1:23

Alexey

**This week, we'll talk about extracting space resources from asteroids. We have a special guest today, Daynan. Daynan and his co-founders created Karman+ to mine near-Earth asteroids. Previously, he worked as a data scientist, so I thought it would be a great idea to invite him to talk about applications of data science for mining asteroids, specifically and for astronomy in general. Welcome.**

1:48

Daynan

Thank you. Good to be here.

# Daynan’s background

1:51

Alexey

**Before we go into our main topic of astronomy, data, and mining asteroids, let's start with your background. Can you tell us about your career journey so far?**

2:00

Daynan

Yeah. When I was younger, I was very interested in astronomy and science. As an undergrad, I took a heavy course load in physics and astronomy – and cosmology actually was one of my favorite classes. I also took some courses in neuropsychology and cognitive perception. I was interested in research and thought very heavily about going into a career in academics. But for a number of reasons, I just didn't feel that that was a path for me. I was also interested in public service. So when I graduated, I ended up doing work with a nonprofit here in New York City, where I live.

That kind of spun my career into a combination of public service, finance, politics – I ran a congressional campaign, that's a whole other podcast I'll spare you. But I ended up working for the New York City Mayor's office under Michael Bloomberg about 10 years ago, after Hurricane Sandy hit the city. It was a very impactful job and I love the people I worked with. It was a trajectory where I could see a lot of possibilities in the future and interesting work, but my heart wasn't in it. I realized, when I thought back to the fork in the road – for me, that was leaving science. I wanted to get back to that. I wanted to work and problem-solve with data.

3:13

Daynan

And I went back to school at NYU to pivot my career and I got a degree in informatics. A couple of very close advisers of mine were actually astrophysicists that were practicing academics. Much of what I learned as a data science scientist, I learned from them and really appreciated it. After I graduated, I ended up working. I thought about getting a PhD, but at that point, it was just not feasible economically or otherwise. And I also wanted to jump in and get my hands dirty. So I ended up working in industry. I worked for a great company called GeoPhy, I worked with The World Bank, I worked with the United States Federal Emergency Management Agency, with a company called New Light Technologies – mostly working with remote sensing. And so over the last several years, I've been able to do a lot of work with data analysis and machine learning related to remote sensing and making sense of that data.

About a year ago, I got a call from a former colleague of mine who said, “Hey, I'm starting an asteroid mining company. Resources are important. We want to go find them on asteroids. Do you want to help me build the science and engineering team?” And I said, “Sure. I’m in.”[chuckles] It was maybe a little bit of a longer conversation than that. But it basically kind of came down to that. So a few months ago, I quit my job full time and have been doing this, and building the team out, and helping design our strategy. Some things are the same – on my best days, I still work with data. It's a big messy soup and I'm looking for patterns. Some things are different. As a co-founder, I now have to think about the business a lot more. And we're looking for asteroids – that's a new thing for me. So yeah, that's what brings us here.

# Astronomy vs cosmology

4:52

Alexey

**You mentioned a couple of interesting things. I want to go way back when you said you studied astronomy and *also* cosmology. Are there any differences between the two? Aren’t they the same thing?**

5:04

Daynan

Oh, good question. Well, I guess one big difference is that the timescales in astronomy are pretty massive – you can even think about it in millions of years. Whereas with cosmology, we are talking billions of years. My sister is an academic – she's a philosopher and in physics, and she looks at things like quantum field theory in cosmology. There, you're really pushing the edge of theory because you're thinking about things that… we don't even know how well our physical laws hold up in certain states or what we can know about them. It gets interesting. You certainly aren't able to do experiments in the same way. Whereas astronomy is at least a little bit more focused (in my mind, I guess) on what you can observe immediately. So they're very closely related. Cosmology is, I don't know – it’s fascinating to think about our universe in that broad perspective. Very related, of course, and one drives the other, for sure.

# Applications of data science and machine learning in astronomy

6:03

Alexey

**The next question I wanted to ask you was about applications of data science and machine learning for astronomy. But before that – are there applications for cosmology in data science and machine learning?**

6:18

Daynan

Well, I think whenever you're looking at massive amounts of data – and we're getting more data – the answer is always ‘yes’. If you have a lot of data that you need to understand, I think data science is directly applicable there and cosmology is probably in need of it. I love talking about LIGO (Light Interferometer Gravitational-Wave Observatory) and the detection of gravitational waves, which are traveling across billions of light years. That is pretty fascinating. There's so much we can learn about the creation of our universe, I believe, and how it exists, and how it's moving. Obviously, parsing that data, as with the Large Hadron Collider, you're dealing with a pretty massive amount of data information. How much of that is noise? How much of that is signal? What is the signal telling us? I think from a cosmological perspective, there is certainly a need for data science and machine learning. I'm sure someone could speak more intelligently about that than I. But yeah, it's there.

# Determining signal vs noise

7:20

Alexey

**Yeah, so gravitational waves. I'm wondering how much we (people from Earth) can observe with these gravitational waves. How much signal is there? How much noise is there?**

7:33

Daynan

Well there's enough to actually make some pretty interesting discoveries. So the fact that we're now, with some regularity, detecting not only black holes crashing into each other, (which is a thing that happens, by the way, if you didn't know that) but neutron stars. So it's pretty fascinating. There was even an anecdote. I know a few years ago, in 2017, they had detected what they thought were neutron stars colliding. There are two detectors in the US – in Washington State and in Louisiana – one of them detected this unmistakable signal, or so they thought. The other detector did not. So by their own rules, they thought, “Well, this is an anomaly or it's not a true detection.” But it was confusing, because it was so clear. There's a third detector in Italy that also didn't detect it. Well, it turned out, the one in Louisiana had what they call a ‘glitch’, which is a noise.

They’re very sensitive instruments and so maybe a jackhammer was going off or something. And they scrub the noise at just the right time that the signal happened – which is a low frequency probability, but it can happen. Sure enough, when a postdoc went back and looked at the data, they saw that that glitch was there. When they were able to mask around it, they were able to see the signal. So indeed, it had been there. And I use this as an anecdote for why you need science in addition to machine learning – the automated pipeline obviously didn't detect it and didn't understand that that was a signal.

What's also fascinating about the story is that third detected Italy also didn't detect it. And the reason there is likely because, as gravity is oscillating through space in three dimensions (four if you include time) – the detectors are actually two lasers that are optimized for two dimensions. So if they're oriented in just the right spot, there's a blind spot. So it *just* so happened that this Italian detector was in the blind spot of where the gravity waves were coming from. So you have this combination of coincidences and physics and all kinds of things around the story that I find quite fascinating. Not to go on a tangent, but to answer your question – yes, as humans, we can detect gravity. That's pretty cool.

9:39

Alexey

**Going back to astronomy, there are applications of machine learning and data science there as well. Can you tell us how it's used?**

9:49

Daynan

Yeah. So, data science, I look at it like data analysis, information theory – there's a lot of things related to that. And so generative processes are an important part of how I think about data science. If you're an astronomer, your proximate generative processes usually involve some detector – a sensor, a camera or maybe a radio antenna. You're measuring energy that's been collected and how that energy is measured and put it into numerical arrays is critically important for you to understand. So understanding those generative processes – even beyond that, what are you observing? What natural phenomenon? Obviously, deep space astronomers are looking at quasars, galaxies, black holes, pulsars – all kinds of things – exoplanets.

In Solar System astronomy, we're looking at objects moving around the Sun, so we're seeing light reflected from the sun off an object and that can tell us a great deal about that object. Certainly for asteroids, that's a big part of what we're interested in. How you pull these things together across different platforms takes on more of the data engineering need, because there are a lot of complex science and data pipelines built around these observatories that you have to really understand and structure.

10:55

Daynan

Signal processing is a major component, certainly, of the data science side. Exploratory data analysis, I think, is pretty underrated. I love doing exploratory data analysis – you can see patterns and things and in the data you're looking at. Then, in machine learning, there are so many different tasks that are applicable, but I think a lot of it can come down to – machine learning labels things. There are a lot of things to label in astronomy. If you're looking at computer vision, you want to find sources, “Is this pixel part of a source? Is the adjacent pixel part of the same source? How do you know?” Exercises related to that. Signal processing, again, when you're trying to figure out how to denoise an image and make some choices about what noise is – because everything's information, some information you can ignore.

Machine learning is great for scaling those tasks out. Time series analysis is a major one. Clustering of all kinds of things – spectral data, you name it. Dimension reduction is something I've used quite a bit. Probably a lot of people, data scientists at least, are familiar with principal component analysis. It's a bit finicky – you have to have data that's fairly well normalized, that's well-conditioned. I like looking at auto-encoders, which have been very popular in terms of techniques for dealing with dimension reduction, among other things. So I think there's a lot of opportunity there.

Super resolution using generative adversarial networks, I haven't worked with myself directly, but I see a lot of promise there to artificially sharpen low-res images and see some structure. Any tasks where you need to scale things out. You're talking about petabytes worth of data, so you need to scale a lot of tasks out. It could be a task where an astronomer might be able to identify some artifact or something that's of interest, can that be scaled out? And I think that's a call for machine learning and deep learning. So there's a lot of promise there.

# What the data looks like in astronomy

12:45

Alexey

**I'm just wondering what the data looks like. Is it not the usual JPEG images with three channels? RGB? It's something different, right?**

12:57

Daynan

Great question. Yeah, if you think about even just an image, right? What is an image? Well, it's usually organized in pixels. Or if you want to think of it as a matrix – you have cells of data. As you mentioned, most images used to have three channels – red, green, and blue. If you just look at an image, the combination of those, at a pixel level, will give you a lot of information visually. If you were doing image analysis, you also know that looking at the histogram of those channels, that is incredibly helpful information – seeing what the distribution of red, green and blue is. Well, if you're thinking about hyperspectral imagery, which is fairly common in Earth observation that I've been working on, as well as in astronomy, you're looking across a number of channels.

It's hard for us as humans to visualize something like that. That's where data visualization takes on an importance. Again, you might look at things like histograms, or other ways to show that data – that are not just image data. That takes on a lot of importance and then that might be something that you're trying to optimize with some machine learning algorithm to look for patterns in that sense. At the end of the day, they're really just numerical matrices – if you want to think of it that way. But the amount of information packed into these images is pretty intense. And don't forget, time is a factor here as well. So special-temporal images are a pretty big factor in astronomy.

14:24

Alexey

**And with these hyperspectral Imager images, (I hope I pronounced that correctly) I guess there are more than three channels, right? So what are these channels usually? Are they infrared channels, or?**

14:40

Daynan

Yeah! They certainly can be. When I was in grad school, we worked with a camera that had 172 channels across mostly the visible spectrum and a little bit into the infrared. So you're getting into 800-900 nanometers. The visible spectrum – the colors that we see – I think are something like 350 or 400 nanometers to about 800, or something in that range. Infrared then goes beyond that. If you have spatial resolution, you can think of a high-res image as having pixels that represent a much, much more precise amount of information spatially. Well, you can also have that with spectral images. So, a red, green, and blue resolute image is obviously not very resolute – you just have three general channels. If you have a hyperspectral camera, you can divide that electromagnetic spectrum, (which is continuous, of course) into much, much smaller segments. In which case, you have a lot more information that you're looking at, just as you would with a high-res image. That certainly can extend into the near-mid or even long infrared. That's actually fascinating for us in astronomy, because we're looking (I’ll get into this more later), but we're looking for water, actually.

It turns out, if you understand the physics of it, that hydroxyl bonds (hydrogen, oxygen) absorb water at about three microns, which is *way* beyond what we can see visually. If you look at the spectra of an object that's got water in it, you'll see an absorption line or some feature at about three microns. That's certainly useful for us, if we're looking at spectral signatures to say, “Hey, there's potentially water in the material that's reflecting sunlight that we're looking at.” And that's useful information. Of course, stellar astronomers are interested in stars and all kinds of things, where you’re looking at emission lines and all manner of information that you're seeing across the spectrum. Certainly, radio waves are in the spectrum as well, they’re just *much* larger – way down the road. But yeah, it's fascinating.

# Determining the features of an object in space

16:44

Alexey

**Maybe we'll get into that a bit later. But from what I understood – there is an object in space, like an asteroid, and it reflects the Sun. The sun is shining on it, the light is there, and then it reflects light. So the light goes towards the Earth, where we capture the light. And different objects in space reflect the light differently. Depending on how they reflect, you can understand what this thing is made of, and if there is water or if there is something else. Is that correct?**

17:25

Daynan

Exactly. Absolutely. Yeah. Photometry is basically the analysis of the flux of light – photons. That can help give you a sense of brightness and potentially size. But if you think about it, size is important because it's variable. You could have a very small, but a very bright object. You could also have an object that's very bright that's very close to you. Or if it's very dim – is it far away? Or is it dim because it's small? Or is it dim because it's just darker material? You have to know these things. So if we're looking at asteroids, it's really, really important for us to know where that asteroid is – meaning we have an understanding of its orbit. If we know where it is relative to us, as the observer, and we know the angle of the Sun and us, that gives us some sort of constant at least. And then if we look at the brightness of the object, we can start to derive things like the size of the object, as well as the reflectiveness of the material of the object. Because we don't know one of those for sure, there has to be some level of estimation that goes into that. And that can be interesting.

You also can see even, in some instances, the rotational speed of the asteroid as it's tumbling based on the light curve. If you look at the changes in the light over a period of time – and now we're talking minutes or hours, really – you can see that light curve and infer rotational speed and potentially even the shape of the object. The polarization of light polarimetry is another potentially underestimated tool, I think, in terms of understanding how the reflected light can tell us about the shape of the object. And, of course, spectroscopy, obviously, if you're looking at what the material is, that's being reflected. Of course, that doesn't necessarily mean that's the material *inside* the objects. To do that, you might want to look at muons or magnetoscopes, or all kinds of interesting things that are possible, but potentially expensive. [chuckles] We're not necessarily doing that right off the bat, but those are things that we might look at down the road.

19:35

Alexey

**So the features in this case are the rotational speed, how light reflects from the object also over time, polarization of light, and all these things. I guess, if we think about classical machine learning, these are the features and then the target is – one if there is water, zero if there is no water.**

19:56

Daynan

Yeah. I think there's another angle to this that's also interesting from the machine learning context. When you're looking for water, you can't really look at it from a telescope on the surface of the Earth. The reason is, you have to look through the atmosphere, which, as we know, is full of water. Therefore most of the telescopes don't even have a filter at that wavelength. If they did, you would just see a bunch of white noise. So unfortunately, there's not a lot of data we can get in the three micron range from Earth. However, we've noticed that there's a high correlation of features around 700 nanometers that may represent iron oxides or something (I don't know if the top of my head) but there's a high correlation of those with water-bearing objects.

There's a potential opportunity to think about classifying an object based on features of its spectral signature – its spectral signature is sequential data, similar to a time series. So you do have to be careful because you're not looking at individual dimensions that are independent, as you would maybe in other scenarios. If you know about time series, you know that there's a lot you have to think about in terms of how one component relates to the preceding and following components. But if you do that, to some degree, you can classify objects using spectral signatures.

That's a big area that we're looking into. Because there's only a small relative sample size of objects that we have spectral signals for, and some of those are sparse. Depending on the object's location relative to us, they might also look differently. So there's an opportunity for us to use machine learning to help identify and classify those patterns in a way that may be more challenging for humans to do. And can we extrapolate that against objects that we don't have spectral data for? But we may have other data in terms of their albedo or even their orbits will give us information about where they're coming from. So pulling that all together and making some inference around it is a big part of the job. And I think a lot of it can be supplemented and enhanced quite a bit by some machine learning techniques.

# Ground truth for space objects

22:00

Alexey

**Where do the labels come from? Do you actually go there and check if there is water?**

22:07

Daynan

Another great question. So ground truth – very difficult to find in astronomy. If you're looking at objects that are measured in millions of kilometers, which is a pretty short distance when you're in a domain that measures things in light years. So yeah, ground truth is not something we have a lot of, but it's not zero. We do have some. As a matter of fact – and some folks don't know this – the Japanese Space Agency has sent a couple of missions (the Hayabusa missions) that have collected samples from asteroids. Recently, an asteroid called Ryugu – they brought back samples and they published some analysis on that. So we've actually been able to look at material they brought back from an asteroid. And it's really fascinating, because you can see some subtle differences in the material, not only from what we're looking at from telescopes on the ground, but even the spacecraft itself that was hovering around the asteroid for a couple of weeks and doing remote characterization work, there were even things that we're finding from the sample that were slightly different than what we would have anticipated there. So there *is* ground truth. There's also the Osiris Rex, which is a NASA mission that's returning, and we'll get that sample sometime later this year. So there actually is data that we can look at to compare, but a sample size of 2-3 is pretty small.

The other thing that we can actually look at are meteorites. Meteorites are, of course, asteroids that made it through and impacted the Earth. There's a lot we can learn from that. Of course, if you have a physical sample – and there are tens of thousands of samples that we have so far – you can analyze that in any manner of ways with very, very intense laboratory instruments. There's a lot of that work that's been done. You can potentially use that to extrapolate to the families of asteroids that we think are out there in space. We have to be aware of some very important things like, if it came through the atmosphere, that intense process has changed the chemistry of that object or potentially changed it, so it's not a direct comparison. But there is some way that you can look at that detailed data on a meteorite and link it to asteroids. Actually now that more people are watching the skies, we're able to, in fact, see trajectories of meteorites as they're entering the atmosphere. And then if we actually find that object, because we have videotaped it, we can actually see its orbit and that gives us more information about where this thing is coming from.

24:33

Daynan

So yeah, ground truth is very difficult. Also, a big part of my job is really thinking about how we can validate our models when we are absent that information. Debiasing the models based on what observational data we have is critical. But, candidly, understanding bias in our models is a major challenge. We can find consistency in terms of spectral classification, for example. We may find a classifier that very consistently produces results and therefore reduces variance. Well, that's great, but as many know, reducing variance is only okay, as long as you're not increasing bias – or if you're not paying attention to bias. We don't want to be more certain about something that's wrong. That is not an easy thing to answer. It's also an emphasis why we're thinking about cheaper ways to send spacecraft to analyze these objects – to get more ground truth data – because that's really the best way we think to understand these mysterious beasts that we're after.

# Why water is an important resource in the space economy

25:42

Alexey

**I think many of us are wondering now – why would you care about water in asteroids? The first time I heard about what you do, I thought, “Okay, mining asteroids. You probably want to extract some precious metals and then sell them here on Earth.” Why water? We have a lot of water on Earth, why should we care about this?**

26:07

Daynan

Well, very good question. Well, scientifically, water is really interesting. Water is so important, obviously, to life and may contain other things, including organic compounds or other things that tell us about how the solar system is generated. But from a commercial standpoint, water is important for a concept that I think is really important that not many people have heard about, which is ISRU (in situ resource utilization). The idea with this concept is that the space economy itself is real. It exists and it is growing. So when we think of space, we think satellites, which are of course important. If the satellites failed, our life would come to a halt. The infrastructure that they produce in terms of communication and infrastructure of the internet is immense, among other things.

These satellites are starting to potentially crash into each other, figuring out how to clear the debris, how to service the satellites – this is becoming an industry in and of itself that's incredibly important. But even related to that is the idea of how to refuel these satellites, how to do more manufacturing and activities and research and space. There are currently, I believe, four private space stations being planned in the coming years. They will be up there. As we know, the International Space Station's end of life is in the next few years, and there's no government plan to replace it.

All these commercial activities are in process and manufacturing of certain things – fiber optic cables, you name it – may benefit from low gravity. Not even to mention activity on the Moon or Mars, if we think about expanding as a species. None of this happens without resources – resources that we currently have to bring from the Earth. This includes water, which is a terrible cargo to lift from the ground. It's heavy, it takes up a lot of space. Rocket fuel is dependent on oxygen – of course, water can be decomposed into hydrogen and oxygen. I don't know the exact numbers, but something like for every one part of rocket fuel, you need several parts of oxygen.

26:15

Daynan

These rockets that are launching – the weight – a lot of it is oxygen itself. If we can find a way to get water from asteroids – that are actually pretty accessible in terms of gravity – if we can find them and go to them, the delta V (the change in velocity) required to do that, is maybe less so than even going to the Moon and coming back from the surface of the Moon, just given gravity. Then, we think we have a viable option in terms of finding water and bringing it back. It costs several thousands of dollars to get a liter of water up into orbit right now. Way more to get it beyond that. So if we're able to bring that to the space economy and deliver it to customers who use it for various things – that's the business model we're working towards.

Of course, the long term view is certainly resources writ large. If we come up with the flywheel that allows us to find these asteroids, bring back material, refine it, and deliver it – we can do that scalability – yes, now we're looking at other materials. Rare earth elements, precious metals. There's something that we can think about there. The sensation is, “Oh, we're gonna find an asteroid full of platinum.” Yeah, they certainly have those resources, we believe. Most of the material we're digging up in the Earth's crust came from asteroids, we also believe. Or at least that's a heavy source of that. But I think the infrastructure has to be in place, there has to be an ability to actually sustainably get that stuff, not to mention how you would bring that back to Earth.

There are big, big challenges – even in the economics of it. What's the cost of platinum now versus what it'll be in 15 years? I don't know. It's a good question. I know diesel engines are a big driver of platinum. Are we going to be driving diesel engines as much? Maybe, maybe not. So there's sort of an economics question there that we don't want to have to tackle if we know that there's an immediate need for certain resources, like water, that are more attractive to customers that currently exist, actually. So that's kind of our focus.

# Other useful resources that can be found in asteroids

30:18

Alexey

**What else can you find on asteroids? You mentioned platinum, but I think it was a good example of an expensive thing on Earth that can potentially sell, but it's just too difficult to extract. But what is there in addition to water that could immediately be useful?**

30:38

Daynan

Yeah. There's a lot of unknowns there. I think that's the big question – what materials really are available. This isn't an asteroid that we're gonna go after – it's in the main belt, so it's too far away for us. But there's a mission that's going after Psyche, which we believe is potentially a planetesimal core – it might be the core of a planet, or a planetesimal object, where the rocky layer was smashed away and it's full of metal. So iron and all kinds of metal that you can think of. There may be other asteroids that are components of that. So any type of metal that you're interested in, including precious metals, are certainly to be found there.

From a scientific perspective, the idea of organic compounds is pretty intriguing. You can think of these asteroids as essentially fossil records of the solar system. They're essentially as old as the solar system. But unlike the Earth and other planets, they have not been… Well, they've experienced space weathering and collisions and other things, but they're probably more of a pure record of compounds. That is more of a scientific value than anything else. But I think rare earth elements and precious metals, which are all things that we’re excavating from the earth right now, are assumed to be embedded in these objects. But we just don't know for sure. So that's part of the impetus to us going and finding out.

# Sources of asteroids

32:12

Alexey

**Where do these asteroids come from? Do they just appear? Then you see them and you think, “Okay, we're going to mine that one.” Is that how it happens?**

32:22

Daynan

[chuckles] Yeah. One of the first things we did was hire a planetary scientist – an astronomer. One of the reasons we did is that he's been working with the modeling that answers that question. Most people, I think, are familiar with the main belt of asteroids - it's beyond Mars. There are a lot of theories about the source of that material – was it a planet that was broken up? Is it just a collection of stuff? If you think about the solar system being a sort of chaos for much of its early state, then eventually, as things collected into planets and they found their groove – I love to think about the solar system in terms of resonances. Everything is just sort of resonating and as the planets are going around, they're creating these resonance patterns, which are critically important for us. Jupiter is a massive animal – as it's revolving around the Sun, it's creating resonances with the other planets.

Every so often, in this asteroid belt there may be a collision or something, or maybe some something will get perturbed, but an asteroid will get thrown into an area of resonance, where, because of the gravity of the planets, it will basically continue to shape the orbit of this asteroid such that it'll bring it into a near-Earth orbit. So when we're looking at near-Earth asteroids, our assumption is that their original source was from elsewhere in the solar system – the main belt or even beyond that. In the outer regions, there's a region around Jupiter called the Trojans, based on Lagrange points. There may be a collection of items there that we think for some reason or other were thrown into the near-Earth space. We think there may be 300-400 million asteroids in near-Earth space, ranging in the very small (so that people don’t get concerned about that). But we think the source of those is all coming from the main belt. So our modeling is basically trying to understand those gravitational dynamics. There are other dynamics too – solar wind is a dynamic to pay attention to, and collisions of other asteroids, of course. But we think that's actually the source.

NASA's JPL has cataloged something like 29,000 near-Earth asteroids. If we think there are 300-400 million of them, we're only scratching the surface. Thankfully, I think most people should feel assured that anything that's large enough to maybe do any damage, we've mostly seen those. There is a pretty intense effort to document more of the largest one, so most of the ones I’m referring to are fairly small objects. But yeah, finding them exactly – statistical modeling is really helpful for us, but as I like to tell our team, we're not able to statistically mine an asteroid. So we do need to find an actual anomaly that we can get.

We think probabilistically and I think we're comfortable with that. That helps us, I think, make choices and decisions about where we're sending spacecraft and what we're going to find when we get there. But the real question is, “What are we going to find? Is it going to be there? What is it going to be?” We'll have to start with a sample of asteroids that we have observed and that we do have some information on. That's probably the easiest way to do this. But there's hopefully going to be, in the near future, more telescopes. We know of a couple that are going to be coming online that will be able to see a lot more of these objects. That will be very useful for everybody, including ourselves.

35:48

Alexey

**So as I understood, your business model is as follows. There is this main belt of asteroid asteroids and then Jupiter passes by and pushes some of the asteroids towards the Earth. So you want to detect “On this asteroid that is coming towards us or somewhere in the area…”**

36:09

Daynan

The process that I described – that can happen over many, probably millions of years. The timescales of astronomy are much larger. So we're looking at the asteroids that that's already happened to. They're already here – they're in the neighborhood. Now, if you think about the Earth going around the Sun, and it's following its orbit, a certain velocity finds a groove, like in a record groove, where it's going around. We have reason to believe that there are a lot of other asteroids that are also in that same groove, and they're going around the Sun in that orbit. But we don't see them because we're on the surface of the Earth.

If you think about and if you can visualize the orbit coming around the Earth – if we were to look at an object that's on the line of that orbit, that means you're looking at an object, that is… you've got the sun right here. So in order to see that object, if you're looking at it dead on, it's going to have to be during dusk or dawn, and the light conditions are not good for seeing faint objects. Or if you're looking in the middle of the night when it's a lot darker, you're gonna have to look really near the horizon just because of the angles. Then, of course, you're looking through a much larger column of air and that's also not great for when you're trying to find things. So it's hard to see these objects, but for gravitational reasons, we think that there's a lot of them in that orbit going around the Earth. Most of the things that ended up making it to Earth are probably in this sort of debris pile floating around.

I think the question is not necessarily that we're looking at asteroids coming to us from the main belt – we're looking for asteroids that are already here. It's helpful for us to look at the main belt to understand those asteroids and what they're made out of. Because a lot of them are larger, we have more data on that. And because we theorized that a lot of the smaller asteroids are just fragments or pieces of the larger ones, if we know about the larger ones that are in the main belt, can we extrapolate what we think the smaller ones in near-Earth space are? That's maybe an open hypothesis in some senses. But that's the working hypothesis that we're dealing with.

# The data team at an asteroid mining company

38:13

Alexey

**I assume that to be able to do all that, you don't just keep your data in Excel spreadsheets and analyze that. You need to collect, store, process all this data. So we are actually talking about a data department, right? A team that does all that. Can we talk about that a little bit? What does your data team look like? Who is there in the team? And what do they do?**

38:39

Daynan

Yeah. We're just starting out, so we're a young, small team and people actively working with the data. We have two research scientists. One, as I mentioned, is an astronomer or planetary scientist focusing on asteroid characterization. So “Where are these? What do they look like?” Then we have another research scientist who's a mission design specialist. As an engineer, she's thinking about the spacecraft mission that we will need to help develop to actually access these. And then myself, of course, looking at the data modeling overall. One of the reasons that we're looking to hire a data engineer is because – I've been in many organizations where I've lamented that we didn't put enough emphasis on data engineering. That should really be one of the first data team members hired. I'm in a position now to do something about that. So we are looking to bring someone in. We have someone who's helping us with some really important work, but we're still open for interested folks. I think this person hopefully will help us think about the architecture and the vision of how we're doing this.

39:46

Daynan

There is a lot of data that exists. Some of it is in catalogs – the Minor Planet Center, NASA JPL, ESA (European Space Agency), there's a lot of public agencies that have done an incredible job organizing this data, even doing some derived attributes, including orbit determination and other characteristics. But it still is a field of a lot of specialists. You have astronomers who are focusing on one component of astronomy, and they're working with data that’s related to that, in contrast with, say, Earth observation, which is the world that I was in before. You've seen the market explode in terms of consumers of the data, which has helped make a lot of that data a little bit more usable – more analysis-ready, if you will. I haven't seen that in the astronomical community yet. Not that there's not a lot of open collaboration, but a lot of the data that is provided is in fairly complicated file structures, produced by fairly complicated processes, as I've mentioned before.

If we're going to look at analyzing things across these platforms, there's a real need for some folks who are used to working with data in more complicated file structures. Obviously, image data is a pretty big part of that. But there's a lot of catalog data where the statistics of images and other detectors has been pretty methodically and comprehensively laid out as table data. So thinking about the architecture for how this comes together, and how we can have a really big view. I mean, we're building a Bayesian engine, really. The more information we can get, helps improve our understanding of the asteroids that we're looking at. And all the pipelines related to that – the analytical pipelines related to getting that data and being able to analyze it, as well as, of course, the infrastructure to support the machine learning modeling that we're going to need to be doing and already have started doing. It's all part of that.

I guess the answer to your question is “the data team” is in a future tense right now. That's what I'm hoping to change very quickly. I mean, we're all working with data, but someone who's really focused on helping us build the pipelines for that. And Karmanplus.com is our website and we have a job posting there. So obviously, I'll make a pitch for people who are interested as data engineers. We’ll also be looking for data scientists and machine learning engineers in the future, for sure. Everyone interested, please let me know. I’ll give you my contact info. But data engineering, specifically, is the key focus.

42:23

Alexey

**I'll make sure to include the link to it in the description. I was wondering – do you use Cloud? Because I think cloud here may... [chuckles]**

42:33

Daynan

Yes. I think we're probably going to set up in an AWS ecosystem. I've used Google Cloud. I'm not as familiar with Microsoft. But yeah, we are going to be – I want to say, as close to cloud native as possible. We'll probably still be doing some prototyping locally. It's just easier to play around with some of the exploratory stuff. But the big emphasis is on Cloud for many, many good reasons. We don't want to have to do infrastructure that we don't need to do. We don't want to have to deal with the security InfoSec issues that are going to be very important to understand. So if we're able to use the infrastructure that is provided by a service company like that, that's very useful. On the other hand, we need people and we'll need DevOps and our cloud architects who can help us when we're implementing to make sure that this is set up correctly.

Especially when you're dealing with image data, which is similar to the Earth observation scenario, the scientific mindset that I've seen, at least in astronomy, is – you're doing some kind of analysis, you download all the data that you might potentially need locally. So that's a massive amount of data. It takes forever and now you're in the data management business on your laptop or your computer, if you can even do that. Unless you're lucky enough to have access to a high powered computing cluster, you're going to have to be doing some kind of analysis with that. Or if that data is in the cloud, then you still have communication costs where you're having to move that data when you need it. I think Cloud-Optimized Geo tips in the Earth observation domain are incredibly useful, I think. Watch for that to grow. And STACC (spatio temporal asset catalog convention) makes it a lot more user-friendly for analysts to query certain tiles of satellite data without having to download the whole file. That way, they can do analysis in a much more precise manner.

I have not seen anything like that in the astronomical community yet. Maybe the incentive isn't there, but the ability to query some of these dense files, as opposed to having to process this in a server where the images are cut to spec by some user in a process that's queued up and may take quite some time and it also puts a pretty big burden on the data provider. Or you download all this information and do it yourself, which is a challenge too. Hopefully we will see more conventions like COGs and STACC take off in the astronomical community. But for our perspective as a company, absolutely – cloud infrastructure is critical.

45:09

Alexey

**I was just going to make a silly joke about Cloud being closer to the asteroids because it's in the sky. [chuckles]**

45:16

Daynan

[chuckles] And I took it off on a ramble on the whole thing? They’re good for computing, but they're terrible for astronomy. Yeah.

# Open datasets for hobbyists

45:26

Alexey

**[laughs] We have quite a few questions. The first question is, “Are there open datasets that are interesting to explore as a hobbyist?” Actually, the question is asking specifically about tabular datasets. But maybe we can talk about datasets in general.**

45:42

Daynan

Absolutely. I mentioned the Minor Planet Center. I can send you a link or how to spell this or anything, but it's minorplanetcenter.net. And they have what I believe is probably the database of records of asteroids –and comets too. So if there's an observation that's reported, that's the place that people report it to, and it’s tabular data. They have records of all the asteroids that we know of – there's over a million of them, including, as I mentioned, the near-Earth asteroids, of which there's about 30,000. Then there's some attributes based on the orbital elements that we know from those objects, as well as the brightness and absolute magnitude. There's also an archive of all the observations. For some of these asteroids, we have observations going back to the 1800s. Those are being updated hourly – every day, there's people finding more observations on these objects.

So if you're interested in just taking a look at that, and there's if you tour around in Python, which is a popular language for astronomers because of the open source nature. There are even a number of packages that are developed. I'll follow up with you, Alexey, just to link some of these. There are packages that allow you to query the Minor Planet Center – you can even propagate orbits and you can do all kinds of interesting stuff with them. If you're interested in digging into the data itself, there's a satellite called WISE (Wide-field Infrared Survey Explorer). It was a satellite platform that has infrared bands that it's capturing. It was reactivated several years ago to actually look for near-Earth optics. So the NEOWISE dataset is cataloged in something called the IRSA (Infrared Science Archive) . I'll send these links. [cross-talk If you want to go and query the source table for single exposures of known objects – not just asteroids, but a lot of other stuff, too – that's a pretty fun catalog to play around with.

Boy, there's a lot. I could spend a lot of time doing that. But yeah, I would start with the Minor Planet Center, if you're interested in understanding specifically asteroids, and just getting a feel for it. The JPL (Jet Propulsion Laboratory) Horizons System – they've recently spun up a few API's that make interacting with that data very useful. They are also providing additional information, including diameter, or maybe some other attributes that they're deriving about these asteroids. Their source data is largely still the Minor Planet Center, but there's additional information that you can get that people have provided. Again, they have a nice API and it's very well-documented. I know they’ve put a lot of work into that – the solar system dynamics group at JPL did a tremendous job pulling a lot of stuff together there.

48:46

Alexey

**I also know that TopCoder is a website, where one of the things they do is host data science competitions. Recently, maybe two-three months ago, they hosted two competitions from NASA. I think both were related to asteroids. So, check it out. Yeah. Maybe just look up TopCoder NASA competitions and I think you'll find some stuff there.**

49:16

Daynan

Yeah, if you're interested in getting into some of the more interesting machine learning stuff, tracking orbit determination is a challenge. If you think about looking at observations of an object over time, being able to link them together and say, “Well, these aren't separate objects. These are the same objects I'm seeing. Not only is it the same object, but I can now determine what its orbit is.” The process of linking those, traditionally you need several observations and you can verify it's the same object, but there are synthetic tracking techniques. There are things that people are looking at and exploring where you're basically shifting images in different directions and different magnitudes and trying to see if you can identify objects that have shifted in a consistent way that indicates that they're moving in a direction. As you can imagine, doing that over the scale of many, many, many millions, or even billions of images is not a human problem. That's a machine problem.

There are a lot of things that I think are pretty interesting in that whole arena. There are a lot of astronomers and scientific researchers who are doing elements of this. The nice thing about the scientific community is, if you're interested in playing around with this data, Google “scholar”, put in some terms and see what's being done. I personally find it fascinating to look up papers of things that people have done and try to replicate their methods and then try to experiment with maybe some more enhanced probabilistic models or throw in some interesting ways to do what they did and see if that changes anything. That's a great place to start. If you find anything, let me know – maybe we'll hire you. [chuckles]

# Mission and hardware design for asteroid mining

50:54

Alexey

**You also mentioned that one of the people on your team is taking care of mission design. There is a question from Matt, “How do you plan to interact with the folks going to get asteroids? Are you going to do this yourself? Is the person doing the mission design going to come up with a rocket? Or are you going to partner with somebody? What are the options there?”**

51:21

Daynan

Yeah, that's a very good question. Thankfully, there has been such an interest in space activity in general, that there are quite a few commercial companies, in addition to the public agencies, that are building spacecraft. We know famous examples like SpaceX and Blue Origin building rockets that are launching things into space. They're doing other things as well. But there's a whole ecosystem of companies that are building. We're pretty interested in Cubesats. They're usually fairly small satellites. Those are being deployed to do a number of different things. But everything from the instrumentation, to the propulsion, to the computing systems – there's a lot of companies that are assembling and building that.

We don't find that we're going to need to build anything ourselves in the sense that we don't want to be in the business of creating hardware. But we would certainly want to work with suppliers and vendors who are. What we need to do is set up our mission profiles so that we understand what it is that we need to develop. Some things will be off the shelf. So COTS (commercial off the shelf) technology is our philosophy as much as we can use that. It's cheaper, flexible. There's probably a lot you can do by flying your cell phone through space. But there are certain things, especially related to the idea of actually extracting material from an object.

We certainly have some very smart people at JAXA, NASA and ISAS who have done that, and we want to learn from them as much as we can. But there are also some aspects that we're going to have to R&D ourselves and we’re working with a couple of research universities to do just that and develop that. But, in general, if it exists as somebody making something or doing something that we can partner with or use, that's our preference. And in that sense, we're developing the architecture for the mission that's needed. If we have to, we'll look at developing ourselves, but that's the idea.

53:22

Alexey

**This mission would not only involve thinking about how to get to the asteroid, but also how to actually drill the hole in the asteroid and extract the water, right?**

53:32

Daynan

Absolutely, yeah. I would say that really, for most of what we're looking at, drilling is a problematic way to look at it because you don't really have gravity. You also don't have a very solid material that you're drilling into. You really need to think of a lot of these asteroids as being rubble pots. So if any of you are familiar with McDonald's restaurants, at least in the US, they used to have ball pits, where they had just these plastic balls that you could jump into and play around. You might need to think of these asteroids as being somewhat like that. It'd be more of a matter of scooping or scraping or somehow collecting that material. That's a major engineering challenge. It's something that we're really trying to think about. Maybe we'll find a larger boulder that we’ll pluck. I don't know, there's a lot of things that we're looking at that are related to that. But yeah, that's part of the overall R&D processes – how we get the material and return it. It's a major challenge.

54:31

Alexey

**I guess you cannot just travel there with a ship and then use a tractor beam to pull it.**

54:40

Daynan

We're not there yet, Alexey. I would like to think that we can be. If we are doing our job to build space infrastructure, God willing, we'll be doing things like that at some point in the future. [chuckles]

54:52

Alexey

**So what century is Space Trek? 24th century? Okay. We still got time, right?**

55:03

Daynan

[chuckles] We still got some time, yeah. You know, warp engines is a big one. We're already behind the ball on that one, I think. If I'm not mistaken, we're only a couple decades from needing to actually pioneer some early spaceflight according to Star Trek lore. So we gotta get on it.

# Partnerships and hires

55:23

Alexey

**[Laughs] There is a comment from Daniel. “Interesting conversation. I'm an astronomer currently researching a multi-wavelength counterpart.” There are some words I don't understand. “Good to know that there is a business case for asteroids.”**

55:39

Daynan

Absolutely. One of the things that I find really most satisfying about this work is that there are so many people, including the people we've hired, who are really interested in bringing the skills to use. Not to deride other domains. If you're looking at energy trading, or other things that are very, very quantitative fields, sometimes people that I spoke to have found that their skill sets in working with derivatives or doing this modeling – they don't have a home in the applied science areas. They apply them in other areas. Hopefully, we can be a home for people that are looking to put their skills to use in astrophysics and scientific computing, looking at some of these challenges. It's a need that we have, for sure.

56:23

Alexey

**I don't know if the name New light Technologies Inc. tells you anything [cross-talk].**

56:28

Daynan

Yeah. That’s a partner of mine. I knew he was going to join. Yeah, I used to work for New Light Technologies. Great, very smart, wonderful people.

56:36

Alexey

**They’re asking if you have a partner program in the works.**

56:44

Daynan

Good question. We should talk about that. We obviously are doing a lot of collaborations. At this point, it’s mostly with research universities and there are a couple of commercial organizations that we're partnering with, in terms of mission design. New Light, I'll make a pitch – you have a lot of really incredible DevOps data engineers, and some folks that are really used to working with enterprise data modeling. So yeah, let's talk. There may be an opportunity there.

57:16

Alexey

**There are quite a few things we didn't cover, so I'm just taking what else we missed. Here’s an interesting one, “What kind of mathematical models do you build?” I remember we talked a bit about that before this.**

57:36

Daynan

Yeah. I mentioned, I would say, that our overall approach to understanding asteroids is taking on the form of a Bayesian engine of sorts. That's a very general way to describe it. But we’re trying to understand how each additional piece of data or new methods that we use to improve – whether we're trying to understand the albedo, the reflective nature of these asteroids, or their orbital elements, or the spectral classification – there might be independent models that we're using to derive that. But overall, we want to be able to understand “Well, how does each new piece of information affect our overall understanding of this?” And the Bayesian framework is really well-suited for that. That can also help us quantify things a little bit more.

In terms of the individual modeling, some things are simple. If you look at the size/frequency distribution of asteroids, it generally seems to follow a power law pretty closely. With basic polynomial linear regression, you can model the size/frequency distribution fairly well. Other things like looking at albedo, for example, are more complicated because you're looking at sunlight being radiated from a body that is not necessarily a sphere. So thermal models are being used to do that and there are certainly folks that are doing that research and thinking about that. Some things there are getting more complex. People don't know this – there's something called the Yarkovsky effect, which essentially states that the thermal radiation from an asteroid, based on reflected sunlight, is enough to actually push that asteroid. It’s over the course of many years, of course, but if it's a small enough asteroid, that's a meaningful force that's being pushed on that asteroid. How the asteroid’s rotating and what its surface looks like has a lot to do with that.

I touched on a couple of deep learning and machine learning techniques, certainly in terms of vision and object detection. Spectral analysis is really, really big. It's gonna be very important for us to try and understand the composition of these. And as I mentioned, there's just not a lot of observed data. We're really going to need to think about where we can extrapolate the known data and apply it to objects that we may know something about, but we don't know their spectroscopy, for example. There's certainly some modeling that's going to need to go into that. Yeah. We're dealing in probabilities here. A lot of it is just how we can define those probabilities in ways that make enough sense for us to make decisions. That’s the name of the game, right?

60:11

Alexey

**Yeah. I prepared another silly joke about Jupyter Notebooks.**

60:19

Daynan

[laughs] Jupiter, yeah. Takes on a new meaning when you're using Jupyter notebooks to model Jupiter orbits, right? Yeah. Well, I don't know if you have a joke there. I mean, I use notebooks. They're useful for telling the story of data. I don't develop in them because I think it teaches bad habits for developers, myself included. It takes on a little bit of different meaning.

60:46

Alexey

**[chuckles] Yeah. I think we should be wrapping up. Thanks a lot for joining us today, for sharing your expertise and a couple of stories. So thanks for sharing all that you shared with us today. Thanks, everyone else, for joining us today, for watching us, for asking questions. I guess you will give me some links and I will post them in the description.**

61:08

Daynan

Yeah, I'll do that. Definitely including my dcrull@karmenplus.com. And I'll send you links. But yeah, certainly if you're doing research in this area, I want to talk to you. As we're growing our team, there are a lot of areas there. Certainly we're looking for data engineers, as I mentioned. We'll be looking for some other data-focused folks, for sure. But, hey – let's talk and see what happens.

61:35

Alexey

**Yeah. Thanks a lot.**

61:37

Daynan

Yeah, thanks a lot. Have a great weekend.

61:39

Alexey

**You too. Good bye.**