EPISODE 1190

[INTRODUCTION]

[00:00:00] JM: Companies can have a negative impact on the environment by outputting excess carbon. Many companies want to reduce their net carbon impact to zero, which can be done by investing in forests. Pachama is a marketplace for forest investments. Pachama uses satellites, imaging, machine learning and other techniques to determine how much carbon is being absorbed by different forests. Diego Saez Gil is a founder of Pachama and joins the show to talk through how Pachama works and the long-term goals for the company.

If you'd like to support Software Daily, go to softwaredaily.com and become a paid subscriber. You can get access to ad-free episodes and it costs ten dollars a month or a hundred dollars per.

[INTERVIEW]

[00:00:49] JM: Diego, welcome to the show.

[00:00:50] DSG: Thank you, Jeffrey. Excited to be here.

[00:00:53] JM: You work at Pachama. Pachama is a company that helps with carbon offsetting. Can you explain what carbon offsetting is?

[00:01:01] DSG: Sure. This is an idea that was invented at the United Nations in the Kyoto Agreement, which was a predecessor to the famous Paris Agreement. And the idea is that as we move away from fossil fuels or as a way to incentivize the move away from fossil fuels, we need to make polluters, companies and countries, to compensate for the carbon emissions they're putting in the atmosphere that are causing climate change, right? So there is a system by which then projects that either reduce emissions or remove carbon from the atmosphere can receive current credits equivalent to the amount of carbon that these projects are

effectively reducing or removing. And then these companies can purchase those carbon credit certificates and use them to compensate their emissions. And in doing so, that would benefit in one side. Now they have an economic incentive to move away from fossil fuels so that they don't have to be spending money on compensating for those emissions. And secondly, that money goes to fund very important projects such as renewable energy. And in our case, forest restoration and conservation.

As you know, forest remove carbon from the atmosphere as trees grow. And thanks to this framework. If you have a piece of land in which you can plant trees or conserve an existing forest that otherwise would be cut down, you can get paid through carbon credits to continue doing that work. So that is the framework. And this framework has existed for many years. But unfortunately, until today, there wasn't a lot of technology, a lot of software powering the certification and exchange of those current credits. And that's what we decided to focus on at Pachana.

[00:02:59] JM: What kind of technology could be useful in the workflow of managing carbon offsets?

[00:03:05] DSG: Yeah. So the first part is validating and monitoring how much carbon is being sequestered by a forest. And to do that we use remote sensing data and machine learning algorithms that basically aim to predict how much carbon is there on a forest. In particular, for example, one of the models that we built train on a combination of satellite data coming from NASA, Landsat data, with LiDAR data coming from companies that collect that data from airplanes, and it's basically a three-dimensional cloud of points that gives you the shape of the structure of the forest. And then we have ground truth that is coming from forest services from around the world that send people to the field to count trees and measure trees with tapes and know exactly how much carbon each of these trees have.

So then we run deep learning algorithms, convolutional neural networks, to train on the data and then be able to predict how much carbon is there on a forest just using satellite. That is one of the parts of the stack that we're building. Being able to replicate the tests that are

needed to validate how much carbon is being sequestered by a forest. Then the second part is really on the exchange of these credits. Today that is done by brokers in a very manual way. And we built an online marketplace that basically allows the different parties to connect online and then we have an exchange and an accounting system on the cloud that allows for the transactions to happen online and therefore is [inaudible 00:04:52] to be retired. And then through an API we connect to the public registries to make sure that everything gets settled on this public carbon accounting.

And then there're a lot of other cool things that can be done with technology to make the system more efficient and more transparent.

[00:05:10] JM: Are you running a business or a non-profit?

[00:05:14] DSG: This is a business. We when we started the company, we thought about whether this should be a non-profit or a business. But we had a clear business model, which is that if we could help increase the flow of capital going from corporations and governments towards projects that are doing reforestation and conservation and we could take a commission out of the transactions that we facilitate, then we have a business model that is not that different from Airbnb or Uber or Zillow. And if we could have a scalable business model that would allow us to grow faster and invest more in our technology and in growing our team, that makes sense. So it's a business model that is aligned with a mission that is not about profits, but it's about protecting the planet.

And also by being a for-profit company, we could tap into venture capital and be able to just grow faster. And the climate crisis is so urgent that we need to go fast. And that's why we decided to go as a for-profit business.

[00:06:20] JM: Before we get into engineering more, can you speak a little bit more about where the profit margins lie? Like where do you actually make money along this process?

[00:06:29] DSG: Yeah. So on the online marketplace where companies meet project developers and transact on the exchange of these current credits. We take a 20% commission. Again, similar to other online marketplaces out there. And all the data validation and monitoring that we do, that's something that we kind of offer for free or as part of the information that is given to buyers when they purchase carbon credits. In the future, when we use that data to help originate and certify projects, then there is a business model there too. But for now we're just focusing on the transactions.

[00:07:09] JM: So walk me through a traditional or a typical transaction that a corporation might make on Pachama.

[00:07:16] DSG: Yeah. Let me use an example that is exciting, Shopify, for example. Shopify is a company that as you may know powers a lot of e-commerce sites, right? And Shopify came to us and they not only wanted to compensate their own emissions and purchase current credits to compensate on emissions, but they also wanted to offer their merchants the ability to offer carbon neutral shipping.

So with them, of course, the first step for them was to navigate through the marketplace and to study all the projects that we represent on the marketplace and to ask a lot of questions about them. But then after they made the initial purchase, they asked us, “Can we do an integration with you guys?”

And so we started building an API so that microtransactions could happen and they could make these microtransactions on behalf of the merchants. So that was the case of not only them coming and selecting projects and purchasing projects in the marketplace, but also integrating a little bit deeper.

[00:08:21] JM: So let's say I select carbon neutral shipping on Shopify. What is actually happening on the back end when I'm checking out?

[00:08:31] DSG: Yeah. So in one hand there's going to be a calculation of what is the current footprint of the shipping. And to do that, basically there is a formula that calculates how many miles the package needs to travel and then there is a coefficient that converts that mileage into carbon footprint. So once the estimation of how much carbon is to be compensated, then a microtransaction for the purchase of current credits gets triggered. And then that basically retires credits from the public registries. And retiring a credit means that that credit cannot be sold again. And that money ends up going to a project developer in Brazil or Peru or the US or Canada that is effectively conserving a forest or restoring a forest.

[00:09:26] JM: How do you find the forest projects that are on the carbon reduction side of Pachama?

[00:09:34] DSG: Yeah, good question. So today there are about four main certification bodies that issue those carbon credits. And generally when a forest project developer or a forest owner wants to participate in this market, they go to some of these certification bodies. They go through a very thorough process of due diligence and then they get listed publicly on public registries. So what we did was initially go into public registries and reaching out to all these project developers, inviting them to list on our platform.

And now that we have a good number of the existing projects and we want to help incentivize the creation of new forest projects, we really are partnering with non-profit organizations and just doing a lot of press to let people know that this market exists. And what's happening now is that they are starting to come to us and say, “Hey, I have 5,000 hectares of forest in the southeast of the United States and would love to work with you guys to bring this forest to market and conserve and restore those lands.”

So it's starting to happen. Initially we went outbound, if you want, and now it's happening. They're coming to us. But of course there's still a very long way to go. There's a lot of land that is available for foreign restoration. So we have a long way ahead.

[00:11:03] JM: Tell me more about the software underlying Pachama in terms of forest analysis.

[00:11:09] DSG: Yeah. So I can talk a little bit about the data pipeline. Measuring forest biomass is a hard problem. You need to basically be able to pull data from a lot of different sources. We pull data from NASA. We pull data from the European Space Agency. We pull data from private satellite companies such as Planet and Maxar and we need to be able to ingest all that data and then be able to run that on a production-ready way, right? So it's been really lots of work for our engineering team.

To mention some of the tools that we use, we run everything on Google Cloud and use Kubernetes, Postgres, SQL, and then of course TensorFlow, NumPy, Python in general. And then yeah, so we have a team that is working on building the data pipeline and then another team that is working more on the models and basically running a lot of different experiments and then training these models on the data that we have.

And then there's another small team that is working on the exchange and the marketplace that pulls data through an API coming from that machine learning pipeline and then feeds into the exchange in the marketplace, maps or charts or different ways in which we display the data that we produce from the machine learning pipeline.

[00:12:37] JM: What kinds of features of a forest are used to estimate the amount of carbon reduction that might be done by improving that forest?

[00:12:47] DSG: Yeah, good question. So the big question we need to answer is how much carbon is there on the forest? And carbon is a function of the size of the trees, and then of course the species of trees, because each species has a different carbon concentration on their trunks under their wood. So basically the algorithms start to correlate the level of greenness that is observed on the satellite images, the height of the trees, which is taken from the LiDAR data. With the carbon concentration that we have on the field plots that we collect, right? So the algorithm starts to correlate greenness, height and carbon and then ends up

coming up with these predictions. And so you feed a satellite image and you can predict that, carbon concentration height. We can also be able to take a species type, conifers and non-conifers, by, again, detecting the shape of the trees and the greenness of the trees.

[00:13:54] JM: And how do you have your set of forests that are on the Pachama platform? Are you getting other people to submit the forest or do you manually select and find the forests that you're looking for investments in?

[00:14:10] DSG: Yeah. Until today is people submitting forest to us and then they have to give us their KMC file, their shape file that we use to then analyze those forest. But it's a good point that you're making, because we're starting to think how can we use the data? Because now we have global data of where is there a standing forest? Where is there deforestation expanding? And therefore putting those forests under risk. And where is there potential for reforestation where the land would allow for trees to grow given the topography, the climate, the water flows and so forth?

So we are thinking, “How can we use this global data that we have to go more proactively to say, “Hey, in this area, there should be projects.” And then reach out to the landowners of those areas inviting them to list on our platform. So that's something we haven't started yet, but we're preparing for that to have a data-driven approach to project origination.

[00:15:21] JM: You've mentioned a couple elements of your stack. I think the first one to dive into is GCP. So you said you predominantly use GCP or you're entirely on GCP. What do you like about Google Cloud?

[00:15:32] DSG: Yeah, I think it was the integration with Kubernetes and Docker. The ease of use was what got us excited. We do have some experiments that are running also on Azure and we also like a lot Azure. But yeah, I mean Google Cloud seems a very easy to use scalable product the engineering team loved.

[00:15:57] JM: And tell me a little bit more about the technology choices. So you use Kubernetes and using TensorFlow for your machine learning pipeline?

[00:16:04] DSG: That's correct, yes. Yeah, TensorFlow, again, it's been evolving a lot recently. And, yeah, recently we also had some Google team members uh coming and helping us out with our data pipeline, our machine learning models. So one of them work on designing and prototyping a new data processing layer based on Apache Beam. And this engineer actually comes from the Apache Beam team inside of Google. So he was able to help us on that front.

And then, yes. I mean modeling improvement is all the time something important to do. So that's what the team is working at at the moment.

[00:16:45] JM: And what are the datasets that you're looking at? Is it satellite imagery mostly?

[00:16:51] DSG: Yeah. So, satellite. Again, satellite, we tap into Landsat, which is NASA's satellite that has been on orbit since 1984 and that have a good coverage of biomass globally. We also tap into radar data coming from the European Space Agency. There is a satellite called Sentinel, Sentinel 1 and 2 that produce radar data. And the good thing about that data set is that radar can look through clouds. And so in tropical forests where there's a lot of clouds, that is important. And then we also pull data from Planet, for example, which is a private satellite company. They produce high-definition satellite images that can be used to go very deep into understanding what's going on in a forest project. And then as I was saying, LiDAR. LiDAR is data that's generally collected by airplanes. And then we purchase this data. And there are some open data sets out there, but we incorporate that data into our pipeline.

One exciting data set that we're tapping into is Jedi the. International Space Station recently installed a LiDAR there on the ISS and is collecting data from the entire planet. And so that's pretty cool because we're going to have LiDAR coverage of the entire – All forests globally. And we're starting to use that data to also train some algorithms.

[00:18:23] JM: Tell me about working with these satellite images. Like what kinds of processing is necessary to derive meaningful features and data from the satellite imagery?

[00:18:33] DSG: Yeah. I mean I guess that the key is on combining these different data sets. And we have to train models for entire regions, for entire landscapes. So for example, we have to train a model for the entire Amazon rainforest. And we're talking about lots of data. So in that sense, it was a challenge to decide what are we going to store locally? What are we going to keep outside of our own storage? And yeah, it's really a work in progress when you're dealing with such big data sets.

[00:19:10] JM: And what about some of the internal tooling that you've built to manage the different projects, the forest projects, and to manage the carbon estimation?

[00:19:21] DSG: Yeah. So we built an internal exchange platform that basically keeps account. It's almost like an accounting system that keeps account for each single project. How many credits have they certified? And then what is the carbon sequestration of a project? And we have to basically also account for the transactions. How much money are these projects receiving? What is the price upper ton? What's the availability of these projects? So yeah, that was also quite exciting project of building an entire accounting and exchange internal system. But now with that in place, we can really start to scale.

[00:20:07] JM: Where are the best forests for carbon capture?

[00:20:10] DSG: Yeah, great question. The forests that have the biggest carbon concentration are tropical forests, right? So the Amazon Rainforest in Brazil, Peru, Colombia, Venezuela. And in Africa, you have the Congo Rainforest. In Asia, you have forest in Indonesia, in Borneo. In Southeast Asia in general there are many areas with lush rainforest. But boreal forests are also great, and here in the US and Canada and in Russia and in Europe, there're a lot of boreal forest that sequesters a lot of carbon. And then there are certain type of trees that grow faster, like eucalyptus and pines which historically have been used for timber, that they can effectively also be used to sequester carbon. So the short answer is everywhere we are putting a big

focus on the tropical forest, because there is multiple benefits. Not only these are the biggest carbon sinks, but they are also the forest with the highest biodiversity, they habitat for millions of species. And climate is one of the crisis that we are facing, but biodiversity loss is another one that has big implications for health on the planet. So that's what we're putting big focus on the Amazon and other tropical forests.

[00:21:28] JM: So now that we've covered this in little fractures of how the overall platform works, I'd like to revisit the higher level view. So from what I understand, I'm a corporation. I want to offer a product that lowers carbon emissions, like a Shopify check box that makes a higher, a more expensive shipping option that does carbon emissions. And then the additional dollars gets sent into Pachama and gets allocated to a Pachama project. And then that eventually leads to more plants being planted at those locations, for example. Am I understanding things correctly?

[00:22:09] DSG: Yeah. Yeah, that's correct. That's correct. So as more and more companies pledge to be carbon neutral and move into purchasing current credits through our platform and through other means, that means that more money is going to be flowing to projects that are either planting trees or conserving standing forests and basically not cutting down those trees that they would otherwise.

And in the planet, there is about a billion hectares available for forest restoration and three billion hectares of standing forest. So yeah, that flow of capital hopefully will help restore massive amount of forests and sequestering a lot of carbon, which is key to solving climate change. So yeah, you got it right there.

[00:22:56] JM: What are the hardest engineering problems you've had to solve recently?

[00:22:59] DSG: Yeah. I think probably the hardest one was building that data pipeline that could scale ingestion of multiple datasets coming from lots of different sources and then be able to run machine learning modeling in production at scale globally. And yeah, we had to –

We did the first version. Then we did a refactoring of that, and that is almost in production now and quite excited about having a strong foundation on the front.

[00:23:33] JM: How you see the company evolving in the next five to ten years?

[00:23:38] DSG: Yeah. I hope that Pachama would become the leader of the carbon offsets market in which all the corporations and governments that want to have an impact compensating their current footprint use and come to us to help them not only purchase carbon credits, but hopefully we can expand into helping them track their emissions. Plan how to reduce those emissions and then invest on carbon sequestration. And then on the side of the forest, we hope to not only help them get paid for restoring their forests, but also build software tools for them to plan those activities, for them to track those activities, for them to manage internally those activities, which is quite complex. And no one is building software for them. So those are some of the areas where we see our self expanding.

[00:24:37] JM: Just to revisit this. Why is this so important for large corporations? Why do large corporations care about carbon offset?

[00:24:44] DSG: Well, they are receiving pressure from a lot of different stakeholders, from employees who care about their companies having a positive impact on the climate, to investors that care a lot about ESG compliance, which is a framework by which corporations get assessed on public markets, from governments and regulations. There's an increasing pressure from governments in different countries about companies emissions.

So today the more forward-looking corporations are moving ahead on achieving net zero and having a neutral climate impact. In the future, all corporations will have to do it either through regulations or because that's going to be the norm, right? So I think it's a matter of corporate responsibility and responding to other stakeholders.

[00:25:37] JM: How effective is reforestation in terms of preventing global warming from advancing? How does it compare to other approaches like electrification of automobile fleets or – I don't know. Carbon sequestration? Stuff like that?

[00:25:52] DSG: Yeah. I mean we have to do all of that, all of the above. We don't think that forest is the only answer. But we do know is that, again, there is the potential in the planet for planting one trillion new trees and there is an estimation that those trees could sequester about 200 Giga tons of carbon. And it's not super clear how much carbon humans have put in the atmosphere since industrial revolution, but there is an estimation that that number is around 300 Giga tons. So we're talking about the possibility of removing two thirds of what we put out there. And of course it's going to take a couple of decades, but that could really reverse the cost of climate change on a meaningful way. And if in addition to doing that carbon sequestration by forest, we move away from fossil fuels, electrifying transportation and industry. And if in addition to that, we can also ramp up some of this industrial current sequestration to go faster. I think that we really have a shot at reversing catastrophic climate change. And there's going to be some climate change that is already underway. We're raising it with the fires in forests worldwide with so many natural disasters that are happening and with temperatures rising. So while some climate change is already happening, we think that this is a very key piece and that we can effectively reverse the worst of this crisis.

[00:27:26] JM: With the donations to various projects on Pachama, how do you ensure the proper flow of dollars? How do you prevent, for example, kind of evil characters from removing the dollars and just taking them away from the NGO?

[00:27:44] DSG: Yeah, that is super important in this market. We make sure that – It's about making sure that all the credits get retired from the public registries. Public registries are the mechanisms that we have to be able to audit. The decrees have been retired. And then, yes, we have an auditable system that transfer payments from the buyers of carbon credits to the project developers that then do the work on the field.

So yeah, for us, it's super important that those systems are auditable and that the information is kept with a lot of integrity. And in the past, everything was done in a more manual way. I think that the fact that now everything can get recorded on cloud-based platforms I think it's a big advancement for the auditability of these systems.

[00:28:40] JM: I'd like to go back to the engineering. Can you tell me some of the defaults for how code is written at Pachama and just the engineering processes?

[00:28:51] DSG: Sure. I think we use pretty standard tools and practices. Of course, we use GitHub. We use Notion, which is this knowledge sharing platform in which we write down our engineering principles. And then we use Asana for project management. So yeah, again, trying to bring the best practices, but also keep it simple. Each engineer has a set of challenges and runs on their own. We generally use OKRs to set objectives for ourselves and then each engineer will take on a particular challenge and run and then we have a weekly sync-up in which everybody comes together and makes a demo of what they're working on and make sure that everything comes together in a nice way.

[00:29:43] JM: Do you have any struggles with the machine learning deployment process, like the model training and deployment process?

[00:29:50] DSG: Yes. It's always a challenge, again. As I said, in our case, we're talking about thousands of weird shapes on a map that is part of our analysis, right? And, again, using some of the tools that I mentioned before is that we were able to improve the deployment and be able to have a production ready global system that could train on the different data sets that we have. But yeah, it's always a work full of challenges.

[00:30:25] JM: Are there any data sets that you wish you had access to? Are there any data sets that would let you have greater insight into what is actually improving the carbon reduction through forests?

[00:30:37] DSG: Yes, good question. So each country, as part of the commitments towards the United Nations and the Paris Agreement is supposed to build a forest inventory. And a forest inventory is basically sending foresters to their lands and collecting a bunch of field plot data. And so those data sets exist on the forest services or the agricultural ministries of every country who signed the Paris Agreement. And some countries are more open than others about sharing the data, right? So we've been knocking on the doors of especially the countries with the biggest forest, right? Tropical forest countries. And we were able to obtain some of them, but not all of them. So yes, top of our wish list is having access to this national forest inventories that can help us train our models and improve the accuracy of our predictions.

[00:31:37] JM: All right. Well, as we wind to a close, I'd love to get a little bit more on your perspective of where we're going with global warming. Do you have any predictions for how bad it's going to get? What kinds of natural disasters are going to occur?

[00:31:50] DSG: Yeah. I mean, unfortunately, I think that things will get worse before they get better. I do think that we will continue to see bigger and more serious fire seasons in places like California, Australia, Siberia and the Amazon Rainforest. We will see some level of sea rise that will cause problems in places like Miami or any city that is at sea level. And unfortunately we will see more hurricanes and big storms and typhoons in places like Southeast Asia and the Caribbean.

Now, another thing that is possible that we will see is the certification of certain areas where agriculture happens today. But as I said, I am optimistic. I do think that this is not a lost battle. And the next two decades, the 20s and the 30s, are the key ones. So if I had to make a prediction I would say that the 20s, the decade ahead, is going to be a decade that's going to see several natural disasters. And in the 30s, hopefully we're going to have a lot of mitigation in place. And by the 40s, we’re hopefully going to be on our way to really a full de-carbonization and hopefully you know even more mitigation frameworks and mitigation solutions in place. But yeah, it's definitely a crisis and we have to get started right now to turn the ship around.

[00:33:23] JM: Well, Diego, that seems like a good place to close off. Do you have anything else to add about Pachama or what you're working on?

[00:33:28] DSG: Sure. Sure. I mean we are a small team. We are 12 team members and we went fully remote. And since we started the company, we wanted to be a fully remote company. And then, of course, COVID, sending everybody to work from home accelerated our intention to be a remote company. So in the last six months, we've been learning a lot of practices to coordinate ourselves remotely. And it's also been fun to see team members move around and be wherever they want to be. One of our engineers is actually living in a van and traveling in the coast of California. Another engineer is in Missouri, another in Utah. Another engineer actually moved to a farm in the Napa area of California. So it's just great to see people going and working from wherever they are the happiest.

And I would say that we are growing our team. We're going to be hiring – we're looking for full stack software engineers, machine learning engineers, data scientists, and we'll continue to grow our team. So if anybody listening to this podcast wants to work on the climate crisis, please come and talk to us on our website. You can find our contact information.

[00:34:37] JM: Diego, thanks for coming to the show.

[00:34:39] DSG: Thank you so much, Jeffrey. Have a good day.

[END]