EPISODE 1302

[INTRODUCTION]

[00:00:00] JM: Governments, consumers, and companies across the world are becoming more aware and attentive to the risks and causes of climate change. From recycling to using solar power, people are looking for ways to reduce their carbon footprint. Markets like the financial sector, governments and consulting are looking for ways to understand climate data to make smart decisions and manage risk. The company Sust Global was founded as a way to deliver sustainable change and climate-resilient action. Sust Global uses an AI-powered platform that combines climate science, satellite-derived data, and geospatial datasets to quantify climate change. Companies can use this analysis to evaluate risk to assets, better understand future commodities like metal, and plan for future supply chain challenges and climate perils. In this episode, we talked to Josh Gilbert, CEO at Sust Global. Josh explains Sust Global's mission and product, and discusses how companies use Sust Global to prepare and respond to climate change.

[INTERVIEW]

[00:01:01] JM: Josh, welcome to the show.

[00:01:04] JG: Thanks so much. It's really great to be here.

[00:01:06] JM: The domain that you work in is sustainability-based, climate-based financial predictions is what I would say. What your company does is, essentially, you look at climatology and derive financial indicators from the projected direction of the climate. Does that summarize your business?

[00:01:31] JG: I think it's a good start. I think there are kind of core groups within that. So you have effectively a few different tribes, right? You have the climatology and, us, system sciences tribe, who really into monitoring things that actually go on on planet Earth.You then have a community of global climate modelers effectively projecting to the future saying, “Hey, at a global scale, what are things going to look like in 2050? 2100? How do these models kind of fit

together?" And then you have the data science community, and maybe the kind of finance community who are saying, "Hey, how do we turn this all into a usable signal in actual day-to-day work?"

Effectively, Sust Global sits at those different tribes of people, and translates all of this complex climate data, all of these forward-looking global climate models into a usable signal for everyday businesses, whether that's large corporations, whether that's the financial sector. Kind of compressing, I guess, all of these different sources of data into a usable signal for real world operations.

[00:02:34] JM: And the company has been around since 2019, right?

[00:02:38] JG: Yeah, the company was founded in 2019, although the company is founded in 2018. I went full time in September 2019. So we are, as of today on the recording day, I reckon we're about just under two years into the full time journey as Sust.

[00:02:54] JM: So given that 2018 era infrastructure, what does your stack look like? I'm sure it's bigger than you can explain in a single answer, but sketch your architecture for me verbally.

[00:03:09] JG: Sure. Yeah. So the architecture begins with these global climate models. And it's effectively an ensemble of over 70 different forward-looking global climate models. So these are climate models on average precipitation, extreme precipitation. How much warming will occur? And then down to specific perils, which is a second set of input data. So that's looking very specifically at models around flooding models around wildfires, models around sea level rise, which kind of bring a deeper level of granularity to those initial macro level climate data sets.

Then we have processing layers, where we fundamentally use Python and Jupyter Notebooks and turn all of this into JSON files. Do a lot of machine learning-based techniques around effectively correlating these global climate models, which predict the future with real observed instances of climate events via satellite data and observe things happening on the ground. And we basically get to speak to each other, within bias, correct and validate these global climate models. We put this all through our processing stack, and then the ultimate output is again collapsing all of these broad ranges of complex datasets that don't often speak to each other

into a usable signal, which is effectively a climate risk score at the lat long location of every asset within a portfolio. So we have the building blocks of all of these risks. And then you can aggregate them up to the portfolio level and then produce real insights for the end customer.

[00:04:46] JM: To introduce a comp to what you do, I think it was called The Climate Corporation, the company that was started by David Friedberg that was acquired by Monsanto. Are you familiar with that company?

[00:04:59] JG: Yeah. Yeah.

[00:05:00] JM: So they were started, I believe, what? Around 2007, 2008? Somewhere around then? Do you know anything about what their technology approach was? What their modeling approach was?

[00:05:13] JG: I don't have many insights into their modeling approach. I know that the markets they tend to focus on are more agricultural. So, obviously, they're required by Monsanto, which is a pretty big giveaway on that. But many of the models that they were looking at were trying to integrate things like seasonal variation over a shorter time horizon. Kind of less of the long term global climate models were integrated into their approaches, because from my understanding, many of the end users tended to need this data in a much shorter timeframe. Because obviously, agricultural yields, when you're looking at them, they tend to be much shorter term focus. And obviously, the Monsanto customers tend to focus on agricultural commodities. So I don't know too much about their process and the kind of tech behind the scenes. But I certainly think that their focus is a good way of maybe suggesting some of the approaches they would have in their tech stack.

[00:06:11] JM: How do institutional investors integrate climate science into a portfolio strategy?

[00:06:21] JG: I don't know if I'm allowed to swear on the podcast, but it's freaking hard. I'll use frickin' as the operative term there. And these users very well used to separating signal from noise. They are not people that require really, really basic tools, but they do require the tools to be in their own domain. So if you imagine, many of the global climate models, complex technical forecasting, and modeling tools, many of the observed events in the satellite data are imagery-

based. And often, to date, dashboards have been used in terms of actually kind of being able to scroll and click and point on the different assets. Many of the geospatial data applications tend to have been in government and defense to date. And analysts tend to work in that type of way. In the financial sector, it's a very different user persona, where they don't want a dashboard. They don't want imagery. They want signals that are direct inputs into their day-to-day workflows. So that's why the actual deliverables to them in many ways a glorified CSV file, where you have the lat longs of all of these different assets, and then the corresponding risk scores on a monthly cadence looking out over the next 5, 10, 15, 25 years. And really, it's a layer of validation and showing the modeling, and showing the working, and showing the transparency of the models and the accuracy of the models in the kind of onboarding process if you're working with the kind of core data teams or the kind of quant teams that are integrating this data. But the actual investors and the financial actors that use this data really don't care about the imagery. They don't care about the complexity. They want to know, “Is this asset highly exposed to the climate over the next 25 years?” in the case of a mortgage. Or is this portfolio of assets or mortgage pool actually higher at risk relatively compared to another mortgage pool based upon this data? So it's really the outputs.

And one of the interesting things that we've found through the whole journey at Sust and is kind of core in our approach is saying, “Who cares? Whose problems are you solving with this data?” And I think the tendency for us tech lovers and for kind of technical people building products is we fall in love with the technology solutions, rather than the actual outputs that are serving the end customers. So we try and obfuscate the technology as much as possible, and actually really get it down to a base signal, which is then directly integrated via an API into their existing workflows, rather than having this dash that maybe is a high-friction process. You don't want somebody to download their own data that's relevant to their assets. Let's say it's an RBS pool, or a Fannie Mae, small loans pool, or whatever it is, right? They export it from their day-to-day workflow. They then upload it onto an external platform, where then they analyze the data, they pass the data, then they export the data from that platform back into their day-to- day workflow. One of the big user insights we have is nobody wants to do that. They want to have the data provided directly and simply into their existing operational workflows.

[00:09:30] JM: How do they want data presented to them? Do they want a CSV? Do they want a TSV? Do they want an API? Do they want to stream? What's the ideal data flow from your system to the customer?

[00:09:49] JG: Fundamentally, API-based integration is always going to be optimal for the financial sector, because it provides the lowest friction way of getting this data into the real world operational use cases that they have. I think CSVs often work fine. And as we've been developing the early stages of the product, providing these CSVs is a great starting point. But I think API's and direct integrations are really the end goal. And we're lucky that we have a very productized offering where it's consistent across different customers, which enables this kind of relatively homogenous API-based integration.

But, yeah, I think, again, coming back to my earlier point, dashboards are great for telling a visual story and for maybe doing a product demo for a customer where you're saying, “Hey, this is actually how the data looks. And this is how you can conceptualize the data. But in the day-to-day usage, and if you really want to create a sticky product – Counterintuitively, geospatial analytics wants to get away from satellite data and imagery as much as possible and get to the raw, the raw underlying signal, which is via an API or via a CSV plugged into the operational workflows of the customers in the easiest way possible.

[00:11:06] JM: You and I were connected through Sina, who started Unfolded AI, which is one of the coolest geospatial products I've seen in a while. Full disclosure, I invested in that. But I'd like to know, can you frame the state of geospatial technology for me? Because this is a space that I don't think a lot of people are attacking relative to how much opportunity there is. Like I look at the company – What's the company that makes ArcGIS?

[00:11:35] JG: Esri.

[00:11:36] JM: Esri, right. Esri. Esri is this powerhouse that nobody knows about that makes really good geospatial software, and they've made it for like 25 years. And there's a whole ecosystem of the post-Esri world as far as I can tell. So you've got data visualization solutions. You've got data providers. You've got data platforms, data sharing systems. And this isn't even talking about the revolution in underlying data infrastructure, or the revolution in visualization

tools provided by companies like Unfolded. Tell me about the state of geospatial analytics and why that's relevant to you.

[00:12:21] JG: Yeah. So Gopal and I – Gopal is my cofounder and CTO at Sust Global. We wrote an article on this back in 2018, I believe, and it was called Approaching Geospatial 2.0**:** Unlocking Billions at Scale Across the Globe, or something like that. It's on Medium. And it did some good numbers and seem to strike a chord. So that's kind of a starting point where we fundamentally posited that there are all of these exciting new data sources coming online, many of them from satellites, many of them from ground-based sensors. So you have this abundance of data. And essentially, the upstream data is going to become commoditized. There're more and more satellite constellations being launched. There're more and more sensors on the ground. And there's just so much information. It's really hard to turn that into real insights and separate the signal from the noise.

I think that one of the big problems in geospatial analytics, and you're right in the way that it is truly an incredibly large addressable market. And it's only growing. We're focusing on climate change, which is one of the defining generational issues and opportunities that we face. But there are many other areas. You mentioned the climate service. I'm sorry, you mentioned The Climate Corp in agriculture, and Indigo that are in agriculture. And there're many different companies across many different verticals.

However, the problem that we identified in the Approach in Geospatial 2.0 article, as well as the problem that kind of still persists across the market, is that many of the geospatial companies really haven't been able to find meaningful product market fit, because there're kind of two extremes in the geospatial market. If you go too far towards specifically solving a customer's problems, you end up becoming a geospatial consultancy, where these problems are very unique. They're very specific to the customer. And you end up working on projects. And you can instantly tell with geospatial analytics companies how they're doing just based upon the way they describe the work that they do. A lot of companies that we speak with, and a lot of companies I know, and a lot of competitors we have talk about projects where they're saying, “Hey, we have this one project. We have this next project.” There's no ARR. There's no deep venturable scalability in the solutions.

Then if you go too far the other way, you can end up building a SaaS platform that serves 25 different verticals and you build it and you hope they will come and they don't necessarily come. And there's this inherent imbalance in the geospatial analytics market, and it's very difficult to get this balance right. I think seeing it in Unfolded, effectively – This is I'm sure a massive bastardization. And no disrespect to Sina if I then go on to butcher everything that Unfolded do very well. But they make incredibly beautiful and incredibly complex data visualizations where you can really bring this data to life. Kind of coming from the Uber world where there was this abundance of data. But how do you turn that into something that's easily understandable? Unfolded focused on that problem very well, but it's fundamentally sector agnostic in the way that it's wrapping up data in new and exciting ways.

I think the way that we have gone at Sust Global, and one of the areas that is essentially a geospatial analytics graveyard to date, but also a hugely untapped market, is in being able to provide insights across a broad enough set of markets so you can have vegetable scale, but providing a specific enough set of outputs and insights that can become productized and a relatively homogenous in terms of the underlying processes on the product side. So I think that article that we wrote is a great starting point. And I also want a shout out to a few others in the geospatial analytics space who are writing some great content. I think Joe Morrison on Twitter, he's @mouthofmorrison, has consistently brought out some great articles kind of describing some of these problems, describing the problem how many geospatial analytics companies run into. These issues of scale, and these issues of real product market fit. He writes very actively.

And then Aruvant, who is based in Europe. I think I'm going to pronounce his name wrong, Ravichandra, does some great work. He was formerly at the PwC space practice looking at satellite data, looking at the business models there. He's now doing a lot of consulting working with startups observing the geospatial analytics market. And he's done a really great job of re segmenting some of those early breakdowns of the market that we did three or four years ago, and breaking them down into EO versus upstream and downstream and the analytics and specific sectoral insights. So by all means, I'm not the fount of all knowledge on this area. And I really recommend if listeners are interested in it. Check out some of these blogs. I'm sure there will be my contact details in this pod.

But also, maybe I can send over some of those Twitter handles, because there's just so many interesting people commenting on this exact question on why haven't we seen more of these breakout successes in this hugely attractive, hugely technology-driven, but customer-focused area of innovation, but we haven't necessarily seen the expected returns to date.

[00:17:26] JM: What are the places in your stack where you feel the pain of insufficient technology?

[00:17:34] JG: Honestly, I think that much of the technology in geospatial analytics is sufficient. In the way that we have these big data solutions now and we can host so much data and process so much data. And the cost curve on this type of processing has come down massively over the last five years. And actually, that's the main driver of many of the innovations that we're building on top of that stack. So to be contrarian, I think that much of the technology has outstripped the customer-focused outcomes that should be derived from that technology stack. I think we're at the stage where we have so much data to process. Really deriving that signal from the noise and, again, finding that sweet spot between not going too far becoming a consultancy on the one side, and not becoming a platform that just crunches loads of data because it's neat on the other side.

I think the technology is there. It's really a case of translating that technology into real world business level use cases that new customer sets can actually use outside of government and defense and outside of the traditional community that's used to handling GIS data. So outside of the SREs, outside of many of the large incumbents today, where their bread and butter is kind of saying, “Look, you can chop up this data in many different ways. You can put it on a map. You can then analyze all of this data. You can then, again, have someone like Unfolded that does this incredibly beautifully and create new ways of looking at that data and visualizing that data.” But ultimately, the next frontier is turning all of this noise using all of the technology-driven tools that we have available into real customer outputs that solve real world problems.

[00:19:21] JM: Can we go a little bit deeper into your data pipeline? So my sense of your product at this point is you ingest satellite data, and you output guidance for financially relevant climate-related condensed data reports.

[00:19:44] JM: Maybe you could tell me about how you get from A to B.

[00:19:47] JG: Sure, yeah. And before we start, I just want to re-categorize being specifically, where one of the big pitfalls I think that companies looking at climate risk face is that reports are increasingly important from regulators. Like the Biden administration in the US and are now saying, “Hey, look, climate disclosures are going to be something that everyone needs to do.” In Europe, this has already been happening. But these reports really are non-mandatory today. There's no standardization and there's no real methodological rigor that exists today, which may be then lead to do things, right? It means that the granularity of the data becomes less important. And I think, importantly, in business model terms, you only need these reports once a year to get these regulators off of your back. This is kind of a cynical view of it. But that's fundamentally the way many companies look at this today.

So the real use case is that we aim to provide B in this process are actually frictionless integrations of climate data where we provide the building blocks, the kind of climate data infrastructure, to allow the customers to build their own products, applications and services through using this data, effectively going from a risk mitigation tool into a new product creation tool. So kind of going towards the opportunities around climate data where you say, “Hey, if I could see all of these risks, if I could understand all of these different variables, and this was really validated and trustable data that's piped really simple into my workflows, I can create new financial products in this.” So maybe I can break down the A to the B, and then maybe go a bit more into what that B would actually look like.

So in terms of the processing workflow, we take the open source models from the Pangea Project to the IPCC data, and their collections of hazard data, whether it's like hazard one, like fire; hazard two, flooding. You can keep going on and on and on. We basically have an output simulation and a model ensembling process where we take all of these different collections of environmental data, and then we ensemble the models. We spatio-temporally harmonize all of these different models. Get them all to speak the same language. And then we can bring in higher resolution datasets. So specific, satellite derived datasets, like the NASA MODIS data sets, or the Sentinel-2 European Space Agency datasets, which are basically built around observed instances of stuff going on on the ground. So we have these forward-looking global climate models, plus these historic observed instances.

We basically use these observed instances to super resolve and create 8 to 16x resolution improvements on the existing global climate models. So in layman's terms, that's effectively bringing in the satellite-derived data on wildfire, say. And we're saying, “Hey, here's a global climate model. It runs about a 100 kilometer grid cell size, or one degree spatial resolution. Where if we're looking at your house or my house on a map, we can say, “Hey, there's no real indicators of anything useful in these flows, or these assessments of risk.”

So by bringing in these higher resolution data sets, we can super resolve. Get down to the one kilometer resolution where you say, “Hey, your house is near a forest. It can burn.” Or, “Your house is in a floodplain. This is going to be a really dangerous thing,” or the sea level rising around your house on the coast is going to be really pronounced in this specific area, but less pronounced in this this next area.

Once we've ensemble all of these models, we spatio-temporally harmonized are these models, we can then bring in historic catastrophic event catalogs from FEMA, from Dartmouth Flood Observatory, all these other data sources, and then active near real time events. So that's looking at 24-hour to seven-day forecasts. So you're bringing in all of these different data sources. You then have a second set of gates in terms of spatio-temporal harmonization, and localized spatio-temporal learning engine, which kind of brings in all of these different spatio data sets to effectively then provide asset level location data for a set of custom assets. So say you had a mortgage pool across the continental United States where you’re like, “Hey, here's 5,000 different lat long locations of assets.” We can then say to you, “Hey, based upon all of these different data sets, each one of these assets has a corresponding risk score on a monthly cadence from now to the year 2100.” Obviously, many of these short term data sets are going to be bringing in the active near real time kind of weather alerting data. And then some of the more longer term data sets are going to be relying more heavily on some of these global climate models, which are broad brushstrokes, but kind of hold true in terms of what would happen in the future in the four degree warming scenario in terms of climate change by the year 2030.

We can provide all of this data at the asset level. So then it gets into the hands – And it can be via CSV. It can be via API. It gets into the hands of an analyst or a quant who sits in the front institution and they say, “Hey, we want to know, what is the asset level risk of all of the assets

within this portfolio? How does this portfolio compared to another one? We want to make sure that if we are making an investment decision, or even in wholesale banking, making a loan decision, we have credible data, validated and trusted data that we can make these inferences from," because the global climate models that are open source today and from world-leading institutions are kind of a good yardstick, but they really don't have any granularity or validation in terms of bringing in bias correction to the models, which this kind of tech stack provides to them.

One of the really interesting outputs in terms of – I talked earlier about how do you get to value creation opportunities rather than risk reduction for the end customer. So say you were using this tooling to assess a portfolio of loans that you want to provide for customers. You can say, "Hey, look, we can now assess the climate-related risks in these loans. If over the next 15 years, there's a lot more risk or a lot less risk, we can then bring that into our model in terms of how we decide what and where to issue loans.

The really cool thing is then if you have tried and trusted data, you can then group these loans into new financial vehicles, and then effectively have climate resilient products, where you say, "Hey, we're going to turn this into a financial product and then sell it on capital markets where it's securitized," right? Where you say, "This is tried and trusted data that goes into the building of these models." You as a proxy the SPVs in the global financial crisis where everyone thought these SPVs were good, and they were full of really, really crappy, risky loads. And then the whole system collapsed in on itself. It's a very real fear that we might have like a Green Swan event, and huge amounts of financial value are wiped out because there isn't tried and trusted and validated data on these types of assets. If you have that, you can then create new securitized products with the belief and with the knowledge that you say, "This is backed by the latest science, which is turned into a financial signal, which is then turned into a financial product, which ultimately provides value in terms of the cap market side and the real users at the end of this kind of A to B tech stack."

[00:27:16] JM: What you're doing, I think it's a novel company category. I've done a number of shows with SafeGraph, the company by Oren Hoffman. You've probably seen that company. But I think the world of sort of commercializing and SaaSifying curated data streams is a little bit nascent. Do you agree with that?

[00:27:43] JG: Yeah. And I know Oren has done a lot of work recently on – Or he set up a new podcast on data as a service.

[00:27:52] JM: Mm-hmm. Yes. He did a show with Jack Dangermond, the Esri CEO.

[00:27:56] JM: Yeah. Yeah, yeah, right. Yeah. Yeah. Yeah. And I think data as a service is a really important – I don't want to use niche in the market, because I don't want to sound pejorative. It's like it's an important part of the market where you're solving problems, right? And then you may be have from data as a service on the one side where you're providing data for then people to do new things. You then almost have like the last mile, which is maybe a consultancy or a user of something like Esri, or something like SafeGraph, where they are then using very specific customer-driven outcomes. There's kind of this messy middle, or this gray area between the two, which is Sust we call the penultimate mile, right? Where we're saying, "Hey, there are all of these different data streams. You can't just provide this data as a service, because fundamentally, it requires finessing to serve a specific subset of customers." In our case, it's customers who are interested in the ways that climate change will impact on physical assets in the future. It's a subset, but it's still a huge and growing market, right? It's a huge problem that we need to solve not just in terms of a business opportunity like as an existential threat to many of the companies and people around the globe full stop. But there's this additional layer on top of data as a service where you need to harmonize spatio-temporarily many different types of data and many different geospatial data sets where some of it is satellite imagery, some of its ground sensor data, some of its historic data, some of it straight from an Excel spreadsheet, some of it is siloed data from the kind of customers that we have. They didn't even know they had this data, right? It could be asset locations that they have. It could be any number of different data sets that don't speak to each other. You need to have this smart way and this this smart set of models to actually transform all of this data into a relatively homogenous set of outputs that then the last mile, i.e. the consultancies, or the real kind of providers to this data can then begin to use.

I guess maybe an interesting analogy is Planet Labs, right? They kind of provide all of this imagery and they then focused on pre-processed imagery, where it's kind of like data as a service plus, where they're saying, "Hey, look, we've done a lot of the complicated stuff with this

data. So you have a really clean signal with which to build your kind of outcomes on as a customer." And ultimately, they're selling data, right? That's the thing they care about.

This is a similar play in terms of the geospatial analytics space, and it probably sits further downstream. But it is kind of taking all of these different data sources and turning them into some type of pre-processed layer of usable data, which can then be used by the last mile across a very specific set of services. So like some of our customers use our data to look at prepayment risk and probability of default in residential mortgage-backed securities markets. Or they use our data in phase one assurance in North America where they're saying, "Hey, how can we build in four different climate risk data for environmental consultants?" These are very broadly different markets and broadly different user personas. But they still rely upon a basic level of pre-processed and easy to use data. So it's turning all of the complexity into a usable signal, which often is a very, very different signal to someone like unfolded or Esri. So if you imagine Planet at the top, where it's just the pre-process data from one data source. And then Esri is able to kind of bring in all of these different data sources. And then the SafeGraphs, the kind of data service place, we kind of sit between them and the end user as this kind of pre-process data as a service stream, I guess. I don't know. I just made that term up on the spot. But I think that is a really interestingly exciting middle space between the end users and then the data as a service providers.

[00:31:41] JM: Your mention of Planet Labs jogged my memory. I did a show with them a few years ago at GoogleIO. And the reason they were at GoogleIO is because they were, if I recall, all-in on Google Cloud. And the way they structure their business, if you are in the business of – Okay, just to reframe what they actually do. So my understanding of what they do is take satellite data, raw satellite data, which there's plenty of raw satellite data providers, and process it, and make it much easier to use, refine it, and then make it available as an API or some kind of data sharing system. Is that accurate impression of what they do? They're kind of like munged satellite data, right?

[00:32:29] JG: It's an accurate description of part of what they do. And the other part is both what gives them their competitive advantage and is also a real – Is albatross around their neck the right phrase? Something around that, a weight around their neck. I don't know if it's an albatross. I now I have a bizarre image in my mind of somebody with an albatross around their

neck. So Planet have their own satellites. They have the DAV sats, which are this – They have the largest commercial fleet of satellites that are monitoring the globe, and they have all of this amazing data. And then what they found was, "Hey, look, we can monitor the whole world using this data. But there's a jump between having raw data and then having clean outputs for the customer. Because like with satellite imagery, what if there's too much cloud cover? Or what if there's variance in imagery based upon atmospheric conditions?

So the pre-processing is effectively then tidying up their own data to then a level which then geospatial analysts can use in a much cleaner type of way. The reason why I described it as an albatross around their neck is they also have a huge hardware cost where they are launching real physical things into the sky into the lower Earth, or the geostationary Earth orbit. And that costs a lot of money. And then they're kind of trying to create this end-to-end solution across multiple parts of the value chain. But it really means that you have to be good at multiple different parts of it.

I think companies like Sust Global and other geospatial analytics companies, maybe someone like Orbital Insight, they choose to not play in the launching satellites and operating satellites part of the business. They choose to exist purely in the analytics part. But one of the problems is if you use too much of the planet data, or the Airbus **[inaudible 00:34:21]** data, or the Digital Globe data, these are all kind of satellite constellation owners, right? These are all commercial datasets. And if you want to have global scale, it's going to cost you a hell of a lot of money to actually bring in all of that data at a high-cadence. And the unit economics then become really, really wild as you're then getting to the end customer. So again, it comes to the idea that like the tech stack and the ability – So Planet use GCS, or if someone else is using AWS or Clarity or whatever it is, the processing costs now are manageable. They're still expensive. But the processing power in terms of cloud native compute is definitely there now.

The biggest problem is not on the technology side. It's in the kind of penultimate mile and the last mile in actually turning that data into something that's usable for the end customers. And I think it needs to be usable is the first part. And it needs to be affordable, which is the second part. So we have a partnership with the European Space Agency where we basically have a commercialization project with them. It's a quarter of a million euro commercialization project where we can use some of their cutting edge satellites and some of their new satellites coming

out. And all of these are free and globally available, right? The smarts that we have in the tech stack is super resolving some of these lower resolution datasets and making them usable by effectively teaching computer vision-based algorithms to kind of say, “Hey, look, a spade is a spade. This is this. That's that.” And be able to teach it and become more granular.

That's a really important part of the stack, because it keeps the unit economics down. And it means we can avoid using commercial application. So we've literally won customers, because they said to us, “
“Yeah, we've seen other companies in the geospatial analytics space.” And this ties back to that kind of approaching geospatial 2.0, where we have the technology and the ability to kind of derive all of these insights from the data. But we can't do it affordably at a commercially viable point for the end customer a lot of the time, because these customers say, “Yeah, we saw these amazing products using all of these incredible tech stacks. We thought it look great.” And then the companies would charge a million dollars for a POC. And it's a really difficult jump where it's a million dollars for POC, and you're still relying on imagery, which then the customer has to work out how to transform imagery into real world insights in their industry. So we have focused on that specific part of the stack. We say, “Hey, we can reduce the unit economics by 10, 20x. And also, we can turn all of this complex geospatial data into a usable signal that fits into your existing operational workflows.” In our case, it's these financial workflows where quants don't want to have an extra layer of friction. And we've heard from them saying, “Yeah, the only way we're going to use this data is if it reduces the friction in our day-to-day work rather than increases it.” It doesn't matter how much better, how much more accurate the data is. If it adds minutes on to their day, they're not going to use it, because their days are already stacked and doing all the things they need to do. So it really comes back to that frictionless and affordable data, which can actually be used in real world use cases.

[00:37:16] JM: I loved your answer there. So something you mentioned, the unification of modern satellite data imagery together with modern computer vision software is a powerful idea, because modern computer vision software is really good. The API's are really good. And satellite image data is as complex of a data set as really as you can get. And so doing things like entity resolution and categorization and other things that machine learning has gotten really good at can give you a lot of utility in a standardized way.

So when I talked to Planet Labs, it seemed like managing infrastructure and managing the costs was pretty important, because when you're a company like that, you basically have infinite that you could do. You could spend all your time, you could spend all your money running machine learning and analytics across satellite data, and you could probably go broke that way. And you could correct me if I'm wrong here. And I'd love to know how you think about cost management. How you think about data pipelines, and just sort of what the actual implementation details of the pairing of satellite data imagery together with modern computer vision looks like.

[00:38:43] JG: Yeah, I think it's a really, really interesting and complex one, right? And I guess I don't know that the planet unit economics inside out. But my assumption would be that owning and operating satellites is always going to be more expensive than Google Cloud compute, right? Like however much data they're processing, it's probably not going to be to the order of X tens of millions of dollars that they use to launch satellites. However, I think the unit economics point still holds true as a principal, right? And I guess sliding down the cost curve as we have. And one of the assumptions that I have, and it's maybe a bit of a kind of cuts against the grain in some ways, is that data is rapidly becoming commoditized, right? Planet have a great first mover advantage where they have this kind of system, and then they have customers. There's like hundreds of satellites being launched every year. And that's great for the downstream data applications. And obviously, like you said, the cloud compute and all of the processing now, like it's cheaper than it was, but it's going to become affordable. It's not necessarily yet affordable. You still have to be careful of unit economics. But it's a really exciting opportunity for downstream data science-focused people, and it's actually much more of a headache for some of the upstream data providers, because like Planet has petabytes and petabytes of data that it's then ingesting and kind of putting through Google Cloud servers. And the unit economics are insanely high.

We, as Sust Global downstream have the privilege of saying, “We don't want all of those petabytes of satellite data.” We can use historic catalogs of data to train our algorithms. We have access to loads of training data where we don't need to pay exorbitant costs and blow out our unit economics by saying, “We have to have data that was observed yesterday.” There are some high-frequency trading applications where that is applicable. But like a lot of the time people are counting cars in parking lots and – I don't know. It tends to be a use case that people

love to point to in terms of computer vision-based analytics. It's like whose problems does that actually solve? Or is it a hammer running around trying to find a nail.

Us companies that exist downstream, we have the real privilege of being able to say, "We don't need to process petabytes of data." Or we don't need to have really high cadence near real time data on everything. We can pick and choose that data. And a lot of that can come from ground sensors and existing data sources, which aren't really expensive, right? A lot of that can come from the free to use NASA satellites, or the Sentinel satellites. And then we can find if there are specific areas that we need to improve the resolution of the data, we can sample some of those large commercial operators of satellites data. So we can really pick and choose many of these different data sets. For us, it's less of a problem in terms of the sheer cloud compute costs. It's more a case of us saying, "Out of all of these different data sets that are available, from all of the different satellite operators, from all of the different ground sensor kind of data suppliers, from all of the different regular data sets that exist across the industrial world today, like where are our assets? What's the output? What's the CO2 emissions? What's the scope one? What's the scope two emissions data on all of these assets?" Like how do we select the most relevant data sources that can provide a clean signal to our customers?

So it's a real kind of – It's not an answer, or it is an answer, but it's an inversion of the question, I guess. Where, for us, the downstream part of the market is less about how do we keep the unit economics in check for these petabytes of data that we're processing from space is more? How do we select the right data sources in this area of data abundance to actually provide clean signal to our customers?

[00:42:28] JM: All right. Well, there're a lot of topics that we've touched on here. And we're nearing the end of our time. Could you close off by sharing some counterintuitive notions that you've learned as you built this business?

[00:42:45] JG: Yeah. I mean, firstly, thank you so much for having me on the pod. Jeffrey, it's great to have many, many questions where my technical answers probably are insufficient, but maybe I can kind of counteract that with the counterintuitive answers that I can sometimes spit out.

[00:43:02] JM: And your charisma. Your charisma – Charisma is far more important than technical knowhow.

[00:43:08] JG: Yeah. Who needs smart people when you've got **[inaudible 00:43:10]**, right? Anyways. No, I'm in no way putting myself in that bracket. I think some of the really counterintuitive things that I found, and like having worked at some of the largest geospatial analytics companies and with some of the largest geospatial analytics companies and seen it up close, it's so easy to always focus on the technology stack where we're kind of saying, "Look, what new bits of technology, what new processing flows, what parts of the tech stack can we kind of reduce friction or increase the output or throughput?" whichever one of those technical problems that we're solving. But the real zero to one innovations that we need to see in the geospatial analytics industry, and more broadly, all come on the go to market side and on the commercial side. Like we live in an era of incredible technical capabilities, especially within this space, right? You show Jack Dangermond the things that we can do now with geospatial technology. Show that to him when he started Esri. He would be like, "Holy fuck! This is insane." Like how on earth can you do all of these things? We have those capabilities, but nobody cares outside of technologists.

Really, like technology is the one to N. Technology is me saying to you, "Hey, I can do what you're doing 30% quicker, or 20% cheaper, or 10% more effectively," whatever that output is. The true zero to one opportunity is in turning all of this data and all of this technical power into actual new industries and actual new use cases. For us, it's climate. Climate is the biggest existential threat facing us as mankind. It's also the biggest opportunity. There are so many incredible climate resilient products and services in the financial sector, in the broader corporate world that we can bring to market. And we have the technology foundations today, right? There's amazing open source data. There's amazing visualization data like Esri, like Unfolded. All of these new ways of parsing and processing the data. But whose problems are you solving? And how do you bring new products to market? So really, it is that go to market. It's the idea that nobody wants a dashboard. Nobody cares how powerful our tech stack is, right? Like you look at Google, they had a competitive advantage on the tech side when they came to market. They had the best search algorithm. But that isn't what turned them into this multi-billion dollar business. It is their go to market in their revenue generating strategies around advertising, where we fall in love with technology, and we fall in love with the innovators. But that's only half

of the solution. And we have to have the commercialization of the new go to market and the business model innovation, which the counterintuitive point is that's the real zero to one player.

[00:45:50] JM: Powerful conclusion. Josh, thanks for coming on the show. It's been a real pleasure.

[00:45:54] JG: Yeah, likewise. Thanks so much for having me.

[END]