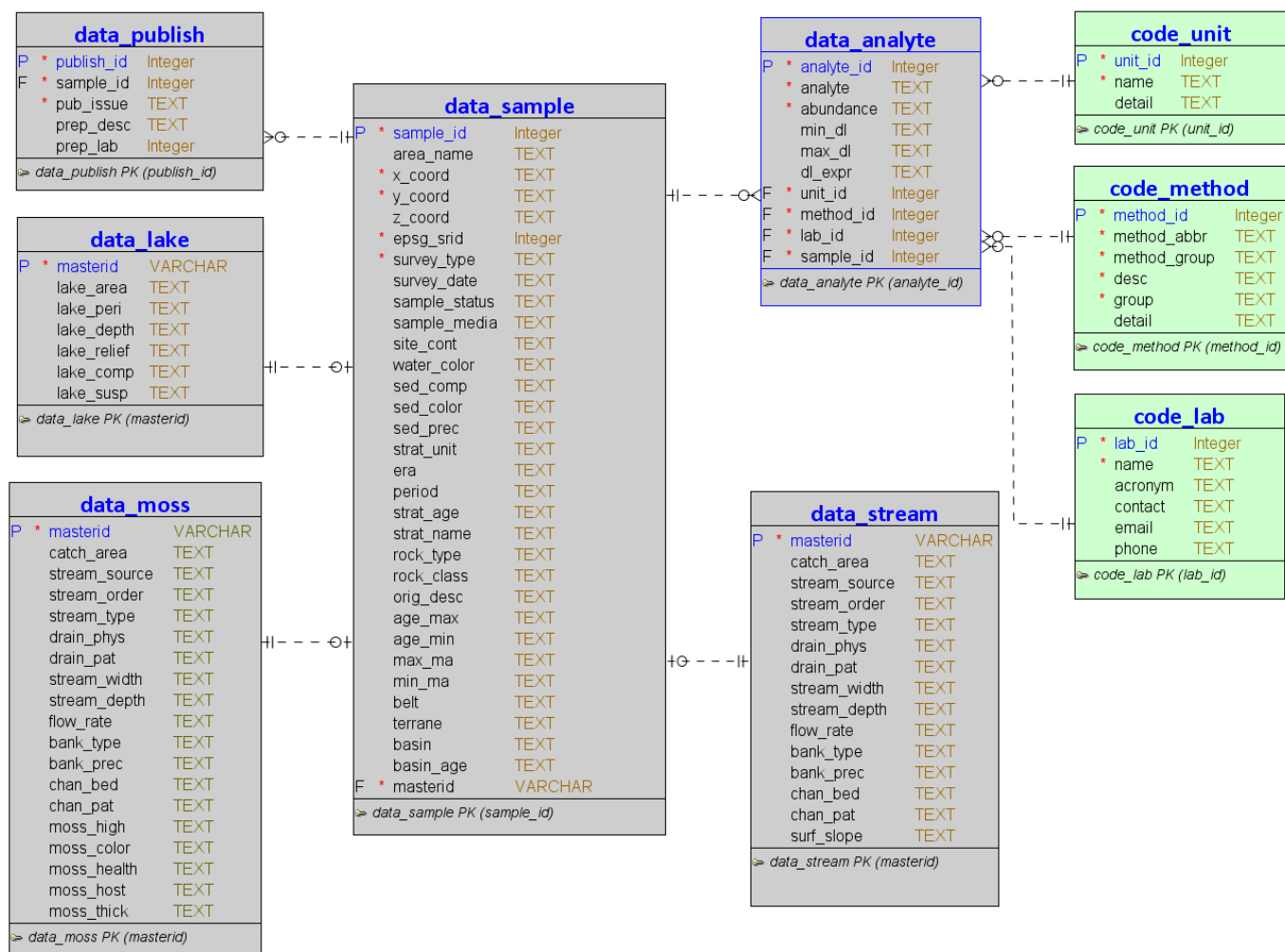
Fig. 3. RGS database logical data model.

model, we constructed the new RGS database using Microsoft Access. Peripheral applications were developed as 4 Python scripts to assist routine data management activities, including data quality control, data loading, data extracting, and product generating. Our model aligns with the Open Geoscience data model (Granitto et al, 2012; Watson and Evans, 2012).

5. Database operation and data management

The development and operation of this database makes it possible for the British Columbia Geological Survey to centrally manage RGS data, which leads to improved data consistency, timely data update, routine delivery of customized data product, and data loss mitigation. The operation and dataflow of the database follows five steps as (Fig. 4): 1) data compilation; 2) data screening for initial QA/QC; 3) data loading; 4) product generating; and 5) Product QA/QC. If errors are found in step 5, steps 1 to 4 are repeated.
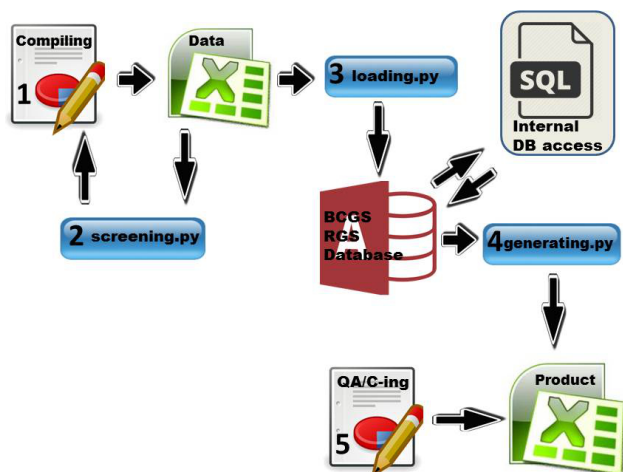

Fig. 4. Dataflow for RGS data management.

During compilation, data are retrieved from different sources and saved to XLS files in a standardized format. Commonly, a single source fails to provide all the data and metadata required by the database, or multiple sources give inconsistent results. Data completion, verification, and correction are commonly needed to deal with these situations. Data QA/QC is then conducted by the corresponding Python script (**screening.py**), which was designed to flag the most common data errors in the source or introduced in the data compilation step. These include errors such as unrealistic determinations and units, improper methods, incorrect sample locations, and redundant samples. The flagged errors are then corrected manually. Following QA/QC, data are loaded into the database. This is an automatic step done by the corresponding Python script (**loading.py**). To avoid generating duplicate entries, the script checks if a sample is already in the database. The step of generating data product, also conducted by a Python script (**generating.py**), retrieves and outputs data in formats suited to the user (e.g., Comma-Separated-Value (CSV) files, ESRI shapefiles, or MS Excel files).

The database and the products derived from the database differ. First, the format of a product is based on data applications and can be changed by the users, whereas the format of the database is decided by the data model and is fixed once database development is finished. Second, data products are commonly derived as the views of a database. As such, they represent only part of the information contained in the database. Separating the database from the product is important, because it helps the user generate multiple tailored products from a single database.

References cited

Carter, N.C., Kalnins, T.E., and Nordin, G.D., 1978. Uranium Reconnaissance Program (82/F, K and 104/N). In Geological Fieldwork 1977, British Columbia Ministry of Energy, Mines and Petroleum Resources, British Columbia Geological Survey, Paper 1978-1, pp. 106-107.

Connolly, T., and Begg, C., 1999. Database Sytems – A Practical Approach to Design, Implementation, and Management, 2nd Edition, Addison-Wesley, 113p.

Cui, Y., Katay, F., Nelson, J.L., Han, T., Desjardins, P.J., Sinclair, L., compilers, 2013. British Columbia Digital Geology. British Columbia Ministry of Energy and Mines, British Columbia Geological Survey, Open File 2013-4.

Darnley, A.G., Cameron, E.M., and Richardson, K.A., 1975. The Federal-Provincial uranium reconnaissance program. In: Darnley, A.J., Ruzick, V., Dyck, W., Cameron, E.M., Richardson, J.A., (Eds.), Uranium 75, Geological Survey of Canada, Paper 75-26, pp. 49-63.

Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, Elizabeth, Montgomery, E.T., Ladino, C.C., Tessler, Steven, and Zolly, L.S., 2013, The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013–1265, 4 p., http://dx.doi.org/10.3133/ofr20131265.

Fersman, A.E., 1939 . Geokhimicheskie i Mineralogicheskie Metody Poiskov Poleznyh Iskopaemyh [Geochemical and Mineralogical Methods of Mineral Exploration]. Izd. Akad. Nauk SSSR, Moscow, 446 p.

Friske, P.W.B., Hornbrook, E.H.W., 1991. Canada's National Geochemical Reconnaissance programme. Transactions of the Institution of Mining and Metallurgy, London, Section B 100, 47–56.

Granitto, M., Schmidt, J. M., Labay, K. A., Shew, N. B., and Gamble, B. M., 2012. Alaska geochemical database - Mineral exploration tool for the 21st century. U.S. Geological Survey Open-File Report 2012-1060, 33p.

Jackaman, W., 2017. RGS Sample reanalysis project (parts of NTS 082G, 082J, 092N, 093E, 093H, 103O, 103P and 104N), GeoScience BC Report 2017-04.

Lett, R.E., 2005. Regional Geochemical Survey Database on CD. British Columbia Ministry of Energy and Mines, British Columbia Geological Survey GeoFile 2005-17.

Lett, R.E., 2011. Regional Geochemical Survey Database 2011. British Columbia Ministry of Energy and Mines, British Columbia Geological Survey GeoFile 2011-7.

Lovering, T.S., Huff, L.C., and Almond, H. 1950: Dispersion of copper from the San Manuel copper deposit, Pinal County, Arizona. Economic Geology, 45, 493-514.

Meyer, W.T., Theobald, P.K., Jr., Bloom, H., 1979. Stream sediment geochemistry. In: Hood, P.J., (Ed.), Geophysics and Geochemistry in the Search for Metallic Ores. Geological Survey of Canada, Economic Geology Report 31, pp. 411-434.

Rukhlov, A. and Naziri, M., 2015. Regional Geochemical Survey Database, British Columbia Ministry of Energy and Mines, British Columbia Geological Survey GeoFile 2015-3.

Vernadsky, V.I., 1924. La géochimie. Librairie Félix Alcan, Paris, 403 p.

British Columbia Geological Survey
Ministry of Energy, Mines and Petroleum Resources
www.em.gov.bc.ca/geology