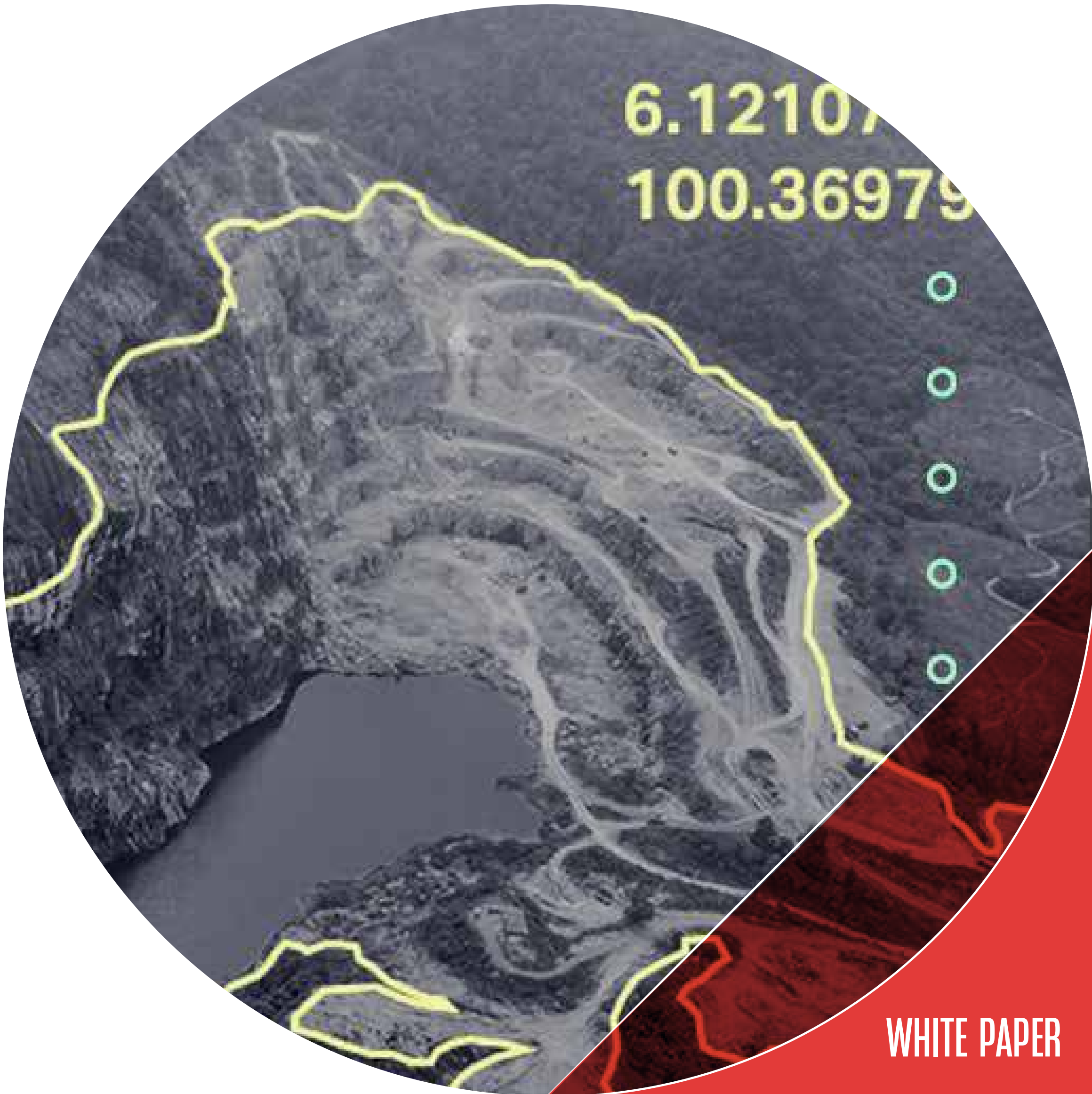
PART 1

GIS DATA ANNOTATION FOR MINING



WHITE PAPER

MINING

AND THE RESOURCE EXTRACTION SECTOR

Summary: The resource extraction sector has become the biggest consumer of GIS and geospatial data, and data annotation specialists have stepped up to fill a key role enriching GIS images and unearthing ground truth data. This is Part 1 of a two-part series.

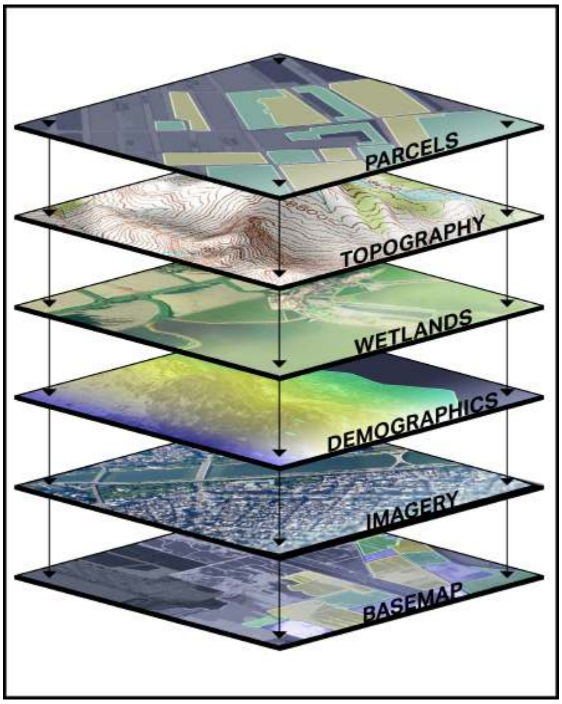
A series of cylindrical objects sit in the Middle East desert as the afternoon sun falls steadily to the horizon. Far overhead a satellite snaps a series of high-resolution images of the ground and facilities below. And somehow, in the days to come, a team of data analysts will examine those images and determine that the objects are oil storage tanks – and, after annotating hundreds of similar objects across the globe, that the world supply of oil within has run unexpectedly low.

It takes a highly trained pair of eyes to look at a series of seemingly abstract objects, discern the shape of storage tanks, take into account the shadowing from the time of day and angle of the sun, and subtle changes in shape that together present a compelling set of clues to the ground truth below. Fed enough images and correct modeling, a machine algorithm can eventually match that performance – but only after that expert human has developed the necessary training data to educate the machine equivalent of a three-year-old child and take it through graduate school.

Oil, gas, minerals, metal ore – that combined resource extraction sector has in recent years become an ever-hungrier consumer of GIS and other geospatial data – in terms of the breadth, depth, and frequency of that data collection and analysis. The insatiable appetite for geospatial data should come as little surprise; few industries are as closely connected to spatial data as mining, where data requirements span both above and below ground data – everything from high resolution satellite imagery to underground sensor data and surface GIS mapping.

A COMPLEX 'COMPUTER VISION' COLLAGE

In contrast to many other forms of Computer Vision data annotation, mining-related analysis typically involves far more varied (and variable) raw data, where automated annotation is a challenge at best. The scope of data requirements encompasses everything from natural topographical features – ground cover and fluctuations in elevation, the presence of wetlands, wildlife migration patterns, rivers and other bodies of water, and even seasonal fluctuations such as snow and cloud cover – to monitoring existing manmade facilities, such as coal, ore, and oil storage terminals, nearby population centers and even transportation access options for proposed sites, as well as land use designations; nearby transportation lines; and parcel ownership. The mix of imagery, map, and usage data is all combined to create a three-dimensional view of a mining site (potential or existing) and all the environmental, regulatory, and economic factors that can affect planning.



That helps explain how mining companies have become among the most frequent users of GIS as well as other forms of geospatial data, all with two primary use cases: exploration and environment protection activities. Exploration applications reflect an ever more scientific approach to pulling minerals and metal ore from the ground, while geospatial data analysis for environmental protection is a factor both of mining companies making more concerted efforts to be stewards of the land – and tighter government regulations borne of some high-profile environmental disasters over the past two decades.

In a typical exploration scenario, teams of geologists will create virtual models of ore layers using exploration drill data. In this example, data derived from exploration drill holes is represented through 3D map visualization enabling geologists to look underground without having to travel to mining sites. That has become even more of a factor in 2020 as the COVID-19 pandemic has placed a premium on remote, digital alternatives to a direct physical presence.

Mining firms generally source raw data – primarily aerial imagery (satellite, drone, and planes) from multiple sources and rely on specialty firms to annotate the data and create appropriate models. But primary usage generally involves large in-house teams of geologists, geoscientists, and geological data specialists, all with their own responsibilities and priorities.

Where human-annotated data is fed into AI-based algorithms, the goal typically is to develop predictive engines that can apply a set of rules on geological features, land formations, and other factors to identify the optimal potential mining sites worth more extensive exploration. In most cases, the process creates multiple-layer views based on different data groupings, including geological and topographical mapping, and even LiDAR and other remote sensor data.

That human/machine partnership develops over the course of thousands of images, sometimes hundreds of thousands for more ambitious projects, all with the goal of transferring human intelligence and reasoning into lines of code that run an algorithm. Analyzing a series of rock core samples to highlight the characteristics that define the richest mining opportunities, can eat up 10,000 images or more – each one of which, has to be annotated by a trained data analyst.

$692B Revenue in 2019 of the top 40 global mining companies which represent a majority of the industry.

Large mining companies often require operations analysis across multiple sites, for example, annotating the different buildings and other structures from one site to another, and even the boundaries of those operations. While some structures are fairly standard, such as equipment and machinery storage, others can vary widely from one site to the next, and sometimes it's far more cost effective to train a human data analyst to identify a series of specific types of structures. If the task includes delineating the boundaries of one mining site apart from any surrounding properties, it can be difficult at best to rely on Artificial Intelligence engines to draw those lines.

Sometimes the analysis – and required analytical skills – extend well beyond the immediate confines of a prospective mining site. An analysis team might need to examine multiple data sources to determine whether there are environmental risks to the site (rather than the other way around). For example: particularly in this time of increasingly severe weather events, does the proposed mine sit within a flood plain? Is there a potential risk from wildfires? That type of project might involve annotation and analysis of ground cover, topography, and geological formations sourced from raw data (satellite and other aerial imagery) and well as seasonal data and historical data.

In the end, human analysis provides the necessary context in order to best understand all the variables that can affect the hundreds of millions of dollars – sometimes billions – invested in the extraction and transportation of oil, coal, iron ore and other raw materials.

The application of geospatial data to environmental-centered activities is arguably a larger – and expanding – sector, particularly in such mining industry hubs as the US, Australia, and Canada where tightening environmental regulations have placed a premium on disaster prevention and remediation. Geospatial data can be organized into dozens of layers of data to account for everything from nearby watersheds, underground aquifers, ground vegetation, property boundaries, wildlife movement, and even archeologically sensitive areas.

Over the past decade, the imperative to stay ahead of any potential environmental impact from mining activities has forced an exponential growth in the use of GIS data – in some cases an information overload that taxes the capabilities of even large mining firms. As one Canadian mining industry executives notes ruefully, “We used to buy geological maps that were updated every 10 years. Now for something like monitoring water levels (in a watershed) we can process real-time data from daily flyovers.” It is perhaps appropriate, that an industry best known for moving massive amounts of rock and soil to generate value from the precious materials hidden within now confronts a parallel task of big data mining to address information overload – and the very real risks hidden within.

80%

Savings while locating new mines using big data, as compared to traditional methods.

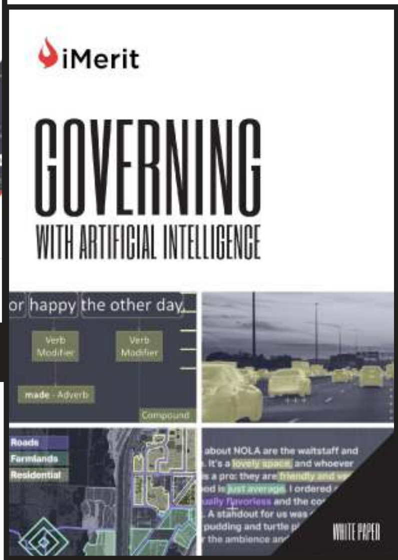
An increasing reliance on real-time geospatial data has created a significant big data challenge as the industry evolves from one formerly focused on periodic data snapshots. Many mining operations now routinely monitor mine output, and corresponding safety threats based on ambient temperature changes, changes in flow rates, weather changes, and even dust propagation. All of that data, flowing from satellite imagery, low altitude drones, and planes, can add up to an overwhelming deluge that – apart from ensuring the jobs of countless in-house technicians, geologists and other data analysts – has put a premium on external resources capable of turning all that data into actionable insights.

Productive geospatial data analysis also relies on the quality as well as the sheer quantity of raw data. Developing training data for mining-related AI software, for example, can be a laborious, human-intensive process, one that does not always lend itself to fully automated systems. Before raw data can be processed, geologists typically want someone to sift through the imagery to remove, for example, images with too much cloud cover or vegetation that obscures geological formations. Someone then needs to review the data, ensure consistency across a given

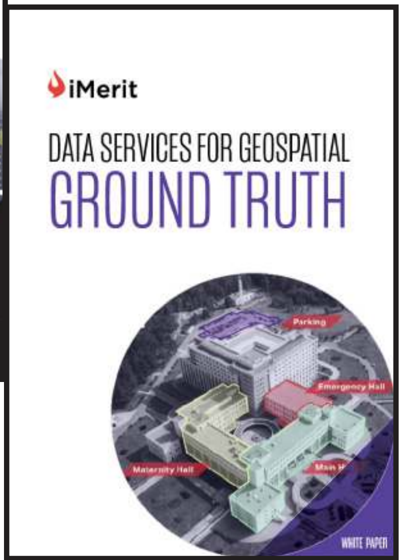
WHITE PAPERS ON APPLICATIONS OF AI

Long-form content, produced by iMerit subject matter experts

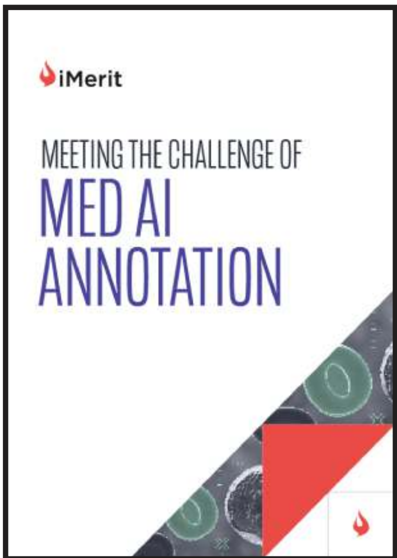
DOWNLOAD



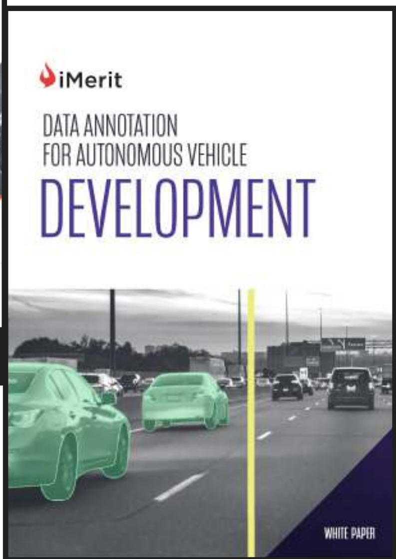
DOWNLOAD



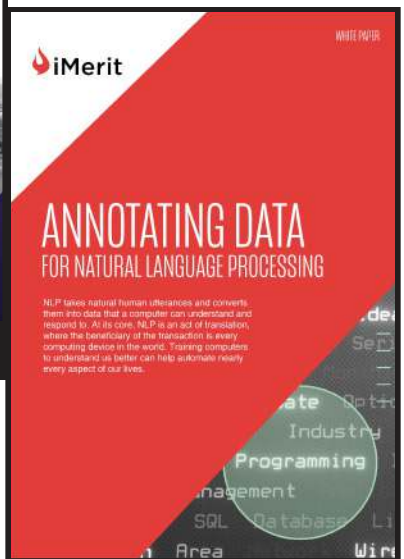
DOWNLOAD



DOWNLOAD



DOWNLOAD



DOWNLOAD

CONTACT US

geographic area, and correct for any errors.

For surface mining in particular, the need to monitor the environmental impact of even mundane activities can create complex data analysis problems requiring multidisciplinary skills. Open pit mining engineers, for example, make daily decisions on when to blast based on everything from wind and rain to the surrounding topography, land use, and wildlife patterns.

Demand for expert data analysis isn't limited to the confines of an open pit mine. Consider the challenge of building near real-time estimates of global, regional and country coal and iron ore stocks purely by analyzing satellite imagery of coal terminals and iron ore storage facilities.

Step one might begin with something as basic as a Google Maps search for all known storage facilities, so a data analyst can mark the coordinates by country and region. Step two might mean zooming in on high res imagery to annotate those facilities, including the use of bounding boxes and semantic segmentation to note the boundaries of the storage structures apart from any other structures or facilities. Step three might involve any indicators of activity and coal or iron ore capacity and its use in order to enable estimates of overall quantities.

10,000

Number of images to be annotated while analyzing a series of rock core samples to highlight the characteristics that define the richest mining opportunities.

A similar process can be applied to other resource estimation projects, for example, a recent iMerit project to develop a census of global oil exploration and storage capacity – and the subsequent development of machine learning algorithms capable of performing real-time estimates of the same. The data annotation and analysis component created massive amounts of training data – tens of thousands of images – of everything from oil rigs and pads, to storage tanks – not only to enable detection through high resolution satellite imagery, but also changes in activity, including fluctuations in production and storage.

Since oil rigs and pads have a generally consistent shape and geometry, well trained eyes can identify them from raw data and then apply polygon bounding boxes and semantic segmentation to define them in ways machine learning can apply to algorithm development. That information can be overlayed on map coordinates for known exploration sites to expand the process by country, region and then globally.

Training an algorithm to help estimate oil volume in a storage tank is a more complicated task but still one where skilled annotation can effectively train a reasoning model.

GIS and other geospatial data can also prove invaluable to the government agencies that oversee the resource extraction industries. GIS data, for example, can be used to track mining activity on a local, regional or national basis, process permits for exploration and extraction, and even enforce land reclamation and remediation regulations.

Some federal and state agencies analyze GIS data to determine government funding priorities for land reclamation projects – everything from returning streams and other bodies of water to their original pre-mining paths, to disposing of toxic waste, and returning open pit mines to something closer to a natural state. As with private industry mining companies doing prospective mining site due diligence, government agencies will analyze legacy site proximity to aquifers, streams, and other bodies of water, ground cover, urban areas, farms and other agriculture sites, and even wildlife migration corridors.

$290 - $390 B

Annual savings for mineral raw materials producers by 2035, achieved through technologies like AI.

Agencies have even acquired LiDAR and other sensor data to help create topographical models, in some cases using data that predates mining activities in order to return the legacy mine and surrounding area to the natural state that existed before mining. The LiDAR data can be used to ascertain the variations in terrain, particularly for open pit mining, as well as vegetation and other land cover.

Australia began a similar initiative in February 2017 when the Australian Senate approved an inquiry into "the rehabilitation of mining and resources projects." The resulting project – really a risk assessment that assigned an environmental danger score to each mining site – reviewed mining geospatial databases for all legacy (inactive) hard rock mining sites to determine the requirements for mining site reclamation projects. The work involved GIS-based spatial analysis that, as in the US, identified and categorized nearby infrastructure and environmentally sensitive areas.

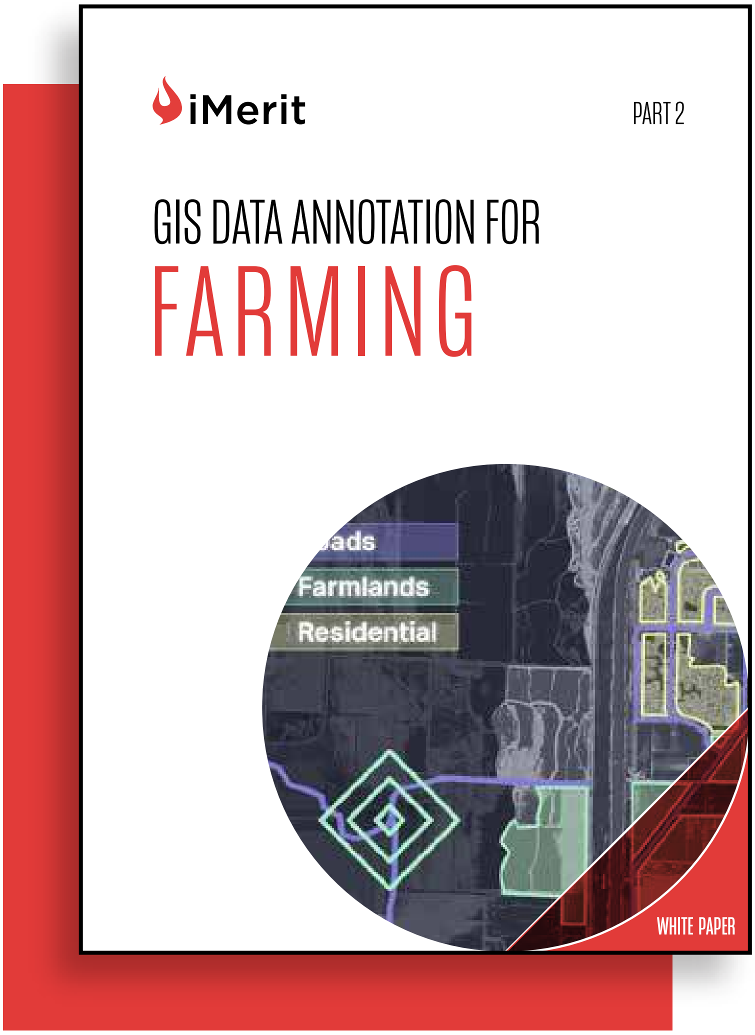
Among the GIS-based factors analysts annotated and reviewed were the proximity of the sites to major urban centers; roadways and rail infrastructure; rivers, lakes and other waterways; environmentaly and culturally sensitive indigenous sites and larger areas; national parks; and the presence of acid sulphate soils – identified by the project team as a crucial environmental factor.

While the basic annotation tool sets were straightforward (the use of bounding boxes to create geographically accurate polygons, lines and GSP-accurate location dots), the work often involved the use of trained personnel able to differentiate between the different classifications, often by apply multiple data layers supplied by a variety of public and private sector sources.

iMerit

READ MORE ON GIS DATA ANNOTATION

GIS and geospatial data, coupled with state-of-the-art farm equipment, precision annotation tool sets and domain-specific, data enrichment specialists will come together to advance farming efforts, making the sector more efficient and effective.



DOWNLOAD NOW

GIS DATA ANNOTATION FOR MINING

imerit.net