

CHoRUS Bridge2AI Standards Office Hours-20250501_181902-Meeting Recording

May 1, 2025, 5:00PM

1h 18m 30s



Alvarez, Marta 2:03

Hey, Jared, real quick question. I just saw your note.

Ken, do you want me to take the recording off of teams, or do you mean on YouTube?



Houghtaling, Jared 2:14

I.

I don't.

I think it should be fine.

We can maybe chat about it offline after I don't.



Alvarez, Marta 2:18

OK.



Houghtaling, Jared 2:19

It was more of an anticipation thing, so we'll see how it goes and if it, if it is going to, if we're gonna need that.



Alvarez, Marta 2:23

OK.

It's OK, just so you know, it's automatically recording you on teams right now.



Houghtaling, Jared 2:30

OK. Perfect. That's OK. Yep. Awesome.



Alvarez, Marta 2:31

And we can change it as needed. OK.



Houghtaling, Jared 2:33

Thank you.



Alvarez, Marta 2:59

Before we get started, there is one new person who I see that just joined the call who might want to do a quick intro from the UCSF team.

Sorry if I get that wrong, but if you hear me, if you want to do it back into that, be great.



Narula, Siddharth 3:23

Yeah, I'm super.

I recently joined UCSF working with Doctor ID.

So like my goal is to like have a nice map between the course data and the imaging.

Right now that's what I'm working on.

So yeah, pretty much.



Alvarez, Marta 3:42

Great welcome.



Houghtaling, Jared 4:07

See, people are still trickling in, but maybe I can already go.

Go ahead and get started.

OK.

So couple things today.

So the first first thing is.

I was in North Carolina this past weekend helping with Koris's contribution to the Duke Datathon, which was a weekend event.

Was quite successful, so there are eight teams.

And four different data sources.

There was a Korean mimic data source.

There was some.

Interesting NSNSTRI.

I wanna say, yeah, total of four data sources of which chorus was one. And out of those eight teams, two teams used the chorus data set in.

In their challenge, let's say so everybody chose datasets or maybe some chose multiple datasets and try to develop models against those sets and then presented at the end. And so the team ended up winning, chose Chorus and I can talk a little bit about what they did and.

How they used the data sets and what kind of the environment looks like because I think.

At least for me as a like coming at this from a data engineering perspective, it's exciting to see.

See how the data is used and then kind of some of the analytic work downstream.

So I was gonna give a little bit of an update on what that looks like.

How people are using it and then I think we'll transition from there into going into more flowsheet focus and kind of a question answer 'cause. I have been getting some questions via e-mail that I haven't had a chance to respond to.

See Eric's on the call.

Eric, does that sound OK? From your perspective? Isn't it awesome?



Rosenthal, Eric S., MD 5:47

It's awesome.

Yeah, I'd love to hear more about the data.
Thon had a long standing vacation plan and I would have otherwise been there.



Houghtaling, Jared 5:54

Awesome.



Rosenthal, Eric S.,MD 5:55

But I had to keep my family together. And so I I couldn't be there.
But I wonder even if if one or two of those, we could help them and maybe facilitate a poster at our at our meeting, even if it's by proxy, it would be really nice too.



Houghtaling, Jared 6:11

Yeah. Yeah. So, so one of the teams that that to the team that wanted, I I I'm also working together with and we're planning to submit an abstracts this summer on it. But there is probably enough to put together a preliminary poster for the conference, so I'll reach.



Rosenthal, Eric S.,MD 6:24

Yeah, yeah.



Houghtaling, Jared 6:26

Out to them.
That's a good idea.
Awesome.



Rosenthal, Eric S.,MD 6:30

Tell smur.



Houghtaling, Jared 6:31

All right, all right.
So I guess I've I've showed a little bit about what we're doing in terms of the onclave.
I'll share my screen and kind of walk through this but.
Yes, everybody at the at the Datathon got a Duke.
External Duke credential.
That credential wasn't added also as an external user to MGB, and with that credential they were able to get to this landing page. They hit get started and that brings them in to a a Firefox browser.
So that's really a jump host.
So each user gets their own little virtual machine, and that jump host directs them into a a machine learning workspace.
So this is something from Azure that facilitates like, yeah, the ability to run machine learning models and develop them and and share things in a kind of in a collaborative

way.

There's lots of different assets and tools and things like this that Microsoft provides. But for you to know every user in this data, Thon, and this is the same for aim ahead and the other things that we're doing.

Has their own compute instance inside of this space, so that compute instance is scheduled to stop after a certain time.

And in this particular case, mine is running and I can access a Jupyter lab instance on that on that compute.

So now the users come in here and they have their own Jupyter lab instance and what we ended up doing and I kind of walked through the workflow because I think this is important to understand how teams are interacting with the course data.

So they they come into the space, they have their own computes, they launch it, they get into Jupyter lab. And then how do they actually do stuff with the data and what ended up being the most useful workflow is these sites had brains, or these teams had brain.

What analytic question?

They wanted to answer what type of of clinical relevant question they wanted to answer.

And so the first step was really, let's grab the data that's relevant to that question and what ended up being useful was was actually Atlas.

More so than than I had initially anticipated because we went into Atlas and in this case of Team aids, they wanted to know.

Patients, patients who are admitted to an ICU and had a a surgical procedure in the ICU.

Following that procedure, if those those patients had to be placed on a ventilator, if they could predict that that kind of event of of a patient post surgery in ICU needing to get gets onto a ventilator, I don't know how.

 **Rosenthal, Eric S.,MD** 9:06
Jared, that's that's a new ventilator.

 **Houghtaling, Jared** 9:08
They yes, yeah.

 **Rosenthal, Eric S.,MD** 9:09
As opposed to continuing on a ventilator.

 **Houghtaling, Jared** 9:11
Yeah, exactly.
So the event starts only after that that that surgery itself.



Rosenthal, Eric S.,MD 9:19

Great.



Houghtaling, Jared 9:20

And so we we what we did for them is we created two cohorts. We created that that kind of outcome cohort of interest where the the patient has to be started on a on a ventilator and we created additionally just this any surgery cohort. So they have their. Sorry, that's this one.

Any surgery yet?

So this is the kind of broad.

Broad any surgery and additionally for some additional comparisons, we hit it in the without vent.

So we actually then also had a a cohort that had no ventilation at all.

But they did have a surgery in the ICU, and so we ran those cohort definitions through Atlas and that then produced in our database a list of the patients that fit those criteria.

And and then what happened is kind of interesting, right?

Because what?

And this is where I think it's important for the data engineers on the call because what what they wanted to have was a set of variables of interest that fit this question.

And most of the people at this datathon were unfortunately not at all familiar with omop.

So it wasn't like they could just come in and write some joins and and get this done.

So what I ended up doing a lot of the the the time was helping the teams get a wide format version of their of the OMOP data set.

So where you then have essentially one row per person, with columns corresponding to variables.

That of of what's happening?

Whether it be a temporal or an average lab value or whatever it might be.

And so this query then produced this flat data set for them that looks something like this.

And in this case, I think that they used a total of 10.

10 difference covariance. Something like that.

So BMI surgery, white blood cell prothrombin, glucose, creatinine, albumin, HG O3 or bicarbonate, I think.



Rosenthal, Eric S.,MD 11:17

Mm hmm.



Houghtaling, Jared 11:18

And then the the label plus plus the age.



Rosenthal, Eric S.,MD 11:23

Jared, so those are those are labs that are going to be available usually every day.



Houghtaling, Jared 11:23

OK.



Rosenthal, Eric S.,MD 11:28

So is it?

How do they select? Is the first one?



Houghtaling, Jared 11:32

Yeah. So they wanted the the labs within. I think the 1st or the 1st, 3 labs of each, each one that they took an average of.

So the first three labs post.

Surgery was the criteria.



Rosenthal, Eric S.,MD 11:43

Yeah.



Houghtaling, Jared 11:45

That they wanted to look at.



Rosenthal, Eric S.,MD 11:46

And would would they have to be before the event?

Before intubation.



Houghtaling, Jared 11:51

Before the event, they didn't.

They didn't include the ventilation as a criteria in that it was it was based off the index event of the surgery. I don't.

I think I don't.

Yeah, but I can double check again.

I I didn't write the the protocol but.



Rosenthal, Eric S.,MD 12:05

No, no, I understand.



Houghtaling, Jared 12:08

But yeah, so so essentially, then they, they they we were able to get them this data, some of which some patients didn't have all the labs things like that and and this is where I wanna emphasize.

The standardization comes into play, right?

Because when you're talking about looking at even the ventilation or or white blood cell counts or things like that, representing these and standardizing these across the various sites makes it much easier to and also more feasible to to actually get them into this format and to be able.

To compare.

Are the appropriate values here.

OK.

So we we we this wide formats and then they're able to run that through.

Through their model train, they actually trained. I think a total of five models on this. I can show you really quickly.

So they had a decision tree.

Random forest logistic regression, a gradient boost and what they call a faster risk model.

And what that those models were then effectively optimizing is given patient a what is the the likelihood of them?

Being on a ventilator after the surgery, so we we all the patients in our in our kind of target group, we know that they're having some surgery.

What's the likely? Have somebody having to go on a ventilator post surgery?

And this is essentially the breakdown of feature importance of their various models.

There's some other figures as well.

There's definitely enough plots and I think information to to put together something like a poster.

And I guess I didn't plan on going too deep into the models.

But I guess what I wanna also add to this is where we left off at the end of the data.

Thon is is I think interesting in that a large chunk of the patients in this I wanna say close to half were MGH patients.

That's that. Have waveform files or that have waveform files accessible for the data phone and so they also in parallel with this they we had another data scientist on that team.

Working on feature extraction from those waveform files.

To try to add those NS features we haven't gotten there yet, but they're actively working on this. I've seen some some action this week.

So the idea would be OK.

We have an omop only version with this very limited set of covariates. As we expand and add covariates and also add waveform features, can we?

Can we start to increase or improve our our performance of our models and what does that look like?

OK.

So that's I'm. I'm happy to go in any direction, Eric.

With that I can show anything or talk about any of it, but I think at a high level.

I think for me it was gratifying to see people able to use the infrastructure to securely

interact with the data, to start to develop some models and and start to do some meaningful things with this data set.

 **Rosenthal, Eric S.,MD** 14:55
Yeah.

 **Houghtaling, Jared** 14:59
So kudos to all of you for getting the data into the cloud and like making the this whole aggregate data set happen.
Because I think it's we're an exciting the kind of a turning point within chorus of like really starting to do some analytics on on this, so.

 **Rosenthal, Eric S.,MD** 15:15
It's credible.
I I love it. You know the.
We see the Nans you know in the in the data you showed.

 **Houghtaling, Jared** 15:21
Yep.

 **Rosenthal, Eric S.,MD** 15:22
So it it suggests that you know, our data's still incomplete and we all we all know that.
But it's really great.
You know, Tom and Jackie are working on something for the face to face regarding the Open House demonstration project and.

 **Houghtaling, Jared** 15:37
Yep.
Yeah.

 **Rosenthal, Eric S.,MD** 15:40
I wonder if we wanna maybe talk to Tom about this as an example. If if you think we can go bimodality for EHR and and waveform data.
He's thinking about more chest X-rays.

 **Houghtaling, Jared** 15:52
Yep.

 **Rosenthal, Eric S.,MD** 15:54
But it could be that you have the same cohort and you you wanna apply his.
You know, plural fusion checker to to the group. That's you know, at the highest risk or something like that.

But it seems like this would be really just a nice demonstration if we could walk people through it in a in an open house, that'd be nice.

You know, we we can't go.

We may not be able to go through everything.

In in super detail.

But I I think it's a really nice example, so it's great to hear what was the other course focus data set project.



Houghtaling, Jared 16:32

Yeah.

Yeah. Sudeep, now, I'm trying to think about that one as you as you ask it, so that I have to pull up the. That was team six and let me take a look and remember what they were doing here.

I spent more time I think with Team 8 on this one, but.



Rosenthal, Eric S.,MD 16:59

And what was the sophistication of the participants in terms of their training?



Houghtaling, Jared 17:03

So they split the team up the teams up into half data science background and half clinician which was I think it ended up being a very good.

A nice split.



Rosenthal, Eric S.,MD 17:12

So each team was, each team was mixed.



Houghtaling, Jared 17:14

Yep. Yep, exactly.

So yeah, it was.

It was.

It varied pretty widely, but it came from people who were at at no experience with, like Jupiter lab to people that were like very well versed in training models and things.

So it's an interesting group.

But yeah, so I guess the so the other so team six, what they wanted to know was they wanted to try to make a classification model looking at the discharge disposition of these ICU patients.

So if they could essentially predict where they were gonna head after using it, either the condition status concept ID, which is can be, you can denote it a diagnosis as a discharge diagnosis.

Or using the visit visit detail.

Tail discharged too.

Concept ID or the visited currents discharged to concept ID like there were like three

different things that we were looking at within omop that might be possible.
To help select that cohort that that was more challenging in the end to get the the even the cohort definition appropriate.
More of a limitation of omop and how it captures that than anything else but but yeah, that was the other one.
So.



Alvarez, Marta 18:33

Mihai joined the call too.
I don't know if he wants to add anything.



Houghtaling, Jared 18:35

Awesome.
I didn't see Mihai around, but we did talk.
He was mentioned in a couple of talks.



Mihai Podgoreanu, M.D. 18:45

Hey, guys.
No, sorry.
This unfortunately coincided with our national meeting, which was in Montreal and I stayed to cover the the Fort so that more junior and brighter faculty can represent Duke at that national meeting.
And so I was working in the ICU throughout that process, but.
Now generating data all along.



Rosenthal, Eric S., MD 19:12

Yeah.



Houghtaling, Jared 19:15

Awesome.



Rosenthal, Eric S., MD 19:16

Cool. Glad to see that it was. It was so successful of a conference.



Houghtaling, Jared 19:23

So one thing I did want to ask Eric is we have this the the resources now set up for Duke datathon like this space. If there are data engineers at sites that would be interested in getting access to this Duke data, Thon space, I think that that should. Be possible.
They already have guest accounts in MGH like does.
Is that something that we could facilitate?



Rosenthal, Eric S.,MD 19:45

Yeah. I mean, I think I really am attracted to the front door concept for. You know, I was emailing with Tom Pollard today about his attending one of the Bridge Ji data dissemination workgroups. They were talking about the new. Executive orders Biden era that were then updated by the the Trump administration. Basically. Stating that we cannot allow controlled access. By countries, institutions at willfully allow controlled access by countries at institutions on a list of concern. So. So it makes me think that, you know, when we register people we we can vet them and so forth. But having more than one layer like when like when you log on to a screen like this, you can imagine having people have to click a an acknowledgement box. To move on. And so I I I am attracted to the front door aspect of it. You might be a question of whether we could talk to. Would it be Ian or or other about embedding you know embedding this somehow? Where where would be the point of access. Would it come from our website? Or you know.



Houghtaling, Jared 21:03

Yeah. So I guess I, I should say this is all still hosted at MGB and it's using the MGB tenant just like. So there's there's actually no difference.



Rosenthal, Eric S.,MD 21:05

I yeah.



Houghtaling, Jared 21:12

The only difference, practically, was Duke for the Datathon gave Duke credentials to all of their guest users, and those credentials were then used to to add them as guest users to MGB. So it was like a guest of a guest and.



Rosenthal, Eric S.,MD 21:19

Yeah.

Yeah, yeah. That was a plan for simplifying the single sign on. Yeah.



Houghtaling, Jared 21:28

But.

Yeah.

And it and in the end it worked actually very well.

It was like very smooth to get everybody coordinated, no.



Rosenthal, Eric S.,MD 21:36

Yeah, very smart.

Yeah.

So yeah, I think there's a lot that we can.



Houghtaling, Jared 21:40

Yeah.



Rosenthal, Eric S.,MD 21:46

Basically, yes we should.

We should try to do so, and I think, oh, there's a lot of people in our group who would wanna have that as a portal to do, some, do some work. Hey, mark.



Houghtaling, Jared 21:55

Awesome.



Wainwright, Mark 22:00

All right. Thanks.

I came a little bit late, but as I joined Jared you were commenting that if I understood correctly that the process of the analysis and the work you did went more quickly.

It was more effective by having a specific set of data or specific problem to focus on, and I'm wondering what you or Eric think about what that applies for.

Prioritizing data collection or specific tool development.

In the future where we have this very, you know, ambitious appropriately scale of this project, but maybe the next phase needs to be more, have a particular focus.



Houghtaling, Jared 22:41

Yeah. So I we had a long discussion about this on Sunday nights and I think one of the interesting things that came out of it was the idea of creating flattened views of.

Either cohorts of interest or variables of interest where going into a datathon instead of having to work with the OMOP data model. As such, there are like a pre pre computed set of views of like OK.

Here's all your sepsis patients with 100 different variables that are flattened out.



Rosenthal, Eric S.,MD 23:12

Yeah, yeah.



Houghtaling, Jared 23:12

Here's all your like to have that already at your fingertips.

That would have made it go much, much faster in the Datathon and I think just both as a characterization tool, understanding what data has been delivered and what needs to

be delivered, and then allowing analysts to like hit the ground running. I think those that spending some time.

On that would be hugely beneficial.

Rosenthal, Eric S.,MD 23:34

RS

So so can I comment on that, which is we had a discussion with one of Leo Chelli's team who's moving to Columbia University after his graduate program, Matt McDermott. And he was talking to us about the the meds framework being complimentary to omop and it's basically this.

Where you you sort of take the omop data and then you say and now I need to do an analysis.

I'm gonna sort of flatten it.

And he what he was saying was sort of omop is is great.

However, when you wanna start to do the analysis, you sort of need the data in a certain framework, and most ml people are gonna take that step anyway.

And I had a follow conversation with Andrew about this and Andrew was sort of like well, you know.

We you also may lose a lot of the metadata in omop by doing so.

And so there's sort of discussion of like what are the pros and cons.

But I I think you're showing a a really important example.

This is how people work in the end.

They want a flat file to run their run their code and run their algorithm on and they we can't lose the metadata.

It's it's incredibly important, but, but you know, just like you said, Mark, there's gonna be someone who eventually wants to build a cohort and run a run. An analysis.

They always are gonna go back to the data, maybe to analyze and lookup patients and so forth in, in the omop framework.

But I do wonder if there's some value there so.

Matt is gonna be doing a KO1 under George Hrip check.

At Columbia, who sort of informatics godfather.

Houghtaling, Jared 25:13



Yep, Yep, that one. Yep.

Yep.

Rosenthal, Eric S.,MD 25:17

RS

And so, you know, I think they're doing real work and might might actually wanna contribute to our project.

He that was his request and so.

We, you know, I think Andrew was a little bit circumspect about, you know, whether it was worth the additional complexity.

My opinion was if this is someone who wants to sort of bring their own work and effort

to the project and it's complementary, not replacing what we're doing, then we could do exactly that, which is we had developed these like canonical use cases.

Like you know, intubation in the ICU neurology.

Deterioration renal, you know, kidney injury, you can go on and on and we I think had it.

I can't remember set of them and it might be really wise for us to start thinking about building these flat, more flat files that that are not meant to replace the the OMOP data, but so.

We can reach out to him and maybe have him come to this meeting.

Or or to our all hands meeting on one of our all hands months.

To maybe talk about the concept.

And make sure that people like yourself. Polina. You know, we feel like we're not doing a side step on our mission.



Houghtaling, Jared 26:39

Absolutely, I think.

I guess I have a couple thoughts, but I would say I think omop's value comes from knowing the ecosystem and the packages that already built. And like if I was able to to put together an omop prediction model very quickly on the data set, but in order.

To like have teams be able to do that.

They have to understand like there's a whole like learning curve of getting into omop and what is it?

What it all means, and if you're already in that space, I think that the omop.

Model is fantastic and you can leverage a lot of cool things and I think that's a big value.

If you do not have any experience with that though, it will be much, much faster for you to get off to the races. If you do have this set of resources of flat files and you can work in the way that you know how to work and kind.

Of push ahead.

So I think that they're complementary and they're gonna. It's gonna depend on the user, which one is gonna make more sense.

To come up.

So I I yeah, yeah.



Mihai Podgoreanu, M.D. 27:30

And you know Mihai here real quick commenting.

I'm sure you're all aware of the fact that, you know, mimic does have this sort of concept of a benchmark data sets that was assembled in the past.

And you know, same concept perhaps in addition to Eric sorta mentioning the canonical use cases.

There the the data set can be processed to fulfill.

A broader if you will, data science.

Once use cases such as a classification problem versus a prediction problem versus a,

you know, a complex phenotyping problem.

And so I think you're right that the amount of effort that teams were were to put into preprocessing the data versus, you know having a flat file that and this sort of benchmarked data set that.

Can then go back together.

Multiple teams can go back together.

To this, to be able to compare performance of various models et cetera.



Houghtaling, Jared 28:39

Awesome. Yeah, 100% agree.

All right.

Well, yeah.

So I think I think, yeah, we there's a whole discussion to be had.

I think we'll talk with probably more at the cloud team about lessons learned from that beyond just kind of this analytic approach.

I.

I do.

I will say it was a very, very useful experience also in preparation for the face to face.

So I have.

I feel very confident now of like how that will go and we've I think already overcome some of the barriers in this process.

So I I think that that should be very smooth and I I agree with you Eric. I think having if we already have some multimodal modeling in the Duke space that we could use that as an as another example.

If if Tom will do it an image based or like a chest X-ray based model together with elmpop, I think that would be nice to showcase both of them or bring them all together. I'm not sure it could be that they that we have full overlap so.



Rosenthal, Eric S.,MD 29:44

Yeah, I mean, we can take it offline, but you can imagine show EHR model, you have your cohorts, you have your positive predictions and your nonpositive predictions and then you could take one example where you go downstream with waveforms.

Sounds like it's in progress. You dig another where you go downstream with one important feature of, you know, radiology maybe the waveform ones in MGH, maybe the chest X-rays ones at Emory or maybe they're they're both at.

Emory, you know, so it's a lot to accomplish an afternoon, but I think that would go along way to showing the this huge multimodal value.



Houghtaling, Jared 30:21

That's OK.



Rosenthal, Eric S.,MD 30:21

Which you know it's it's not really.
That's the thing we we wanna sort of.
Sort of miss for an opportunity to show.



Houghtaling, Jared 30:33

Awesome, yeah.
All right. Amazing.
So I I we're halfway through the call and I did wanna leave time.
I know a few of you have reached out with questions.
About.
Flow sheets and about mapping and some of them are epic specific.
So I don't know if there's anything of of those of you who have reached out.
I believe Jackie had sent the question, Satish.
I I you might have got it sorted.
I'm not sure I know Eddie, who has also some questions with UCSF.
No.
If any of you all want to, oh, awesome.



Amorim De Cerqueira Filho, Edilberto 31:11

Hey folks I just joined.
So I got you halfway through your through your question.



Houghtaling, Jared 31:18

Awesome, I was just setting the stage to see.
I know I've gotten some emails in the last couple days, so I've been trying to get back up to speed.
From folks about some of their either flow sheet mappings or general Omop work.
And I wanted to see if if any of all wanted to share those questions or we can discuss them on this forum.



Amorim De Cerqueira Filho, Edilberto 31:38

Yeah. My question was had to do with time shift best practices and changing everything to UTC, particularly on the point when we lose the hour and we restack things.



Houghtaling, Jared 31:38

For.
Yeah.



Amorim De Cerqueira Filho, Edilberto 31:53

So I was just trying to see if the and maybe there's some guidance specifically about that that I missed, but that was just some things we wanted to confirm. We were doing

it the way we should be doing.
We have a prototype of that going now, but.
That was my question, yeah.



Houghtaling, Jared 32:10

Yeah.

Yeah, it's a great question.

I know this this got brought up like a year ago and I know at least at standards we have not covered either.

Time like we've talked a lot about date shifting in general, but not about shifting to use UTC time or daylight savings and how that's handled.

So I don't think that we have at the moments unless data acquisition has pushed further on this. We don't have any sort of like SOP related to handling time conversion at the moment.

I'm looking a little bit toward Polina or anybody else on the standards team, but as far as I'm aware that that we haven't.

We haven't defined all of that in in very much detail yet.

I would be curious to know what you are doing though, because I think it will.

It could inform how we're we're proceeding.

Yep, Tony, go for it.



Amorim De Cerqueira Filho, Edilberto 32:59

Yeah.



Pan, Tony 33:01

Hi so so I think Jill was developing a sorry time shifting SOP. I I don't where that went because it was sent around and no review by multiple people and provide feedback.



Houghtaling, Jared 33:09

OK.



Pan, Tony 33:18

And related to to UTC versus not versus localized time, he was strongly advocating for, for using local time because.

Otherwise, if we use UTC in their local procedure, says always occurring at roughly the same time of the day.

Then you will be able to derive like know which time zone a particular data set is coming from.

So. So that's, you know, definitely a consideration and and I tend to agree with that. But at the same time, we're right now dealing with some historic data in our clinical data warehouse from the you know pre epic times.

And our ADT table.

The the values that we can easily pull.

Are have been localized with the time zone information.

Stripped out and no, I think.

That that's might be the same question that that no.

That's being implied here that that we have once a year, there's one hour at 2:00 AM that repeats itself.

And that causes problems for for our waveform mapping, because no, we're we're looking for patients as they move from bed to bed.

You know, if granted, it's unlikely to happen at 2:00 AM.

We can't say that it doesn't happen and know when we're sending out. When we're submitting waveform data.

We might be capturing either either missing like an hour worth, or include an extra hour. So so we're trying to resolve this by looking at other places, seeing the CDW for that information.

But this I think we need to have a a standard strategy on on how to represent time.

Tc or.

Or local time and have it be very consistent in this daylight savings. Time's really kind of causing problems.

A

Amorim De Cerqueira Filho, Edilberto 35:37

I put in the chat the latest thing I reviewed and if that's supposed to be shared or not. So if if ask for forgiveness, if this is not supposed to be shared, but that's what I thought. Lately, and it does mention the UDC.

And so again, I was because for us to share the data we need to we can share local time, but now we are dealing with organizing things.

So if it we don't need to shift, then we can share.

I think the problem is.

We need to re identify the data to figure it out when actually this happened in the real world to then figure out if daylight savings to then shift it DE identify again after we we do the UTC.

So that has been one of the.

Painful things that we have to do.

Because you can, if you apply the shift that's going to screw up your daylight savings.

RS

Rosenthal, Eric S.,MD 36:37

Yeah.



Houghtaling, Jared 36:41

Eddie, is that? Is that like the final blocker for you, being able to submit the the code or is that like where does that fall in terms of the things that are preventing the submission?

A Amorim De Cerqueira Filho, Edilberto 36:51

Well, I mean the at least for the EHR data, we need to find a way to transfer the data, which I'm hopeful we're gonna find out in the three hours or so. 'cause, we're gonna have a meeting between the UCSF it and the mghit. So that was because.

 **Houghtaling, Jared** 36:57

Yep.

A Amorim De Cerqueira Filho, Edilberto 37:06

Azure is not considered HIPAA compliant and folks you know more things that I don't really fully understand that we'll talk more today, but.

But once that's clear, which I think we will be able to find a solution, we should be able to share.

The data we just finalized, checking who had the waveform.

So we can make sure that people had waveform, but that seems to be solved now.

So the UTC was the last thing. The visit detail table.

Is working on with us now. Finalizing the update on that.

 **Houghtaling, Jared** 37:41

Awesome. Great.

Thanks. That's great to hear. Thank you.

A Amorim De Cerqueira Filho, Edilberto 37:43

And then if she's in the call.

But yeah, I don't.

Yeah, she's there.

BK Balakrishnan, Kasturi 37:47

Right. Yeah, yeah, I am.

A Amorim De Cerqueira Filho, Edilberto 37:47

Hi Kasturi.

BK Balakrishnan, Kasturi 37:50

I just I I was.

I'm running some QC on the data and.

I we we will have to report the data based on the latest release. So I am in the process of doing that and I should.

You know, based on the last discussion, I thought that based on the patients that you had already pulled, we should be able to get the data set, but.

I was going to reach out to you today.

We I I will have to pull the data again for you.
Before we can submit it.
But it isn't huge.
I I can get that done.

 **Rosenthal, Eric S.,MD** 38:25
We we can definitely.

 **Houghtaling, Jared** 38:28
All right, awesome.
Thank you for that. Take care.
All right, great.
And then, yeah, so I saw Satishi had come off mute.
I can just give you give you a quick update I will.

 **Dandayudhapani, Satish** 38:40
Yeah.

 **Houghtaling, Jared** 38:43
So I will send you after this call.
What you need to be able to to download or redownload your waveform files so you
can make you can edit and reupload. So you'll have that today.
Hopefully that gives you enough time to do what you need to do, but it's been a bit of a
wild couple.

 **Dandayudhapani, Satish** 39:00
Yeah, sounds good.
This is to bill that.
The registry and stuff to print the procedure table.

 **Houghtaling, Jared** 39:05
Yep.

 **Dandayudhapani, Satish** 39:06
Yeah, the mapping table. Yep.

 **Houghtaling, Jared** 39:07
Awesome. Awesome. Thanks.

 **Dandayudhapani, Satish** 39:11
And for the flowsheet I had, yeah. One of the question I sent that was resolved, I was

looking at a different table, but I was able to find that in the epic.
So the other one is for the flow sheets deriving.



Houghtaling, Jared 39:20

OK, perfect.



Dandayudhapani, Satish 39:24

We had two approaches.

One is the exact matches right?

The other one is the bag of words.

I we were able to implement the exact matches. The back of course is one I wanted to review it with you. I have that.



Houghtaling, Jared 39:37

Awesome. All right.



Dandayudhapani, Satish 39:38

Already. But when you get some time, like maybe a quick.

Look, before you know, just to make sure.



Houghtaling, Jared 39:43

Yeah, absolutely.



Dandayudhapani, Satish 39:45

Yeah, that's the only thing.



Houghtaling, Jared 39:46

Alright, I'll I'll send an invite for early next week, if that's OK, then we can go through it.



Dandayudhapani, Satish 39:50

Oh yeah, that that's great.

Yeah. So anytime next week is fine. Yeah. Thank you.



Houghtaling, Jared 39:52

Perfect.

OK. Thanks.

Yeah. Go for Eddie.



Amorim De Cerqueira Filho, Edilberto 39:58

Yeah. The other question that's unrelated to this has to do with the accession number and the imaging.

I forgot the name.

Imaging something on the omop equivalent, but I wasn't sure.

The procedure.

There's a separate procedure thing that we need to also link so some of that wasn't fully clear to me as I'm actually trying to do that bridge now on our data set where I would find that procedure that's not in the accession that's not related to the access.



Houghtaling, Jared 40:24

OK.



Amorim De Cerqueira Filho, Edilberto 40:31

Number what has been the experience from folks who are more? Experts than I am.



Houghtaling, Jared 40:41

Pull that up really quickly. I think I know exactly what you're it's the procedure.

There's there's two IDs that are like requested, let's say in the.

In this multimodal linkage right, you have the procedure, and I'll share my screen as well, so see this.

But you have the procedure occurrence idah and you have the procedure concept ID. I think that's where.

I'll open this in a new tab.

Make it a little bit bigger hopefully.

Right. So you have these two, two procedure things procedure ID.



Amorim De Cerqueira Filho, Edilberto 41:18

Are you sharing something?



Houghtaling, Jared 41:20

And yes, can you see it?



Amorim De Cerqueira Filho, Edilberto 41:21

Oh, no. Now it's coming, OK?

Now it's coming.



Houghtaling, Jared 41:26

All right, let me know when you can see it or if you can still get. OK. Cool. So this is again, this is an old version of this, but it's still, I would say the the current registry version as of now.

You have this procedure ID and I think this this hung up a lot of people, which is this just has to be a unique primary key identifier for in the procedure occurrence table and that there's no other requirement.

It doesn't have to even be the accession number. It doesn't have to be anything.

We just auto generate it in a range that we know is going to be unique and that's just the primary key of that table in omop that you can use to then reference any given row. The procedure concept ID is what Paulina has spent some time going through.

And that is something else.

So that in in this view of the multimodal linkage, this was overly simplified.

So all we said was this blue for image registry, we just map to this kind of generic image imaging concepts and for this yellow which is the waveform registry, we map to a generic waveform concept.

In practice, this requires mapping with the.

Kind of the custom ontologies similar, I think we probably need. It's gone.

Through that, on a couple of prior calls, but this procedure concept that he should correspond to the actual process of capturing your image, whether that is a procedure of of like a CT scan, it's an MRI, something like that.

Some omop coded procedure and in some cases it's going to be already existing in the OMA vocabularies, like these ones are like.

This less sub 2 billion concept and in some cases it's going to have to be a custom 2 billion concept.

That you're gonna need to get from the chorus mapping stage repository in the delta tables so.

That's that's a critical distinction between those two.

But I and I think that it's confusing that they're so close in name.

But the procedure occurrence ID, we have no restrictions on that identifier value other than it cannot conflict with any other procedure occurrence IDs in that table.

And the concept ID should correspond to what actually happens to capture your immature waveform over go for it.

Yep.

 **Amorim De Cerqueira Filho, Edilberto** 43:48

And to read it back to you, the procedure ID is something that we will generate.

So that's what I was hung up on.

Or if there is something that would be already in the omop that we have.

 **Houghtaling, Jared** 43:58

So this procedure occurrence Id'd on the left here.

That's what you're asking about. This one is what you would.

 **Amorim De Cerqueira Filho, Edilberto** 44:02

Yeah, yeah.

 **Houghtaling, Jared** 44:02

Yeah, that one you can generate fully yourself if I we have we give an example that we created as a greater than 2 billion value just because yeah, why not.

But you can just stack this on the like the like, starting at the maximum procedure occurrence ID in your table. You can just kind of increment off of that, or you can put it in some separate range as like. However you would like to do it. We re.

 **Amorim De Cerqueira Filho, Edilberto** 44:27
Mm hmm.

 **Houghtaling, Jared** 44:28
These centrally anyway.
But the the most important thing of this column if you're gonna do the integration is. Like I said it, it needs to be unique.
So every every row needs to have its own unique procedure occurrence ID. And it also can't conflict with existing procedure occurrence idsi think.

 **Amorim De Cerqueira Filho, Edilberto** 44:47
Yeah, and and I and I get that.
But the the the question I was asking is more like in our current omop that field is not gonna exist.
I'm not trying to pull that from a no mod that we have built in our system.
We on my side will create this code for this procedure.

 **Houghtaling, Jared** 45:05
That that column will exist in your in your table.

 **Amorim De Cerqueira Filho, Edilberto** 45:09
OK.
So we should already have the procedure occurrence ID in my current omop instance ACSF and I just need to match it and that has nothing to do with the accession number, because that's a different that's an image occurrence.

 **Houghtaling, Jared** 45:15
Yes. Yep, Yep.
No.

 **Amorim De Cerqueira Filho, Edilberto** 45:23
Yeah. OK. Perfect.

 **Houghtaling, Jared** 45:23
Indeed awesome.
All right.

 **Balakrishnan, Kasturi** 45:27
Chair. Sorry, I I just have a follow up question.

 **Ashe, William (wa6gz)** 45:27
Great.

 **Houghtaling, Jared** 45:29
Yep.

 **Balakrishnan, Kasturi** 45:30
This is Kester.

 **Houghtaling, Jared** 45:31
Yep, that's good.

 **Balakrishnan, Kasturi** 45:31
I just have a follow up question on what Eddie was asking.
So essentially you're saying that these procedure occurrence IDs are like we are taking a set of data and inserting it as procedures into the procedure occurrence table, which is why we take the Max value, right?
Did I understand that correctly?
Umm.

 **Houghtaling, Jared** 45:52
Yeah, you're essentially creating new rows that are seen as procedures by omop.

 **Balakrishnan, Kasturi** 45:55
Well, it's perfect. Yeah, I get it.

 **Houghtaling, Jared** 45:56
Yep. And so they just need to follow the Convention of the procedure table in OMOP, which is this is the primary key and it has to be unique and cannot, yeah, cannot conflict with existing rows in that table at your site.
Everything else is all good.
However, you wanna do it.

 **Balakrishnan, Kasturi** 46:15
Yeah, Eddie, we might have to meet up on this.
And.
Thank you.



Houghtaling, Jared 46:26

If you if you need to include me or or reach out please please do so.

It will be around next week.

All week. So.

Hopefully get that sorted.



Balakrishnan, Kasturi 46:34

Sounds good.

Thanks Jen.



Houghtaling, Jared 46:41

All right.

Oh great.

Great questions. Thank you.

Anybody else with a flow sheet question I don't know.

I didn't see Jackie on the call.

Oh Jackie's here.



Gravitt, Jacqueline A *HS 46:53

Hey, I'm here. Yeah.



Houghtaling, Jared 46:54

Yep. Awesome. Awesome.

I know you had a question.

I I think I saw a question come in about.

Epic stuff, right?

Let me double check.

Or maybe it was pho sheet. Yeah.



Gravitt, Jacqueline A *HS 47:04

Yeah, yeah.

So just kind of in general, the way this was set up, first question was, are you guys loading all of your flow sheets from epic because we have like 30,000 of them and seems like a lot.

So the way that's oh God.



Houghtaling, Jared 47:16

So.

I was going to.

We are not loading all of them.

We are loading only the ones that we have mappings for, so we're not including all 30,000 or however many there are that don't have yet omop mappings in our case.

Gravitt, Jacqueline A *HS 47:20

G

OK.

OK.

Yeah, 'cause the way we built it out and don't know it's the best way, but it's what we what we did.

We essentially just tell it say, hey, you know specifically grab these flow sheets.

So we're restricting it that way. And then we're kind of forcing the mapping to the concepts based on we had somebody say, hey, you know there should be mapped to you know link code 123 or whatever.

So that's how we're forcing it.

And then if we want to add more, we just add more to that table as long as we tell it with mapping to go to so then.

It's kind of a specific one, but like we have height mapped in but I saw that height was in like some of the ones that you guys pulled on from your site.

So I guess like that's a very specific one, but are you mapping to the standard codes instead of the ones you guys provide?

Is it going to both places or how does that work?

Houghtaling, Jared 48:11



Yeah. So the example you gave is awesome. So what?

What you'll find in the Delta vocab is that some of these things end up as non standard concepts with this maps to relationship.

So I would need to double check, but I believe that that 200 or two billion 15 for height and I'll pull this up. Right now, it should have a maps to relationship going to the standard concepts.

Gravitt, Jacqueline A *HS 48:34

G

OK.

Houghtaling, Jared 48:34



And that's the relationship that they would use in your ETL when you would do your mapping.

Gravitt, Jacqueline A *HS 48:35

G

Got it.

Houghtaling, Jared 48:39



So maybe you match on that 2 billion.

Concept. Depending on what logic happens upstream, but then you use that concept

relationship in in order to get to the final kind of preferred standard body height concept.



Gravitt, Jacqueline A *HS 48:52

OK.

That makes sense, yeah, because I just started like the person through the code. I was already hitting. And like, wait, why would we force it to a nonstandard if a standard already exists?

OK, I'll I'll take a look at that and I can follow up if there's more questions.



Houghtaling, Jared 49:02

Awesome. All right. Thanks so much.



Gravitt, Jacqueline A *HS 49:03

Cool. Thanks.



Houghtaling, Jared 49:03

Great questions.



Gravitt, Jacqueline A *HS 49:04

Thanks, rod.



Houghtaling, Jared 49:08

All right.

All right.

Anybody else have any flow sheets related questions or otherwise that that we can? Look at.

Eddie, I see your comment now in the chat.

I assume that that's what you had already asked, or that we covered.

So we're all good there, great.

OK.

Well, if there's there's no additional questions.

Yeah. Thank you all for for attending today.

It was an awesome group and a nice discussion.

Like I said, I I I do think that there is potentially interest in getting access to this data set, so I I will follow up about that about how we can coordinate that. Also we can talk Eric offline about maybe what what the the roots might be there.

But again, congratulations to all of you for some awesome data engineering work for actually getting the data.

Like wrangled and up onto the cloud and making all of this this cool stuff possible because I know I think as well as you know that it is not an easy process going from this kind of source mess of data into a even more mess of kind of.

Omop and everything else so.
Yeah. Thank you all for all of your hard work on this.
It's it's very much appreciated.

 **Ashe, William (wa6gz)** 50:38
Harry, that's really cool.

 **Rosenthal, Eric S.,MD** 50:38
Thank you, chair.

 **Ashe, William (wa6gz)** 50:40
It's it's really cool to see the the the platform in progress now you all and the the cloud team have worked really hard on that.
So it's it's cool.
Cool to see that working.

 **Houghtaling, Jared** 50:48
Thanks.

 **Rosenthal, Eric S.,MD** 50:50
Yeah, and.
Maybe a deeper dive to show the the due Datathon project that you showed at the beginning on to one of our exact or all hands meetings would be really encouraging too.

 **Houghtaling, Jared** 51:03
Awesome. Yeah, I can put something like that together.

 **Rosenthal, Eric S.,MD** 51:07
Alright, bye everybody, talk soon.

 **Houghtaling, Jared** 51:08
Fantastic.
Thanks. I'll talk to you next week. Bye.