Constellations Podcast

Episode 27 – Big Data, Heavy Data and Location Data is the New Database

Speaker: Shay Har-Noy, Vice President, Unified Platform at DigitalGlobe – 26 minutes

John Gilroy: Today on Constellations, we will focus on one of the hottest satellite applications, Earth Observation, or as sometimes referred to as Geospatial Intelligence. The EO business is growing and changing fast. Technological trends such as increased cloud computing and machine learning capability and improved software driven automation are streamlining the image to insights process. At the same time, business trends such as volume based imagery subscriptions, the rise of analytics, and the demand for a higher refresh of global monitoring are expanding opportunities in the value chain.

Today, we'll visit with Dr. Shay Har-Noy, vice president of Unified Platform at DigitalGlobe to discuss EO, its changing technology, new business models, and new market entrants. Shay, one of the DigitalGlobe blogs states that location data is the new database. Instead of building spreadsheets, companies are creating maps to inventory information important to their business. Can you explain that to our listeners? That's quite a concept.

Shay Har-Noy: Absolutely. I would also double down on that and say the map is a new operating system. Take a look at what you spend your time on your phone. Look in your pocket. That device that you're attached to and addicted to, my guess is most of your time is either spent on your email or navigating some place, doing a search about things around you, about things around where you're going. Fundamentally, more and more of our data is being collected, is being tabulated, and much of it has a location, a location component. So how do you make sense of all that?

All that streaming down, it's accelerating. How do you make sense of it to create insights out of it? That's what I spend my time thinking about.

John Gilroy: Yes and actually the question is, so you have all this massive amount of information. I mean, I was watching one of your YouTube videos; you talked about terabytes and gigabytes, and all kinds of different sized information. How do you make sense of all this big, big data? In fact, I think you called it heavy data.

Shay Har-Noy: Yes, exactly. That’s exactly right. A lot of the work that's been done over the last five years has been in the Big Data space. That's like looking at ATM transactions, looking at tweets as a stereotypical use case. There's a lot of really interesting technologies that allow you to pass that information. Imagine billions and billions and billions of small records. Cool. Well, what we're doing

and what DigitalGlobe does, and what the satellite imaging industry does, is it collects imagery around the globe. It's global in nature, but each one of these images is not that big. It's just really heavy. What you end up getting is relatively few, really heavy files.

We collect about 100 terabytes every single day. 100 terabytes is about 400 MacBook Airs, 400 or 500 MacBook Airs. I'm looking at a MacBook Air on my desk. About 500 of those collected every single day.

Now, that's greater than my IT budget typically is for personal computing, but it gives you the sense of scale. It's not an IT problem. Historically, we thought about it as a problem. How many disks? How much bandwidth? How many fiber links? All this nonsense. It's an opportunity. It's a giant opportunity to tap into this information and make sense of it at a scale like never before.

John Gilroy: You're based in Colorado. I know there's a Colorado School of Mines out there. From my understanding, really what that is, is a bunch of statisticians trying to figure out what they have, and this is mining. It can be applied to this data too. It's got all the stuff; what's actionable, what isn't actionable. What can add services, what can't. Is that a good analogy there about School of Mines and the School of Information from satellites?

Shay Har-Noy: That's not bad. Now, one of the things that you mentioned mining, and you mentioned it as an analogy. Actually, one of the spectral bands that we have, on worldview three, one of our high resolution satellites has a collection of spectral bands called SWIR, short wave infrared. Well, if you look at an image, it's cool, it's interesting in the short wave infrared bands. But what's really interesting is when you apply statistical algorithms to it, when you convert those SWIR bands and short wave infrared bands, you can now detect materials from space.

Imagine being able to classify every rock, every bit of soil, every roof, and every building material around the globe, and what that allows you to do. For the mining industry, you can now look for high likelihood areas that you might want to explore. You might be able to do environmental baselining to understand what the impacts on the environments are, your insurance company. I assure you a clay roof has much different fire profile, fire risk profile, than a cedar roof or than an asphalt shingle roof. But tapping into this information at a great scale, you could always look from space and try to guess what's down there. But now when you're able to do that scale you're able to ask totally different questions and get totally different insights. Absolutely, I think the innovations around statistical processing that's been going on over the last 15, 20 years in mining, in energy exploration, in geophysical data is absolutely ... The specific techniques might not be relevant, but it's directly remote sets data, but it's definitely similar motivation. Definitely motivating.

John Gilroy: We get the analysis reports, we get analytics services, we even get crowdsourcing, I would assume, huh?

Shay Har-Noy: Yep, absolutely. My background, I actually came from a company called Tomnod, that I was co-founder and CEO of, which we joined DigitalGlobe in 2013. One of the amazing things is that back then, machines were totally dumb. In order to classify large collections of imagery; terabytes and terabytes of imagery, you use humans and we're crowdsourcing the way to scale humans. And now we're seeing the opportunities to use the crowd in order to augment the machine, using the crowd in order to create training data and using the crowds in order to clean up the outputs for the machine.

For the more paranoid of us, machines are not coming after us. Machines are still really dumb. There's still a lot of work to be done, my training them and having them work in conjunction with really smart people and really smart analysts allows us to focus on the problems where we can have an impact on rather than just the mundane chore of sorting through statistical data.

John Gilroy: Let me try to come up with one of my fracture analogies here. You could say that Earth Observation satellites might be computers floating in the sky, and Satcom satellites might be nodes in a network. Does this hold up anywhere, or does this fall apart, this analogy I'm trying to put together?

Shay Har-Noy: DigitalGlobe is now part of Maxar, at the forefront of a new space economy. What that really means is that we're making space way, way, way, more accessible. Meaning once you put something into space, you can retrieve it, you can repair it, you can beam down data, you can make sense of the data that's coming down. Some of the highest resolution satellites, actually the highest resolution satellites are Maxar owned, the highest throughput satellite communication satellites were built by SSL out in Palo Alto, one of our sister business units. And so we spent a lot of time thinking about how to do more within space.

Now getting a little bit technical about it, high resolution or high capacity communication satellites are definitely nodes in a network, but they're actually fairly simple in their operation. They're very sophisticated in their design but what they're doing is they're relaying message from the ground up to space and back down and transmitting it in the lowest cost per bit per second as possible and doing it at a crazy scale.

Awesome. That’s a theme here, the Maxar theme is operating at scale. Anybody can look at a picture, anybody can make a phone call, but how do you connect an entire civilization?

John Gilroy: Earlier, I talked about Earth Observation satellites. Help me contrast this phrase Earth Observation with something called Geospatial Intelligence. Two different concepts completely, or different sides of a coin, or how do you compare these two phrases; Earth Observation and Geospatial Intelligence?

Shay Har-Noy: Earth Observation is… this is just my take on it… Earth observation allows you to be to a large extent much more context independent. Geospatial Intelligence allows you to figure out why something is happening and to what is happening beyond what you just see in an image. For example, if you know that there's a pipeline traveling in this area through additional information and you use remotely sense imagery and detect changes in vegetation, it's not just a random flower, it's not just a random change in brush. It can actually indicate something that's going on under the ground if you have that preconceived conceptual context of what's going on there.

A lot of times what we see with using satellite imagery is that you're not able to see what's going on directly, you're looking at second order effects, and a lot of that is largely context dependent. You're going to look at a lot of Geospatial Intelligence applications where fundamentally you combine multiple sources of information and preconceived information in order to know what's going on.

I like to use the example of the Apple Campus. Apple campus, beautiful campus, a spaceship campus in Cupertino on the intersection of Wolfe and Homestead roads. I know because I grew up down the street from there. If you were looking at it from space without any context of where you're looking at a random image, it looks like a stadium, like a World Cup Stadium, Mundial if you're in Europe or Los Angeles. But only by knowing, hey, it's actually in Cupertino. Hey, this is actually Steve Jobs' last thing that he decided to build when going from the Cupertino city council. Only if you've been following and understanding what the trends are, all of a sudden make sense.

You can see where people are parking, you can understand where people are going to eat their lunch, you can understand what the overall capacity of the building is, what are the effects of transportation, how the city services are responding to it, what's going on to the environment around it. What do you think the real estate markets going to do around there? Well, intuitively you know that there's going to be 13,000 people stationed there. But by looking at quantitative effects from space, you're able to actually be much more rigorous in your analysis.

John Gilroy: I want to jump back to this human intelligence and solving problems looking at large data sets here. Everyone seems to be talking about machine learning these days, and about artificial intelligence, it would seem that there's a combination where artificial intelligence can be gleaned from these observations as humans

make or ... What came first here? ML, artificial intelligence, and how does this fit in with Earth Observation?

Shay Har-Noy: What's interesting is that Earth Observation applications have been around for the last 20 years. Actually, if I may have permission, I'd love to educate your listeners a little bit on what machine learning means and why now. What's different? Machine learning, AKA, artificial intelligence nowadays, has really taken off because of a number of things. One of which is cloud computing, the availability of cheap computing. If you look at Nvidia and what they're pumping out, it means that there are a lot more computational capabilities out there in the world that's much more cost effective.

Cool. So you can have very analytically rigorous techniques. That's one. Two, there's a lot more data out there, a lot more training data as evidenced by DigitalGlobe's collection of data. Twenty years ago, we weren't collecting data at nearly the same pace. I think we've increased. I don't actually predict 10 x or 100 x even relative to 20 years ago. And then finally, three is what people are talking about as Convolutional Neural Network, CNN, or deep learning. A lot of people are thinking about that. Let me just get very specific.

Machine Learning has been around for a while, but historically, what you've had to do is teach the computer what to look for. For example, an airplane, "Hey computer. Look for something that's white, oftentimes on a tarmac, so on a gray background, that has straight lines and a round cone connected at… it's not 90 degrees, it's going to be at around 65 degree angle." Okay. You're basically explaining to the computer what to look for, and then you are able to tell it to go find these things and you tell it where it's right where it's wrong. Well, that's great, and it's oftentimes very effective.

What deep learning allows you to do is to not articulate heuristically what to look for. You just throw a number of examples. You say, "Listen, I have 10,000 examples of what airplanes look like. I'm going to have a training step, right? Teach the machine to identify things that are like this." Now, all of a sudden, you don't need a different heuristic detector for a Cessna versus a Boeing. You don't need one for a Concorde. Even though they don't fly anymore but you get the point, versus the B-2 bomber. All of a sudden you can just have a training set and if it's large enough and sufficiently diverse, the machine will learn to identify patterns. It will tell you what's significant about an airplane. You don't have to articulate to it the 65 degree angle of a particular model of aircraft. I hope that makes sense.

John Gilroy: That makes sense to me, and that ties in perfectly with your historical library there. That's the pool you can draw from to draw these conclusions I would think.

Shay Har-Noy: Yeah, absolutely. One of the things we've seen over the last few years ... We have an environment called GBDX, geospatial Big Data X. X is because it's extreme, X is for its extract, X is for Maxar, X is because who let the engineers name their own product? In other news ... one of the things that we're seeing there is the very first thing people do when they're creating new algorithms is tapping into the library of imagery that we have… so going back in time to extract examples. If you want to identify UPS trucks, you go back in time and try to collect all the UPS trucks that have been imaged, and then teach the machine to find more like it. We've seen that this library, this rich repository of information, is fundamentally game changing in the world of AI.

John Gilroy: Who knew when you started, but it's there!

Shay Har-Noy: It's actually a credit to my boss, to our founder, Walter, Walter Scott. He always had the vision of, "Hey, machines are going to be able to sort through all this data someday." For many years we kept the camera on, we kept the sensors on collecting imagery at a furious pace. A lot of other places, a lot of other satellites, image when they're told and they do a good job at that, but if nobody's telling you to image something specific, why don't you image anyway?

We have a lot of really clever people working on collection planning to try to collect imagery that has the life likelihood of being relevant in the future.

John Gilroy: Let's shift to some of the new entrants in the market today. We talked about high capacity earlier, and people today are talking about small satellite constellations, and increasingly they're being focused on EO. Are they complementary, or are they competitive to what you folks do at DigitalGlobe?

Shay Har-Noy: I think it's all fit for purpose. Forget about space, forget about satellite, forget about imagery and line scanners, and spectral bands. What problem are we trying to solve and what's the best tool for the job? We spend a lot of time thinking about potential markets with commercial as well as defense markets, international defense markets, and looking at where the industry is going, where the opportunity is. We choose where along the curve we need to be, where we can show the most value based on revisit, resolution, spectral resolution. For example, it's very possible ... I think it was the Apollo 11 mission where they shot an image, the beautiful blue marble. Do you remember that shot? When NASA shot a blue marble?

John Gilroy: Yeah, from the moon, I think.

Shay Har-Noy: It was beautiful. A beautiful shot. I think it comes from the shuttle. The point is that was hundreds of millions of square kilometers of imagery in one shot, in one minute, in one snap of a shutter. Cool. It collected a large area, but you

can't zoom in very far, right? If you're looking at the topography of mountains, at what the continents are, if you're just trying to inspire the next generation of scientists and explorers, that's fit for you. That's the perfect shot. I can't deliver a better shot than blue marble. It's what inspired me to sit here today. Okay, but if you're looking at identifying aircraft, if you're looking at identifying vehicles or finding scooters, at identifying delivery trucks, at looking at manmade objects that are changing on a day to day basis, we believe the manmade objects require a certain level of resolution. That's the trade space.

Shay Har-Noy: If the small sat players are complimentary, I would say absolutely. I would say that the best of them are going to have adjacent market applications that maybe we don't serve today, and I wish them the best of luck in achieving that, and I see there being an opportunity in combining the two having both the low resolution as well as the high resolution. That's how we're shaping our strategy.

John Gilroy: Can I Walter Scott what's going to happen next 20 years? He's got a good batting record here. Doesn't he?

Shay Har-Noy: He sure does. If you look at some of the announcements we've made about ways of merging both high revisits, as well as high resolution, we think that's the killer app. We have, I think, a fairly rigorous plan to making that a reality.

John Gilroy: Jumping around in small sats here, there's something else that's not growing as fast, but the drone industry is growing pretty fast. A lot of VC funding there, I'm sure you know about that. Do you see space base remote sensing companies like yours partnering with drone companies and end to end services for large industry customers? Is that not in your strike zone?

Shay Har-Noy: I'm glad you brought up drones. Kind of piggybacking off the last question, you talked about small sat constellations. I think small sat constellations are a target of agricultural use cases. We need to look at what drones are doing in the agricultural market. Just because it's on space doesn't mean it's not relevant to the same set of customers. We're seeing amazing capabilities in drops and prices from folks like DJI and PrecisionHawk, so we're absolutely partnering with them. We have an ongoing engagement with PrecisionHawk. We're helping to make drone navigation safe.

There are a lot of opportunities in looking at as complimentary data sets. And across Maxar as a whole, there's a number of work we're doing with our MDA sister business unit in Canada on supporting missions associated with drones, drones navigation, drone piloting, and drone data collection processing.

John Gilroy: Even with concepts like water conservation, bang! Drones can help with that combined with your skills or crop field estimation. All kinds of agricultural applications I see here.

Shay Har-Noy: It's all about fit for use, right? If you're looking for the entire seaboard, or eastern seaboard, drones are probably the worst tool for the job. If you're looking at flying a waterway to understand its impact along an agricultural ditch, I can't think of a better tool for it. It's really about what are you trying to do, and how are you trying to solve it. The world is really big. The world is remarkably big. Next time you get in an airplane, look out the window.

John Gilroy: I notice that. It's big. So we got small sats, we got drones. Now we got to talk about Amazon Web Services, which is really a significant player in this game here. It seems to me what they're doing is they're offering freely available data and just letting it available to whoever wants to process it. It seems like DigitalGlobe's offering, and is this the future of Earth Observation data? Just give it away and let people try to figure out?

Shay Har-Noy: I don't know about giving it away, but reducing the technical hurdles for people using it. We're great partners of AWS. We're very early adopters there. The trick is when you collect 100 terabytes a day, when it's 400, 500 MacBook Airs worth, how do I get it to you? Even if I wanted to give it to you, how do I give it to you? How do I ship it to you? It's not trivial, and so what we found is a good compromise is to meet in the cloud. Let's meet somewhere neutral where our data is there, and we bring the processing to the data, rather than data to the processing. This way you're able to reduce your input-outputs, you're able to save time, money, and resources, in order to fulfill the mission.

For example, we recently had a press release of GBDX notebooks where we showed how image processing residing next to the data can be used for drone navigation, for delivery. If you're going to be delivering packages to urban environments, right? We're talking about it, Amazon's talking about it, Google's talking about it, with the project being able to make deliveries on people's doorsteps requires a whole new level of resolution as far as mapping, and a whole new level of resolution as far as the situational awareness. Our system, our GBDX notebooks are being used to help navigate the buildup environment safely.

John Gilroy: I went to YouTube this morning and typed in your name. It's a very distinctive name, easy to find, you had a real, a terrific 15-minute talk there, and you talked about, "Bringing the compute to the data." I thought that was just really insightful. It's like why worry about FedExing 20 terabyte hard drives to Denver when bringing the compute to you. I think that's a ... That's what Amazon allows.

Shay Har-Noy: USB is where data goes to die.

John Gilroy: It's like the second page of Google on search.

Shay Har-Noy: I'm so over it.

John Gilroy: Yeah. Well, we got EO constellations, we got drones, we got Amazon services. Looks like there's going to be a huge spike in the amount of space imagery here. What kind of impact is that going to have on ground infrastructures and distribution platforms?

Shay Har-Noy: One of the big hypothesis of the Maxar combination is because of the necessity of having that antenna system of being able to own and optimize your ground links. I see there being increased pressure on the ground networks and increased opportunity to help make them better and stronger. We're seeing a lot of new ideas, space ideas, space relays, higher capacity, higher throughput satellites, downlinks. There's a lot of opportunity in order to optimize that.

What's interesting is that Earth Observation imagery gets high resolution. What that means is it's actually getting heavier, which means it needs more storage, it needs more backhaul, it needs more downlink, it needs more uplink, it needs more information transfer from place to place. I see there being an opportunity in the space and the space sector, and in the grounds systems sector.

John Gilroy: It goes back to what you said earlier about big data, and then heavy data as well. When I look at some of these value added services, some of them are pretty easy to name, and some of them aren't easy to name. Where do you see the next few years with this increased amount of data and a more sophisticated way to understand? What kind of value added services do you think maybe companies out there can provide?

Shay Har-Noy: Yeah, it's very interesting, but I see there being a big opportunity to expand the market of people using Earth Observation imagery. We've been able to, we call it "Show me there and show me where". We've historically been able to satisfy the show me there use case. Show me that intersection, show me my house, show me my place of business. But what we have the ability to do now is show me where. Show me where there's new construction I should know about. Show me where you think real estate markets are going to be going up. Show me where there's a new Apple campus being built up.

There's a huge opportunity for the "Show me where" costs of use cases, and fundamentally to move into new markets we have to move beyond the pixel. We have to move beyond just an image, or a collection of images, or a bunch of images, or a ton of images, right? We need to allow people to get at the

answers that they're looking for. If you're trying to optimize your supply chain, and you have 300 sites across… and mind you, I've talked to some fairly large manufacturers that have 300 supply chains or more. 300 choke points in their supply chain or more, and sending them 300 images a day is not the way that they're heard, not the way through efficiency. Especially since sometimes it's cloudy, just kidding, maybe 300 images on some days.

What we really need to do is get them the answer, "Hey, here's a chain service." "Hey, here are things that you might want to watch out for," "Hey, here's how the areas are being impacted by neighboring construction." That's where we need to get to. There's a huge opportunity for value added partners. We have an entire ecosystem of strategy built around this allowing people to build products, to build market expertise, and to build on top of our data, on top of our analytics, on top of our platform, to allow people to build products for new market that we're not addressing.

But even in the defense sector, even in the IDI sector, even in the US government sector, there's still an opportunity to move beyond the Geospatial Intel analyst. The person that's been trained to look at imagery and to understand and notice every change of shadow and angle; there's an opportunity to go beyond that and to allow people to jump towards answers or situational awareness. Even there, I would say that we have a 10 X opportunity of increasing the number of people that touch our content, and touch our information, and are able to derive insights from it. That's what I'm excited about.

John Gilroy: Well, this has been a great conversation. Unfortunately, we're running out of time. I'd like to thank our guest, Dr. Shay Har-Noy, vice president of Unified platform at DigitalGlobe.