**5\_31\_2020\_data\_team\_pdap\_kickoff**

[00:00:00] **Goose:** [00:00:00] And , so Sid has been helping me put together an agenda for this, but, basically what you'll see here is we have two high level items we'd like to do intros and discuss member's skill sets because, though, this is an open source project and we are probably going to see a lot of people come and go.

[00:00:21] The fact that right now, in the early stages, we have folks who are making introductions and, getting to know each other go a long way towards capturing some of that momentum. So, I'll be taking notes during this meeting. So, I'll mute myself. I do that cause my keyboard is loud, but what we'll do is just, go down the list that I have here in front of me, based on Discord and we'll do introductions.

[00:00:43] I'll start with myself. I have roughly 10 years in the IT industry. Mostly in systems administration, design architecting, with the last couple of years being focused in cloud development and, data and AI architecture. And, the skill sets I [00:01:00] think I bring to the table are going to be some development skillsets.

[00:01:03] C# no JS. So, a little bit of Python as well as a cloud design architecture when it relates to, systems and data. So why don't we go down the list and go with Austin next?

[00:01:18] **Austin:** [00:01:18] Sure. Thank you. Can you guys hear me?

[00:01:21] **Sid:** [00:01:21] Yeah. Yeah. Sorry. (unintelligible) Putting the loading icon on. Is anyone else getting on the screenshare?

[00:01:32] **Nick:** [00:01:32] It sounds like your connection's not very good.

[00:01:35] **Sid:** [00:01:35] Let me, let me see what I can do. Sorry. Continue.

[00:01:42] **Austin:** [00:01:42] Yeah, my connections. Good to hear. I had to wait a little bit for the loading icon. My name is Austin. I'm a senior undergraduate student in software engineering. I'm hopefully going to end up graduating at the end of next year.

[00:01:54] So I just have some university experience as well as, some basic internship experience with [00:02:00] relational databases, some algorithm experience. And in the past, I have some web scraping and data analysis. I did undergraduate research for the past couple of years, and we've done some machine learning, et cetera, but my skillset specifically for this group, it's probably going to relate more to data transformation and, as well as some relational database stuff, because of my lack of actual work experience, I think I'd be more suited as a worker for this project.

[00:02:27] I think I’m a creative person, but I'm going to leave a lot of the creative stuff up to other people. And if somebody needs something done, put it on me and I will get that done for you.

[00:02:40] **Goose:** [00:02:40] Excellent. Great to hear, Blake, let's go with you next.

[00:02:50] **Christina Blake:** [00:02:50] Hey, can everyone hear me alright? Great. Christina Blake, I currently work as a marketing strategist. So, a lot of my experience [00:03:00] is going to be limited to data visualization, but I do have some intermediate SQL, and data transformation. I also currently work as a project manager bridging the gap between my data engineering team at my current company.

[00:03:14] As well as the end users for our internal software. So, I can probably fill into a lot of different spaces for this project, just depending on where the needs are.

[00:03:29] **Goose:** [00:03:29] Excellent. That sounds great. And Joe Good.

[00:03:45] Yeah. If you don't, if you don't feel like going with an introduction, you could always just say "pass". We'll go ahead and move on to Joshua L.

[00:03:55] **Joshua L.:** [00:03:55] Hey guys. This is Josh. I'm currently leading the [00:04:00] Operations team. I do have about six years’ worth of industry experience, doing a lot of backend work, but I do have full stack capability. I've done a little bit of everything, from stuff with Kubernetes things. Yeah, there's a lot of stuff.

[00:04:16] I don't have a full, like depth in a lot of different technologies. But I have touched things and I know what questions to ask. So, I don’t know if I can really contribute to the product team in general at the moment, but I would like to one day eventually, so.

[00:04:36] **Goose:** [00:04:36] Cool. Yeah, Nick.

[00:04:41] **Nick:** [00:04:41] Yeah everyone. I am a web developer. So, I've been freelancing since 2012 for variety of different local businesses. Until around 2017, I got a full-time position doing web development. And that's a full stack [00:05:00] position. So, we have gone through a lot of transitions there.

[00:05:03] We started using PHP. We've transitioned to C#. Some Java, recently started looking into Rust and Golang, and always like frontend work. (unintelligible), but I have a more general, kind of like Josh does, just generic experience with a lot of different technologies. A few of them, I would consider myself fully proficient in, but I'm not saying that I know everything I'm saying that I know enough to go by.

[00:05:29] I’ve done some fielding for data types and data architecture that needs to be in place for projects, for a few different clients. And full time every day at work. Yeah, that's about it for me.

[00:05:43] **Goose:** [00:05:43] And Nick is the product team lead, correct? Yeah.

[00:05:47] **Nick:** [00:05:47] Sorry. I'm the product team lead should have mentioned that.

[00:05:53] **Goose:** [00:05:53] I'm sure just based on- organically people will just start to pick up who's doing what. [00:06:00] Cool. And then, Shredding.

[00:06:06] **Casey Shredding:** [00:06:06] Okay. Hi, I am Casey Shredding is my name I go by in the roller derby circles and I forgot I had as my username here, but stick with that. I am a web developer at work, usually doing things with various types of JavaScript, sometimes interacting with various APIs and databases.

[00:06:24] I've been doing that for a couple of years now. I have an education in physics, so I'm pretty good with abstract and out data models and the like, and I support IT in dev office type things as well, but exactly where I fit into what we're talking about today and how much time I have to put into that, I don't know. We'll see. But I have a pretty comprehensive technology and general background and tend to be pretty good at picking stuff up. That's who I am.

[00:06:53] **Goose:** [00:06:53] Cool, welcome. We got a local roller derby team. That's fun. Sid [00:07:00] we can't hear you.

[00:07:09] **Sid:** [00:07:09] Right. My name is Sid. I have been working in IT consulting for two years. Before that, had more of an engineering background in college. So, more on the chemical and petroleum engineering side. More to Casey, I think I can think pretty abstractly about data models and things like that.

[00:07:32] And for the majority of my time at this IT consulting company, I've been working with some like adjacency with sharing and data modeling. So pretty comfortable with SQL server, Python. And most of my work has been on Azure. So, data fact, three blocks. [00:08:00] Details, Azure-based workflows for data transferring. So that's a little bit about me.

[00:08:12] **Goose:** [00:08:12] Awesome. And Watchman, can't hear you.

[00:08:31] Yeah, we're not picking you up.

[00:08:39] Joe, could we offer you another opportunity? If you want to introduce yourself? You don't have to, but sorry to throw it out there.

[00:08:46] **Nick:** [00:08:46] Yeah, it looks like he's having mic issues.

[00:08:49] **Ray:** [00:08:49] Yeah. Okay. So, I'm Ray. I am in software development in (unintelligible). Now I would say it [00:09:00] will be more than 15 years and have been doing pretty much everything that has done: IT, software architecture, mostly doing architecture-type work today and have done backend work, frontend. And add PowerScript, C ++, PostGres as well. So recently we've been playing around with, gene protocols, like Kafka. I've been involved in the automotive industry for a while, and that's what I do for a living.

[00:09:36] **Goose:** [00:09:36] Awesome.

[00:09:40] I scrambled really quickly. So, if anyone, this isn't the formal introduction and I know that there is a formal process starting for organizing members of the project and what their skill sets are, how they can contribute. So, it seems like any of this has kind of like incomplete or just poorly captured by me I [00:10:00] apologize. Feel free to hit me up. I won't commit it right away if you want to make some minor edits, but really this isn't the formal document of who's who and what they can do. So, don't worry too much about it.

[00:10:13] **Nick:** [00:10:13] No, it was good. Thank you for taking notes.

[00:10:16] **Goose:** [00:10:16] Sure. So, I'm going to take an official roll call, which doesn't have to be done cause, everyone's here and present. We could probably formalize that process, but I've captured it off to the side. So, I'll add this to document afterwards. Just think that's important as participation is important in a project like this. And then, so let me move on to the technical discussions. So, Sid and I-go ahead.

[00:10:43] **Sid:** [00:10:43] I missed like the first 30 seconds due to technical problems, but we are recording?

[00:10:49] **Goose:** [00:10:49] We are recording. I didn't say anything. Nothing of substance. Great. I will definitely [00:11:00] pass the mic around to individuals who want to speak as this is a small group. So, I think this should be pretty easy to do.

[00:11:06] I was expecting a little bit of a greater turnout, especially as there were a lot of strong voices on data architecting and design early on, like Wednesday and Thursday. But we don't have a ton of information today, right. So, one thing I didn't notice: is there someone from the scraping, you know, "scraping team" on this call?

[00:11:28] **Ray:** [00:11:28] I think I would be that, I don't know.

[00:11:31] **Goose:** [00:11:31] Okay.

[00:11:32] **Nick:** [00:11:32] I think there is going to be a lot of people. Right now, we have no formal recognition other than leadership; or who is on what team; or how the teams are being necessarily conducted. So as time goes on, that'll be solidified. But for the most part, I see that maybe we would have a core team of people that would be doing scraping. And once like a definitive guideline for contribution is in place, I think we would actually get a lot of open-source [00:12:00] contributions instead, or people just issuing PRS essentially. so, for the time being, I guess we're just kind of solidifying, "what are we looking for", as opposed to "who are we delegating" that these tasks are being assigned to.

[00:12:13] **Goose:** [00:12:13] Sure. And that makes sense. And with the current structure with leaders above certain groups, I suppose something like this is going to maybe filter up to you then in the short term, just as things that are of concern that need addressed.

[00:12:28] **Nick:** [00:12:28] Yeah. So, for right now, the dust is still settling and we're still getting everything in place like as an organization is concerned. I think the most important thing we can take away from this meeting is what do people actually want to store? Like what data is actually important to those of us in this meeting, and leadership going forward, and community feedback about what people want to see, within our data set.

[00:12:54] **Goose:** [00:12:54] Sure.

[00:13:04] [00:13:00] **Nick:** [00:13:04] But yes. In addition to that, we do need to discuss some of the specifics and what we think are best practices to doing some of these things as regarding distribution storage or, maybe redundancy or more complex topics as well.

[00:13:19] **Goose:** [00:13:19] Sure.

[00:13:22] Yeah. Well, you know, Sid put together some of these bullet points and I can address them as to how they are impactful for the overall vision of what direction we ended up going with and laying down a data landscape. And some of the top three here are definitely important and we'll need clarity on that. And I know that may be difficult until we start getting data collected. And maybe we could discuss on this call where would be a good place in the short term as folks start producing data where we can land that.

[00:13:56] Probably likely be very messy operation from the [00:14:00] start, but it can be polished. But really, we're focusing on like Nick said, what are we storing? What's the content of the data, the fields and the types of data, the speed and size of data? Right now, we know that there are these systems out there where we can access this data.

[00:14:16] How far are we looking to go back? How far are we able to go back? What's the volume of data we're actually talking about here and the philosophy that the data will be entering into the data landscape? And then the delivery medium or the delivery mechanism for the scraping folks. And on the periphery-on the other side, whether it's the APIs for programmatic access to the data or, APIs for a app.

[00:14:46] And just to kind of highlight this piece again, just because I think this is really important with, as far as information that we've gotten so far out of leadership, which the video you guys posted was awesome; was really [00:15:00] transparent in , the document that came out of that. So, this really picks up on a couple of key points for me, which is, you know, we want publicly available library of records.

[00:15:14] We want to make sure it's easy to use and to use open source software analytics against that data set; provide an interface for reporting and analysis. So, based on what some other folks have said in chat, that does sound like maybe a web front end with some visuals, some of those items, APIs for programmatic access to the data.

[00:15:38] So that's, that's big. There's a lot there for an MVP. So hopefully we can, at least with the data landscape, focus in on what would be the quickest wins for us and considering the level of funding we have right now, which is zero. Which hopefully that'll get better over time. [00:16:00] So we don't again, without having anybody to speak to the data itself and we haven't seen it yet we have to consider a high-level data architecture. I know Sid and myself are fans of the modern three-tiered architecture. You know, we have a data lake that we're ingesting data into. We're performing those ETL operations against the data. That way we can manicure data sets with different schemas, different data, maybe we're excluding data. Maybe we decide we want a flexible schema. So, we're accounting for that flexible schema in our transformation pipelines. And a data house, a data warehouse that can support those flexible schemes.

[00:16:41] **Austin:** [00:16:41] Just if I can cut in, I'm an advocate for this, just with my university experience here, I think that this is a good way to go for everyone to be able to get involved.

[00:16:51] I think if we had anybody here speaking from the graph database perspective, we could theorize more on how the data warehouse would work, but [00:17:00] are you guys also just looking relational for the data warehouse?

[00:17:05] **Sid:** [00:17:05] Yeah, I think it's-and correct me if I'm wrong-but I think it's pretty accepted across the industry if you're gathering all this from a bunch of different sources, the advantage of having this data warehouse is that you get to normalize it, sorry, not normalize it. create a usable semantic layer on top of the data, which is easily accessible.

[00:17:39] Oh, very familiar with graph databases. But that might be something that we can have towards the end. Like if there's a different way of storing them, we might make that available to the end users through an API.

[00:17:58] **Goose:** [00:17:58] Yeah. I think you raised a good point, [00:18:00] which is, especially when you start talking about closer to the analytics engine that we'll have. There are a lot of options there that we'll need to have more concrete facts in front of us before we can make that determination. Like I, myself, I'm envisioning every one of these web interfaces producing some sort of comma separated data and who knows, I guarantee they all have different schemas.

[00:18:28]So then you start getting into the discussion of all right: do we want to attempt to define that schema in a common model through our ETL pipelines, or do we want to allow the database and then what data warehouse to grow into a semi-structured state where, folks who, when-first of all, if we decided to add new fields later, new columns later, we aren't stuck with restructuring the schema and we allow flexibility within our analytics.

[00:18:56] If we start creating data sets off to the side of the primary dataset, [00:19:00] it depends on-I mean, I think once we release this data out to the "scientists", the data science community-and there are a few communities out there where if we were to have this data ready to go, that people are going to jump on it- I think we could find that all sorts of interesting use cases pop up. Does anyone else have any ideas or talking points on the high-level data architecture, maybe big data sets or big data stores that they've managed or interacted with, or even just an idea of what it should look like to them.

[00:19:41] **Ray:** [00:19:41] That's in data ID of sources, is that stored and then processed in real time and or large (unintelligible). It's good for that because it has a lot of storage space, (unintelligible). As for being, [00:20:00] having a homogenous dataset or maybe fields that-there are some things which are going to be common across all of these cases or cases have a way to be represented and all that data is going from across all of them. It can have that data in defined. So then everything else is linked to something else like in blob somewhere like that would make it easier for processing at some other end.

[00:20:38] **Nick:** [00:20:38] I'm in favor of the tiered data architecture or structure. We have concerns right now about what we would store, even in a private matter. If we can store anything in a private matter, as far as our eligibility for nonprofit status. Not that I personally believe that there's anything we could store because it's all [00:21:00] already publicly accessible.

[00:21:01] But I am not a lawyer and I am not allowed to make those legal decisions or make these decisions without having a firm legal advice or legal understanding of the implications. So, I think at the very least we could get something in place that is storing non-PII. I know we would inevitably not want to manage or modify schema too heavily going forward, but I think to some extent it might be inevitable. If anybody has thoughts on that? I'd love to hear it.

[00:21:33] **Ray:** [00:21:33] Names of the parties will be there. That is the personally identifiable information, which is very hard to take out.

[00:21:43] **Goose:** [00:21:43] Well in the world I imagined would, the initial ingestion actually would occur in a private data set and that we're only publishing the public dataset after we've performed ETL transformations to remove that PII [00:22:00] data. Cause I do agree. We don't want to put that. I definitely wrestle with the pros and cons of both. On one side, you want to be completely open and this is publicly available data, but on the downside, there are going to be third parties who use this data for means in which we did not intend.

[00:22:19] **Austin:** [00:22:19] Well, I don't remember who I got in a conversation with on Slack on last week, but there were some FOIA requests being sent in for Arizona public court records. And, I think there are stipulations stated, we are not allowed to release on data that we receive through a FOIA request. So that system in general, so I think there's some legal perspective into it, but I don't think-as long as we remove PII, I don't think that should impact our schema.

[00:22:50] **Sid:** [00:22:50] One of the advantages of having this multi-tiered architecture is we can decide at which point these rules start to apply. Going [00:23:00] back to what Goose just said. Like, basically I view the data lake as just, you know, the, the scrapers scrape their data and they just dump everything into the non-structured or whatever native format they have.

[00:23:14] And then we can parse through and decide whether or not we want to strip out certain things. But by the time it gets to the delivering mechanism, it's going to be highly processed and we'll have control over that. But if from a PII perspective, we are not even allowed to have that data on our servers or whatever. And that's also a different issue.

[00:23:37] **Joshua L.:** [00:23:37] I was actually just going to mention that. I don't know what the legality is around even just having that on our data warehouse.

[00:23:46] **Sid:** [00:23:46] I mean, not even on the data warehouse. The data lake, or if we're not even want to have it. And yeah, we're just going to have to figure that [00:24:00] out.

[00:24:03] **Austin:** [00:24:03] I think it comes between scraping how we receive the data though it's scraping and getting started through something like a FOIA request. If we scrape it, I can see that, but if we get a FOIA request for a mass data set, we are, I think if we can keep it in a data lake that's not publicly accessible with no repercussions.

[00:24:25] And obviously some legal representation would have to verify that. But what I'm thinking is even with the data lake, like if we have some highly normalized set of data, or a highly normalized schema, I think that we can limit access to PII. As long as we're legally allowed to have that on our, in our warehouse in general, we can limit that through API calls.

[00:24:52] So. I really just want to make sure we're not throwing away data here.

[00:25:01] [00:25:00] **Goose:** [00:25:01] Yeah. It's interesting. There are a lot of, there are so many angles to what we're trying to accomplish here, even just from the data storage perspective, whether it's from the legal perspective, the security perspective. That's a concern I have down the line, which is: okay if we open up API access to folks publicly, how do we prevent being DDOSed? I mean, those API calls, those queries to our data sets are going to end up costing money. And as a nonprofit organization, like that's going to be, those are going to be precious resources, especially in the beginning, which is coming back to the three-tiered architecture.

[00:25:35] Why I like it so much is because we can create the initial, data lake with several areas of just raw data and curated data that's cleansed of PII. We're just offering flat files at that point. And that will be, and that's how you see if you Google, public datasets. Most time they're being offered in flat files and ends at files.

[00:25:58] You know, if you take this one, [00:26:00] for instance, you know, they tell you a little bit about the data. And then if you go into the data folder, they're actually just giving you a zip file. So again, we're going to have a ton of data, but, you know, this would be an easy way to get started that we're providing for all flat files to the community of scrub data that they can go through and grab and perform their data science activities against. But I do think from what I've heard, the end vision, the grander vision is something polished, something that has an impact on the community outside of just providing the data, via public APIs and maybe even a

[00:26:37] web front end.

[00:26:40] **Sid:** [00:26:40] So I'm thinking there's probably a couple of different consumers here, right? So, one is the public APIs, who are just downloading CSVs or whatever. And in that respect, I'm thinking [00:27:00] like, why do they actually need a data warehouse? A fully fledged data warehouse if they can just get all the stuff that we dump into the data lake.

[00:27:12] **Austin:** [00:27:12] I'm sorry. If, if we're just dumping flat files to everyone, this is kind of why I jumped on board. There's no issue with just releasing flat files from a perspective of making it easier for us, but it doesn't make it easier for them. And if we are the people with the skills and our role is to be helping them, helping the data science community, we should do what we can, right?

[00:27:40] **Sid:** [00:27:40] The data science community. I mean like, so the way I've used data warehouses, you can hook them up to Tableau or power thing or whatever. And then the data warehouse becomes more useful because the queries that you were [00:28:00] executed getting against the data warehouse, (unintelligible). So that's one example. And then the other is if you have active direct connections to the data where they're running queries against it. And so, there are physically joining all these fact and dimension tables together to provide, to (unintelligible). So, are we going to be, or are we going to have those types of users?

[00:28:25] Are there going to be people just hitting our database and creating those joints and things? Or is it more going to be people who are consuming processed data in the form of (cutout), because right now, I don't think we have many of the former, right. But if there was like a separate entity, like the analyst group that were doing their own analysis and AI and modeling off of our data, then it would be really [00:29:00] warehouse.

[00:29:01] I don't know if that's in the, in the landscape anywhere if that's been discussed.

[00:29:06] **Nick:** [00:29:06] Well, I think that's definitely a possibility, with, with us, either having like a tiered access, I'm assuming we're comfortable, depending on the dataset included and where, or what the data set is. We may, allow access to maybe other nonprofit groups, to be able to view the data lake or things like that.

[00:29:25] I, I can't say with any certainty that we would do that, but I'm saying right now, nothing is certain. And that has maybe a possibility that we would have users that do want this raw data as it is for ingestion, for whatever way they see fit. But I think you are correct that the majority of people would want to see sifted through data and data that's been curated to, to exactly what they want to see out of it, instead of all the extra stuff.

[00:29:52] **Goose:** [00:29:52] And I think that's why the three-tiered architecture is so powerful because it allows you to grow into from simple, [00:30:00] simple, raw storage access all the way into advanced analytic workloads. So, I think it services the whole gamut, I think, to Austin's point, I definitely agree that we need to focus on making things easy. But what I would say is, I don't know if you guys have ever been on Kaggle, but this is kind of the de-facto social, social forum for a lot of data scientists who are looking for public data sets to working on public data sets.

[00:30:26] So, I think when we start talking about personas, there are plenty of people out there who will be just happy with accessing our raw data. And when I say raw data again, I mean, curated, no PII, like data that we're, we believe is ready for public consumption, but there is going to be - the three tier architecture allows you room to grow into, APIs for analysts seeking a different experience.

[00:30:52] Maybe people trying to add data sets into their own processes or their own projects or a web front end that this [00:31:00] project wants to put out there to be, you know, the face on the web for what we're doing and to show stats that scientists involved internally with the project have a surface to individuals.

[00:31:12] So I guess what I'm saying is the three-tier architecture allows for all of it. And I think it leaves room for us as a community to determine how we want to surface that data. And it gives us, all the options and how to, how to surface that data.

[00:31:29] **Sid:** [00:31:29] The other issue is access and how that jives with our mission statement being of like, you know, nonpolitical and open, right? The idea is we just, we just give you the data, you can do what you want. The data for us, we just give you the data. If that's the case, then does that mean we are not doing our own analysis, right. And if we're [00:32:00] not doing our own analysis, then we should be mindful or guiding ourselves towards the, the persona of someone who's just downloading processed, not raw data, but like processed data from our (cut out) , and again, I completely understand your original point, which is the, the architecture gives us flexibility to do whatever, but I don't know. I just thought I'd make that point. So maybe the focus shouldn't be on the data warehouse and storing it in like Azure data warehouse or Amazon Redshift or something, but process that data, clean it up and get it ready to use. And then, you know, we can put in the S three bucket or Azure blob or, or an actual data warehouse. The point is that the data is clean.

[00:33:01] [00:33:00] **Goose:** [00:33:01] Agreed, and I think that's a great first step for us when we start talking about, working towards that MVP, I mean, step one is going to be, get those raw files somewhere and step two is going to be okay. How do we want to curate them? And a really quick win as this project is getting off the ground, a quick win, if, if things appear stagnant from the outside is to provide those data sets publicly. And I think the whole idea, at least when I see, you know, police state accessibility project is we want to make sure that this data can get out there. That is definitely in, in this, you know, in this statement here.

[00:33:38] And I definitely agree with ease of consumption. But step one, low hanging fruit. We can get that data out there for consumption by starting with flat files. Okay. Right.

[00:33:50] **Austin:** [00:33:50] I did. I did just, I wanted to clarify here on, so, I mean, this is mostly just the data architecture team, right? This is what we're working with at the moment.

[00:34:01] [00:34:00] **Goose:** [00:34:01] That's a good question. I think there's still some ambiguity around teams and structure of teams. I think it's going to be one of those things that kind of solidifies as time goes on.

[00:34:15] **Austin:** [00:34:15] I think if we can just work towards solidifying that now, though it will be helpful in the end, I believe is that we need to focus on what exactly we are doing as a, as people who are involved with data architecture here. Is that while I agree with, with, Sid's saying you know, there are going to be different personas here involved, and we need to, you know, as first steps or just get the first two tiers set up pretty much, is that we, I think as data architecture, we just need to get things clean and organized.

[00:34:52] We need to have certain schemas set up and our process set up. And what other, whatever other teams in [00:35:00] police data accountability project, organization, what to do with the data, you know, statistical analysis, if we can create methods for them to use the data, that's great. And if we can create ways to release, you know, raw data without PII, and in a cost-effective way for us, then that's great and that's what we need to do. But as we got to make sure we focus on data architecture here; you know what I mean?

[00:35:25] **Ray:** [00:35:25] What does it say right here that RC and the foreman about the PII is that what an auditor looks. Yes.

[00:35:37] **Austin:** [00:35:37] It depends where it comes from. if we're doing FOIA requests, like many, many people are starting to do, I think there could be a mix of CSVs and PDFs actually. You know, and then whatever scrapers tend to use to scrape their data.

[00:35:52] I don't think we're starting from ground zero. I think, you know, we have the data that [00:36:00] Kristen put together for South Florida, whatever jurisdiction. So, for instance, the MVP word, it does get that into it. They want to model it and I don't know, create some sort of backend delivery server. You know, it's something (unintelligible) go over the specifics. I think Austin's correct that we can make progress without knowing like all the specifics.

[00:36:32] Right, right. And my worry is that we, we, if we create something where it's not going to be extendable on, I'm not very good with words sometimes. So, I'm sorry if that wasn't the right word there.

[00:36:41] But, all of these different organizations that, you know, the different portals and different counties and public jurisdictions are going to have different formats for data. So if we can create something that is, you know, I think our biggest problem is going to be transformation, is we can get, we can get a great [00:37:00] warehouse set up as long as we create a heavily normalized schema that works for a vast amount of different, jurisdictions data. And I think we can do that without their data. if we just, you know, think of what we could possibly get and if so, maybe some notable values that some places won't have.

[00:37:19] **Goose:** [00:37:19] Yeah. I think we're all talking about the same things here, which is we want to get started with grabbing the raw data, curating the data so that it's ready for consumption. and then I think where it's starting to get a little foggy is where we think the next step is. And. We definitely don't have to make a determination today.

[00:37:38] there’s still a lot of work to be done to clarify that picture, I think as far as what we're going to be doing on the periphery. But, yeah, I don't know, from Nick or Josh's perspective on next steps for this project as a whole, or maybe for product team members as a whole, what you see those being. [00:38:00]

[00:38:00] **Nick:** [00:38:00] for the product team in general, I initially posted what I thought at least, a rough draft of what the teams would be. And I really had a time envisioning exactly how, data people fit into everything. And I'm glad we're here having this discussion because it really solidifies my, I don't know what the word for it is, intuition that we need a team dedicated to data and architecture and structure, within, within legal reason and within, bounds of the scope of the project in general.

[00:38:33]so I, I would like to at least make a formal team of people that are going to be denoted as data engineers. and if you, if you're here listening right now, if you're listening to the recording of this, if you could send me, Nick Frost, a DM and Slack, I can put you down. when we have formal member intake, set up and make sure that I know that you're interested in helping [00:39:00] with this aspect of this project going forward.

[00:39:05] **Joshua L.:** [00:39:05] Yeah. And, I think some aspect of this too, is sort of just maybe starting out with something and, you know, through that exploration, we'll find out more of what we need. cause yeah, it is a little harder to sort of grasp exactly all that we need to, you know, get everything going. But I think a little bit at a time will help a lot.

[00:39:25] but yeah, that's really all my comment is.

[00:39:28] **Nick:** [00:39:28] Yeah. so, we're. Yeah. We're going to get people together and we're going to try stuff out and we're going to fail. We're going to make mistakes and we're not going to know everything right away. And that's okay. what matters is that we're trying and we're learning and we're going forward.

[00:39:44] **Goose:** [00:39:44] Yeah. And I agree. And I think, the fact that people join the call and that we're talking shows that people are really. You know, these are the kind of steps that I look at as, okay. Somebody dedicated to putting forth some time and effort towards something. and I've really liked [00:40:00] what, what Sid said in the database general group, which is, you know, a lot of us are going to be new to this decentralized collaboration format, but I an attitude of generosity, respect, and humility will get us really far.

[00:40:12] And I think as we continue to have technical discussions about, how these things can be implemented and. you know, there are going to be disagreements on things, but, I think we're all here, operating the same spirit that, we're doing something cool and something for the greater good with our spare time.

[00:40:29] but

[00:40:32] **Ray:** [00:40:32] any of you see that we would have to be in sync with the data that is available on the sites. Example, certain records are gone and we used to have them do we also. I'm off. Considering the historic nature of this, for example, we did, we start today. somehow, we can get access to 1908 and same goes [00:41:00] by say 2020. And there's only 19 data are available on website. Hold on. Can we still keep it?

[00:41:08] **Nick:** [00:41:08] I would presume. So that might be a legal issue, but if I understand what you're asking correctly is that like, let's say there is a record from 1990 or so, and we have records of it and then the local jurisdiction or government decides to get rid of that data because it's not in whatever policy of theirs to maintain records older than 20 years or whatever.

[00:41:29] Right?

[00:41:32] **Ray:** [00:41:32] Something like that. Also say for example, there have been some changes. In the record that was altered and then be scraped it how often do we have to keep the (unintelligible) backed up. Are we doing like a snapshot of today?

[00:41:48] Well, that's where the Sid?

[00:41:52] **Sid:** [00:41:52] Yeah. I mean, it depends on it all depends on my incoming data source. Right. Like if there's a reason they took [00:42:00] that data point out. Right. Is it just because, you know, they don't have space on their database, so, you know, they only get the last three years where the rec (unintelligible) back is? Yeah, we should just keep it right.

[00:42:11] Or is it, you know, they, they changed something, right? Like they were like, Oh, we released this guy. So, he's no longer in our books. And again, I'm wildly speculating here, so that might not be applicable. Then we would make the corresponding change in our data warehouse. Right. So, we would say, okay, this person X was released.

[00:42:34] So we would update that. there was something else, but I forgot.

[00:42:39] **Goose:** [00:42:39] And without knowing how reliable the process of ingesting scraping ingesting this data is that is something we could certainly account for. And we can also create versioning with our datasets to say, as we change the data set, we have versions that exist previously. Now, obviously the more changes that we [00:43:00] commit and the more versions of those datasets, we create the cost of storage and as well as cost of, the, you know, the warehousing is going to increase substantially,

[00:43:09] **Sid:** [00:43:09] yeah, and this is a very good question for the data scrapers. Like what is their strategy for scraping data?

[00:43:16] Are they going to scrape all the data every day for every website and. Are they just going to rely on us to update the things or are they going to somehow filter, just new data? or is it going to be just...what if it's not as scraping, right? Like, it looks like there's FOIA, right, which is more manual.

[00:43:41] Okay. We just have like batch loads of data. And so there there's a lot of variables here.

[00:43:50] **Austin:** [00:43:50] Sorry. Sid, thanks. so, I don't know if something that anybody else has experienced with here, but if we're worrying about changing data, which is something we should be worried about, there's a, there was a guy who [00:44:00] was in the database channel on Slack.

[00:44:02] I don't think he's there anymore. His name's Eric w if you go look at some of those messages, they actually make sense. And it fits in with the data warehouse perspective. It's called the data vault, perspective. I don't know if anyone's heard of that. it just gives us some satellite tables to record changes to data.

[00:44:16] So we don't, again, we're not throwing away any data. but we're maintaining changes. It does make, the data scientist job a little bit more difficult because they have to look for changes in time as well. but it is something if we are worried about changing data, it is a good perspective to have.

[00:44:34] **Sid:** [00:44:34] Yeah, I took a look over that, that article, it looked, it sounded like a slowly changing dimension. And then I went into like some of the related links and there was a huge kerfuffle about data vault versus dimensional modeling and Ralph Kimball and all those sort of stuff

[00:44:53] **Austin:** [00:44:53] yeah, it is. It's, it definitely does not make our database prettier.

[00:44:57] I will say that.

[00:44:58] **Sid:** [00:44:58] Yeah.

[00:45:00] [00:44:59] **Goose:** [00:44:59] And I personally not to get into. Go down this rabbit hole, but I think that's a little bit of an outdated method for data versioning. We have things now called Delta lakes where we're applying metadata directly to the file. and it's just, it's smooth and you're still existing within the flat files of your raw data.

[00:45:20]so. As far as tracking changes to the data the data within the, if we go with a relational database system or a semi structured system, there's umpteen ways to track the versioning within those. But when we come down to the raw file level, there's, there are a lot of good options for us now. And Delta Lake is one of the leading ones in the industry.

[00:45:44] **Austin:** [00:45:44] Great. Thank you. I'll take a look into that.

[00:45:49] So to clarify, do we have example data there is that reference or to that Florida scraping, is that data actually available to look at?

[00:46:01] [00:46:00] **Sid:** [00:46:01] That's a good question. And that was my, that was going to be my question. Like it's got to be somewhere really. Kristen has access to it, but I don't know if we, as a group have access to it.

[00:46:12] **Austin:** [00:46:12] We have access to the raw criminal data the, the felonies CSV and it does not include everything that was displayed in the original lawsuit.org posts that many of us probably saw. but there is some example data available. Yeah. As well as there's the Bruxer, Texas County data website, where they make it all available in the first place, and you can download CSVs for any date.

[00:46:35] So my point with verifying that I know it's not going to be exact, but it means we do have something we can play with and start with to get a concept together, proof of concept.

[00:46:47] Right. And it's also something that other teams are working with. Maybe Nick can verify that, but I know that the data analysis team, the people who are looking forward to working with data analysis [00:47:00] are looking at that example data and running some example projects. And I believe that scrapers are looking at it as well.

[00:47:09] **Nick:** [00:47:09] at the time I can't clarify what everyone else is doing. I can only clarify what I'm doing.

[00:47:21] **Ray:** [00:47:21] there's a question about scraping as I can see a task to be a scheduled job. It's going, files in the queue. we might not be scraping the whole site. We might be scraping, (unintelligible) looking things, just a dime. So, there is a possibility we might have some decayed (?) data which might get removed. It's not going to be like being scraped every day.

[00:47:55] three days. Or four days,

[00:48:00] [00:48:00] **Goose:** [00:48:00] I think you bring up an interesting point, which I had not thought of, which is. And I did make a note of it here, but it's like, what is a record, what if a record is updated and that changes the outcome of the case. So that would come with how somebody would perceive the data. And I think that's probably something, a worthwhile takeaway, to make sure we have discussed.

[00:48:19] Because it goes right into what Watchman is saying. how often are we going to scrape the data? Do we need to go back because there's something could have changed with the data set that would impact the quality of our data, the way that, outsiders are viewing the data that we're providing?

[00:48:35] **Sid:** [00:48:35] I think it, I think we're speculating. I think the most important part here is we need to have some sort of connection with the data team so that the scrapers (unintelligible) and the sooner that happens, the. The sooner we can get started and get answers to some of these questions that we have.

[00:48:59] I agree [00:49:00] they would be the most intimate with the data that they're handling.

[00:49:02] And I understand the challenge and sort of trying to make all of this fit into a singular sort of schema. But I think as like a, and I'm sort of pushing towards the MVP (?) conversation, you know, I think this would be an easier conversation if we had, if we started out with a single data set for now. And as we move forward and build out our infrastructure, I think this'll be an easier conversation.

[00:49:35] **Goose:** [00:49:35] Agreed. Yeah, it does feel like a strong, next step is data and the data team and the scraping team coming together and having a chat.

[00:49:45] **Austin:** [00:49:45] I think as well as someone from management there, to be honest, while I also agree with Sid, but I think there was some speculating, our, the timeliness of it. And how often should we be scraping and what should we do with, with data like this, doesn't only matter to us and the scraping team.

[00:49:58] It also matters to the [00:50:00] analysis team, as well as probably the people doing FOIA requests. So, I think some management perspective of, okay, with this organization we have set up here, how often are we updating our usable information, would be even more beneficial.

[00:50:15] **Nick:** [00:50:15] Yeah. so, I'm going to try to be getting a data engineer team and a scraper team in place, like in a more formal capacity.

[00:50:23]and we can definitely schedule time to sit down and really hash all this out, once that's been formed. in my personal opinion, I would think that the project would be, mainly for preservation of record and synchronization of current records. So, if the, if a County omitted or no longer has that, we were just, we're just going to keep whatever record we have of it.

[00:50:48]and update any records that they, they do have on file. and that's just my opinion. I don't know if it's right or wrong, going forward, but I think that's, that would be a noble [00:51:00] cause as far as like making the information accessible, in the case that local jurisdiction fails to do so

[00:51:08] **Ray:** [00:51:08] there might be cases where we shouldn't be doing that because there have been companies that have come up, which are doing some star off, Social scoring these days in America as well.

[00:51:19] They are using court records and public records to do that.

[00:51:22] **Nick:** [00:51:22] That's a good point would

[00:51:25] **Ray:** [00:51:25] we shouldn't have all the data available all the time. Should follow what the code on the County is doing, whether it's data. Also, I don't know if the product team has considered this. There might be opportunity. (unintelligible) let’s not just data?

[00:51:44] So our responsibility and our liability of storage, retrieval, and production obligation of their data, doesn't go wide to others who really are interested in doing it on a regular basis. To me, might not as an [00:52:00] organization one all the time and do it.

[00:52:07] **Sid:** [00:52:07] I liked what you said and agree. I just wanted to put that out there.

[00:52:13] **Goose:** [00:52:13] We're making some good points and I think we've captured most of everything we want to touch on. Is there anything else somebody wants to bring up in front of this group, or just want to close this out, we've been at it about an hour.

[00:52:28] **Sid:** [00:52:28] Do we want to go into more detail about the roles and responsibilities? Are we still premature for that?

[00:52:40] **Nick:** [00:52:40] I think it's a little bit premature at this moment in time. everybody here has definitely shown initiative and will definitely be part of the core team of data engineers. in my opinion, well, Josh has got his own stuff to deal with, but everyone here that's been actually participating. If you want to continue [00:53:00] to participate, I have no objection to that.

[00:53:08] **Ray:** [00:53:08] Sorry, one last thing. can we have like a, some sort of even a raw idea on what data would look (unintelligible) would be even accessible in this project. Again, the account, the data that was already (unintelligible) available on GitHub. If you take that data, how will the final accessibility data.

[00:53:34] **Goose:** [00:53:34] I would say we've put that in the parking lot for a question we have later with, either it's what's once the management team has solidified their vision, or once we have like a front end dev team or maybe it's the data visualization folks. but yeah, I think, I don't think we'll be able to answer, answer that today.

[00:53:53] **Ray:** [00:53:53] Oh, I was a putting it out there for any questions as management has time to do that? They can probably out this, [00:54:00] no. These questions go to leadership team,

[00:54:04] **Goose:** [00:54:04] which speaking of leadership team, while we're closing this call out on logistics. so, we use discord today because, we thought that made the most sense right off the bat.

[00:54:13] But, we didn't know if there was going to be some sort of a formalization to how, I mean, this was kind of janky I've like OBS recording my screen right now to capture this meeting that we'll post it. But, I know, if we want to steer everyone towards zoom, I know in order to get a call over 40 minutes, you have to have paid zoom.

[00:54:33] And I don't know if there are any discussions around those types of like planning things with those who are engaging.

[00:54:41] **Joshua L.:** [00:54:41] Yeah. Just to answer that question. I, we used to zoom yesterday, just to have a longer meeting. I ended up paying for a personal account because I didn't want to miss the opportunity of having a good conversation going.

[00:54:55] it’s harder to answer that question, cause right now, at least. On my end. I'm trying [00:55:00] to look for, really easy to use tools and also tools that fit our needs to do that. So far, I haven't found anything within the free range of, digital communication tools that will work completely. So, yeah, that's pending

[00:55:16] **Nick:** [00:55:16] is a, did they make Google meet free? Do You know?

[00:55:25] **Joshua L.:** [00:55:25] I'm not sure of the complete limitations of Google Meet, no I was looking at creating like a Google or Gmail business account and do stuff with, to sort of facilitate a little bit of that, but I wouldn't. Yeah, I it's literally just, it's just been me sort of looking into that. So, I was hoping to, as soon as possible stand up the team to help me sort of source., A lot of the open source tools or any tooling that could be useful for our organization.

[00:55:54] Okay.

[00:55:54] **Ray:** [00:55:54] Jitsi is a pretty good tool, which works well, video conferencing and audio conferencing it's [00:56:00] open source available for free. Yeah. And the free version can handle what making the open source version. You can do 102.

[00:56:08] **Casey Shredding:** [00:56:08] I was going to say the same one.

[00:56:10] I'm also in the code for America Slack channels. And that is one that gets posted about regularly. So, it's definitely been used for similar and large-scale groups before. Well, I will say Jitsi, I believe, J-I-T-S-I,

[00:56:31] I think so. I haven't personally used it. I just know I've seen it in a lot of Slack channels if you really do. I think as long as the host has a pro account can do the pro features. And I would imagine a lot of us have that through work and the likes,

[00:56:47] **Nick:** [00:56:47] I think it's J I TSI.

[00:56:50] **Joshua L.:** [00:56:50] Yeah, I actually have that on my list. I just haven't gotten through to it. I think I was going to look into WebEx next and Jitsi was the next thing. Like some, [00:57:00] some companies have opened up their free licensing, to have more features, and I'm trying to just distill the ones that are offering more stuff. Like I think WebEx, instead of the 15-minute limit, they made it unlimited.

[00:57:16] but I'm not sure, like, I I'm. Pulling stuff out of my ass right now. But as far as I'm,

[00:57:21] I just want to say I'm really proud that no, one's mentioned Skype so far, so thank you guys.

[00:57:27] I forget that Skype's still a thing.

[00:57:29] **Nick:** [00:57:29] Don't worry, Josh. I'm just pulling stuff out of my ass too. Nobody knows what they're doing

[00:57:35] **Ray:** [00:57:35] I've used Jitsi in the past, and it's pretty good.

[00:57:37] Instead of hosted version, (unintelligible) well across a continent and

[00:57:43] **Joshua L.:** [00:57:43] that will pop it to the top of my list and look at it soon.

[00:57:53] **Sid:** [00:57:53] It's like a rundown of action.

[00:57:59] **Goose:** [00:57:59] Well, the big [00:58:00] one is, the meeting cross team meeting for data and scrapers that Nick is going to, complete. Outside of that. I don't know if, I mean, definitely let me know. You guys think, I don't know if we came out with any real big formal ones other than continued discussion within the team, which doesn't have to happen over a meeting, about which direction we want to go with the information that we have

[00:58:25] **Joshua L.:** [00:58:25] Right,

[00:58:25] **Nick:** [00:58:25] yeah (unintelligible)

[00:58:26] oh, go ahead Josh.

[00:58:28] **Joshua L.:** [00:58:28] Oh, I was going to say, I did think there were some really good questions that came out of this meeting and it'd be good to transcribe them somewhere so that we keep track of them. and maybe post it somewhere in general and see if, you know, people would, can come up with good solutions or approximation to good solutions and you can look up.

[00:58:52] **Sid:** [00:58:52] Maybe I can add one for myself. Sorry. Do you need to make that then? Hi. I was hoping I could [00:59:00] sort of put together a high-level architecture diagram. Not saying that that's what we're doing, but as a, a conversation starter, right? So here are the pros and cons of doing this. do we really want to do that?

[00:59:13] What if we get data in a different way? Like, how's that going to impact us just as a starting point so we can refer to, or rather refer to it.

[00:59:22] **Nick:** [00:59:22] And I, I posted a DB diagram dot IO, it's like a relational database, essentially, like a GUI to make one. What it is, is not perfect and it's not solidified.

[00:59:35] I just tried to get a, just a little jumpstart on it and everybody else can make their own and we can kind of come together, if you want to, say, Hey, this is what I think the data should look like. this is how you should do it in this manner. I don't care what tools you use to make something like that.

[00:59:47] And I'm all, I'm all ears and all eyes to see what, what you guys come up with and what you guys think is right to do.

[00:59:54] **Joshua L.:** [00:59:54] So that was more of like the database schema, right?

[00:59:58] **Nick:** [00:59:58] Yeah. Essentially.

[01:00:00] [00:59:59] **Joshua L.:** [00:59:59] Okay.

[01:00:02] **Goose:** [01:00:02] Yeah, I think, what I heard at least listening to that management recording was that right now it's anyone who's taking action.

[01:00:10] Just go for it. Cause you're right. Sid, I think if you start that it will start conversation and discussion around what technologies we want to use and why, and have those Slack conversations.

[01:00:24] **Austin:** [01:00:24] Yep and Sid, I'm all, I'm all in for helping you on that. And I'd like to have some, some eyes and ears with that as well, so.

[01:00:31] **Sid:** [01:00:31] Okay, cool.

[01:00:33] **Nick:** [01:00:33] Yeah. So obviously, like I can't delegate tasks. We don't have anything formal set up and this is, this is like relying on contribution from everyone out of their own Goodwill. I can't tell you, "you need to do this" or "you need to do that." No, that's not how it's going to work. Like if you want to take initiative to do something, you can absolutely do that.

[01:00:51] And I'm going to try to help you and encourage you in any way that I can. as long as it's within legal reason

[01:00:58] **Austin:** [01:00:58] and on the other side of [01:01:00] that, if you have something that you don't want to do, but it needs to get done message me.

[01:01:04] **Nick:** [01:01:04] Okay.

[01:01:11] **Goose:** [01:01:11] I'll take-aways an actionable, which is to research landing zones. I like to say I'm an Azure guy, but I will, I will do some homework because they're, outside of, you know, the big cloud providers, there may be some smaller providers that we're maybe overlooking who may have a good cost, cost of entry or who knows.

[01:01:29] Maybe there's even some out there who support, Do gooders like this group?

[01:01:36] **Sid:** [01:01:36] Yeah. I saw some rumbling. I don't know. Is it the dev ops channel or something that they were leaning towards, like Lambda functions and elastic Beanstalk on AWS?

[01:01:48] **Goose:** [01:01:48] yeah, that all felt interesting. And premature. I saw a lot of that as well

[01:01:53] **Sid:** [01:01:53] from

[01:01:54] **Goose:** [01:01:54] a

[01:01:55] **Sid:** [01:01:55] selfish perspective. I got my AWS certainly two years ago, but I never used it.

[01:02:00] [01:01:59] So I would like to use it.

[01:02:02] **Goose:** [01:02:02] I don't have problem with either. I mean, I think having an understanding of what needs to be done is greater than, opening up the docs on AWS and figuring out how to put your code over to Lambda function. I mean, it's not really a big deal, so,

[01:02:16] **Nick:** [01:02:16] right, right now there's a lot of, we need to have a lot of these discussions about what we're supposed to be doing.

[01:02:22]as opposed to like getting a lot of things set up and in place in order, because we'd have to redact them or make modifications to them. So, for our own sanity, I think it's good that we're, approaching this at a, at a reasonable pace. and I know everyone's eager and wants to get started, and there's a lot of stuff happening, especially in the United States that make it even more urgent, but we have to approach it from.

[01:02:43] a steady and, good, good pace. so, we don't get ahead of ourselves.

[01:02:49] **Goose:** [01:02:49] Yeah. Some people just want to drop into a channel and flex their e-muscles and then pop out and never contribute to.

[01:02:59] Cool.

[01:02:59] **Sid:** [01:02:59] I [01:03:00] think one thing we can do pretty quickly after the, the data scraping meeting, like we can get started on the data model, right? The data warehouse scheme or whatever, how we're going to, like the data is going to come in and all sorts of different ways. But they're generally going to have some sort of pattern.

[01:03:20] Here's what that pattern is going to be. Here's how we're going to clean that up. Here's what we think is going to be important for consumers of this data. So, we're not going to be super blocked on decisions about any particular architecture or any particular cloud provider?

[01:03:42] **Austin:** [01:03:42] I agree.

[01:03:43] **Goose:** [01:03:43] Just a word of warning for all you relational guys out there.

[01:03:47] I'll be a strong voice for semi-structured in the chat. So, come prepared.

[01:03:54] **Sid:** [01:03:54] I will want to establish a cadence.

[01:03:57] **Goose:** [01:03:57] No, I think it's good. Those are the discussions we need to [01:04:00] have.

[01:04:02] **Sid:** [01:04:02] Do we want to establish a cadence for these meetings

[01:04:10] **Casey Shredding:** [01:04:10] for step two, that's going to depend entirely on the data scrapers. That's what I was understanding was this thing that needs to happen.

[01:04:20] **Sid:** [01:04:20] Okay.

[01:04:24] **Austin:** [01:04:24] something that I would like to bring up, and I don't know if it's just my inner ADHD working here, but, is there, is there any organizational setup for, or at least anybody talking about kind of some kind of calendar set up? yet there's a lot of Slack channels and there's a lot of messages going on a Slack channels.

[01:04:40] And I know most of us have jobs here. so you guys think that there's any reason that why we shouldn't set up maybe an organizational calendar for when people do have team meetings that we can set them up on their and... cause even I would like to keep my mouth shut, but come listen to like a data analysis meetings, but it's [01:05:00] not as when I pay attention to all the time

[01:05:02] **Nick:** [01:05:02] as things get more solidified, McKenzie just set us up with the Calendly.

[01:05:08] And that kind of hooks into like your Google calendar. that's like this morning happened, so it's, it's relatively new. And once we have more people, set up formally, within each team, I'm sure we'll be able to send out calendar invites and get at least a little bit more synchronized as far as schedule's concerned.

[01:05:35] **Goose:** [01:05:35] Alright, well, I'm going to make an effort to get some of the things down into this note, which I may have, not been able to while we were speaking on things, but the recording is going to be posted. And of course, this is going right into the gilt repo, so anyone can make amendments to this document. if I forgot anything or they think there's something critical that should be addressed in the note.

[01:05:56] but I'll go ahead and, I'll go ahead and end the recording. [01:06:00]