A role for data richness mapping in exploration decision making.

Alan R.A. Aitken1*, Sandra A Occhipinti1, Mark D Lindsay1, Allan Trench2

1 - Centre for Exploration Targeting, School of Earth Sciences, the University of Western Australia, Perth Western Australia, Australia
2 – Business School, the University of Western Australia, Perth, Western Australia, Australia

*alan.aitken@uwa.edu.au

Abstract

Exploration programs involve long, multi-branched decision pathways with many potential outcomes from a single starting point. Critical decisions along this pathway include the choice of the area(s) in which to explore, and the choice to acquire new data sets to assist in exploration. Included in this is a factoring of risks, including the technical risks associated with imperfect data and the value relative to cost of collecting more data. In our experience it is not common practice to explicitly assess and quantify the impact of data on exploration decisions. A better understanding of the opportunity provided by data has the potential to improve the quality and/or ease of decision making. In this study we explore the potential role of data richness mapping as part of a decision making workflow. We outline a method to quantitatively assess data richness following a four level hierarchy, namely presence of data (L1), quality of data (L2) and attributes of data (L3). Measures of these are combined across several data sets to give an overall estimate of data richness (L4). Using this data richness map, along with past exploration records and a prospectivity model, we conduct an illustrative case study for gold exploration in the Bryah Sub-basin (Yerrida Basin) region of Western Australia. The first application is to produce a simple data based classification of regions with different technical risk characteristics. The second application is to combine this with a prospectivity model to better understand exploration potential. The third application uses data richness mapping to understand which data sets are most likely to have an impact on regional exploration. In each case we show how an understanding of data richness can help support exploration decision making.

Introduction:

Modern exploration often demands the integration of concepts that operate across a variety of scales (Hronsky and Groves, 2008, McCuaig and Hronsky, 2014, McCuaig et al., 2010) which leads to a long and multi-branched decision pathway (Kreuzer et al., 2008). To facilitate this process, the number of geoscience data sets used in a typical program has increased substantially. Programs grow in complexity as attempts are made to combine these data sets meaningfully (Czarnota et al., 2010, Joly et al., 2012, Lindsay et al., 2016, Lindsay et al., 2014, Occhipinti et al., 2016, Payne et al., 2015). Compounding the increase in data...
types, the method(s) used to generate individual data sets have typically become more advanced over time, often increasing the complexity of each information source. Data sets have become larger as digital collection and storage barriers are less significant (Perrons and Jensen, 2015), increasing the volume of information to be processed. Finally, each data set also possesses a level of uncertainty, including for example, the relative strength of the contained data signal to local background levels, thus all subsequent decisions are inherently uncertain. The propagation of these uncertainties along a multi-branched decision pathway further reduces confidence in the final decision (Kreuzer, Etheridge, 2008, McCuaig et al., 2007, Yousefi and Carranza, 2015a).

As a consequence, decisions surrounding the geoscientific data sources and analyses used in an exploration workflow have become more complex. Other fields have also faced problems with decision making in the face of increasing data complexity, including petroleum exploration (Binns and Corbett, 2012, Perrons and Jensen, 2015) and environmental science (Bagstad et al., 2013, Crossman et al., 2013, Rounsevell et al., 2012). Under such conditions, traditional decision making based on heuristics can lose touch with the capability of modern data to inform the decision. One approach to assist in this is the development of semi-automated systems which seek to aggregate and refine data into a reduced model. As an example, a variety of techniques can be applied in order to generate models of mineral resource potential (Knox-Robinson, 2000, Porwal et al., 2003, Rodriguez-Galiano et al., 2015, Yousefi and Carranza, 2015b).

In this work we consider the role of geoscientific data in exploration decision making processes. We neglect here all other aspects of decision making, for example, infrastructure accessibility, commodity prices and organisational factors, whilst recognising that they are all very substantial (Penney et al., 2004, Tulcanaza and Ferguson, 2001). We define a multi-level analytical approach to understand spatial variations in the "richness" of geoscience data sets. Combining this with historical exploration data and a prospectivity model, we apply this in three illustrative examples. In the first application we define a simple data based classification approach for ground selection. In the second we incorporate a prospectivity model, to further delimit exploration potential. In the third application we use data richness to assess the value of several data types in both relatively mature and relatively immature exploration settings, with a view to guiding new data acquisition.

The exploration decision cycle

When making any decision, explorers must consider, either implicitly or explicitly, their knowledge of the level of risk and likely reward associated with different exploration strategies or target areas. These risks may include a variety of economic, political, logistical, social and environmental factors (Trench et al., 2013), but here we focus on the risk associated with imperfect geoscientific data, in particular its distribution and usefulness. These data are associated with risk in the sense that better surveyed regions provide, in the first
place, a greater opportunity to understand what mineral resources may be available and, in the second
place, a better understanding of where to find them and how. However, in many cases, better surveyed
regions possess fewer opportunities for new discoveries (Guj et al., 2011, Mamuse and Guj, 2011), and so
the potential for substantial reward is less than in little surveyed areas. In an exploration context, this
difference in opportunity and risk is often reduced to “greenfields” versus “brownfields”. The former is
classified by areas lacking deposits with little supporting infrastructure and many unknowns, whereas
the latter is characterised by areas possessing deposits with better known geology and more supporting
infrastructure.

Good ground selection underpins all successful exploration (Hronsky and Groves, 2008). Consequently, the
most fundamental decision is the decision to take up or abandon areas for exploration, or to seek to manage
risk through a joint venture with another organisation (Guj, 2011). Fundamentally, the decision maker must
separate regions into favourable and unfavourable regions. These regions may vary considerably depending
on the needs of the organisation (Whiting and Schodde, 2006). As an example, one company may be seeking
relatively low technical risk opportunities in a "brownfields" region, while another company may be willing
to accept higher technical risk in a "greenfields" region in the hope of higher reward. In either case,
geoscience data define the degree to which the region can be considered “known”, with associated
implications for technical risk and reward.

The next level of decision, once a region is chosen (or before it is abandoned) is to decide if more data are
required, and if so, which data set(s) to acquire. Implicit in this decision is, firstly, a knowledge of the degree
to which geology is known, and secondly, an understanding of which data types are most efficient for
progressing to the next phase of exploration. Although less fundamental than ground selection, poor choices
at this stage are a substantial barrier to exploration success (Kreuzer, Etheridge, 2008). In this case, a
judgement must be made regarding the potential for new data to add value relative to the cost of its
acquisition (Eidsvik and Ellefmo, 2013, Kreuzer, Etheridge, 2008, Perrons and McAuley, 2015, Vearncombe et
al., 2017).

The next stage of the cycle follows any new data collection, and the decision is whether this data has
provided enough new information to proceed towards the next exploration phase, or to prompt a change of
exploration strategy, or to change the region of exploration focus (Hronsky and Groves, 2008, Kreuzer,
Etheridge, 2008). The impact of new data on knowledge is always positive, but depending on the aims of the
organisation, the value of the data may vary considerably, and it may of course have either positive or
negative implications for the specific exploration program.

From the above, it is clear that this decision cycle carries a high level of implied knowledge about both the
value of data that has been or will be collected and also of the optimum approach to gain knowledge
through new data collection. For robust decision making, it is important that a degree of rigor is applied to
defining the level of data richness and the potential to improve it.

Methodology

Data richness concepts

For our purposes we define data richness as the opportunity provided by data to gain useful knowledge that
can help make exploration decisions better, or easier to make. We note that data richness does not define
the quality of the decision itself, which may be influenced by a wide variety of other factors.

Numerous studies have investigated decision making processes in the face of complex data sets. Task
complexity can be defined in several ways, considering a number of dimensions, such as the number of task
components, ambiguity in task components and relationships between task components (Liu and Li, 2012,
Wood, 1986). A simple overarching definition with relevance to geographic problems may be the number of
knowledge states that are possible in the task (Swink and Speier, 1999). A complex task, like finding an ore
deposit, will be characterised by a long and multi-branched decision tree, with considerable uncertainty
associated with each decision (Kreuzer, Etheridge, 2008, McCuaig, Kreuzer, 2007), leading to many possible
knowledge states.

Exploration relevant data are fundamental to selecting which branches to follow, and we consider that good
targeting data will, in general, reduce task complexity by eliminating unpromising branches (Kreuzer,
Etheridge, 2008). On occasion, new data will prompt a "new" branch to be required. This apparent increase
in task complexity comes from the uncovering of considerations that were previously unknown, and so
contributes positively to the ability to make a sound decision, although not necessarily an easier one. An
abundance of data is an increasingly common problem that can increase the perception of task complexity,
or so called "information overload" (Li, 2016, Perrons and McAuley, 2015). However, this is not universal (Li,
2016), and can often be overcome through appropriate data selection or aggregation techniques (Perrons
and McAuley, 2015, Swink and Speier, 1999) and good presentation (Li, 2016, Speier, 2006). Consequently,
we consider the contribution of new data to decision making to be always positive, although it may be very
small.

Swink and Speier (1999) explored, using spatial data sets, the influence on decision making of problem size,
data aggregation and data dispersal. In their study, problem size encompasses the inherent complexity of
the problem, with a constant level of data richness and no uncertainty. In the case that these are non-
uniform, we can reframe the concept of problem size to include also the amount of data that are available to
solve the problem, and the applicability of those data to the problem. Their study indicates that increasing
problem size negatively affects both decision time and decision quality. Data aggregation reflects the degree
to which raw data are collated and clustered for analysis (e.g. by spatial averaging or binning). In their study,
higher levels of data aggregation did not strongly influence the quality of the decision, but it had a significant
effect in reducing the decision time. Data dispersal indicates the degree of spread in the data values, with
the notion that more dispersed data provides more recognizable patterns. Their results suggest that decision
quality was superior with highly varied data, and that there was little effect on decision time.

The results of Swink and Speier (1999) suggest that sound decisions are more likely where clear data signals
are present for problems that are simple relative to the richness of the data. In the presence of time limited
decision making, the ability to aggregate the data meaningfully is of use. However, their study did not assess
the problem of data quality, which is fundamental to resource exploration due to the typically imperfect
nature of geoscientific data.

We consider data to possess a hierarchy of levels that describe its value to a process (Figure 1). For each
data set, three levels of value exist that are separate in concept, although in some cases the data does not
permit clear separation of the levels. These levels are hierarchical, in that L3 is not relevant without L2, and
L2 is not relevant without L1. Not all data sets possess quantifiable descriptions of all three levels, however,
so the hierarchy may be truncated at L1 or L2. A fourth level (L4) considers the integration of these data sets,
accounting for redundancies and interdependencies.

Figure 1: Four levels of data richness. Level 1 considers the presence of nearby data, level 2 considers the
quality of those data, and level 3 considers the data's attributes. Level 4 considers how different data sets
interact. These form the basis of the analyses in this study. Level 4 is considered here on a conceptual basis.

Our first level (L1) describes the presence of data nearby. This component will typically be defined by a
numerical measure of either data density or the distance to data. On occasion, presence versus absence
may be a good descriptor. At this stage no information is included on the value of the data, merely their
existence. However closely the data may be spaced, the likely presence of "bad" or irrelevant data in the
data set means that L1 mappings are not yet an adequate descriptor of data richness.
Our second level (L2) considers the intrinsic quality of the data, i.e. their precision and relevance to the purpose, but without considering the properties of the data themselves. Data richness as we define it considers the data solely as it is applied to a particular process, and so high quality data can be low richness if it does not help support a decision. Typical descriptors at L2 will include numerical or categorical descriptors of precision or confidence, as well as definitions of applicability or usefulness. For example, we may wish to downgrade observations from geological units of unsuitable age or lithotype regardless of their intrinsic quality. If we consider "problem size" to be defined as task complexity relative to the availability of useful data, L1 and L2 together can be considered the inverse of problem size, so higher data richness at L1 and L2 can be expected to lead to more sound decisions, and also to easier decisions in most cases.

At L2 the content of the data is not considered. Therefore, our third level (L3) is a consideration of the degree to which the data has the right attributes to be of use in solving our problem. This is similar in concept to the "dispersion" quality of Swink and Speier (1999) in that we promote clear and strong signals over weak or indistinct signals. We do not consider directly the difficulty of the problem that the data content may present, for example due to high geological complexity; we consider only the qualities of the data. High values of data richness at L3 can be expected to contribute to sound decisions, but not necessarily easier decisions.

Finally, in a typical exploration decision process several data types are combined, and so we must consider how these data sets interact to form the final decision, including, for example, redundancies and interdependencies between data sets. This constitutes our fourth conceptual level (L4) and is accounted for here in the different weightings of each data set, and form the framework for an overall estimate of data richness. Several methods exist to analyse interdependencies between datasets rigorously (Journel, 2002, Sun et al., 2013), and we do not explore this element further here. Level 4 is akin to data aggregation, and based on the work of Swink and Speier (1999) we can expect appropriate data aggregation to reduce perceived task complexity and decision time.

Mapping approach

We use standard Geographic Information System (GIS) tools to quantitatively map data richness, following previously developed data richness mapping methods (Aitken et al., 2013). Key operators include the following: Proximity is defined using the Euclidean distance operator while spatial data density is defined using the kernel density operator (Silverman, 1986), often weighted by a descriptor of data quality (L2). In addition, raster data sets are combined and modified using a variety of mathematical operators, including arithmetic, logarithmic and trigonometric operators. Most of these operators are contained within the Spatial Analyst module of ArcGIS and many are available as plugins for the open-source Quantum Geographic Information System (QGIS) package. The analyses here were performed in an automated fashion.
in ArcGIS using the Python-based geoprocessing functionality, which allows the process to be easily repeated in full or in part as new data are obtained. The specific operations for each data type used in our case study are described below and in Table 1. For this analysis we use public domain precompetitive data that are sourced from governmental data providers, including the Geological Survey of Western Australia and Geoscience Australia.

Our work in the case study is focused on constraining interpretations of near surface geology for a regional mapping exercise. The resulting maps were used as part of a prospectivity modelling exercise (Occhipinti et al., in prep). As a practical case study, we focused on methods that provide good spatial knowledge of lithological and stratigraphic variations across the region, and also are commonly used in regional exploration. The main data sets used were outcropping geology, drillcore and geochemistry data, as well as airborne radiometric and magnetic data and regional ground gravity data. Any other data sets of relevance could be included as long as a descriptor of at least level 1 data richness could be defined.

For outcrop geology we consider two descriptors of quality (L2), which in this case we base on the geological age of the outcropping rocks, and the scale of the best quality mapping available (Table 1). In this case we are interested in Precambrian geology, and so we weight Archean and Proterozoic rocks to 1. Phanerozoic rocks are lower weighted (0.5), but non-zero due to their potential to provide information of relevance to the older rocks. Cenozoic rocks are predominantly "cover" and are weighted zero (the value of regolith as a geochemical repository is considered separately). Map scale makes a substantial difference to understanding geological complexity (Ford and McCuaig, 2010) and so 1:250 000 scale maps are weighted at 0.4 relative to 1:100 000 scale maps (weighted 1). These are combined to give an Outcrop Quality Factor (OQF) for each outcrop (Aitken, Holden, 2013). We do not apply a level 3 descriptor to outcrop geology, as this is hard to define quantitatively at a regional scale. At smaller scales, or for more specific applications, L3 could be defined. For example, a numerical ranking of the lithological difference between a polygon and its neighbours could be used. To define the spatial relevance of surface geological data, we choose the spatial density of outcrop within a certain radius. This is defined using a kernel density function operator, each point weighted by its OQF. In this case a 5 km radius is chosen, reflecting the scale of the mapping and the resolution of the desired result.

For drillcore, only publicly available cores are used, and not exploration holes from companies, although at appropriate scales these could easily be included. As a suitable L2 descriptor we consider the maximum depth of the core. We do not apply a level 3 descriptor here as this is hard to define numerically without accessing the core, but there is obviously scope to define such a measure based on the contents of the drillcore log. As with outcrop we define combined L1/L2 data richness using a weighted kernel density operator with a radius of 5 km.
For geochemistry, we consider both hard rock and regolith geochemistry samples. A level 2 descriptor is defined based on the type of sample, and on the likelihood that the sample is representative of the underlying region. Regolith (excluding sand) is weighted 2, as in this region it has largely been found to form as a product of in-situ regolith weathering processes (Thorne et al., 2018). These data are considered more likely to be representative of a substantial area than the "spot" data from hard rock samples (weighted 1). Sand is assigned 0.25 as transported cover. No L3 descriptor is used here, but in less general cases this could involve, for example, particular element ratios that define a characteristic of specific value. We define combined L1/L2 data richness using a weighted kernel density operator with a radius of 5 km.

Data richness for airborne radiometric data is defined from L1 to L3. The properties of the data itself (L3) are considered through two measures: First we define the radiometric slope, which indicates the spatial rate of change of each of the three channels (U, Th, K). Each channel is normalised, the slope calculated, and the three slopes added together. This measure provides high values for what might be perceived in a ternary image as an "edge" in colour contrast, perhaps associated with a geological boundary. Another L3 measure is defined as the radiometric distance. For this measure we define, in U-Th-K space, the distance of a measurement from the centroid of the normalised data space. This measure provides high values for what might be perceived in a ternary image as a "peak" or "trough" in one or more channels, perhaps associated with a linear geological feature, such as a dyke or fault alteration zone. These normalised slope and distance measures are added, focusing on edges in this case (Table 1). For L2, a sample quality descriptor is not directly considered, as sampling methods and technologies are similar throughout and detection limits are low. However, the radiometric signal is only as relevant as the rocks it traverses. Therefore we multiply the L3 products by the OQF (excluding the mapping scale adjustment). Level 1 data richness is defined, on a cell-by-cell basis, by the inverse of the line spacing of the highest resolution survey to intersect that cell. This is multiplied by the combined L2/L3 data richness to yield total radiometric data richness.

Data richness for magnetic data is defined using similar principles to those applied to the radiometric data. L3 descriptors describe the total gradient magnitude using the analytic signal, and its arc-tangent. The former accounts for the use of amplitude dependent images (e.g. vertical derivative (Hood, 1965)), whereas the latter accounts for the use of amplitude independent images (e.g. tilt (Miller and Singh, 1994)). These are combined with equal weighting in this case. For L2, a sample quality descriptor is not directly considered, as sampling methods and technologies are similar throughout and detection limits are low. Level 1 data richness is defined by the inverse of the line spacing of the highest resolution survey to intersect that cell. This is multiplied by the combined L2/L3 data richness to yield total magnetic data richness.

Finally, we consider regional ground gravity data. For L3 we use the analytic signal as in magnetic data but omit the amplitude independent arctangent operator. For L2 we must consider the accuracy of the gravity
data which have been collected over several decades and with differing technologies. The Geophysical Archive Data Delivery System (GADDS) database includes a station error (Wynne and Bacchin, 2009), and we use this to describe the quality of individual measurements. For L1 we use the kernel density operator, weighted by station error, with a radius of 20 km. The data density of the gravity data shows a log-normal distribution, with very detailed surveys around some mining areas, and > 10 km spaced data in others, and so the logarithm was taken. This composite L1/L2 product was multiplied by the L3 gradient magnitude to provide the full gravity data richness map.

<table>
<thead>
<tr>
<th>Category</th>
<th>Proxy Map</th>
<th>Basis</th>
<th>Quality Assignment</th>
<th>Process</th>
<th>Sub-weight</th>
<th>Overall weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcrop Geology</td>
<td>DQF</td>
<td>Presence and relevance of outcrop</td>
<td>DQF: Archean/Proterozoic = 1; Paleozoic/Mesozoic = 0.5; Cenozoic = 0</td>
<td>Kernel density of DQF points R = 5 km</td>
<td>1</td>
<td>0.45</td>
</tr>
<tr>
<td>Drillcore Geology</td>
<td>Depth-proportional</td>
<td>Presence and usefulness of drillcores</td>
<td></td>
<td>Kernel density of core locations R = 5 km</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>Geochemistry</td>
<td>Presence and usefulness of gravity data</td>
<td>By sample type: Regolith (ex.Sand) = 1; Hard-Rock = 0.5; Sand = 0.25</td>
<td></td>
<td>Kernel density of sample locations R = 5 km</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>Radiometrics</td>
<td>Slope</td>
<td>Overall rates of radiometric changes</td>
<td>&quot;Slope&quot; function applied to normalised U, Th and K channels and added together</td>
<td></td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance</td>
<td>Deviation from mean</td>
<td>Cumulative “distance” from the U, Th-K centroid</td>
<td></td>
<td>0.25</td>
<td>0.1</td>
</tr>
<tr>
<td>Magentics</td>
<td>Gradient Magnitude</td>
<td>High gradient magnitude indicates interpretable signal</td>
<td>Analytic signal describes gradient content (amplitude dependent)</td>
<td>Analytic signal added according to weighting factor</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Line Spacing</td>
<td>Amount of data per unit area</td>
<td>Select highest resolution survey at each location</td>
<td>Inverse of Line Spacing</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outcrop Mask</td>
<td>Null values where no outcrop</td>
<td>Outcrop as per DQF, without map scale component</td>
<td>Multiply by DQF</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Gravity</td>
<td>Data Density</td>
<td>Amount of data per unit area</td>
<td>Quality assigned by station error from GADDS</td>
<td>Kernel Density of station locations, R = 20 km</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Gradient Magnitude</td>
<td>High gradient magnitude indicates interpretable signal</td>
<td>Analytic Signal describes gradient content (amplitude dependent)</td>
<td>Multiply by Data Density</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Basis, assignment, and processing of data richness products for each data type used in this study.

The final process is to aggregate all of these data richness indicators for individual data streams into a single image. This aggregation could use methods such as fuzzy logic or machine learning should the data sets and purpose warrant these methods. For example, redundancies between data sets might warrant a fuzzy OR operator whereas interdependent data sets might warrant a fuzzy AND operator. Following Aitken (2013), we consider in this case that our data sets are sufficiently independent that a weighted sum was adequate.

Weightings were defined on the level of usefulness of these data to define near surface geology. Outcrop, where it exists, is considered the premium data set and is weighted at 0.45. Magnetic data is considered very valuable in mapping beneath cover, and is assigned a weighting of 0.25, whereas radiometric data is reduced to 0.1 due to its limitations in areas lacking outcrop, and partial redundancy with outcrop data. Gravity is given a weighting of 0.1 due to low resolution through most of the study area. For this regional mapping exercise, drillcore and geochemistry were given low weightings (0.05) due to the sparseness of core in the former case, and the limited correlation with "geology" in the latter. Of course these may be weighted differently in studies undertaken with a different purpose.
Data Richness in the Capricorn Orogen

The mapping process described above provides an image of data richness within the Capricorn Orogen of central Western Australia (Figure 2). This map is targeted to the process of producing a revised regional map of near surface geology (Figure 3), and describes, to an extent, the opportunity provided to the interpreters to make correct interpretations. Low data richness implies that the map was made from few relevant data, or data of low quality, whereas high data richness indicates an abundance of good and useful data.

Figure 2: Relative data richness in the Capricorn Orogen region, including richness estimates for outcrop geology, geochemistry, magnetic, radiometric, gravity and drillcore data (see Table 1). The white line indicates the boundary of the remapped region (Fig. 3). The black box indicates the Bryah Sub-basin case study area (Fig 4).

The Pilbara Craton, Gascoyne Province and Bryah Sub-basin regions are all relatively data rich, whereas the Yilgarn Craton and the Ashburton and Eararheedy Basins have more variable data richness, ranging from quite low to locally high. The Bangemall Basin in the central orogen is less well served by current data sets, especially in the east. The Phanerozoic Officer and Carnarvon Basins also have low data richness in this analysis due to the lack of information pertaining to Proterozoic basement geology although this does not imply low data richness in general.
Applications: Data Richness in exploration decision making

Based on the exploration decision process and the concepts outlined above, exploration success demands, in the first phase, sound judgements of the level of risk and potential reward associated with areas for exploration and, in the second phase, a sound judgement regarding which data types are most likely to lead to improved knowledge.

In this section we demonstrate an approach that combines data richness with past exploration data and also a prospectivity model to assist in these two phases of decision making. We apply this mapping method in a case study of the Bryah Sub-basin region, which has relatively high data richness. We focus on gold exploration for which there is a suitable history of past and current exploration (Fig. 4). Although we use real data, our purpose is to illustrate the method, rather than to provide a genuine assessment of exploration potential, for which our data sets are not sufficiently detailed.
Figure 4. Reclassified past exploration data in the Bryah Sub-basin Region. Interpreted bedrock geology is coloured using the same legend as Figure 3. MINEDEX entries (gold only) were selected with the 1:100 000 map sheets shown (red boxes) and are reclassified as described in the text. Sizes of all symbols indicate known endowment, including current resource and reserve estimates and past production. Italicised letters indicate particular projects: L - Labouchere; F - Fortnum; PH - Peak Hill; NS - North Star; H - Hermes.

Past exploration data in the Bryah Sub-basin region

A common approach to understanding belt-scale mineral endowment is to conduct a power-law analysis to predict whether there is "missing" endowment relative to the largest deposit, which is often among the first discovered (Fallon et al., 2010, Guj, Fallon, 2011, Lisitsin et al., 2014, Lisitsin et al., 2010, Yigit, 2012). In our case we are mapping spatial variations in endowment within a belt, and so a different approach is needed.

To conduct this analysis, we use the public domain mines and mineral deposits database, MINEDEX (GSWA, 2018) to define resources within the belt. We select only entries listed with gold as the target commodity and limit the area to the following 1:100 000 scale map sheets: Moorarie, Padbury, Milgun, Jamindi, Bryah, Glengarry, Doolgunna (Fig. 4)

MINEDEX separates the record of resources and their development into several categories, including occurrences, prospects, deposits and mines. Occurrences and prospects represent two stages of exploration; an occurrence being recorded where economic minimum cut-off limits for minerals are exceeded, and a prospect being recorded in the case of a number of significant intersections of mineralisation. Once a
A maiden resource has been released by a company, the prospect is upgraded to a deposit. Mines include a number of stages, including proposed, under development, operating, care and maintenance, and shut.

Figure 5 Resource Rate within the Bryah Sub-basin Region as defined from MINEDEX data. MINEDEX symbols are coloured as per Figure 4. This measure indicates the rate of resource discovery (E) relative to exploration activity (n).

For our analysis, we reclassify the MINEDEX entries into four categories that describe the intensity of past exploration and the level of success recorded within any particular area. The first category, "Fire", represents evidence of ongoing exploration success, whereas the second, "Smoke", indicates evidence of potential future success, but not current success. The third, "Ashes", represents evidence of past exploration success, whereas the fourth, "Cinders", represents evidence of past activity but not at substantial levels.

Thus, we consider "Fire" and "Ashes" representative of recent and past success, respectively, whereas "Smoke" and "Cinders" are representative of exploration activity (recent and past, respectively).

"Fire" includes deposits and mines with an unmined resource estimate exceeding 200 kg contained Au (approximately 7000 oz). Resource estimates include gold contained within resources and reserves from MINEDEX. The cut-off value is defined by the deposit with the lowest resource estimate (the Blake deposit, part of the Hermes Project), and excludes the small leftover resources associated with some shut mines. This yields 31 "Fire" sites including at Hermes, Peak Hill, and Fortnum (Fig 4). These have a cumulative
endowment of 106 071 kg Au. "Smoke" includes all prospects and occurrences, as well as deposits with no resource estimate. A total of 190 smoke records are generated (Fig 4).

"Ashes" includes all significant past production from care and maintenance or shut mines. A cut-off of significant production is imposed at 1400 kg Au (50,000 oz). Within all of MINEDEX, this cut-off includes the top 210 mines (thus excluding the remaining 9421 locations), from which 93% of past production was derived. In the case study, four records are included in the ashes category, namely Fortnum (18 477 kg Au), Labouchere (6 949 kg Au), Peak Hill (20 572 kg Au) and North Star (7 017 kg Au). "Cinders" includes all other past production from care and maintenance or shut mines, with 227 records returned (Fig 4). The cinders category has a cumulative past production of 3 430 kg Au.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Conceptual Basis</th>
<th>Derivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource Rate</td>
<td>Resource endowment relative to activity.</td>
<td>$RR = \frac{E}{n}$</td>
</tr>
<tr>
<td>Strike Rate</td>
<td>Ratio of success to activity</td>
<td>$SR = \frac{F + A}{n}$</td>
</tr>
<tr>
<td>Currency</td>
<td>Ratio of recent activity to total activity</td>
<td>$C = \frac{F + S}{n}$</td>
</tr>
<tr>
<td>Positivity</td>
<td>Ratio of recent success to recent activity</td>
<td>$P = \frac{F}{F + S}$</td>
</tr>
<tr>
<td>Freshness</td>
<td>Ratio of “in ground” resources relative to all resources</td>
<td>$Fr = \frac{EF}{E}$</td>
</tr>
<tr>
<td>Resource Potential Rate</td>
<td>Ratio of “in ground” resources relative to activity</td>
<td>$RPR = \frac{EF}{n}$</td>
</tr>
</tbody>
</table>

Table 2: Descriptors of resource endowment and past exploration history from reclassified MINEDEX data. $E$ - resource endowment from all sites, $n$ – number of sites, $F$ – number of Fire sites, $A$ – number of Ashes sites, $S$ – number of Smoke sites, $EF$ – resource endowment within Fire sites.

From these reclassified MINEDEX entries, we derive several descriptors of resource potential (Table 2). These are calculated for defined areas, and in this case we use a 5 km by 5 km graticule. Firstly, we define the Resource Rate (Figure 5) as the total endowment (E - including all production, plus resources and reserves; see Supplementary Figure 1) divided by the number of MINEDEX entries (hereafter, $n$; see Supplementary Figure 2). This parameter describes the overall rate of resource discovery relative to activity, and is highest in frontier regions with few entries.

In addition we define several measures that describe the state of past exploration in the area. Strike Rate (Supplementary Figure 3) is defined by the number of successes ("Fire" and "Ashes" entries) divided by $n$. This describes the overall rate of success relative to activity. Currency (Supplementary Figure 4) is defined by
the number of recent sites ("Fire" and "Smoke" entries), divided by \( n \). This is intended to describe the amount of recent exploration activity relative to past activity. \textit{Positivity} (Supplementary Figure 5) is defined by the number of "Fire" entries divided by the number of recent sites ("Fire" and "Smoke" entries) indicating the rate of success in recent times. \textit{Freshness} (Supplementary Figure 6) is defined by the resources contained within "Fire" entries divided by the total resources, \( E \). This descriptor indicates the degree to which significant resources remain, with potential to be mined. Finally, we define \textit{Resource Potential Rate} (Supplementary Figure 7) as the \textit{Resource Rate} multiplied by the \textit{Freshness}, to indicate the areas with the greatest amount of unmined resources relative to overall activity.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Overall data richness in the Bryah Sub-basin Region, also showing MINEDEX data, coloured as per Figure 4.}
\end{figure}

These descriptors of past exploration activity define only what has been done, and it is important to recognise the degree to which this may have been enabled or inhibited by the availability of useful data. In our first application of data richness mapping, we consider how combining the record of past exploration with data richness may provide a meaningful tool to assist in ground selection.

\textbf{Application 1: A simple data based classifier for ground selection}

As a simple decision making tool we propose the following classification scheme for ground selection, based on a) data richness and b) the degree of past success relative to exploration activity (the \textit{Resource Rate}) (Fig
7). Although many schemes could be used to classify this, we define four conceptual zones, defined as follows:

1. Little explored: Low data richness and low resource rate (or no exploration data)
2. Under explored: Low data richness and high resource rate
3. Well explored: High data richness and high resource rate
4. Over explored: High data richness and low resource rate (or no exploration data)

Figure 7 A simple 2D ground classification based on resource endowment rate and data richness

The first category has high potential for new data to aid discovery, and may be highly rewarding should there be a large deposit. However the probability of success is low, and low knowledge levels lead to an exploration environment containing high technical risk. The second category is potentially the most promising, with a large opportunity for new data to aid discovery, but also with some demonstrated success. This can be used to establish an exploration strategy characterised by lower technical risk of failure, although the largest deposits are commonly discovered in the early stages (Guj, Fallon, 2011). The third category has more limited potential for new data to aid decision making and discovery, although it may provide low technical risk opportunities with good data already available, and so may provide success over a relatively short term. The last provides limited opportunity for new data to aid discovery, and has not demonstrated success in the past. In this last case, it is possible that, despite high data richness, the region may not have been explored thoroughly for the commodity of interest, and so may preserve potential. This may occur as a consequence of a focus on a narrow mineralisation style or a specific commodity, or due to a cultural or administrative restriction on exploration. We do not consider these influences here, although we note that they could be accounted for in most cases.
Data richness in the Bryah Sub-basin region (Fig 6) has a positive skewed and weakly bimodal distribution, and we set the cut-off for high data richness at the higher modal peak (0.5). This limits the "data rich" areas to quite a small proportion of the map area (16%). For Resource Rate, the distribution is close to a power-law distribution and the cut-off was set at the onset of significant upward curvature (~ 40). This limits the resource rich areas to a very small proportion of the map area (2%).

The resulting classification (Fig 8) indicates much of the map is classified as "little explored". It is appropriate for such regions to consider new data collection to support predictive targeting approaches as part of a "greenfields" exploration strategy. Along the main ranges, which have substantial outcrop, the region is predominantly classified as "over explored". New data collection for such regions is perhaps not warranted and it may be preferable to release such ground. Any data collection should focus on supporting a predictive approach (Hronseny and Groves, 2008).

Figure 8: A simple classification of resource potential and data richness in the Bryah Sub-basin Region. MINEDEX data are shown with the same colouring as Figure 4.

A few small areas are classified as "well explored", notably the regions around Fortnum and Peak Hill. Opportunities for new gold deposits in these regions may be limited, but it is appropriate to consider new data collection to support a "brownfields" exploration strategy. Finally the classification highlights a few regions which are "under explored" including around Hermes, in the Horseshoe Range and in the extended...
regions of Labouchere, Fortnum, and Peak Hill. It is appropriate for such regions to consider new data collection to support the transition from prediction based to detection based targeting (Hronsky and Groves, 2008, McCuaig and Hronsky, 2014, McCuaig, Beresford, 2010).

Application 2: Coupling data richness with prospectivity analysis for enhanced classification

It is becoming common practice to include quantitative models of mineral prospectivity within the exploration workflow, and so we consider here how this might be combined with the data based classification above. Mineral resources data, data richness and prospectivity modelling are complementary methods and should be analysed in a coupled fashion, but with clear separation. We achieve this by considering each an axis of a 3 dimensional model space (Fig 10). Although there is an implicit orthogonality between these axes, interdependencies and redundancies will exist between them. Not least is the requirement that good data is needed for an accurate prospectivity analysis, with the consequence that low resolution data will underestimate variations in prospectivity, and show less correlation with mineralisation (Ford and McCuaig, 2010, Lindsay, Aitken, 2016). In addition, data is likely to be clustered around existing deposits, and knowledge from those deposits is often used in defining the prospectivity model workflow (Porwal and Kreuzer, 2010).

Figure 9: 3D classification based on data richness, modelled prospectivity and mineral resources data. Each face defines a knowledge plane with four classes. The data plane classes are LE – Little Explored, OE – Over Explored, UE- Under Explored, WE – Well Explored (see Fig 7); the prospectivity plane classes are LP – Low Prospective, URP – Unrealised Prospectivity, UXP – Unexplained Prospectivity, RP – Realised Prospectivity (Supplementary Figure 8); finally, the exploration plane classes are KL – Knowledge Limited; DL – Data Limited, PL – Prospectivity Limited, and EL – Exploration Limited (Supplementary Figure 9). These combine to give eight categories of exploration potential. See text for discussion of these and the categories defined.
Adding a prospectivity model to the classification, leads to three distinct planes (Fig 9), which we define as the data plane, the prospectivity plane and the exploration plane. The data plane describes the extent to which the region is known from data, and it is defined and classified as in the previous application (Figure 7). The prospectivity plane describes prospectivity from models and data, and is comparable to conventional methods for interpreting prospectivity models relative to endowment. Classes describe the relationship between the prospectivity model and the exploration data (Supplementary Figure 8). The exploration plane describes the limiting factors for further exploration success. The four states defined are knowledge limited, where knowledge level is too low to define possibilities; data limited, where the availability of exploration relevant data is the dominant limiting factor; prospectivity limited, where lack of apparent potential (or poor model knowledge) is the dominant limiting factor; and exploration limited, where neither data nor model prospectivity limits potential, but exploration opportunities may be limited by other factors (Supplementary Figure 9).

From this, eight classes can be defined for a more thorough description of exploration potential (Fig 9). The lower tier lacks mines and deposits (at least relative to prospects/occurrences) and can be categorised as follows:

1. Undefined potential describes regions where little is known, and basic data collection may be needed to better understand the mineral system and its opportunities.
2. Unrealised potential describes regions where, despite low data richness and few known resources, high potential is suggested by the prospectivity model. This region may present good opportunities for new discoveries.
3. Low potential describes regions where, despite high data richness, prospectivity is low and known resources are few. Unconventional opportunities may exist, however.
4. Undeveloped potential indicates regions with high data richness and high model prospectivity, but few known resources. Exploration activity may be restricted, or these regions may be characterised by subeconomic mineralisation.

In the upper tier, substantial deposits exist. Categories in the upper tier are as follows:

5. Unexplained potential is where existing deposits are not well described by the prospectivity model, but are in areas with low data richness. In this case data collection should focus on delineating the components of the existing mineral system.
6. Unsupported potential is where high prospectivity is associated with deposits, but data richness is low. In these cases new data are needed to build support for the discovery of new deposits.
7. Anomalous potential indicates deposits that exist despite low prospectivity and high data richness. In these cases a better model of the mineral system is needed.
8. Realised potential indicates where all elements are high, indicating a mature exploration region. Opportunities may exist for brownfields exploration and satellite deposits.

Recent studies have undertaken prospectivity analyses for the Bryah Basin region (Occhipinti, Lindsay, in prep), using a knowledge based method similar to previous works (Lindsay, Aitken, 2016, Occhipinti, Metelka, 2016). From this study we extract a gold prospectivity analysis suitable for the Bryah Sub-Basin gold mineral system. We consider this analysis to be representative of the potential of the region to have experienced ore forming processes, but not necessarily the capacity for economic deposits to be present.

Within each 5 km cell, we take the maximum value of the prospectivity model found in that cell to represent its value. Then, for classification, we apply a cut-off at 0.14, representing half the maximum observed value within the analysis, and classify into high and low prospectivity on that basis.

Application of this 3D classification to the Bryah Sub-Basin region (Figure 10) provides further definition of the exploration opportunities within the region. Mineral exploration potential over three quarters of the region is classified as “undefined” with “low potential” also common in several parts of the region.

Unrealised potential is found in some areas, often in regions with existing deposits, but also in some new areas. Undeveloped potential is also found, associated with better exposed regions with many prospects and occurrences, but few deposits (Fig 10).

In areas with deposits, realised potential is found at Peak Hill and at Nathans, whereas unsupported potential is found in regions around Hermes, Fortnum, Horshoe Range, Wilgeena, Harmony and Harrods (Fig 10). A few unexplained and anomalous potential results are returned, mostly associated with small deposits (e.g. Homestead, Bloodstone, Durack) but also the large mine at Labouchere. All these are located on the fringes of the basin, close to large regions of extensive cover, low data richness, and undefined potential (Fig 10). These anomalous results may therefore be explained by the limited resolution and accuracy of the prospectivity model in low data richness areas, rather than any systemic deficiency in the model.
Figure 10: 3D classification applied to the Bryah Basin region, including a prospectivity model. Eight categories are defined showing exploration opportunities within the region, see text for definitions. L – Labouchere, N – Nathans, F – Fortnum, HR - Horseshoe Range, PH – Peak Hill, He – Hermes, Ho – Homestead, Hd – Harrods, Ha - Harmony, D – Durack, W – Wilgeena, B – Bloodstone. The Yellow Box indicates the Peak-Hill/Hermes Region (see Figure 11)

Application 3: A tool for data value comparison

A key question remains as to the types of new data that are most likely to lead to further success. The data richness mapping approach can be used to define, within potential target areas, which data sets are below the regional standard. As an example, we examine the components of data richness in the Peak Hill – Hermes region (Fig 11), which preserves a contrast between the historic Peak Hill exploration area and some frontier regions, including the sizeable deposit at Hermes. With our public domain data, drillcore data richness is limited to one cell. The impact of new drilling is not discussed further here, although drilling programs are of course a substantial component of undercover exploration. Gravity data have relatively consistent spacing (2 km), and most of the variability is due to contrasts in L3 data richness. New data would potentially be of benefit in high signal strength regions in the south and north.

The Peak Hill region contains the largest historic gold production within the region, from a main mine group and a number of smaller satellite mines and deposits. The data richness around Peak Hill is high, owing to
abundant outcrop and high quality airborne geophysical data with strong signals. Consequently, the immediate surrounds of the Peak Hill region are classified as "well explored" (Figure 8) and have “realised potential” (Figure 10), consistent with an exploration history extending back to 1897 (Armstrong et al., 1970).

Figure 11: Ensemble of data richness elements in the Peak Hill/Hermes region. Top Row: a) overall data richness with a colour scale as per Figure 6 b) Resource Rate with a colour scale as per figure 5 c) 3D classification, with colour scale as per Figure 10. MINEDEX data are shown on a), b) and c) with colours as per Figure 4. Middle Row: d) outcrop data richness, with outcrop outlines in black e) geochemistry data richness with sample locations shown including weighting (W), f) drillcore data richness with collar locations (stars). The relative magnitudes of d, e, and f each use the colour scale at the right hand side. Bottom Row: g) radiometric data richness with outcrop outlines in black, h) magnetic data richness, i) gravity data richness. Classified magnitudes of g,h, and i each use the colour scale at the right hand side. See Fig 10 for location. Ha - Harmony Mine (shut), E - Enigma deposit, D - Durack deposit, PH - Peak Hill Group, W - Wilgeena Mine (shut), H - Hermes Group

Both data richness and exploration success reduce away from Peak Hill leading to areas that are "little explored", with “unrealised potential” and for which new data might benefit exploration. Most of the reduction in data richness reflects reducing outcrop, and a corresponding increase in Cenozoic cover, perhaps causing the reduction in exploration success. The region to the south and west hosts the Harmony
mine (Resource – 3 566 kg Au), the Enigma deposit (Resource – 2 069 kg Au) and the Durack deposit
(Resource – 3 484 kg Au). Harmony is situated within an "under explored" region with “unsupported
potential”. Durack and Enigma occur in regions classified as "well explored" and with “anomalous potential”
and “realised potential” respectively. It should be noted that with data richness values of 0.52 and 0.55,
respectively, they are very close to the cut-off. Data collection within the immediate vicinity of deposits may
focus on more directed targeting datasets, not considered here, consistent with their classification (Fig 11).
The large region to the southwest is classified as “little explored” and with “undefined potential”. This
southwestern frontier is fairly well served by geophysical surveys but has in places low data richness due to
weak signals (Fig 9h,i). Low data richness exists for geochemistry data due to poorly distributed sampling,
and improvements to that data set are likely to be most beneficial to regional exploration in this area.
The region north and east of Peak Hill possesses no deposits and low data richness, but is in part classed with
“unrealised potential”. This region has little outcrop, and is served only by low resolution magnetic and
radiometric data (400m), although geochemical and gravity data richness are each relatively high. Therefore,
for this region, improvements to the airborne geophysical data coverage may be of most benefit to regional
exploration.
In the south of the region, Wilgeena mine (Resource - 1543 kg Au) sits in a region classified as "under
explored" with “unsupported potential”. Data richness is 0.49 so, like Durack and Enigma, this is close to the
threshold. The surrounding region is classified with “undeveloped potential” to the east and west, but
“unrealised potential” to the north. Data richness is moderate for outcrop and is mixed for magnetic and
radiometric data, with 50m spacing in the south but just 400m spacing in the north. Geochemistry data
richness is low due to a lack of samples. Further regional exploration in this region, in particular the region
to the north, might benefit most from upgrading the airborne geophysical and geochemistry data coverage.
The Hermes region is different to Peak Hill in that it there is no significant historic production, but there is a
significant group of deposits, with a combined MINEDEX resource estimate exceeding 19 000 kg Au (approx.
670 000 oz). Data richness is high in the southeast, where substantial outcrops and high resolution airborne
geophysical surveys (50 m) both exist, and the region is classified with “undeveloped potential”. Hermes
itself is situated within an area of low data richness, classified as “under explored” and with “unsupported
potential”. In the context of new data, the deposit proximal region has good geophysical coverage, but
overall low magnetic data richness due to fairly weak gradients, and low radiometric data richness due to
lack of outcrop. Chemistry data richness is also low, with few data within the region. It is likely, therefore,
that the most valuable data to improve data richness within the deposit proximal region is additional
geochemistry.
The region to the west is classified as “little explored” with “unrealised potential” and is perhaps prospective. This region is strongly deficient in several data types, including geochemistry, for which there is a distinct data gap, and high resolution airborne geophysical data are not available (existing data are 400 m spaced). Improvements to these data types are likely to be of most benefit to exploration in this region.

The decision to collect new data includes, as outlined above, the option to collect more detailed, more extensive or higher accuracy surveys of the same data types already analysed. In these cases, the existing data richness map can be used to guide the collection of the new survey. Moreover, the mapping process can be repeated to quantify the influence of new data collection. Alternatively, options exist to collect other data types that were not used in the original analysis. Strictly, the current data richness map provides no information on these new data types, and to include them the mapping process must be reconfigured at level 4 to take account of the new data, their value, and relationships with other datasets. In practice, redundancies and interdependencies between data types may allow data richness to be used to avoid redundancy and maximise benefit during planning.

Conclusion

Successful exploration is contingent upon some key decisions including ground selection and the choice of data to collect in an attempt to improve baseline knowledge, to develop predictive capacity, and/or to improve detection capability. Each of these decisions depends on a knowledge of existing data and its quality, as well as what data might be most useful to build further knowledge. In our experience of exploration practice such matters are implicitly considered from a heuristic standpoint, but it is not common to approach this process in an explicit and numerically rigorous manner.

Here we have demonstrated an approach for semi-automated mapping of data richness that can be applied without undue difficulty and using commonly available GIS tools. Using a case study of gold exploration in the Bryah Sub-basin, Western Australia, we demonstrate three ways in which this approach can be used to support decision making. The first application is as a simple data based classifier for ground selection, considering both data richness and past exploration data. The second application combines this with a prospectivity model to further characterise exploration potential in the region. The third is a data value comparison that highlights those data sets that are relatively deficient, with a view to collecting the most effective additional data.

Due to the static natures of the regional data sets used here, we were not able to map changes in data richness and endowment through time. However, the data richness mapping process could easily be used in a time-lapse manner, with repeated analyses following data collection programs, so long as the data set types remained constant. This would allow the growth of knowledge surrounding the project to be mapped through time, and might help identify data streams that are, or are not, proving effective. Indeed, at least
for L1 and L2 data richness, this step could be applied prior to a data collection program to help determine
the optimum spacing and target precision. Adding completely new data types, or combining or splitting data
richness elements, require a reconfiguration of the mapping process based on the value of the new data and
its interdependencies and redundancies with other data types.

In each of these applications, a role is presented for data richness mapping that can support decision
making. The first exploration phase involves selecting favourable and unfavourable regions according to the
requirements of the organisation. Our case study demonstrates how this process can be aided by comparing
the potential benefit of new data collection against the past record of resource discovery and models of
prospectivity. The second phase involves collecting the data most likely to be of benefit to exploration,
perhaps emphasising prediction or detection depending on the scale of the study and the maturity of the
region. With our approach, deficiencies in data can be recognised and, if deemed worthwhile, rectified.

Following or even preceding data collection programs, the impact of these data on data richness can be
compared with the cost of collection, and the effectiveness of different methods can be enumerated. When
combined with other indicators of exploration risk and potential reward these measures have a significant
role to play in making sound exploration decisions more easily.

Acknowledgements

This work was supported by the Science and Industry Endowment Fund (SIEF) project, Distal Footprints of
Giant Ore Systems - UNCOVER Australia (RP04-063). ARAA received additional support from the Goodeve
Foundation. MDL thanks the GSWA and the WA State Government Exploration Incentive Scheme, Royalties
for Regions for funding. The authors are indebted to Roger Cooper of GSWA for his help with MINEDEX data
and to the reviewers for their constructive comments. The data and modelling scripts associated with this
work are available from the authors on reasonable request.

References

Aitken ARA, Holden EJ, Dentith MC. Semiautomated quantification of the influence of data richness on

Goldfields, Western Australia; and, Early history of Peak Hill.1970.

Bagstad KJ, Semmens DJ, Waage S, Winthrop R. A comparative assessment of decision-support tools for


Czarnota K, Blewett RS, Goscombe B. Predictive mineral discovery in the eastern Yilgarn Craton, Western
Australia: An example of district scale targeting of an orogenic gold mineral system. Precambrian research.
2010;183:356-77.


Supplementary Figure 1 - Gold endowment, E (kg), for the 5 km graticule over the Bryah Sub-basin region. Includes past mine production, as well as estimated resources and reserves for mines and deposits.
Supplementary Figure 2 – MINEDEX entries (n) for the 5 km graticule over the Bryah Sub-basin region. Includes all entries with “gold” as target commodity. Gold resources contained in other styles of deposits are not considered.
Supplementary Figure 3 – Strike Rate (see Table 2 for definition) for the 5 km graticule over the Bryah Sub-basin region.
Supplementary Figure 4 – Currency (see Table 2 for definition) for the 5 km graticule over the Bryah Sub-basin region.
Supplementary Figure 5 – Positivity (see Table 2 for definition) for the 5 km graticule over the Bryah Sub-basin region.
Supplementary Figure 6 – Freshness (see Table 2 for definition) for the 5 km graticule over the Bryah Sub-basin region.
Supplementary Figure 7 – Resource Potential Rate (see Table 2 for definition) for the 5 km graticule over the Bryah Sub-basin region.
Supplementary Figure 9 – Four way classification in the prospectivity plane for the 5 km graticule over the Bryah Sub-basin region. Red shading shows region with model prospectivity above the cut off value, set at half the maximum value.
Supplementary Figure 9 – Four way classification in the exploration plane for the 5 km graticule over the Bryah Sub-basin region. Red shading shows region with model prospectivity above the cut off value, set at half the maximum value.