USING EXISTING, PUBLICLY-AVAILABLE DATA TO GENERATE NEW EXPLORATION PROJECTS

Philip M. Baker[1], Paul A. Agnew[2] and M. Hooper[1]

[1]Rio Tinto Exploration, 37 Belmont Ave, Belmont WA 6104. Australia

[2]Rio Tinto Exploration, 1 Research Ave, Bundoora VIC 3083. Australia

*Corresponding author (e-mail: philip.baker@riotinto.com)

Objectives

To generate new exploration projects, using publicly-available data. Compiling geochemical results for ten million surface samples from publicly-available data over a period of six years has generated eight exploration projects.

Introduction

The goal to systematically collate, interpret and deliver publicly-available, surface geochemical sample data to geoscientists has required a pragmatic approach to grouping geochemically-similar samples with similar analysis. Having all the surface geochemical data in one database, means the data is easy to find and spatially view; allowing the surface sample data to test various geological hypotheses. However, these advantages have been extended to enable direct interpretation of data across whole terranes. To facilitate the interpretation of the data, the raw data has been compiled and categorised, so it can be extracted from the database in layers of data with similar geochemical properties. About 90% of the compiled geochemistry can be extracted into one of nine different layers (e.g., rocks analysed with a total analytical method or fine-fraction stream sediments analysed with a strong digest). At the time of writing, a total of about ten million surface geochemical samples had been compiled (Figure 1) from which eight exploration projects have been directly generated.

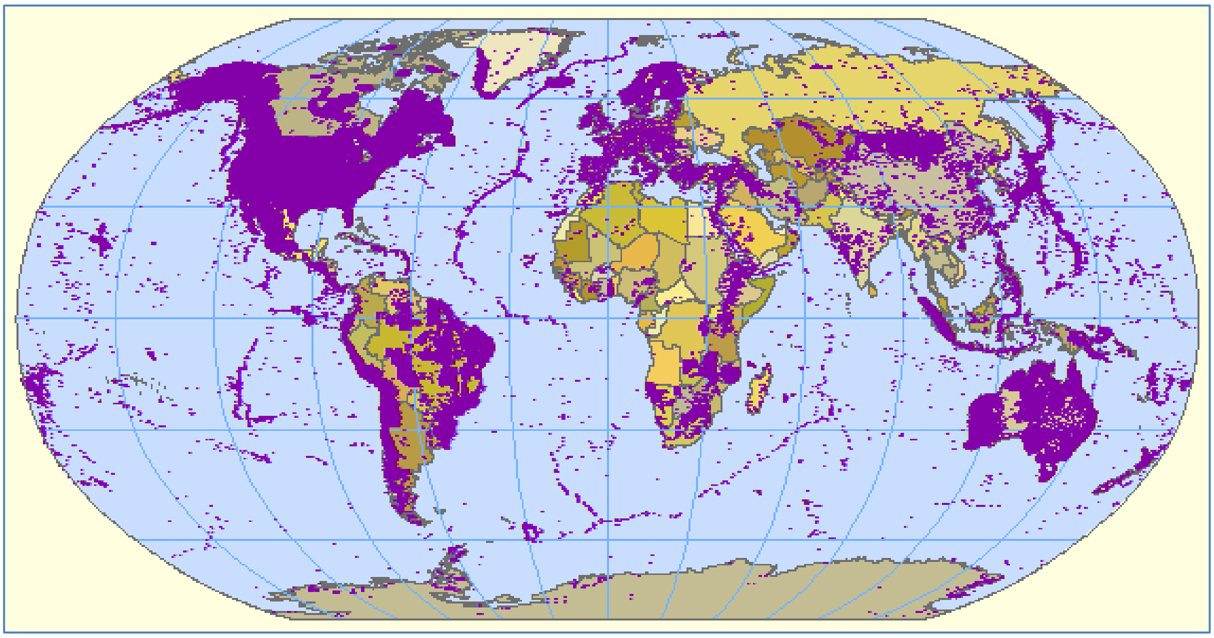

Figure 1. The global database of publicly-available surface geochemical data showing the current ten million sample points. At this point, a total of eight exploration projects have been directly generated from interpretations of the data.

Methodology

The data is obtained from various public sources. The contact information for these sources is stored in the database and is revisited every two years for updates or new data. The web data and web-served databases have been downloaded. For larger, web-delivered databases, we have often requested their entire database on DVD. The Georoc database (Georoc, 2013) provided >350,000 rock sample data in this way. The publicly-available data is not always free and a significant amount has been purchased; such as the Council for Geoscience in South Africa soil database for South Africa, covering the Bushveld Intrusive Complex (Figure 2). This dataset had to be compiled from the 110 separate files, using the two, incompatible coordinate systems from which it was provided. However, on the positive side, this data has unusually consistent analysis (Figure 3). The data from publicly-available, scanned reports has also been reconstructed by keypunching the data to digital format and grouping for important areas, e.g., India and Saskatchewan. Whilst the data is publicly available, the ownership of the data does not change and the licensing is collected and stored with each sample. Where the data licence is unclear, the owners are contacted and their permission for us to store and distribute the data is sought. In some cases, the data has restricted access or is not stored in the database, although this is a rare issue.

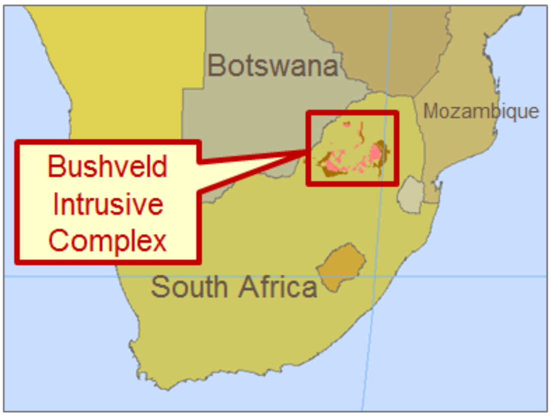

Figure 2. Location of the Bushveld Intrusive Complex in South Africa.

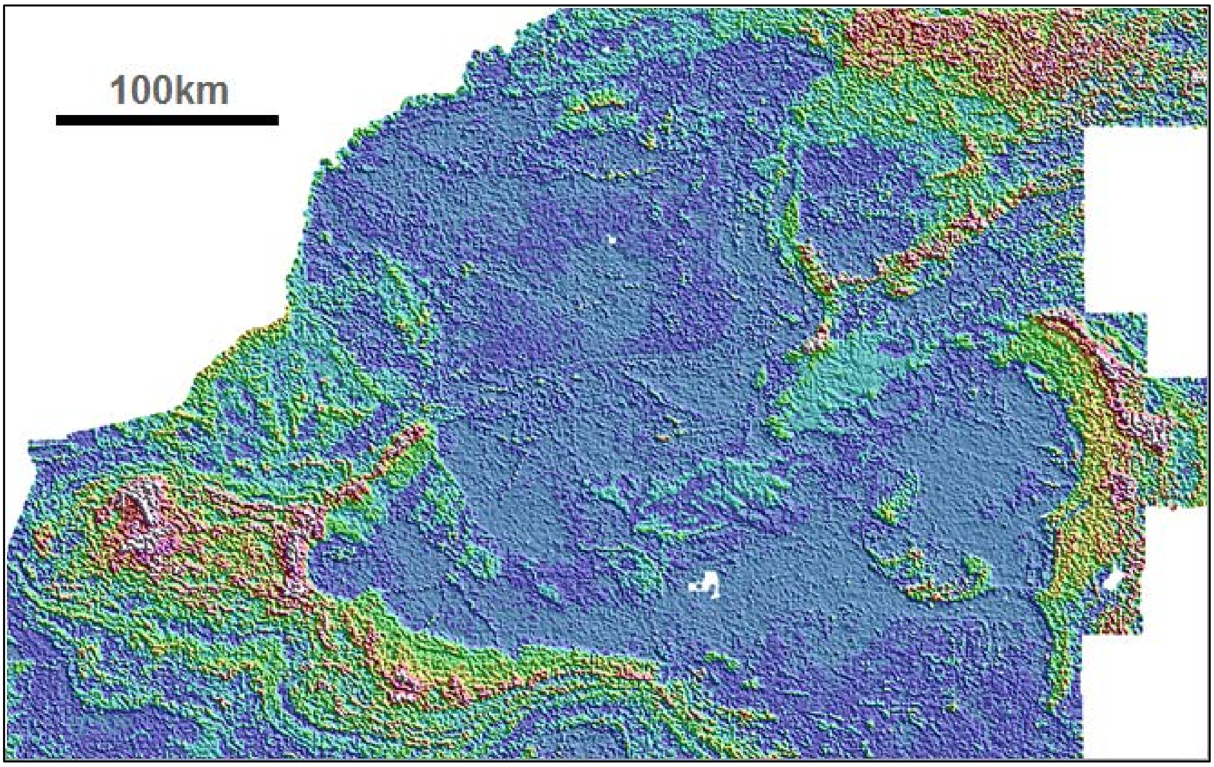

Figure 3. Council for Geoscience in South Africa nickel in soil data compiled from 110 files and two incomptible coordinate systems. The consistency in the analysis from the 167,000 samples over the Bushveld Intrusive Complex (shown here) allows superb detail of the lower stratigrapy of the lobes to be seen in this image.

The data sets are prioritised for loading into the database, depending on the prospectivity of the area; quality and type of sample data. All of the elemental data is categorised and loaded into the database. The meta data is managed in four categories:

1. **Important Raw Data Fields**. E.g., original sample number, original sample type, coordinates and sample comments.

2. **Fields to be Translated**. Yet to be translated into a category field (often hierarchical), e.g., the translated sample type, mesh group and translated lithology.
3. **New Fields**. To help administer the database, e.g., Sample ID, Source ID, global coordinates, data compiler's comments and batch information.
4. **Fields Not Stored.** (In the database.) These are fields that are mostly not consistent between projects and are not used in the first stage of geochemical interpretation, e.g., sample colour. These data are always accessible by following a hyperlink to the raw data.

A number of fields are critical for the data to be categorised and output correctly. The acceptable error rate in these fields must be less than 0.01% to be functional; and the error rate includes both errors and unknown data. The compilation follows a pragmatic process to reduce errors and unknown data in the critical fields, allowing estimates and coding them as estimates. An example of estimating unknown data would be if the sample type is missing but when the data are plotted on a map the samples are all located just above the confluence of the streams, then the sample type would be estimated as 'stream sediment'. A few of the fields are used to group samples based on the geochemical behaviour. The mesh size of the sample collected is essential in deciding which samples are compatible with fine (<100µm), medium (100µm to 400µm) and coarse (>400µm) fraction samples behaving geochemically differently in a number of environments.

The analytical data follows a similar scheme for categorising the analytical technique. The analytical technique is separated into the digest degree and the determination method. There are many digest degree and determination methods that produce similar results for different sample types and these can be output together in a single layer. An illustration to the problem is depicted in (Figure 4) where a database from the Western Australian Department of Mining and Petroleum (WADMP) contains 20,000 columns of data, if output as a flat table with one column per assay label. The Cu assays alone would be stored in >500 columns. Some of these columns contain the same type or compatible Cu assays, whereas others are quite different. These >500 columns are translated to one of ten Cu columns in the final database. These 10 columns are then further grouped for output. How the assays are grouped for output depends on the sample type and sometimes, mesh size. As an example: for fine fraction stream sediments or oxidised soils, the Cu from strong acid digests (two- and three- and four- strong acids) and whole-rock methods such as XRF and fusion-based analysis, can often be interpreted together. The weak, proprietary and specialised, non-total digests are not compatible and are generally difficult to use across multiple projects. For rock sample data, the strong, acid-based digests are reported separately from the fusion and XRF total analysis data.

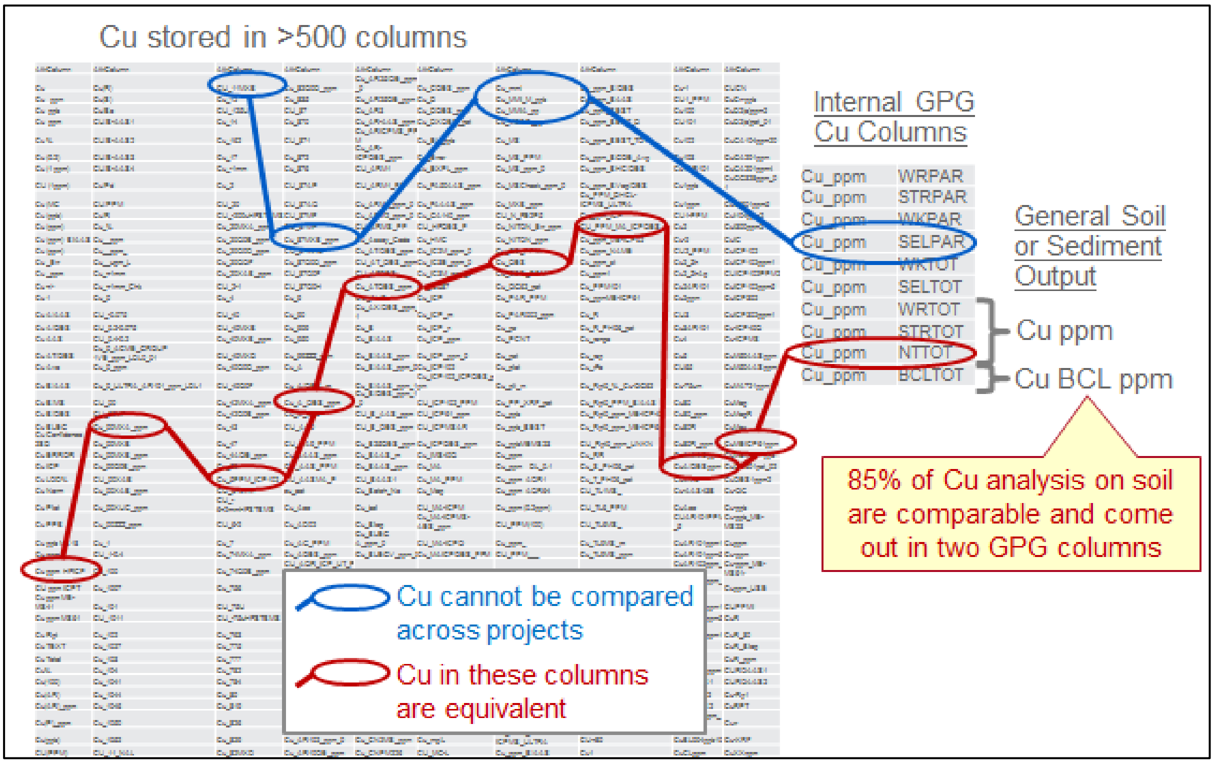

Figure 4. A database from the Western Australian Department of Mining and Petroleum (WADMP) showing the >500 column Cu assays labels stored. These >500 are translated into one of ten Cu column labels in the database, which can then be grouped for output e.g., for soil or stream sediment, the Cu assays can be combined into one column for strong acid digests (two, three and four strong acids) along with whole rock methods such as XRF and fusion-based analysis. Weak, proprietary and specialised non-total digests do not appear in this column.

The compilation process moves the data from a staging area, through a working stage to a final resting place for loaded data (Figure 5). After the data has been loaded, it is routinely interpreted for lithological information or interesting anomalies, as part of the error checking process, prior to being released. At this point, the geoscientists may require more comprehensive or specialised interpretation, depending on the terrane.

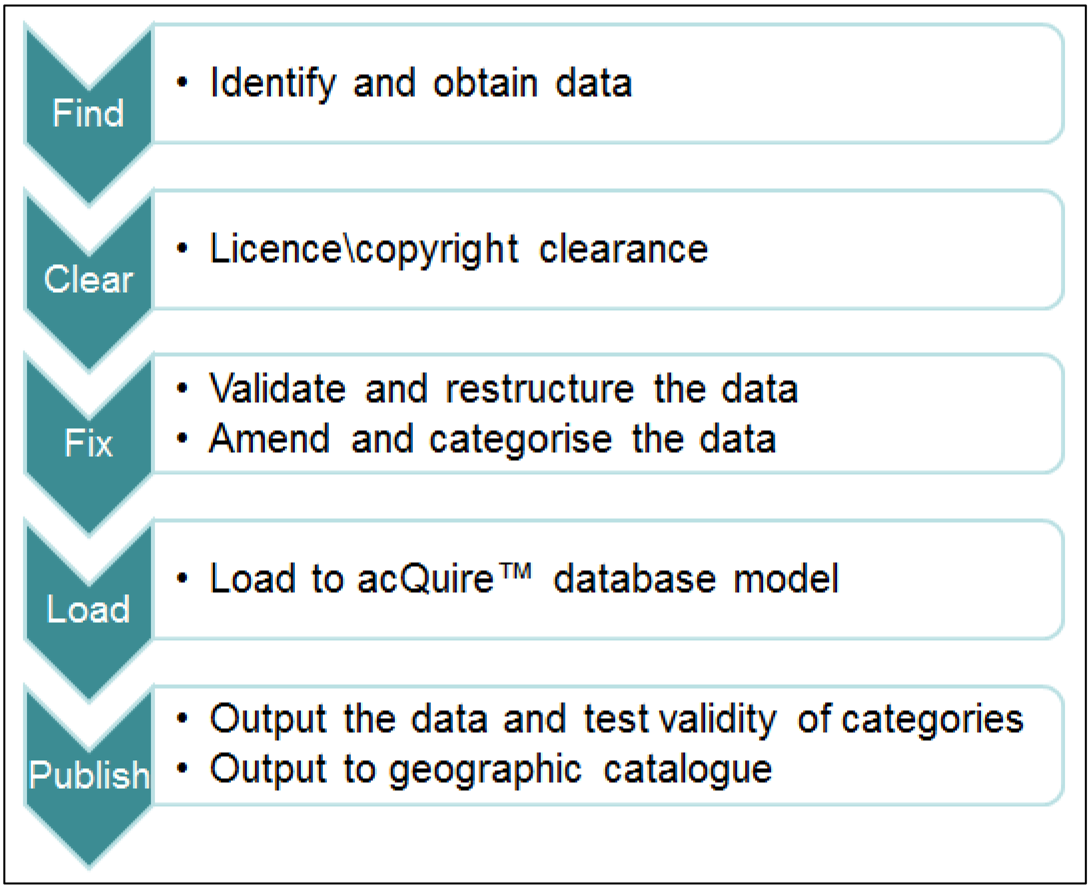

Figure 5. Simplified work-flow chart depicting the process from identifying the data to publishing the data to geoscientists.

Case History

The following is a case history of one of the projects generated from the compiled, public data which project resulted in drilling the magmatic nickel sulphide mineralisation. The target hypothesis was to search for a magmatic conduit feeder to the Bushveld Intrusive Complex, South Africa (Figure 2) containing disseminated nickel sulphides. The location of the feeder could be within the intrusion or below or at the base of the intrusion. The presence of disseminated sulphides could cause the target to be recessively weathered and hidden from surface prospecting but still identifiable in soil geochemistry. The spatial gridding of images for the compiled, Council for Geoscience in South Africa, 1km^2 spaced soil data (Figure 6) is useful for mapping the lithology (Figure 3) and stratigraphy. However, soil samples related to magmatic nickel sulphides often have a unique geochemical signature when the data is plotted in chemical space. A unique signature to mineralisation is important to make a successful assessment of the 167,239, Council for Geoscience in South Africa soil samples. The unique signature depends on Ni being a compatible element incorporated in early-formed minerals in a differentiating intrusion, whilst Cu is an incompatible element only being taken up later. Due to this contradictory behaviour, the soil samples over most rocks tend to be either high in Ni or high in Cu - but not both (Baker, 2004). The samples that are high in both Ni and Cu are related to sulphide accumulations which concentrate Ni, Cu and other chalcophile elements. A scatter plot of Ni vs. Cu (Figure 7) can clearly separate these samples. Only 65 samples of the 167,239 have this signature (Figure 9, Figure 8). However, the sulphide can be related to either magmatic or sedimentary sulphides and to discriminate the two, the general association of relatively low Zn compared to Cu in magmatic Ni sulphides is used. A second plot of Cu vs. Zn (Figure 10) discriminates the samples that are high in both Cu and Zn, which are mostly related to sedimentary sulphide accumulations. The samples associated with the platinum reef-type deposits are also highlighted, along with soil samples related to magmatic Ni sulphide because there is a small amount of Ni sulphide within these reef deposits. About half the 65 samples selected have been demonstrated to be related to magmatic Ni sulphide, the other half have not been tested. Follow-up soil sampling and drilling on one project confirmed magmatic Ni sulphide with intersections of 7.9m at 1.14% Ni and 0.43% Cu and within a longer intersection of 41.6m, at 0.51% Ni and 0.18% Cu.

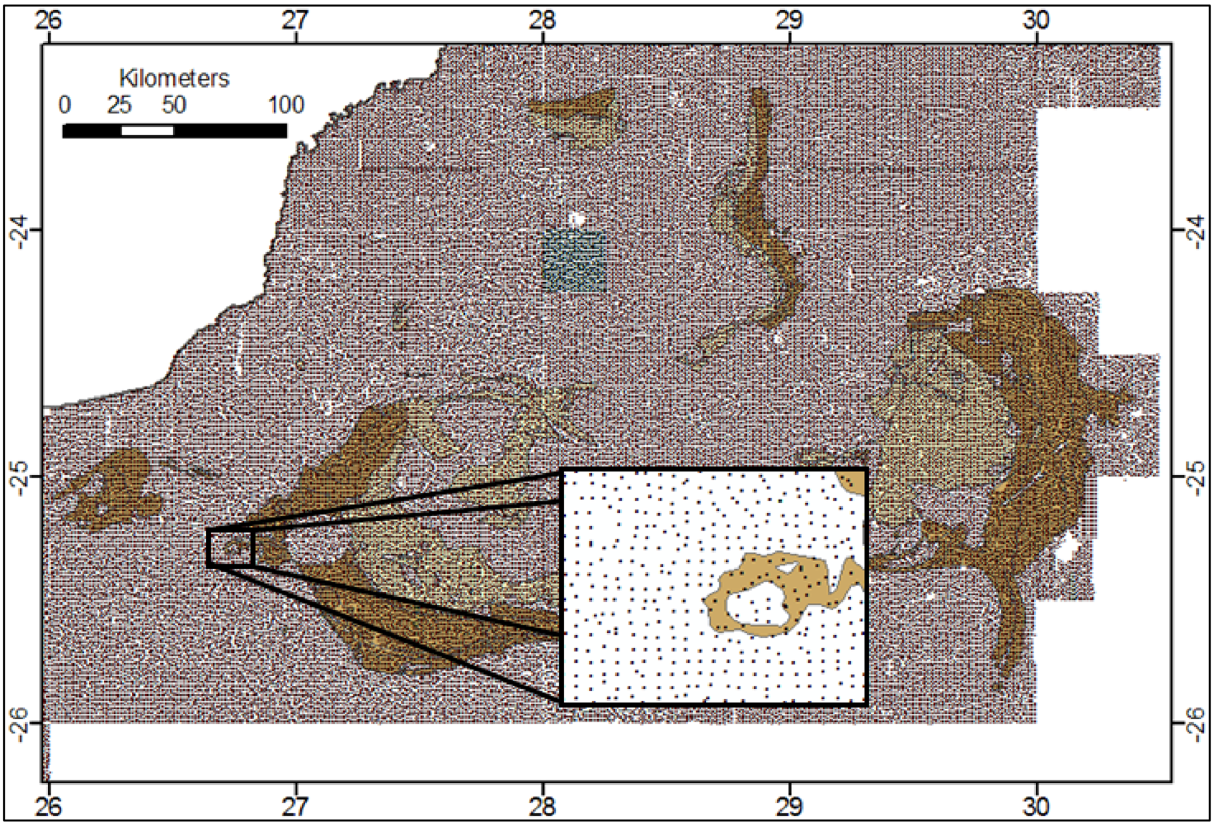

Figure 6. The sampling grid of approximately 1km spaced soil samples collected by the Council for Geoscience in South Africa.

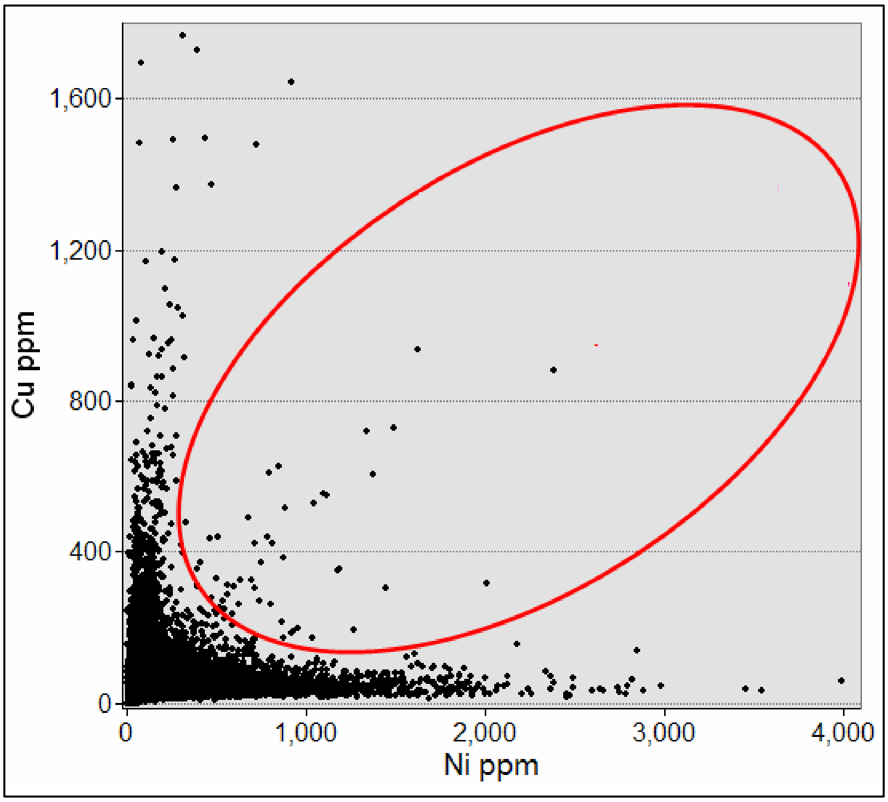

Figure 7. Ni vs. Cu for 167,239, Council for Geoscience in South Africa soil samples, mostly from over the Bushveld Intrusive Complex. Ni is a compatible element in early-formed minerals in a differentiating intrusion and Cu is an incompatible element. Due to this contradictory behaviour, the soil samples tend to be either high in Ni or high in Cu but not both. The samples that are high in both Ni and Cu are related to sulphide accumulations which concentrate Ni, Cu and other chalcophile elements related to either magmatic or sedimentary sulphides.

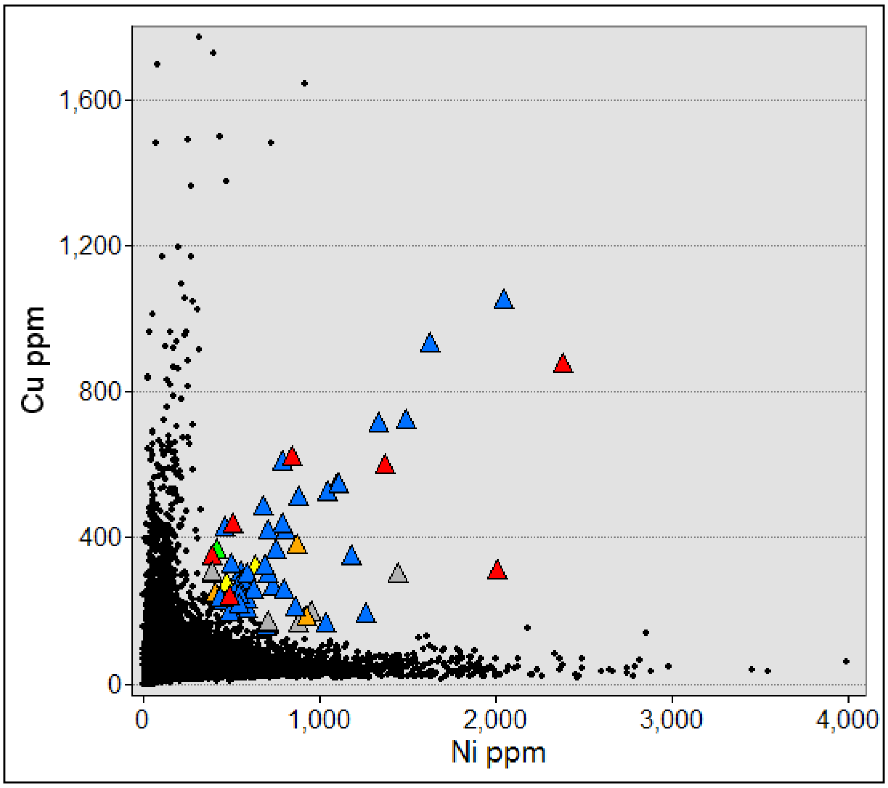

Figure 8. The samples that are high in both Ni and Cu are related to sulphide accumulations which concentrate Ni, Cu and other chalcophile elements related to either magmatic or sedimentary sulphides. The coloured triangles depict in which geological unit the sample is located. Triangles are associated thus: blue with Platinum deposits; red with the base of the intrusion; orange not with intrusive rocks but potentially with footwall rocks; green outside of intrusion on line with the Nkomati Ni deposit; grey with Archean intrusions and sediments outside of Bushveld Complex Intrusion; and yellow with upper Bushveld Intrusive rocks.

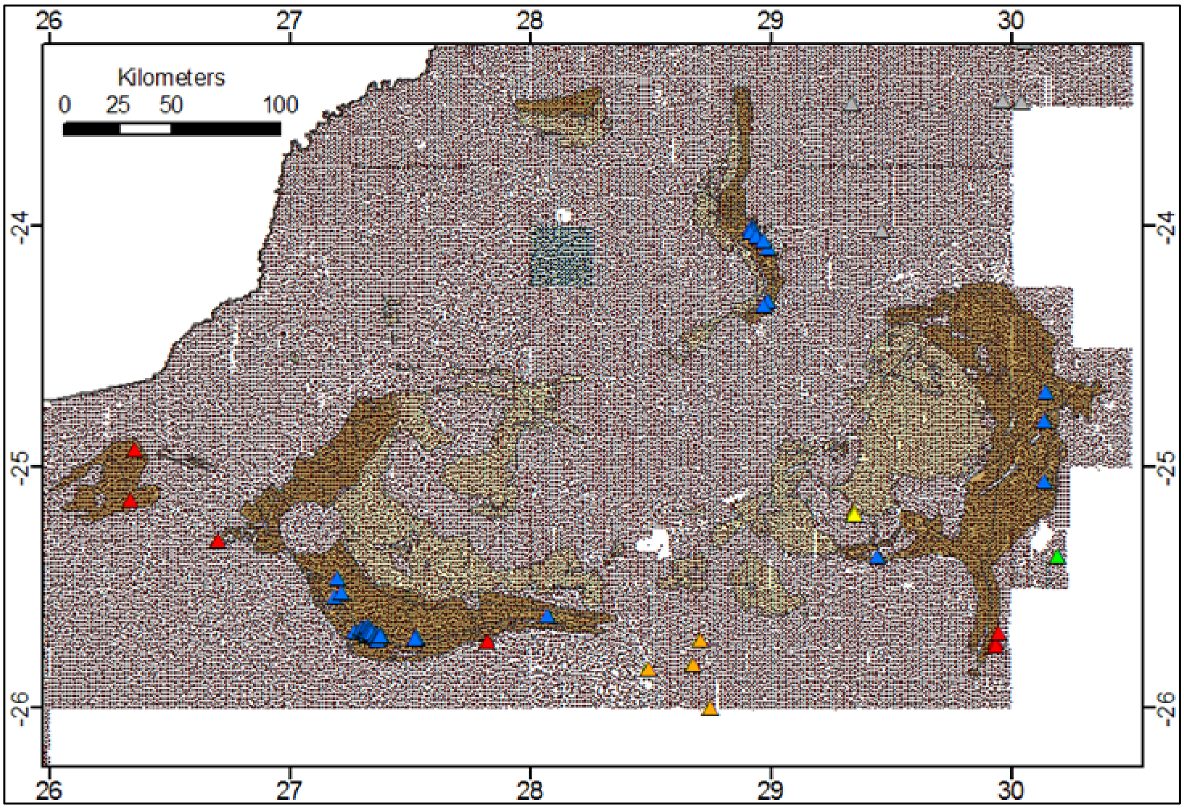

Figure 9. A plot of soil sample locations (Council of Geoscience South Africa) with samples relatively high in both Ni and Cu denoted in Figure 8 as coloured triangles, depicting in which geological unit the triangle is located. Triangles are associated thus: blue with Platinum deposits; red with the base of the intrusion; orange not with intrusive rocks but potentially with footwall rocks; green outside of intrusion on line with the Nkomati Ni deposit; Grey with Archean intrusions and sediments outside of Bushveld Complex Intrusion; and yellow with upper Bushveld Intrusive rocks.

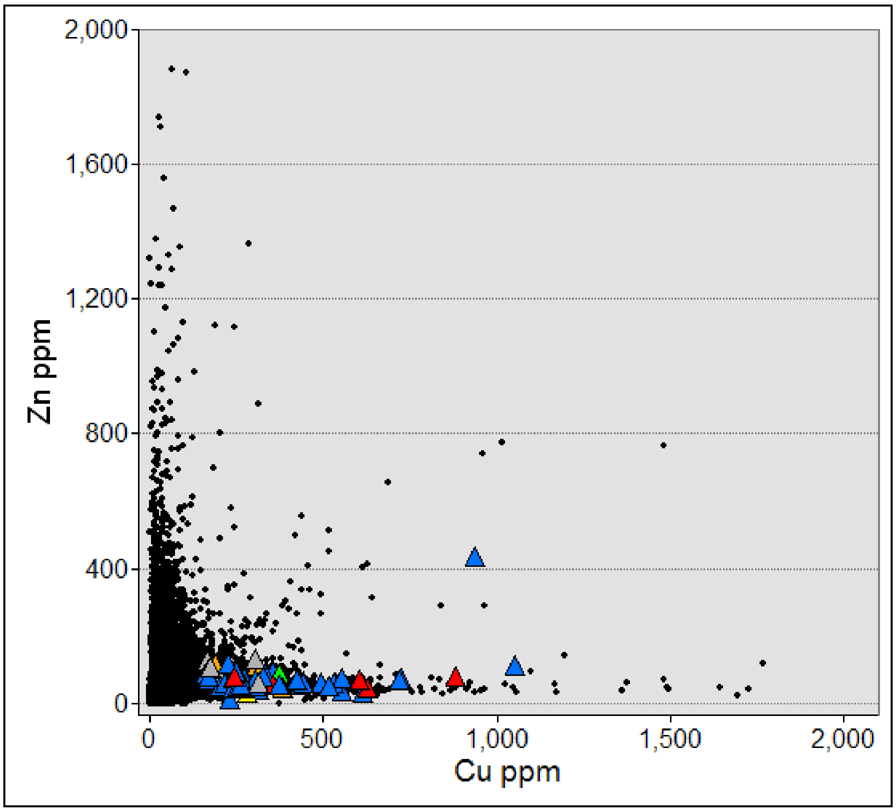

Figure 10. The samples that are high in Ni and both Cu and Zn are related to sedimentary sulphide accumulations. As can be seen here, only one sample appears high in Ni, Cu and Zn. However, this sample is related to a Pt deposit. The coloured triangles depict in which geological unit the sample is located. Triangles are associated thus: blue with Platinum deposits; red with the base of the intrusion; orange not with intrusive rocks but potentially with footwall rocks; green outside of intrusion on line with the Nkomati Ni deposit; grey with Archean intrusions and sediments outside of Bushveld Complex Intrusion; and yellow with upper Bushveld Intrusive rocks.

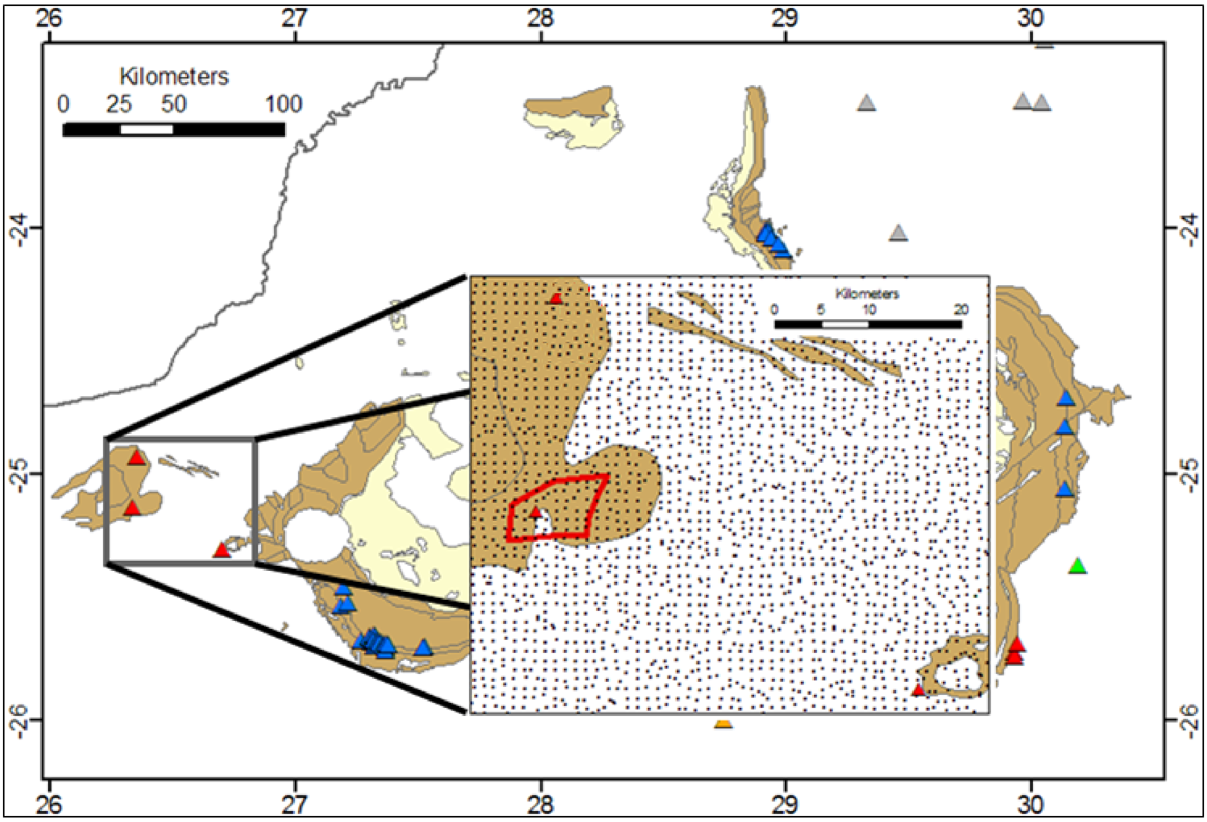

Figure 11. Sample locations and location of one area followed up with soil sampling and drilling. The base of intrusion samples followed up demonstrated Ni sulphide mineralisation.

There is an explosion of publicly-available data coming in the next few years, with massive quantities being released by governments from exploration reports. The drill-hole data reports contain perhaps five times the surface sample data if the data recently released by the WADMP for exploration from 2007-2013 is repeated elsewhere. The data is often released as scanned reports (not digital tabular data) which present an additional challenge. A test area in Saskatchewan's Athabasca Basin (Figure 12, Figure 13) demonstrated the data can be recovered as digital data at an approximate cost of USD 0.50 per sample. In important terranes, this data is essential prior to further exploration.

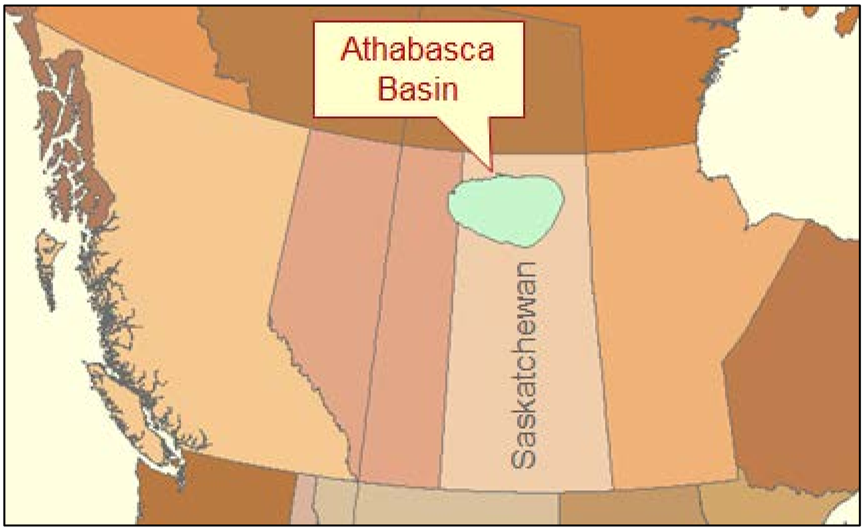

Figure 12. Location of the Athabasca Basin in Saskatchewan, Canada

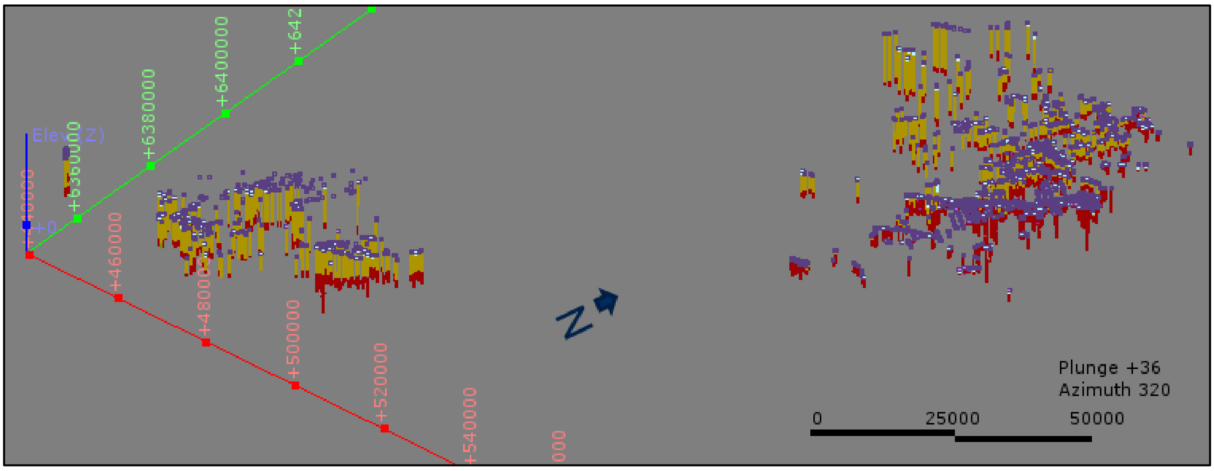

Figure 13. The results of a test data capture of two small areas of the Athabasca Basin in Saskatchewan, Canada. The scanned reports had geochemical results from 4,646 drill-holes with 69,947 samples for about 20% of the available data.

The goal is to systematically collate, interpret and deliver publicly-available surface geochemical data to geoscientists. At the time of writing, the data for a total of about ten million surface geochemical samples had been compiled globally Figure 1) from which eight exploration projects have been directly generated (Figure 14). It is expected this database will grow at the rate of one to three million samples per year.

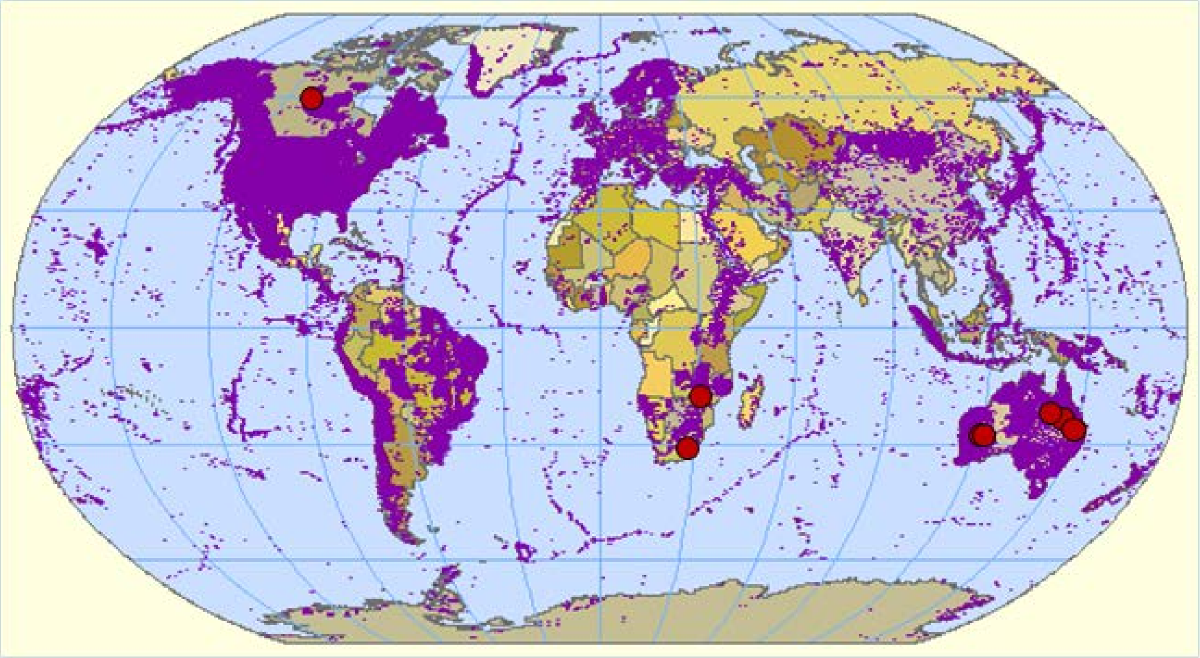

Figure 14. Exploration Projects directly generated from publicly-available geochemical data from 2007-2014.

References

BAKER, P. M. & WAUGH, R. S. 2004. The role of surface geochemistry in the discovery of the Babel and Nebo magmatic nickel–copper–PGE deposits. *Geochemistry: Exploration, Environment, Analysis*, **4**, 1–7.

Data downloaded from http://georoc.mpch-mainz.gwdg.de/georoc/Start.asp

Data purchased from http://www.geoscience.org.za/ with the acknowledgement: "*Portions of this work include intellectual property of the COUNCIL FOR GEOSCIENCE and are used herein by permission. Copyright and all rights reserved by the said COUNCIL*."