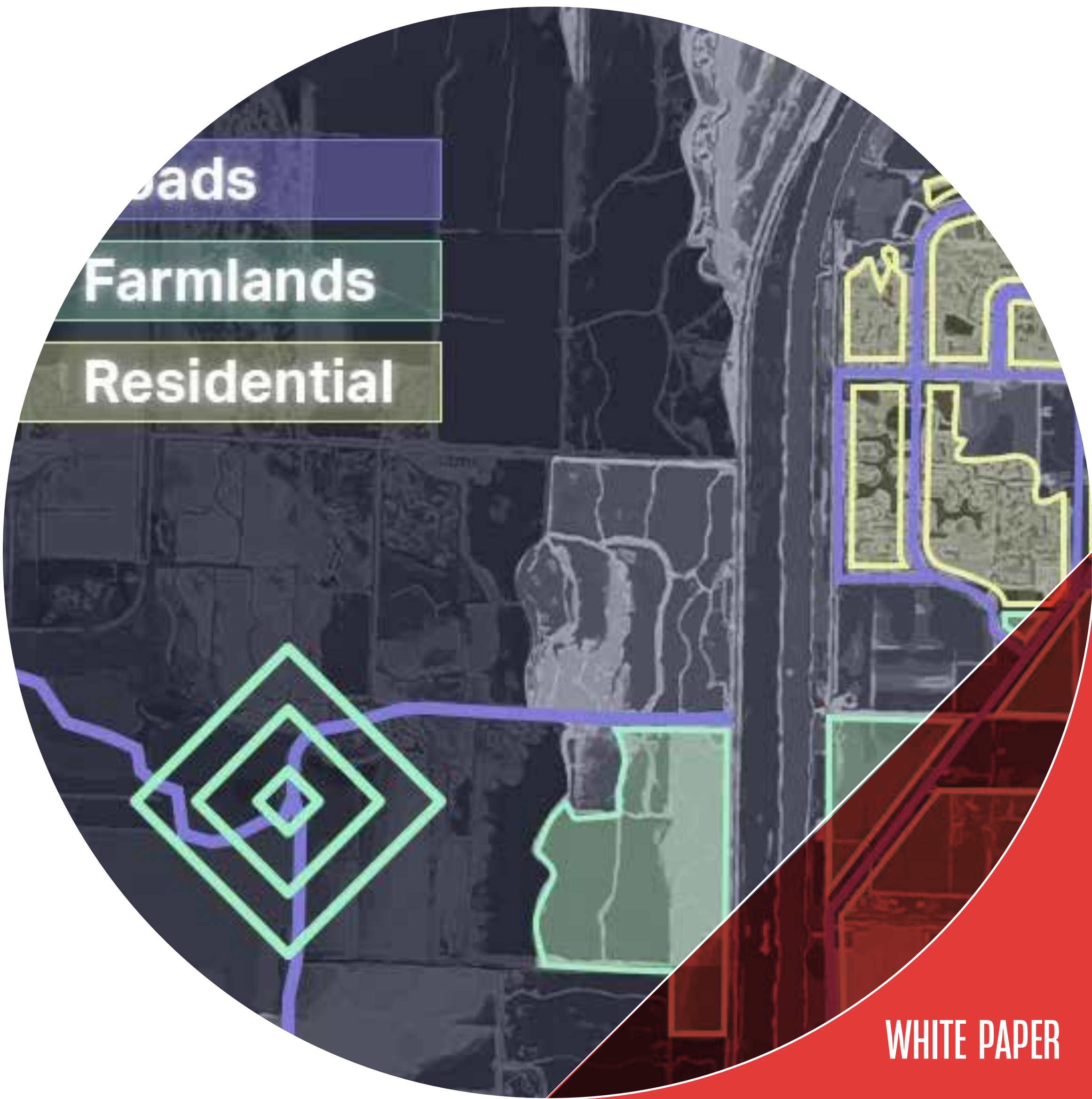
PART 2

GIS DATA ANNOTATION FOR FARMING



FARMING

AND THE USE OF DATA AND DATA ANALYTICS

Summary: GIS and geospatial data, coupled with cutting-edge farm equipment, precision annotation tools, and data enrichment specialists, will combine to advance farming efforts, making them more efficient and effective. This is Part 2 of a two-part series.

Farmers around the world have been cultivating the land for thousands of years, and throughout those millennia, there have always been the visionaries searching for new methods, ways to make the tilled earth of their fields more bountiful. From new irrigation processes to more informed crop management, farming has become a steadily more scientific enterprise, one that in the 21st Century increasingly benefits from a bird's eye view of all those tilled acres.

Geospatial data, in fact, has become the lifeblood of advanced farming. Nowhere is that more apparent than in the evolution of so-called "precision agriculture" – the use of data and data analytics to develop more efficient farming techniques that reduce costs for everything from irrigation to the use of fertilizers, pesticides and herbicides through "as-needed" models heavily reliant on GIS and other geospatial data. The aims are the two sides of the coin for any business: increase productivity and reduce costs.

Precision agriculture's foundational concept is that farmland is anything but a homogenous commodity, and that critical variations in soil composition, the nutrients contained in that soil – even moisture levels and quality. All of that requires consistent measurement through all of the toolsets available to geospatial data analysts, and the consequent use of all available means of acquiring that data (remote sensors, satellite imagery, drone and other aerial imagery, GPS reckoning, and the like.

The use of remote sensors has accelerated over the past few years, as even small farms have adopted more scientific methods, including zone-based field management that splits farm fields into different sectors based on the type(s) of crops, the type of soil composition and a multitude of other factors. Precision agriculture techniques require constant monitoring and analysis of such factors as pH rates, the presence of pest infestation, the types of nutrients found in the soil (and consequent need for any intervention with specific types of fertilizers), the crops – and their characteristics, including health and density – and even overlaying weather forecasts.

The GIS and other geospatial data used in precision agriculture can run the gamut from historical land use and characteristics, often supplied by government agencies in the form of survey maps and records, as well as more current data such as infrared and photographic imagery culled from both public and private sector sources.

Much of the satellite imagery used by the U.S. agriculture sector comes from Landsat 8, a joint initiative of the United States Geological Survey and NASA. The satellite orbits the earth every 99 minutes and has onboard sensors that encompass nine spectral bands, from visible, to red to near-infrared, and two spectral bands for infrared data. The sensors can be used to measure a broad spectrum of agriculture sector factors, from crop disease, nutrient deficiencies, even the presence of too much or too little moisture in the soil. The sensors are sensitive enough for even highly specialized tasks – such as evaluating the maturity of fruit hanging from trees and bushes on farms hundreds of miles below.

1B Pounds of pesticides are used in the United States (US) each year.

Data acquisition often, also, has to account for seasonal and situational variances that add substantially to the requirements for acquisition and data analysis. Over time, the data can paint a precise picture of crop productivity on a granular level, as well as growth and crop yield patterns over the years, all adjusted for seasonal fluctuations.

That process, heavily dependent on both the quantity and quality of the data acquired from various sources, also places a premium on the ability of data analysts to accurately label and organize data in order to deliver accurate results. Whether the end consumer (and customer) is a farm itself, or AI algorithm developers, both groups need both accurate annotation of the data itself, as well as the consistency crucial to pattern recognition.

An individual farm, for example, might be focused on crop potential based on the nutrients present in one field versus another. Or, conversely, the threat posed by insect or other pest infestation that could threaten one type of crop versus another. An autonomous tractor developer, or an automated crop sprayer developer, on the other hand, might require tens of thousands of images delineating everything from different types of crops, non-crop vegetation, and even the presence of objects that have little to do with farming – but everything to do with the ability of an autonomous farm machine to effectively, efficiently, and, most importantly, safely, navigate within a farm environment.

Autonomous farm vehicles benefit, to a degree, from the research and development work conducted by autonomous vehicle developers. The algorithm guiding an automated thresher, for example, would need to perform many of the same tasks as a high-end passenger sedan, avoiding collisions with animate and inanimate objects, such as farm animals and tree stumps, albeit within a more controlled and potentially simpler environment.

In other respects, though, the task becomes far more complex, as the thresher (or any other piece of specialized farm equipment), needs to perform its appointed tasks as if there was a keen sighted and experienced human farm worker manning the controls. And it needs to navigate without the benefit of set roadways, lane markers and the other physical cues that are guiding the early generations of autonomous cars and autonomous trucks.

9.7B

People To Feed Globally, by 2050.

Autonomous farm equipment relies on a combination of GPS, GIS, and visual sensor data to guide it around a farm field, and all of that data must be categorized by both source and nature. And there is often a second set of data tied to a specific task or task that requires a next level of expertise. Accurately segmenting different types of crops is a fairly straightforward (if laborious) task when the crops in question are as different as melons and wheat. But farms often plant different type of wheat, or melons. And it takes a keen, experienced set of eyes (see below) to tell the difference from one to the other.

The increasing reliance on automated farming techniques, and the need to increase productivity from the finite resource of farmland, helps explain the steadily increasing R&D investments in the sector, particularly for data collection and data analysis.

That evolution in focus and investment has been more than a decade in the making. A 2009 paper by a team of Brazilian researchers, "A Framework for Semantic Annotation of Geospatial Data for Agriculture," noted the compelling use cases for everything from farming production planning to public policy (one reason federal government agencies throughout the world are so heavily invested in geospatial data collection for agriculture). "Spatio-temporal factors vary widely (from region to region) and are crucial in decision making."

The researchers' primary contribution to the state of the art (at the time) was a proposed extension to the United States Federal Geographic Data Committee's specifications for geospatial metadata. Noting that the FGDC standard was designed for general applications, the Brazilian researchers proposed a series of metadata labels specifically for agriculture.

The research paper was one of the first to develop a framework for geospatial data annotation specifically for agriculture, including a set of annotation tools that could be used by data scientists. Ironically, the researchers focused on automating the process, while in the intervening years the industry has recognized the essential processes of expert human-in-the-loop data annotation combined with very precise software tools.

While the private sector – farmers, farm equipment manufacturers, and the broader ecosystem that serves both – has driven much of the R&D and commercialization of everything from precision farming techniques to autonomous machinery development, little of that work would have progressed to the current state of the art without equally aggressive initiatives on the part of government agencies and other entities around the world.

The U.S. agriculture industry, for example, was an early adopter of GPS technologies, to enable smarter farming and as increasingly more precise civilian GPS tools became available many farms began to embrace geospatial data gathering both as consumers and producers of that data. Imagery from aerial drones, for example, can be used to develop more precise watering patterns, or even to check for plant disease before an outbreak can cause widespread devastation.

27.4 M

Number of IoT connections used in global agricultural production by 2021.

That geospatial data-driven focus has only intensified in recent years through increasing Research and Development expenditures and, in at least some cases, a shift away from low-cost, often crowd-sourced data annotation, in favor of teams of trained, domain-specific, data analysts who bring specific agriculture subject matter expertise to the engagement.

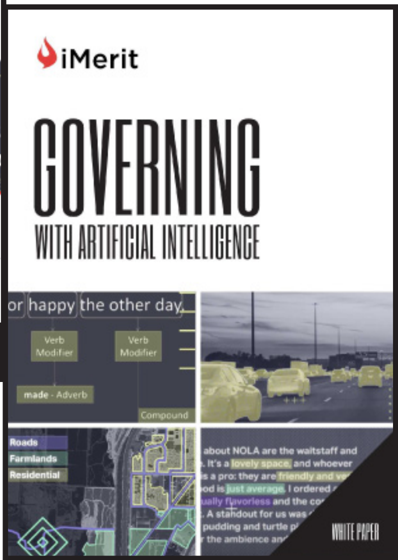
The Internet of Things is arguably best known for connected in-home appliances and personal devices, but IoT is every bit a growing factor in agriculture – all driven by a geospatial data collection and analysis model. IoT-based farming processes include everything from crop monitoring sensors to the farming equivalent of autonomous vehicles: tractors and other heavy farm equipment guided by geospatial algorithms.

iMerit

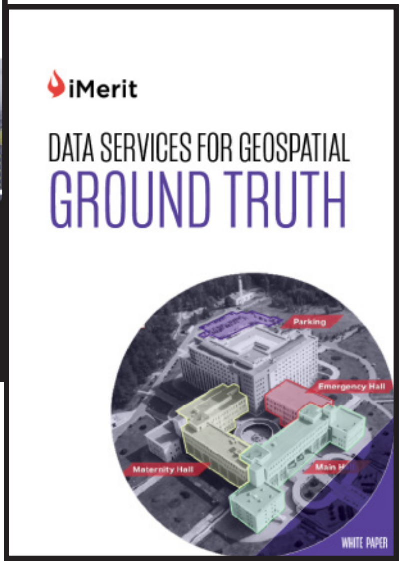
WHITE PAPERS ON APPLICATIONS OF AI

Long-form content, produced by iMerit subject matter experts

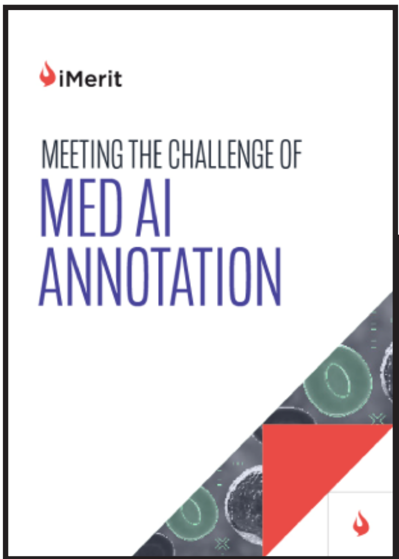
DOWNLOAD



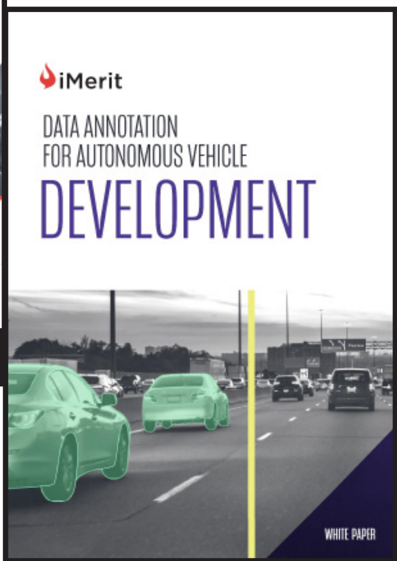
DOWNLOAD



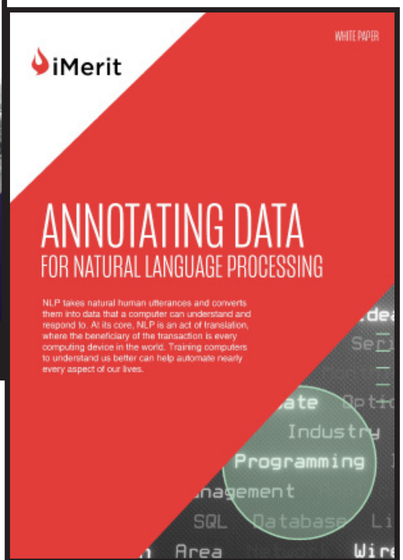
DOWNLOAD



DOWNLOAD



DOWNLOAD



DOWNLOAD

CONTACT US

In some cases, those autonomous tractors are collecting data as well as harvesting, and work in conjunction with in-ground sensors to identify high-yield areas, determine use of pest and fertilizers, and even where to direct field irrigation. A combine harvester equipped with a GPS tracking device, for example, can capture both quantitative and qualitative data – crop yields, and such real-time crop quality measurements as plant water content and chlorophyll levels – as well as the exact location where the measurements were taken.

The average American farm this year will generate hundreds of millions of data points – approaching 500,000 data points daily, according to the market research firm BI Intelligence. Within 10 years that number is expected to quadruple, approaching a whopping two million new data points each day. In the more traditional sense, those farm vehicles are almost as voracious consumers of data; they're guided by algorithms trained through satellite and drone imagery and sometimes even weather feeds.

Self-guided farm tractors, plows, and other vehicles represent a subset of the autonomous vehicle sector, one with the same general objectives but with – perhaps – somewhat fewer safety concerns. Navigating a farm field involves fewer hazards, and generally fewer surprises, than downtown city streets, but a self-driving plow still needs to understand the difference between a section of field that needs tilling and a bordering road. And even a farm field has people, animals, and a tree or two that must be avoided from time to time.

TEACHING A TRACTOR TO DRIVE

As with AV projects, self-guided farm vehicle development initiatives rely on the same general set of tools to create bounding boxes around discrete objects and defined ground cover, and semantic segmentation of entire fields to educate the algorithm that controls the vehicle. Ground cover designations could be as basic as noting the difference between tilled and untilled land so that a mechanical plow understands where to till the field.

But the complexity of the exercise expands from there as the self-guided plow developer begins to address all of the scenarios and different types of tasks that will emerge on a typical farm. Depending on the vehicle, it might need to recognize and understand the significance of different types of crops, crops at different growth stages, even the difference between crops that are thriving and those sections in need of human attention. And in all cases, the vehicle must understand objects and boundaries that define where it can and cannot go: farm buildings, people working in the fields, farm animals, roads, and even subtle variations in terrain, such as ditches and hills.

A data annotation and analysis company such as iMerit will generally assemble a team of dedicated data analysts to take one or more of a series of annotation-related tasks: annotating raw images, conducting QC on pre-annotated images, or even some secondary QC. Combining annotation and verification as a client service can cut costs significantly, up to 50 percent or more from what developers face from other options.

Self-guided farm vehicles designed to apply herbicides – or even targeted herbicides and growth enhancing sprays – need an even more sophisticated understanding of their surroundings than other automated equipment. So called "see and spray" technology requires the system to understand different crop types, for example, such as soybeans, corn, or cotton, that might be planted in adjacent fields. Simply drawing bounding boxes around a section of farm field and labeling it "crops" doesn't begin to get the job done.

Annotating crops, in fact, can be surprisingly challenging. Many variations of wheat, for example, can, to a relatively untrained eye, look like weeds – or vice versa. Consequently, training an algorithm to approximate the knowledge and trained eyes of an experienced farmer calls for tens of thousands of images, sometimes quite a bit more. There's a huge advantage to diversity – different ground cover, plant height at different stages of growth, even crop density – and more is always better so that data scientists can select the imagery that can best accelerate algorithm training.

25.5%

Projected Compound Annual Growth Rate (CAGR) for AI in agriculture market from 2020 to 2026.

It has become obvious to many in the sector that, a multi-year development project to provide training data to an algorithm for field machinery, could require semantic segmentation on hundreds of thousands of images before a farming operation would be able to move from development to live production technology. Clearly, no small task.

Ironically, complex imagery is not always an advantage. Just as children learn to walk before they can run, algorithms begin their education by incorporating fairly simple, straightforward statements in the form of annotated images. Data analysts will use (images of) stalks of wheat, outcroppings of weeds, a field of wheat, rather than feeding it the whole farm.

That training process resembles a feedback loop of sorts. The model will "look" at an image and then makes a judgment call on what it believes it sees. Data analysts will score the results and the team of data scientists developing the algorithm will adjust the imagery diet to correct errors. Over time, as the process adjusts to model's progress (or lack thereof) the algorithm's assessments – and speed – will begin to mirror those of the annotators.

As with any effort to educate a machine there is inevitably a point of diminishing returns, where the learning curve begins to level off. No individual person working a field will be correct 100 percent of the time, nor will a machine – but there is a generally accepted (albeit subjective) criteria for bringing the training process to a conclusion: would the average human farm worker outperform the machine?

Most autonomous far vehicles rely on a series of onboard cameras to guide them, so the training data required for development is generally similar: low-level images that provide the same views the vehicle will encounter at work. Those sensor arrays will almost certainly evolve over time, with LiDAR a logical extension to provide more detailed and arguably more reliable 3D views. If needed, drones and other aircraft can supply aerial imagery to complete the picture.

CRUCIAL GOVERNMENT EYES IN THE SKY

The voracious appetite of the global agriculture industry for GIS and other geospatial data long ago spurred both the public and private sectors to satisfy that hunger. Multiple United States federal agencies have spent decades compiling data on everything from ground cover to weather patterns, and property boundaries. In addition to the Federal Geographic Data Committee, and its work on geospatial metadata, a number of US federal agencies actively collect and make available a rich set of geospatial data.

The US Farm Production and Conservation Business Center Geospatial Enterprise Operations Branch (formerly the Aerial Photography Field Office), for example, holds an archive of more than 10 million aerial images dating from 1955 to the present. The images have been used in support of environmental assessments, change detection, and even property disputes.

Geospatial data also plays a purely public sector role in the administration of federal agriculture support programs. The latest such initiative: the National Agriculture Imagery Program (NAIP) supplies digital imagery to the United States department of Agriculture (USDA) Service Center Agencies to help administer federal farm programs throughout the country.

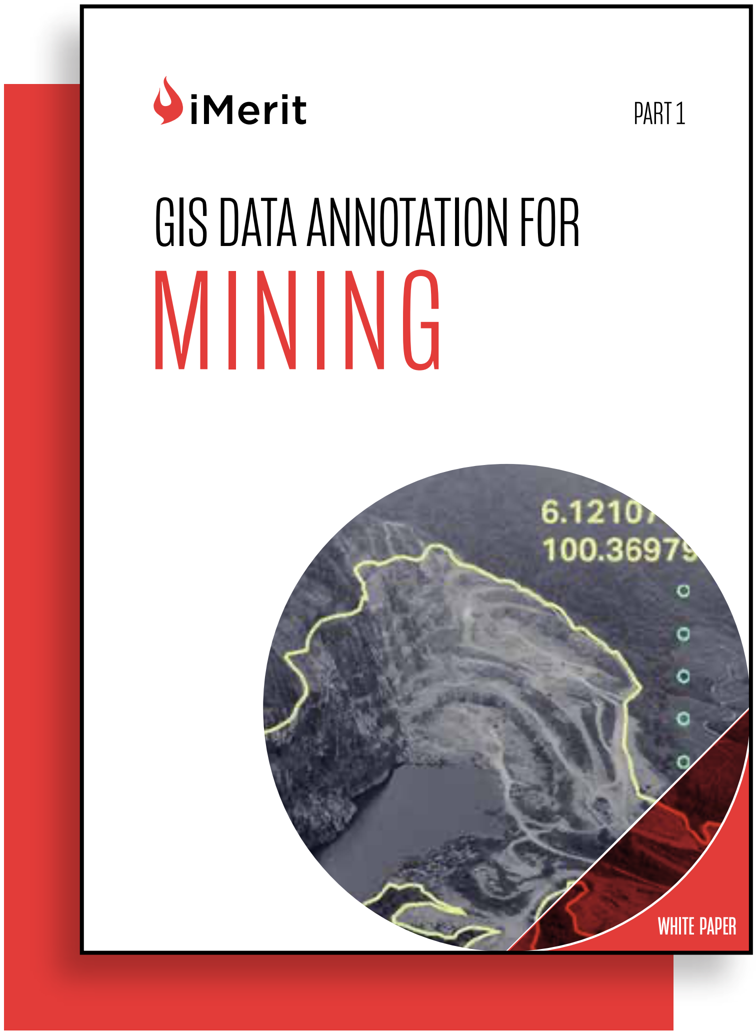
NAIP provides what it has termed "leaf on" aerial imagery to USDA County Service Centers for eventual use on behalf of local commercial farms. Among its uses is maintaining so-called Common Land Unit (CLU) boundaries between farms, as well as for the application of farm assistance programs. In the latter case, the USDA determines farmer benefits provided through its Farm Service Agency through GIS data, and associated data analysis, on crop acreage, other land use factors, and even conservation compliance.

NAIP is in the midst of an ambitious program to collect 1-meter (or even more detailed) imagery for all farmland with the contiguous 48 states. The imagery will be provided in both so-called 4-band (red, green, blue and near-infrared) natural color, and infrared versions, so that it could solve for a variety of use cases. All of the data is publicly available.

iMerit

READ MORE ON GIS DATA ANNOTATION

The resource extraction sector has become the biggest consumer of GIS and geospatial data, and data annotation specialists have stepped up to fill a pivotal role enriching GIS images and unearthing ground truth data for mining related analysis.



DOWNLOAD NOW

GIS DATA ANNOTATION FOR FARMING

imerit.net