

CHoRUS B2AI Standards Module Office Hours Transcript

February 15, 2024

0:0:0.0 --> 0:0:0.830

Talapova, Polina

Hello everyone.

0:0:4.190 --> 0:0:4.330

Schmidt, Heidi E.

Hello.

0:0:4.150 --> 0:0:4.470

Alvarez, Marta

Alina.

0:0:7.20 --> 0:0:7.300

Williams, Andrew E

No.

0:1:16.190 --> 0:1:28.430

Williams, Andrew E

Marty, as part of the canoe ongoing feedback about work we're doing to help streamline and consolidate information and make what needs to be done clear to data contributing sites.

0:1:28.980 --> 0:1:45.60

Williams, Andrew E

I like to take a minute to show a draft version of a flow chart of our high level understanding of what data contributing sites are going to be doing that I'm going to be promoting for abuse and.

0:1:46.50 --> 0:2:0.940

Williams, Andrew E

Umm, broader context as a way to organize kind of the relevant activities associated with the different steps for acquiring data, standardizing it, umm and all the other things that data contributing sites are associated, you know?

0:2:1.270 --> 0:2:13.260

Williams, Andrew E

Asked being asked to do just a really high level thing and that would help understand that basic order of operations and steps and be tagged with the links to the documents.

0:2:13.270 --> 0:2:34.810

Williams, Andrew E

That kind of specify what what this different activities involve and I'd like to pass that by people on the call to get their sense of whether it's a useful step in the right direction to help solve some of the issues that people have been, uh, faced with in trying to understand what to do and how to do it.

0:2:35.560 --> 0:2:45.690

Williams, Andrew E

So in addition to that, constructing the areas index, so plan to only take a few minutes to kind of go over that and get a quick temperature taking like this seem like it's worth going forward.

0:2:45.700 --> 0:2:53.730

Williams, Andrew E

It seems like it to us, but it'll be helpful to have other data contributors kind of weigh in and say is this the kind of thing?

0:2:54.580 --> 0:2:57.720

Williams, Andrew E

Once it was finalized, because it's only a draft form that would be useful.

0:3:4.690 --> 0:3:7.790

Alvarez, Marta

And everyone that's on board right now is everyone's familiar.

0:3:7.800 --> 0:3:9.480

Alvarez, Marta

There's one new name I thought I'd give.

0:3:9.490 --> 0:3:11.780

Alvarez, Marta

We could give them a quick chance to introduce themselves.

0:3:12.680 --> 0:3:13.820

Alvarez, Marta

Looks like maybe Zoe.

0:3:15.210 --> 0:3:16.510

Zhou, Zoe

Hi, this is Zoe.

0:3:16.560 --> 0:3:21.790

Zhou, Zoe

I'm the data analyst at Doctor Parks Lab and it's nice to meet you all.

0:3:24.240 --> 0:3:24.730

Alvarez, Marta

Welcome.

0:3:24.800 --> 0:3:25.420

Alvarez, Marta

Nice to meet you too.

0:3:40.140 --> 0:3:40.360

Houghtaling, Jared

Right.

0:3:42.610 --> 0:3:44.260

Houghtaling, Jared

Andrew, so do you want me to kick it off?

0:3:47.460 --> 0:3:47.590

Williams, Andrew E

What?

0:3:50.520 --> 0:3:50.710

Houghtaling, Jared

OK.

0:3:47.990 --> 0:3:57.200

Williams, Andrew E

How do I do the the flow chart 1st and and I'll make sure we don't go beyond 115 and that either enough time for the rest of what you wanna do.

0:3:57.240 --> 0:3:57.630

Houghtaling, Jared

Perfect.

0:3:57.740 --> 0:3:58.400

Houghtaling, Jared

Yep, sounds great.

0:3:59.550 --> 0:4:1.470

Williams, Andrew E

Alright, let's see.

0:4:2.480 --> 0:4:3.780

Williams, Andrew E

Uh, let me.

0:4:7.260 --> 0:4:11.390

Williams, Andrew E

Bring up something fresh off the press.

0:4:12.730 --> 0:4:14.630

Williams, Andrew E

It is a.

0:4:16.510 --> 0:4:27.430

Williams, Andrew E

Version of a document that we think might help to clarify what the overall high level activities are sociated with.

0:4:28.590 --> 0:4:28.990

Williams, Andrew E

Umm.

0:4:30.460 --> 0:4:32.410

Williams, Andrew E

And it's not displaying.

0:4:32.580 --> 0:4:33.40

Williams, Andrew E

Let's see.

0:4:35.80 --> 0:4:37.290

Williams, Andrew E

Let's see if I put something new in.

0:4:38.220 --> 0:4:40.650

Williams, Andrew E

Sorry, I should have had this ready, Bill.

0:4:40.660 --> 0:4:41.180

Williams, Andrew E

Just take a SEC.

0:4:56.480 --> 0:4:56.810

Williams, Andrew E

Hmm.

0:5:4.330 --> 0:5:13.690

Williams, Andrew E

You know, I'm going to show a different, slightly different version of it just because that's the one that's easy to display, but it's got all the same information in it.

0:5:17.20 --> 0:5:17.470

Williams, Andrew E

And.

0:5:20.400 --> 0:5:27.230

Williams, Andrew E

I will increase the font and other size so that it's easily viewed and then share.

0:5:41.670 --> 0:5:42.680

Williams, Andrew E

Can people see that?

0:5:42.690 --> 0:5:44.30

Williams, Andrew E

Is it big enough?

0:5:48.870 --> 0:5:49.10

Schmidt, Heidi E.

The.

0:5:51.280 --> 0:5:51.870

Schmidt, Heidi E.

What is?

0:5:51.920 --> 0:5:52.620

Schmidt, Heidi E.

Oh, that's big.

0:5:55.640 --> 0:5:56.250

Williams, Andrew E

Took a minute.

0:5:56.260 --> 0:5:59.570

Williams, Andrew E

My machine was not behaving the way I want it.

0:5:59.620 --> 0:5:59.990

Williams, Andrew E

Let's see.

0:6:9.150 --> 0:6:21.820

Williams, Andrew E

So this is meant to be just very simple because I get the impression that we don't all have the same consistent understanding or we have not communicated in a way that allows for consistent understanding of justice.

0:6:21.830 --> 0:6:40.30

Williams, Andrew E

The main steps of acquiring the different kinds of data are going to need to be contributed and where that happens, and the approximate order in which it happens, and this is really just meant to cover kind of workflow at data contributing sites is a lot that isn't specified that will happen in the cloud.

0:6:40.40 --> 0:6:43.590

Williams, Andrew E

It isn't meant to specify that it isn't targeting that.

0:6:43.600 --> 0:6:45.220

Williams, Andrew E

We probably need separate things.

0:6:46.190 --> 0:6:46.410

Williams, Andrew E

Uh.

0:6:46.670 --> 0:7:21.520

Williams, Andrew E

That that actually clarify that this is really just for folks at data contributing sites are just saying I'm not sure what all those steps are and what's happening where and in what order and it's meant to not only clarify that, but if we agree to it to have tags for these numbered steps that say you know what's

involved in getting things, what's involved in standardizing and what's involved in joining them to ensure the integrity of them and the identifying for each of the different data source types that are being requested that are required.

0:7:22.110 --> 0:7:56.470

Williams, Andrew E

And so it clearly just lists that we're getting these three different types of things from UHRS the structured flow sheet and unstructured data, the wave form and monitor data and the imaging data that are going to go in raw form somewhere locally and that they're gonna undergo some kind of standardization process that maps concepts, EHR concepts or features of things and the information needed to link them in such a way that they can be used in the integrated fashion we want them to be used for analytics at the end of the day.

0:7:57.380 --> 0:8:13.380

Williams, Andrew E

A additional identity identification will be done on things and that will be submitted to the cloud where it will exist in a draft form and be assessed by people, including people on this team on this call and and others.

0:8:14.980 --> 0:8:17.650

Williams, Andrew E

And feedback will be given to improve.

0:8:17.660 --> 0:8:34.770

Williams, Andrew E

So maybe some things that are needed weren't acquired or they weren't standardized in the way that was intended or there's referential integrity problems or they some aspects of the deidentification wasn't done in a way that we're we're requesting and that needs to be done.

0:8:34.780 --> 0:8:39.120

Williams, Andrew E

And that feedback will be provided and then some improvement.

0:8:39.350 --> 0:8:46.920

Williams, Andrew E

Either we got the data, improve the standardization or some other aspect, and then you'll have another version of this deidentified data set will get submitted, and so on.

0:8:47.20 --> 0:8:52.750

Williams, Andrew E

And and then it'll once it passes all of the different criteria, it will be approved and that will be the final data set.

0:8:52.760 --> 0:9:23.140

Williams, Andrew E

So it's meant to be just this really high, uh level view of where things happen and the kind of things that are happening and to be a tagged, as I said with standard operating procedures that will specify how are

you standardizing your EHR data in more detail, how are you doing all the work to involve involved and joining the different kinds of data to share referential integrity, et cetera.

0:9:23.850 --> 0:9:25.140

Williams, Andrew E

So that's the gist.

0:9:25.150 --> 0:9:26.640

Williams, Andrew E

It's meant to be that clear.

0:9:26.650 --> 0:9:32.360

Williams, Andrew E

And so the green things are steps that would be done by folks at the data contributing sites.

0:9:33.70 --> 0:9:34.560

Williams, Andrew E

Blue things are not.

0:9:34.600 --> 0:9:43.630

Williams, Andrew E

There are things that are being done by people like people in standards module or other other engineers working on things in the cloud to do different kinds of curation.

0:9:45.370 --> 0:9:45.890

Williams, Andrew E

Uh.

0:9:45.930 --> 0:9:53.130

Williams, Andrew E

And so I'll just stop there and ask if folks think that is.

0:9:55.520 --> 0:9:56.130

Williams, Andrew E

Useful.

0:9:56.140 --> 0:10:0.10

Williams, Andrew E

Does it jibe with your understanding of kind of the overall flow of things?

0:10:0.20 --> 0:10:1.220

Williams, Andrew E

Is that, umm?

0:10:3.280 --> 0:10:15.40

Williams, Andrew E

Potentially a helpful kind of document to have that would be sort of a main way to go and get at some of the other information they're associated with the different steps.

0:10:20.10 --> 0:10:20.570

Williams, Andrew E

Anybody have?

0:10:22.900 --> 0:10:23.80

Williams, Andrew E

Yeah.

0:10:19.800 --> 0:10:26.500

Daniel Popham

Yeah, there's been a problem from a Duke, so I'm fairly in the standardization would be just converting it to ohmart, right?

0:10:28.740 --> 0:10:34.690

Williams, Andrew E

Yeah, standardization is going to be mostly about the semantic representation of the information.

0:10:34.700 --> 0:10:36.290

Williams, Andrew E

That's there.

0:10:36.300 --> 0:10:40.50

Williams, Andrew E

I think probably we will cover in the sysops for that.

0:10:40.100 --> 0:10:48.140

Williams, Andrew E

Any standardization of file uh formats and other kinds of things that are meant to happen locally?

0:10:48.960 --> 0:10:50.450

Williams, Andrew E

Umm as well.

0:10:58.460 --> 0:10:58.740

Daniel Popham

OK.

0:10:51.30 --> 0:11:0.200

Williams, Andrew E

But I think for the most part, it's just about the semantic representation of things in OMOP and the compiling of OMOP data, yeah.

0:11:0.820 --> 0:11:1.310

Daniel Popham

All right then.

0:11:1.320 --> 0:11:4.30

Daniel Popham

The my main question was the the step three at the join.

0:11:5.820 --> 0:11:9.360

Daniel Popham

I'm guessing there's gonna be some documentation on what exactly that entails them.

0:11:11.40 --> 0:11:11.240

Williams, Andrew E

Yes.

0:11:10.410 --> 0:11:12.890

Daniel Popham

I'm not sure what we would be joining with integrated data.

0:11:14.870 --> 0:11:48.230

Williams, Andrew E

Well, we want to make sure, for example, when you have waveform and monitor data and imaging data on the same patients in the EHR data that they that you have the IDs that allow you to understand not only who we've got those data on, but when it was acquired, how it relates to the other longitudinal data coming from the EHR on those same patients and that the ID and uh, other referential integrity, umm, strategies that we will disseminate are properly implemented.

0:11:48.300 --> 0:11:48.830

Williams, Andrew E

That's the gist.

0:11:49.980 --> 0:11:58.40

Williams, Andrew E

And so that's we're close to coming up with a document that says what we mean by that, that's in process, but it's not.

0:11:59.670 --> 0:12:2.250

Williams, Andrew E

So that's the kind of thing that would be linked there for their join.

0:12:3.140 --> 0:12:3.430

Daniel Popham

OK.

0:12:3.440 --> 0:12:6.260

Daniel Popham

That's basically kind of like supplemental data or whatever, right?

0:12:9.700 --> 0:12:10.240

Williams, Andrew E

Supplement.

0:12:9.250 --> 0:12:14.940

Daniel Popham

And I know you mentioned imaging supplemental data like imaging at that.

0:12:13.970 --> 0:12:15.480

Williams, Andrew E

Not non EHR data.

0:12:15.790 --> 0:12:16.120

Williams, Andrew E

Yeah.

0:12:16.130 --> 0:12:21.680

Williams, Andrew E

So the data sources are all these on the all the data contributing sites have to contribute all of these things.

0:12:30.770 --> 0:12:31.30

Daniel Popham

OK.

0:12:22.230 --> 0:12:31.930

Williams, Andrew E

These three different types of data from the EHR as well as monitor data from bedsides, waveform and other kinds of monitor data as well as imaging, right.

0:12:32.190 --> 0:12:38.600

Williams, Andrew E

And so those are got all going to be contributed and they have to be linked together in the right way in order to be used in the way that we want.

0:12:39.470 --> 0:12:40.400

Daniel Popham

OK, I understand now.

0:12:40.410 --> 0:12:40.990

Daniel Popham

Alright, thank you.

0:12:41.980 --> 0:12:43.600

Williams, Andrew E

Yep, right.

0:12:48.240 --> 0:12:55.100

Schmidt, Heidi E.

The IT would be good to have a table of contents for the terms that are being used so that.

0:12:57.520 --> 0:12:59.810

Schmidt, Heidi E.

You gave a great explanation right now.

0:13:0.280 --> 0:13:7.190

Schmidt, Heidi E.

It would be great if it were, you know, the standardized data is the transformed.

0:13:7.600 --> 0:13:11.470

Schmidt, Heidi E.

Uh first three bits of information.

0:13:11.480 --> 0:13:18.130

Schmidt, Heidi E.

Joining it into integrated data, which is waveform and monitor imaging.

0:13:21.190 --> 0:13:21.370

Williams, Andrew E

Yep.

0:13:20.750 --> 0:13:24.20

Schmidt, Heidi E.

Which is deidentified it feels like.

0:13:26.850 --> 0:13:32.950

Schmidt, Heidi E.

4 might wannabe after the deidentification to be joined?

0:13:34.190 --> 0:13:34.710

Schmidt, Heidi E.

Umm.

0:13:35.240 --> 0:13:49.920

Schmidt, Heidi E.

So I'm not sure where three and four fit in that workflow, but that's that's what I initially thought when I'm looking at this, this is a great initial idea of how the workflow happens.

0:13:49.930 --> 0:13:54.830

Schmidt, Heidi E.

It's just that the the terms are vague and they need more context.

0:13:57.640 --> 0:13:58.280

Williams, Andrew E

Thanks.

0:13:58.340 --> 0:14:9.620

Williams, Andrew E

Yeah, I think this is meant to be a figure that would be embedded and other tasks that gives context, including some of those definitions.

0:14:10.210 --> 0:14:16.420

Williams, Andrew E

And I think you raised issues that are specified in the legend.

0:14:16.490 --> 0:14:29.730

Williams, Andrew E

Jared raised them also that it's not strictly the case that you would always do things in exactly the order here, but that'll be clarified in the linking SOP documents, but I think you're making excellent points as always.

0:14:29.740 --> 0:14:35.70

Williams, Andrew E

And yeah, this isn't meant to be completely sufficient to help guide everybody.

0:14:35.300 --> 0:15:0.480

Williams, Andrew E

It's meant to be a central figure in a uh, a page that provides more context and where some of the other ancillary documents that specify what these not only what these things mean, but how to actually do them are all you know in one place and have a, you know, a common, umm, reference to this basic workflow.

0:15:0.650 --> 0:15:3.230

Williams, Andrew E

So, but absolutely great points.

0:15:7.380 --> 0:15:14.830

Williams, Andrew E

So given that, does it seem if we were to would address some of those things, you think it's potentially useful thing?

0:15:15.920 --> 0:15:24.810

Williams, Andrew E

How about anybody who thinks it's not or it's not needed or not useful, and I can't see all of the folks who might have their hands raised will.

0:15:26.320 --> 0:15:27.110

Ashe, William (wa6gz)

Hey, yeah, I was.

0:15:27.240 --> 0:15:30.470

Ashe, William (wa6gz)

I was coming mostly on on on highest point.

0:15:30.480 --> 0:15:48.520

Ashe, William (wa6gz)

I think this this is a piece that fits into come what you were getting at the beginning, which is you know we we wanna make things accessible and I think this as you know potentially on that home page of GitHub, I think we could really.

0:15:49.560 --> 0:15:49.720

Williams, Andrew E

Yeah.

0:15:49.920 --> 0:15:50.790

Ashe, William (wa6gz)

Reconfigure that.

0:15:50.800 --> 0:15:54.270

Ashe, William (wa6gz)

So it's not just a mess of links, but there's actually some organization.

0:15:54.580 --> 0:15:57.930

Ashe, William (wa6gz)

It's not even really a mess right now, but it's just kind of a page of links.

0:15:58.180 --> 0:16:4.620

Ashe, William (wa6gz)

There's no direction or flow to understanding and getting at the information you need to kind of have to hunt report.

0:16:4.660 --> 0:16:9.260

Ashe, William (wa6gz)

So I think this could help channel not the negative opinion you were looking for.

0:16:12.480 --> 0:16:12.800

Williams, Andrew E

Good.

0:16:9.270 --> 0:16:13.440

Ashe, William (wa6gz)

All positive here but but but yeah it looks good.

0:16:14.440 --> 0:16:14.930

Williams, Andrew E

Thanks.

0:16:15.120 --> 0:16:15.450

Williams, Andrew E

Thanks.

0:16:15.460 --> 0:16:16.150

Williams, Andrew E

Well, I'm.

0:16:16.560 --> 0:16:19.510

Williams, Andrew E

We've gone 2 minutes or when I said we would, so thanks very much for your feedback.

0:16:19.520 --> 0:16:22.640

Williams, Andrew E

If you have any other comments you want to share about this.

0:16:23.480 --> 0:16:28.780

Williams, Andrew E

Umm, please feel free to put them in the chat or send them asynchronously.

0:16:28.790 --> 0:16:29.230

Williams, Andrew E

Thanks so much.

0:16:30.860 --> 0:16:32.680

Williams, Andrew E

Please take it away, Jared.

0:16:35.650 --> 0:16:36.380

Houghtaling, Jared

Alright, awesome.

0:16:36.450 --> 0:16:36.610

Houghtaling, Jared

Yeah.

0:16:36.620 --> 0:16:37.400

Houghtaling, Jared

Thanks again for that.

0:16:39.110 --> 0:16:40.180

Houghtaling, Jared

So I got.

0:16:42.670 --> 0:16:50.250

Houghtaling, Jared

We've kind of been discussing about where or this idea of being able to submit your data quality results.

0:16:50.260 --> 0:16:52.180

Houghtaling, Jared

This has come up, I think, in multiple office hours.

0:16:53.120 --> 0:16:54.360

Houghtaling, Jared

Both how to generate them?

0:16:54.370 --> 0:16:57.190

Houghtaling, Jared

We've talked about what that generation looks like.

0:16:57.200 --> 0:17:2.680

Houghtaling, Jared

Some of the motivations behind things like the data quality dashboard and the keilis and how those feed into areas.

0:17:4.390 --> 0:17:20.510

Houghtaling, Jared

And then we're at the stage where I'm working together with the team at Central kind of MGH cloud to bring up the services that will effectively host berries and host kind of this upload portal for where you will place some of these quality control files.

0:17:21.30 --> 0:17:28.670

Houghtaling, Jared

Uh, things that that have executed checks against your OMOP data and there's been some active discussions on the threads.

0:17:28.680 --> 0:17:39.510

Houghtaling, Jared

So Heidi, I think had kind of walked through her thought process in terms of how to execute these various tools from the Odyssey community that will run these checks.

0:17:40.160 --> 0:17:40.870

Houghtaling, Jared

Umm.

0:17:41.240 --> 0:17:43.0

Houghtaling, Jared

And it's kind of gone back and forth on that.

0:17:43.10 --> 0:18:5.860

Houghtaling, Jared

And I I wanted to also or in that light I wanted to kind of clarify what the process is for creating one of these exports, what this actually looks like practically, what that export then how the, what the structure should look like, how the file should be arranged and then maybe more importantly what those files actually practically contain and why that that's useful for areas and for us in a broader context.

0:18:7.630 --> 0:18:22.610

Houghtaling, Jared

So with that kind of context in mind maybe I'll just get started here on the red on the discussions that that Heidi had posted and in response to that I kind of posted this general structure of an areas index.

0:18:27.820 --> 0:18:38.360

Houghtaling, Jared

And before I dive in, are there any any questions related to that or any topics related to areas that that I didn't mention potentially or that you would like me to discuss?

0:18:40.760 --> 0:18:41.180

Houghtaling, Jared

As well.

0:18:47.480 --> 0:18:49.260

Houghtaling, Jared

Then I'll I'll go ahead and share my screen here.

0:18:49.320 --> 0:18:49.440

Houghtaling, Jared

Yeah.

0:18:52.240 --> 0:18:52.620

Houghtaling, Jared

Let's see.

0:18:54.480 --> 0:18:56.90

Houghtaling, Jared

OK, so.

0:18:59.750 --> 0:19:10.800

Houghtaling, Jared

So this is the discussion page in Chorus and you have this thread that is talks about kind of this idea of the running the data quality dashboard and and areas indexer.

0:19:11.130 --> 0:19:17.400

Houghtaling, Jared

How do you then kind of walked through her thought process and her flow and doing this and then kind of the issues that she encountered?

0:19:18.30 --> 0:19:29.380

Houghtaling, Jared

I tried to answer those, so I'm not gonna go through this in in explicit detail, but take a look if you're interested in that ohm but one of the things that came up was what this actually produces, right?

0:19:29.390 --> 0:19:37.60

Houghtaling, Jared

So if you were to run all of the steps necessary to create your areas index that you will be submitting what it looks like.

0:19:37.110 --> 0:19:48.650

Houghtaling, Jared

And here I tried to and I apologize, I realize probably small, but what what we'll get produced at the end of this is a dated directory.

0:19:48.660 --> 0:20:4.840

Houghtaling, Jared

It's your source release directory, and it'll have a date that looks exactly like this, but for whatever date you decide to run this on, and inside that directory you have kind of a many many files like thousands of files that are all nested together in a subdirectory structure.

0:20:5.620 --> 0:20:10.630

Houghtaling, Jared

Umm, the top of which will be kind of high level overview and aggregates.

0:20:10.640 --> 0:20:28.630

Houghtaling, Jared

So like this Achilles performance file gives us insight into how long each Achilles check took to execute and we can kind of compare those performances across the network to see maybe are there are there things that are slowing down certain sites that that could be improved.

0:20:28.720 --> 0:20:39.320

Houghtaling, Jared

We have the general CDM source information, which if you've transformed your data, you know you feel the CDM source table with information about who owns the the data and who transformed it.

0:20:39.330 --> 0:20:41.880

Houghtaling, Jared

And the documentation location and things like that high level metadata.

0:20:43.50 --> 0:20:43.490

Houghtaling, Jared

Umm.

0:20:43.490 --> 0:20:50.720

Houghtaling, Jared

And then the the actually the bulk of what you're producing in this export is in this concepts directory umm.

0:20:51.430 --> 0:20:51.930

Houghtaling, Jared

And So what?

0:20:53.560 --> 0:20:54.650

Houghtaling, Jared

It's officially it's.

0:20:54.660 --> 0:21:7.970

Houghtaling, Jared

It's an A keyless process that's that's helping generate a lot of this, but so it's in the Achilles package and I can actually, I'll show you this function that's doing that and kind of talked about this in various detail.

0:21:8.270 --> 0:21:25.140

Houghtaling, Jared

But there's this export to Aries function within the Achilles package at that is then going through and it's creating not only kind of these high level reports that I just talked about, but also the very individual concept level and it's creating each of them as a Jason file.

0:21:25.150 --> 0:21:38.750

Houghtaling, Jared

So each concept gets its own nifty Jason that has different distributions of like prevalence over time and stratifications by gender, and all of these cool things that we end up seeing visualized in Aries.

0:21:38.960 --> 0:21:47.580

Houghtaling, Jared

Those are done for every concept in every unique concept I should say across these domains in your database, in your OMOP database.

0:21:50.870 --> 0:21:55.40

Houghtaling, Jared

So the our script that does this and that's that's gonna be here.

0:21:55.50 --> 0:21:59.940

Houghtaling, Jared

It's actually going through and it's running kind of this L apply command.

0:22:0.250 --> 0:22:9.400

Houghtaling, Jared

You can kind of see it's doing it for each domain and it's setting up a report as it's called with different uh, more or less different is it?

0:22:9.410 --> 0:22:12.270

Houghtaling, Jared

It's a JSON report with kind of this key value pair structure.

0:22:13.370 --> 0:22:13.760

Houghtaling, Jared

Umm.

0:22:14.470 --> 0:22:21.540

Houghtaling, Jared

And so it's doing this concept by concept and what we had seen actually on our side at Tufts is this was taking quite some time to execute.

0:22:21.970 --> 0:22:43.710

Houghtaling, Jared

So before I also dive into some of the manual here, I just recently this this week finalized a refactor of this process so that this stuff instead of being done in series umm in our I have SQL queries that will do this in the database and produce then a table that has all these nice JSON reports and it does it vectorized much faster.

0:22:44.0 --> 0:22:51.310

Houghtaling, Jared

So if you're interested in that, if you've gotten to this point, let me know and hopefully I can help alleviate some of the potential slowdown from doing that.

0:22:52.120 --> 0:22:57.740

Houghtaling, Jared

If you only have a few 100 or a few 1000 concepts in your, I'm in your database.

0:22:57.750 --> 0:23:4.730

Houghtaling, Jared

This shouldn't probably be a huge issue, but if you're talking about 10s of thousands or hundreds of thousands of unique concepts, this could be somewhat time consuming.

0:23:6.830 --> 0:23:7.250

Houghtaling, Jared

Umm.

0:23:7.390 --> 0:23:26.370

Houghtaling, Jared

OK so and I can I have I have a query now pulled up of what this looks like so again now now we're into the, the, the kind of Jason reports for each concept and this is just an example of one such report that gets produced and it looks like this.

0:23:26.430 --> 0:23:34.210

Houghtaling, Jared

So you end up with all of these key value pairs and they have, like I said, they have stratifications of like different distributions.

0:23:34.220 --> 0:23:40.950

Houghtaling, Jared

You have min values, 10 percentile, 25 percentile, things like that across the females for this particular drug.

0:23:41.570 --> 0:23:46.710

Houghtaling, Jared

For males in this particular drug you have then like prevalence over time stratified by month.

0:23:47.670 --> 0:23:50.800

Houghtaling, Jared

This is what's being then produced, and this is what's being visualized.

0:23:50.810 --> 0:24:14.970

Houghtaling, Jared

If you go to the actual areas web application here, when we take a look at like a source release or something like this where we look at conditions and you can click on one of these uh and you can view right this this count proportion by month then you have your Pi type and you have your distribution across genders.

0:24:15.320 --> 0:24:23.770

Houghtaling, Jared

So what arisen is doing is it's taking these Jason files that have been produced per concept and it's been rendering them in this nice format.

0:24:24.360 --> 0:24:32.600

Houghtaling, Jared

And the cool thing about all of this is, is that you can get kind of this very nice concept level information without needing a database connection.

0:24:32.610 --> 0:24:40.10

Houghtaling, Jared

It's producing a static set of files for you that give you this very rich overview of of the concepts.

0:24:41.890 --> 0:24:42.140

Houghtaling, Jared

OK.

0:24:42.670 --> 0:24:43.440

Houghtaling, Jared

Ohm.

0:24:44.650 --> 0:24:48.600

Houghtaling, Jared

But that's then that I I guess is the the the meat of what you're going to produce.

0:24:48.610 --> 0:24:57.350

Houghtaling, Jared

And then in addition to those you have again these summary files that are also produced by the same export to Aries process with one exception.

0:24:58.480 --> 0:25:4.330

Houghtaling, Jared

Umm this DQD result dot Jason file and that comes directly out of the data quality dashboard.

0:25:4.600 --> 0:25:8.390

Houghtaling, Jared

So that is not produced by this script.

0:25:8.800 --> 0:25:26.790

Houghtaling, Jared

It needs to be placed in this proper location, but it is produced by your data quality dashboard execution and there I think we also we've gone through what that looks like how to run that in in our our studio wherever you're planning to run it and there's ways to run it.

0:25:26.800 --> 0:25:31.370

Houghtaling, Jared

So it populates the database table and you can export that database table or you can run it directly.

0:25:31.760 --> 0:25:31.990

Houghtaling, Jared

Kind of.

0:25:32.0 --> 0:25:40.910

Houghtaling, Jared

The data frame format within R there's lots of options to do that, but this is coming out of a separate process and that's the main exception.

0:25:41.170 --> 0:25:46.900

Houghtaling, Jared

There's one other file in here, this temporal characterization, which is actually generated by.

0:25:47.950 --> 0:25:48.370

Houghtaling, Jared

Umm.

0:25:48.510 --> 0:25:53.130

Houghtaling, Jared

Another our process in Achilles, you can see this perform temporal characterization.

0:25:53.140 --> 0:25:54.320

Houghtaling, Jared

And here I'm in the.

0:25:54.330 --> 0:25:55.200

Houghtaling, Jared

This is the odyssey.

0:25:55.210 --> 0:26:5.540

Houghtaling, Jared

Achilles, GitHub we're talking about, but there's this performed temporal characterization script and that's actually going through concepts that have at least three years worth of data.

0:26:5.810 --> 0:26:12.520

Houghtaling, Jared

And it's looking at kind of more specific longitudinal metrics to describe those concepts.

0:26:12.680 --> 0:26:28.980

Houghtaling, Jared

I guess at a high level and so this is another process that needs to be run after you do that export to Aries and what that process does is it goes through and it then creates for you this additional supplemental file temporal.

0:26:30.260 --> 0:26:38.290

Houghtaling, Jared

It is not the requirements like so there's certain concepts that you simply don't have three years of longitudinal data and so they're not fulfilled.

0:26:38.360 --> 0:27:2.990

Houghtaling, Jared

So areas doesn't fail if you don't have this file available, but it's likely that you will have that level of longitudinal data and it's it's good to run that process as well to make sure this is umm, but then the rest of these these are just CSV files that kind of give an overview, either per concept or per domain of different counts of relevant information and those counts.

0:27:3.0 --> 0:27:12.210

Houghtaling, Jared

Then, like I said, are either rendered at the concept level, or they're rendered at like the data source level where you have things like domain continuity.

0:27:12.220 --> 0:27:14.430

Houghtaling, Jared

I guess this is the bad example to show you here.

0:27:14.440 --> 0:27:27.60

Houghtaling, Jared

I can show you maybe one that has a bit more the view on things, but they they help generate reports like this that show you the number of concepts over time for a given a given source.

0:27:27.190 --> 0:27:37.80

Houghtaling, Jared

In this case, data source would be something like Columbia or Tufts or MIT, and the data source release would be a dated OMOP instance from that particular site.

0:27:39.960 --> 0:27:40.380

Houghtaling, Jared

OK.

0:27:40.540 --> 0:27:48.440

Houghtaling, Jared

And so once you've executed those processes, so I think here you need to have Achilles executed and it's normal state.

0:27:48.450 --> 0:27:50.660

Houghtaling, Jared

So that's that's running these characterization checks.

0:27:50.800 --> 0:28:10.70

Houghtaling, Jared

You need to then execute DQD that's running the quality metrics like for plausibility, conformance and completeness, and then you run this Aries export, which I've just kind of gone through and and like I said, the the biggest thing to note here is that this process is just generating her concept.

0:28:10.260 --> 0:28:22.360

Houghtaling, Jared

These nifty JSON representations of things like prevalence or stratifications across, uh, different elements of interest, and that's then producing all of your knife.

0:28:22.440 --> 0:28:26.10

Houghtaling, Jared

The files that areas will need to to to render.

0:28:27.140 --> 0:28:32.60

Houghtaling, Jared

Once you run that areas export, then you can run the perform temporal characterization which I just noted.

0:28:33.170 --> 0:28:40.420

Houghtaling, Jared

Umm there's then an areas indexer process called augment concept files and that's in a different repository.

0:28:40.430 --> 0:28:42.450

Houghtaling, Jared

I can actually run to that now just a second.

0:28:45.970 --> 0:28:49.70

Houghtaling, Jared

UM, that's been here and that's that.

0:28:49.80 --> 0:28:50.60

Houghtaling, Jared

Should be a quick process.

0:28:50.70 --> 0:29:0.480

Houghtaling, Jared

It actually parses it steps kind of through each of those concept files and it adds a couple of key value pairs depending on whether there was temporal characterization and like one or two other aspects.

0:29:2.50 --> 0:29:3.410

Houghtaling, Jared

Testing here.

0:29:4.170 --> 0:29:6.820

Houghtaling, Jared

Uh, OK idea is asking.

0:29:6.830 --> 0:29:8.810

Houghtaling, Jared

So I wanna make sure I'm also getting to what's in the chat.

0:29:9.690 --> 0:29:14.80

Houghtaling, Jared

So how does the export to aries.org get executed right?

0:29:14.130 --> 0:29:15.200

Houghtaling, Jared

How do you actually run this?

0:29:17.360 --> 0:29:23.520

Houghtaling, Jared

I believe if we look at the extras in the Achilles repo, it depends.

0:29:25.220 --> 0:29:25.630

Houghtaling, Jared

Yeah.

0:29:25.680 --> 0:29:32.240

Houghtaling, Jared

So here I think they have an example of running export to Aries, not mistake.

0:29:38.610 --> 0:29:41.960

Houghtaling, Jared

The should be either it's here or maybe it's actually.

0:29:41.970 --> 0:29:42.580

Houghtaling, Jared

I believe in that.

0:29:42.590 --> 0:29:43.380

Houghtaling, Jared

Everything, Dexter.

0:29:45.740 --> 0:29:45.990

Houghtaling, Jared

Yeah.

0:29:46.0 --> 0:29:49.870

Houghtaling, Jared

So we have this code to run and and I think this is maybe where some of the confusion came up there.

0:29:49.880 --> 0:29:54.100

Houghtaling, Jared

This is at the moment still outdated.

0:29:54.110 --> 0:30:8.120

Houghtaling, Jared

There's an issue that's not been yet resolved on this repository, but what the areas indexer gives you is kind of this process that I've just described where you're running Achilles first.

0:30:8.590 --> 0:30:10.220

Houghtaling, Jared

You're running the temporal characterization.

0:30:10.230 --> 0:30:12.230

Houghtaling, Jared

In this case, you're running the data quality dashboard.

0:30:13.130 --> 0:30:14.120

Houghtaling, Jared

Umm.

0:30:14.750 --> 0:30:18.170

Houghtaling, Jared

And then you have this export AO process here.

0:30:19.200 --> 0:30:22.360

Houghtaling, Jared

And that's actually currently now called export to Aries.

0:30:22.370 --> 0:30:34.250

Houghtaling, Jared

So this is a misnamed function, but the way that you run it is the same, so you need to provide it with these arguments and the probably the the biggest most important argument is this output path.

0:30:35.800 --> 0:30:37.830

Houghtaling, Jared

That's where all of these files will be generated.

0:30:39.860 --> 0:30:50.80

Houghtaling, Jared

You want to make sure that that's consistent across your different processes, like that this JSON did he result is ending up in the same location as, but this is actually the function that would be called.

0:30:50.390 --> 0:30:57.300

Houghtaling, Jared

Like I said, with a different naming convention instead of export AO it'll be export to Ares which you can find.

0:31:1.120 --> 0:31:6.330

Houghtaling, Jared

And that's been calling you Scroll down in this file through all of the crazy report generation.

0:31:9.210 --> 0:31:9.510

Houghtaling, Jared

Got it.

0:31:11.640 --> 0:31:28.450

Houghtaling, Jared

That's this function that's being called keeper and there's some documentation describing what's the inputs and outputs, but more or less dysfunction is, like I said, creating those those various files that are populating this directory to create this structure.

0:31:29.100 --> 0:31:50.380

Schmidt, Heidi E.

OK, so you're loading the export to Aries as a a library and then you're invoking the export to Aries code in the format that was listed in the areas indexer with the name that is in the code that you loaded.

0:31:51.320 --> 0:31:54.850

Houghtaling, Jared

So the library is just Achilles, so you don't need to reload any additional libraries.

0:31:54.860 --> 0:31:55.910

Houghtaling, Jared

This is a part.

0:31:55.920 --> 0:32:2.230

Houghtaling, Jared

It's a public function in the Achilles library, so as long as you have Achilles accessible in your workspace, you'll be able to run this function.

0:32:3.40 --> 0:32:13.380

Houghtaling, Jared

Umm, you just and and in this case in the code to run that I showed that they have the, they just say this is an A Kiwis function called export AO that's all you need.

0:32:13.440 --> 0:32:15.650

Houghtaling, Jared

There's no additional libraries or other packages.

0:32:15.920 --> 0:32:17.570

Houghtaling, Jared

This comes when you load a keilis.

0:32:17.660 --> 0:32:19.270

Houghtaling, Jared

You have this capability as well.

0:32:23.610 --> 0:32:23.920

Houghtaling, Jared

Thanks.

0:32:28.70 --> 0:32:28.440

Houghtaling, Jared

OK.

0:32:28.930 --> 0:32:32.20

Houghtaling, Jared

But once you've done that, right, this is gonna take some time to run.

0:32:32.30 --> 0:32:33.50

Houghtaling, Jared

It creates a bunch of files.

0:32:33.930 --> 0:32:39.590

Houghtaling, Jared

Uh, and you end up with this this structure with a source release folder.

0:32:41.760 --> 0:33:14.840

Houghtaling, Jared

What you can do if and and I think we've also this is kind of where you start to delineate between if you wanna set up areas locally at your site and use it for iterative development of your ETL and to kind of have a view on your own sites data over time, you do need to then go through an additional step which is shown here that actually goes through and it builds an index of data not only in this directory but also in any other dated release directory that you have and that then gives you kind.

0:33:14.960 --> 0:33:40.370

Houghtaling, Jared

Of an areas these overviews that compare umm this type of thing where you have like different data sources and you have overtime what those sources look like or actually more likely different source releases where you have different sources and and what they look like over time how many people are included how many quality issues you have things like that.

0:33:41.430 --> 0:33:42.150

Houghtaling, Jared

Umm.

0:33:43.280 --> 0:33:45.750

Houghtaling, Jared

So, but that's only if you want to run it locally.

0:33:45.760 --> 0:33:52.550

Houghtaling, Jared

If you don't wanna run it locally and set this up on your own site also fine, you can just stop at this point.

0:33:52.560 --> 0:34:6.960

Houghtaling, Jared

Once you have this directory created, you can zip this up into a single kind of zip file and then you will go to a like upload portal that we've set up in the central location.

0:34:7.650 --> 0:34:17.500

Houghtaling, Jared

What I should also say is that it sounds like and this is still being finalized, but it sounds like the data managers per site at least one designated person per site.

0:34:18.280 --> 0:34:25.800

Houghtaling, Jared

We'll be invited as a guest like and receive kind of a guest account at the MGH.

0:34:27.920 --> 0:34:40.60

Houghtaling, Jared

Uh cloud instance and that will then allow you to access the portal to upload the zipped file and it will also allow you to view the consortium wide Aries view.

0:34:40.70 --> 0:35:0.540

Houghtaling, Jared

Like what I'm showing here for our Tufts data, but across all the various sites, and also the source releases per site, because I guess practically speaking, what will happen is you'll upload that and then we have some automated processes at the central location that will place it into this nice directory structure that has the organization per site and per release.

0:35:0.670 --> 0:35:2.870

Houghtaling, Jared

And then we will run at the central location.

0:35:4.10 --> 0:35:12.970

Houghtaling, Jared

These particular processes, in order to construct kind of this network index and these comparisons that allow visualizations that show.

0:35:13.200 --> 0:35:16.470

Houghtaling, Jared

Here's how much data Columbia has versus MIT.

0:35:16.480 --> 0:35:19.60

Houghtaling, Jared

And here are the data quality issue comparisons and things like that.

0:35:19.960 --> 0:35:42.40

Houghtaling, Jared

Umm so I guess all that to say, this is the this is the structure you're aiming for and you're trying to create a source release directory that has this set of kind of characterization files of your data source that can then be rendered by Aries, Umm and.

0:35:42.50 --> 0:35:46.290

Houghtaling, Jared

The reason I guess why we would ask you to do this.

0:35:48.360 --> 0:36:0.600

Houghtaling, Jared

I guess the initial intent was that this would give us a way to provide feedback to you about the quality of your data without you having to submit all of it at once and have us run this on our side.

0:36:0.650 --> 0:36:5.790

Houghtaling, Jared

So if you imagine it's much quicker and easier to submit a zip file, that's like total.

0:36:5.800 --> 0:36:29.990

Houghtaling, Jared

Maybe this is gonna be like 50 megabytes versus having to send gigabytes of data and have us then run these processes on the central server every time you want some sort of review of the data, and so then I guess looking forward my my understanding is that as you're updating your ETL, you would continue to just have a processor pipeline that produces these source releases.

0:36:30.260 --> 0:36:34.70

Houghtaling, Jared

You upload those source releases, we review them, we kind of give you feedback based on that.

0:36:34.200 --> 0:36:46.400

Houghtaling, Jared

You can also see them in the broader scope of the rest of the sites and like kind of iteratively improve whatever processes you're using to standardize map HR data that we discussed in other.

0:36:50.0 --> 0:36:52.430

Houghtaling, Jared

All right, that was a lot of information in a short time.

0:36:53.0 --> 0:37:5.420

Houghtaling, Jared

Umm, I'm gonna pause there for a second, just to see if if any of you have questions about either this process or any of the kind of offhand comments I've made throughout the way.

0:37:12.210 --> 0:37:13.950

Dandayudhapani, Satish

I'm quick question the Satish.

0:37:14.410 --> 0:37:14.570

Houghtaling, Jared

Yeah.

0:37:14.690 --> 0:37:19.620

Dandayudhapani, Satish

So these reports are executed on the entire data set on our side?

0:37:19.630 --> 0:37:22.250

Dandayudhapani, Satish

Or is it just the the cohort we're gonna send?

0:37:23.430 --> 0:37:26.680

Houghtaling, Jared

Great question and it depends on where you're launching this by.

0:37:26.690 --> 0:37:35.140

Houghtaling, Jared

By default, if you don't specify in both DQD and Achilles, you can specify like a cohort number to run those checks against.

0:37:35.390 --> 0:37:40.120

Houghtaling, Jared

Uh, if you do not specify that, it will run it across all data in your OMOP tables.

0:37:43.330 --> 0:37:57.780

Houghtaling, Jared

One other way that you could go about this would be to to create a separate schema entirely that only contains that cohort that you're of of OMOP thing that you want so you can actually separate that out and then run it against that separate schema against everything.

0:37:58.90 --> 0:37:59.570

Houghtaling, Jared

So I guess those are your two options either.

0:38:1.110 --> 0:38:2.560

Dandayudhapani, Satish
But you're you.

0:38:2.610 --> 0:38:4.780

Dandayudhapani, Satish
You're from the central site of it.

0:38:4.850 --> 0:38:6.180

Dandayudhapani, Satish
You guys are fine either way.

0:38:6.520 --> 0:38:6.990

Dandayudhapani, Satish
Is that right?

0:38:14.650 --> 0:38:14.870

Dandayudhapani, Satish
OK.

0:38:7.890 --> 0:38:18.470

Houghtaling, Jared

I would say that that we're probably most concerned with this on the cohort that you will be delivering if you if you don't wanna run this on all of the other data, I think that that's probably fine.

0:38:18.560 --> 0:38:19.170

Houghtaling, Jared
Think it's?

0:38:19.240 --> 0:38:27.310

Houghtaling, Jared

We wanna have a quality view on specifically the extracts that are coming in to the core of central location.

0:38:28.110 --> 0:38:29.390

Dandayudhapani, Satish
OK. Thanks.

0:38:30.20 --> 0:38:33.170

Houghtaling, Jared

And like I said, there's multiple ways of actually practically doing that.

0:38:33.180 --> 0:38:38.410

Houghtaling, Jared

You can specify a cohort identifier in both of those.

0:38:38.420 --> 0:38:41.350

Houghtaling, Jared

Upstream processes were to do that.

0:38:41.360 --> 0:38:49.780

Houghtaling, Jared

You would even need to like fill your cohort table either with Atlas or some other tool that then it would say here's my cohort number and this person is included.

0:38:49.910 --> 0:39:4.790

Houghtaling, Jared

It's more or less and we've I think we've shown that at one time or another where you can kind of define a cohort in an odyssey specific way and then you can reference that cohort identifier in DQD and Achilles and that will then limit your your quality checks to that scope.

0:39:7.130 --> 0:39:7.410

Dandayudhapani, Satish

Thanks.

0:39:14.510 --> 0:39:14.980

Houghtaling, Jared

All right.

0:39:15.30 --> 0:39:20.580

Houghtaling, Jared

Any any other questions on on what this process flow looks like or how to actually run it?

0:39:24.10 --> 0:39:27.390

Schmidt, Heidi E.

I'm probably doing it incorrectly, but I'm getting an error so.

0:39:31.690 --> 0:39:34.70

Schmidt, Heidi E.

I can follow up in the GitHub discussion.

0:39:33.700 --> 0:39:35.480

Houghtaling, Jared

Yep, that's great.

0:39:35.890 --> 0:39:36.100

Houghtaling, Jared

Yeah.

0:39:36.110 --> 0:39:40.310

Houghtaling, Jared

And I don't know how many of you have actually tried or started to run Achilles and DQD.

0:39:42.160 --> 0:39:48.400

Houghtaling, Jared

I think based on the the status on GitHub, it seems like at least five or six sites have already done this.

0:39:51.350 --> 0:40:5.740

Houghtaling, Jared

But yeah, if if you've been able to successfully execute these packages and you're at this point, UM, sounds like Satish or you've, I think we already went through some of this, you've been able to kind of generate run some of these processes.

0:40:4.890 --> 0:40:6.680

Dandayudhapani, Satish

Yeah, the right.

0:40:6.690 --> 0:40:9.10

Dandayudhapani, Satish

The only thing is the last step, which is the umm.

0:40:9.90 --> 0:40:11.140

Dandayudhapani, Satish

Ohh export to Aries.

0:40:11.550 --> 0:40:13.300

Houghtaling, Jared

Got you. OK.

0:40:11.150 --> 0:40:19.400

Dandayudhapani, Satish

That's the only one I haven't done, but that was previously we have done on the DQD, which is a we've been iterating on the data.

0:40:19.950 --> 0:40:20.130

Houghtaling, Jared

OK.

0:40:19.450 --> 0:40:22.450

Dandayudhapani, Satish

So I had to run the fresh new set of reports, yeah.

0:40:23.150 --> 0:40:23.930

Houghtaling, Jared

OK, awesome.

0:40:25.850 --> 0:40:26.610

Houghtaling, Jared

Alright, good to hear.

0:40:29.370 --> 0:40:29.570

Houghtaling, Jared

Cool.

0:40:32.110 --> 0:40:32.570

Houghtaling, Jared

Alright.

0:40:32.720 --> 0:40:33.150

Houghtaling, Jared

Uh.

0:40:33.510 --> 0:40:58.390

Houghtaling, Jared

Any any further questions on what this looks like or or I guess maybe one thing that I can do is or I will plan to do is I'll take this code to run uh and I will upload a version of this today that is fixed more or less and that also has some comments with the certain things that we have discussed today and also in other points in the office hours.

0:40:58.390 --> 0:41:4.660

Houghtaling, Jared

So I'll make sure that if you want to have the specific code that you can execute, you can do that here.

0:41:4.710 --> 0:41:6.750

Houghtaling, Jared

I know I I talked with Dell as well.

0:41:6.760 --> 0:41:9.780

Houghtaling, Jared

I think she's been trying to run some of these these processes, so.

0:41:11.850 --> 0:41:17.280

Houghtaling, Jared

Yeah, I'll try to make sure that that's accessible and hopefully easy to kind of follow.

0:41:17.290 --> 0:41:20.410

Houghtaling, Jared

What steps are happening in which parts of this script?

0:41:28.120 --> 0:41:28.440

Houghtaling, Jared

All right.

0:41:32.290 --> 0:41:40.330

Houghtaling, Jared

Then I I don't have much more on at least constructing the index, or really what it includes.

0:41:40.340 --> 0:41:56.530

Houghtaling, Jared

I mean I have, like I said, I have scripts that are that execute a lot of these SQL processes or a lot of these are processes to create the jasons they execute them actually in the database or in SQL using kind of different JSON constructors.

0:41:57.300 --> 0:42:1.720

Houghtaling, Jared

Uh, and that just speeds it up considerably on our side.

0:42:1.790 --> 0:42:13.230

Houghtaling, Jared

And like I said, I'll I can also post those scripts that with the resources from today, but I did, maybe it's worth talking about.

0:42:14.20 --> 0:42:38.250

Houghtaling, Jared

So Andrew showed A and I pretty awesome diagram and overview for sites in the first part of this meeting and I don't think that we've talked about this, but I did wanna also mention that I posted umm this particular discussion and I've pinned it in the discussion thread with kind of a similar intention of giving sites a bit of an overview of what.

0:42:40.810 --> 0:42:54.60

Houghtaling, Jared

Kind of the different steps along the way and then some descriptions of those steps as well as the resources both in terms of code itself and office hour videos that might get at some of these, these certain elements.

0:42:54.890 --> 0:43:2.970

Houghtaling, Jared

I realize that there's a lot of detail here, and so my intention is not to like get you lost in the weeds of all of this, but more.

0:43:3.350 --> 0:43:13.920

Houghtaling, Jared

We've talked about a lot of topics and a lot of potentially relevant topics at these office hours and they're kind of now all just in the office Hour recording directory.

0:43:15.390 --> 0:43:33.680

Houghtaling, Jared

But I think it's it's useful to place them in a broader context of like, where does this fall along the cut, let's say, pipeline of my source data and it's different formats down to a data extract that's say curated it in the format that's that chorus has requested.

0:43:35.630 --> 0:43:35.880

Houghtaling, Jared

Uh.

0:43:35.890 --> 0:43:39.0

Houghtaling, Jared

So if you haven't seen it already, I encourage you to take a look at this.

0:43:39.90 --> 0:43:50.730

Houghtaling, Jared

It's likely that we'll the the figure that Andrew showed will also have some sort of equivalent to this, where we have, let's say, documentation or descriptions for each of those steps that he showed.

0:43:51.480 --> 0:43:52.330

Houghtaling, Jared

Umm.

0:43:52.700 --> 0:44:2.870

Houghtaling, Jared

And like I said, this is in this with a slightly different angle to it, but I do think that there there is relevant information and hopefully it can help resolve some of the questions that are outstanding.

0:44:4.260 --> 0:44:12.190

Houghtaling, Jared

If you do have additional questions, if you're going through this, feel free to comment on this discussion and I'll I'll I should be able to get back to you on that relatively quickly.

0:44:12.820 --> 0:44:21.690

Houghtaling, Jared

Umm, but that's just to make you aware that this resource is here and that you do now have a bit more of an overview of of how this actually looks.

0:44:25.0 --> 0:44:25.180

Houghtaling, Jared

Right.

0:44:25.910 --> 0:44:30.620

Houghtaling, Jared

Uh, so I'm well ahead of of of schedule here.

0:44:30.630 --> 0:44:50.30

Houghtaling, Jared

I see and I don't have many additional things to to to touch on, so I can either open it up to see if there are other things that that you all would like to discuss or we can, I guess, call it a day and then give you some big time to time back.

0:44:50.420 --> 0:44:51.860

Houghtaling, Jared

I don't know what people would prefer.

0:44:58.680 --> 0:45:0.490

Daniel Popham

Hey uh, this is Daniel from Duke.

0:45:0.500 --> 0:45:9.280

Daniel Popham

Again, I was just meeting with my team earlier today and they were talking about doing some comparison of the data.

0:45:9.570 --> 0:45:11.480

Daniel Popham

See where gap analysis could occur?

0:45:12.140 --> 0:45:13.840

Daniel Popham

And they were talking about Aries.

0:45:13.940 --> 0:45:15.900

Daniel Popham

I'm in the middle of installing areas right now.

0:45:15.910 --> 0:45:17.510

Daniel Popham

Is that the right tool to do all of that stuff?

0:45:19.350 --> 0:45:19.960

Houghtaling, Jared

Uh, yeah.

0:45:19.970 --> 0:45:29.580

Houghtaling, Jared

So I guess it it depends on on what specifically you're looking at, but I think areas will give you some nice visual overviews of the data quality and the characterization.

0:45:29.650 --> 0:45:37.200

Houghtaling, Jared

And I think for me, what's maybe the the more important part of areas is like this longitudinal view on your data transformation.

0:45:37.210 --> 0:45:45.540

Houghtaling, Jared

So being able to compare both the quality and the coverage of the data from one source release to the next.

0:45:45.540 --> 0:45:46.520

Daniel Popham

Another OK.

0:45:45.550 --> 0:45:48.10

Houghtaling, Jared

So if you transformed it last week versus this week.

0:45:48.740 --> 0:45:49.230

Daniel Popham

OK.

0:45:49.640 --> 0:45:57.650

Daniel Popham

And then I just thought of something else as far as you said, you could separate actually separate out the process between cohort IDs.

0:45:58.110 --> 0:45:58.420

Houghtaling, Jared

Yeah.

0:45:57.920 --> 0:46:4.40

Daniel Popham

Was there a I guess office hours that it was explaining that or is there some manual somewhere?

0:46:4.620 --> 0:46:4.930

Houghtaling, Jared

Yeah.

0:46:4.940 --> 0:46:6.740

Houghtaling, Jared

And one of the the DQD ones, I'm.

0:46:7.410 --> 0:46:8.170

Daniel Popham

But they did get it.

0:46:6.750 --> 0:46:9.100

Houghtaling, Jared

Let me see if I can find it. Uh.

0:46:8.510 --> 0:46:10.100

Daniel Popham

I was going through all of those.

0:46:10.110 --> 0:46:14.60

Daniel Popham

I've got an email that you sent me a few weeks ago with the with the list.

0:46:14.110 --> 0:46:16.90

Daniel Popham

I'll go back through those and see which ones they were.

0:46:16.900 --> 0:46:17.290

Houghtaling, Jared

Awesome.

0:46:17.740 --> 0:46:18.30

Houghtaling, Jared

Great.

0:46:18.460 --> 0:46:18.680

Daniel Popham

Alright.

0:46:18.40 --> 0:46:24.830

Houghtaling, Jared

Yeah, I think it it it, it should be in one of the DQD ones where we went through actually running like the code to run for DTD.

0:46:25.530 --> 0:46:27.990

Daniel Popham

Umm, right.

0:46:24.840 --> 0:46:28.820

Houghtaling, Jared

And there's an argument that you can specify like a cohort ID.

0:46:28.70 --> 0:46:31.520

Daniel Popham

I was just, uh, yeah, my question is, how do I actually create the cohort?

0:46:31.830 --> 0:46:33.40

Houghtaling, Jared

Uh, yeah. Yeah.

0:46:40.560 --> 0:46:40.840

Daniel Popham

OK.

0:46:33.50 --> 0:46:41.230

Houghtaling, Jared

So that's that's a, that's a different topic and actually we haven't really gone through that in detail on these office hours.

0:46:43.480 --> 0:46:46.190

Houghtaling, Jared

Very I can show you quickly.

0:46:47.930 --> 0:46:48.370

Houghtaling, Jared

Let me.

0:46:47.680 --> 0:46:50.50

Daniel Popham

If it if it if it doesn't take too much time.

0:46:50.620 --> 0:46:51.70

Houghtaling, Jared

Yeah, I can.

0:46:51.80 --> 0:46:52.170

Houghtaling, Jared

I'll I'll show that quickly.

0:46:53.10 --> 0:46:53.280

Daniel Popham

Alright.

0:46:52.180 --> 0:46:54.920

Houghtaling, Jared

Here, let me just quickly check if I go here.

0:47:4.730 --> 0:47:7.40

Houghtaling, Jared

OK umm so.

0:47:7.130 --> 0:47:7.800

Houghtaling, Jared

Ohh, we don't have any.

0:47:23.80 --> 0:47:28.150

Houghtaling, Jared

Let me just take a quick look there and see where we have some coolers that we've generated here, but.

0:47:36.670 --> 0:47:41.550

Houghtaling, Jared

Maybe that's so I can show you what the cohort table looks like.

0:47:41.560 --> 0:47:47.230

Houghtaling, Jared

Let me start my screen share real quick and I'll swap over here to was Chris.

0:48:18.10 --> 0:48:20.700

Houghtaling, Jared

All right, so let me share this really quickly here.

0:48:29.50 --> 0:48:33.180

Houghtaling, Jared

All right, so this is, uh, what?

0:48:33.190 --> 0:48:42.930

Houghtaling, Jared

More or less, the cohort looks like are the cohort table looks like once it's it's been populated and there's four columns.

0:48:42.940 --> 0:49:0.30

Houghtaling, Jared

So you have like definition ID as subject ID, a start date and an end date and you have our packages from Odyssey to do this cohort creation and one of the more popular ways to do it is by using a tool called Atlas which I will show very briefly here.

0:49:0.40 --> 0:49:0.820

Houghtaling, Jared

If I can get it to load.

0:49:1.570 --> 0:49:5.290

Houghtaling, Jared

Uh, but more or less what that does, is it?

0:49:5.300 --> 0:49:5.630

Houghtaling, Jared

It has.

0:49:5.640 --> 0:49:7.150

Houghtaling, Jared

It takes a definition of your cohort.

0:49:7.200 --> 0:49:21.60

Houghtaling, Jared

Let's say you want all patients that had an encounter with the ICU, and it translates that definition into a SQL query that can be used then to select those set of patients.

0:49:22.110 --> 0:49:28.470

Houghtaling, Jared

In your case, it sounds like you probably already have a cohort of patients that you'll need.

0:49:36.240 --> 0:49:36.520

Daniel Popham

Right.

0:49:46.20 --> 0:49:46.230

Daniel Popham

Umm.

0:49:29.340 --> 0:49:48.370

Houghtaling, Jared

However you selected them, but you'll have you have a cohort that you would deliver for this particular project, and if that if that is the case, then all you would need to do is write an insert query that inserts all of the person ID S into this second column of the the cohort table with some unique cohort definition ID.

0:49:48.380 --> 0:49:49.150

Houghtaling, Jared

Maybe it's 13.

0:49:49.160 --> 0:49:50.280

Houghtaling, Jared

Maybe it's 150.

0:49:50.290 --> 0:49:59.330

Houghtaling, Jared

Whatever you want as the cohort definition ID and then the start date and end date, I would those are going to be relevant for like exit events and things like that.

0:49:59.970 --> 0:50:00.670

Houghtaling, Jared

Umm.

0:50:01.50 --> 0:50:09.960

Houghtaling, Jared

If you're only concerned about using this for DQD and Achilles, you can set these just to like the birth date and current date, and that should be sufficient.

0:50:10.150 --> 0:50:14.300

Houghtaling, Jared

So like the patient birth date would be then the start date and then current date is end.

0:50:14.750 --> 0:50:17.970

Houghtaling, Jared

If you wanna actually use this for model development and other things, don't do that.

0:50:19.330 --> 0:50:26.830

Houghtaling, Jared

But in this case, if that's if, if that's your cohort definition, and you're assuming that you're gonna take all of the data for these patients, then that should be sufficient.

0:50:27.940 --> 0:50:41.620

Houghtaling, Jared

And then what you will do is whatever ID you decide to place in this first column, you'll reference that exact ID in the covert ID arguments for both the DQD and the Achilles executions.

0:50:42.620 --> 0:50:45.70

Houghtaling, Jared

And in those executions, it's gonna go first.

0:50:45.80 --> 0:50:46.10

Houghtaling, Jared

Look at this cohort table.

0:50:46.20 --> 0:50:57.360

Houghtaling, Jared

It's going to grab all of the person ID's that are associated with that cohort table, as well as the dates, and it's then gonna filter the rest of your event tables based on that and run it on that subset.

0:51:00.130 --> 0:51:00.540

Daniel Popham

Gotcha.

0:50:58.680 --> 0:51:0.590

Houghtaling, Jared

So yeah.

0:51:0.550 --> 0:51:4.0

Daniel Popham

And we're gonna, we're gonna probably even do something fancy.

0:51:4.10 --> 0:51:6.940

Daniel Popham

Where we know there's some issue with some data.

0:51:11.500 --> 0:51:11.700

Houghtaling, Jared

Yep.

0:51:14.90 --> 0:51:14.490

Houghtaling, Jared

Exactly.

0:51:7.150 --> 0:51:15.0

Daniel Popham

Just select those, put them into a specific cohort and then run the DQD process on just that small amount of data.

0:51:15.390 --> 0:51:15.730

Daniel Popham

Perfect.

0:51:15.300 --> 0:51:17.60

Houghtaling, Jared

Yep, and it'll run much faster than as well.

0:51:17.410 --> 0:51:17.940

Daniel Popham

Yes.

0:51:17.950 --> 0:51:22.360

Daniel Popham

Yeah, that was my main thing that's trying to figure out how to debug this thing as quickly as we're good.

0:51:22.870 --> 0:51:23.170

Houghtaling, Jared

Awesome.

0:51:23.210 --> 0:51:24.120

Daniel Popham

Alright, cool.

0:51:24.130 --> 0:51:28.300

Daniel Popham

And you said that other tool was OK or?

0:51:26.850 --> 0:51:33.740

Houghtaling, Jared

Atlas and that this is this is a whole probably office hour to cover this but more or less that's what this allows you to do.

0:51:33.910 --> 0:51:34.520

Houghtaling, Jared

You can go in.

0:51:34.530 --> 0:51:48.440

Houghtaling, Jared

You can define cohorts and then with those cohort definitions, when you hit generate it more or less it populates the cohort table that I'm showing you here with the relevant information per person.

0:51:48.630 --> 0:51:54.690

Houghtaling, Jared

So this is allowing you to kind of go through create kind of a definition that's almost drag and drop and visual.

0:51:55.980 --> 0:51:59.590

Houghtaling, Jared

This then produces with the help of Cersei and a couple of other Odyssey packages.

0:51:59.600 --> 0:52:06.810

Houghtaling, Jared

It produces a SQL query that gets executed to then load this table, but that's maybe outside the scope of this discussion.

0:52:07.790 --> 0:52:10.300

Daniel Popham

No, that, that, that give me a good starting point appreciate it.

0:52:10.800 --> 0:52:11.90

Houghtaling, Jared

Awesome.

0:52:10.670 --> 0:52:11.240

Daniel Popham

That was it for me.

0:52:11.100 --> 0:52:12.320

Houghtaling, Jared

Thanks. Cool.

0:52:14.80 --> 0:52:14.590

Houghtaling, Jared

All right, del.

0:52:16.210 --> 0:52:29.650

Bold, Delgersuren

Well, that there can you just put arbitrary number for that definition ID or does it have like a some sort of dependency or reference to the scheme that that generates from the defining the cohort?

0:52:30.280 --> 0:52:30.610

Houghtaling, Jared

Yeah.

0:52:30.620 --> 0:52:34.710

Houghtaling, Jared

So there is, there is an additional, it depends on if you have your tables all constrained.

0:52:35.630 --> 0:52:36.50

Bold, Delgersuren

Umm.

0:52:35.240 --> 0:52:38.990

Houghtaling, Jared

Ideally you would have also some sort of cohort definition.

0:52:39.0 --> 0:52:40.930

Houghtaling, Jared

I think that I can show that table as well.

0:52:55.820 --> 0:52:58.800

Bold, Delgersuren

So it wouldn't complain if you give some random number.

0:52:59.680 --> 0:52:59.900

Houghtaling, Jared

Yeah.

0:52:59.910 --> 0:53:2.90

Houghtaling, Jared

So right now we actually don't have it constrained, right?

0:53:2.40 --> 0:53:2.460

Bold, Delgersuren

Uh.

0:53:2.100 --> 0:53:4.600

Houghtaling, Jared

So our cohort definition table is is currently empty.

0:53:5.430 --> 0:53:6.150

Houghtaling, Jared

Umm.

0:53:6.250 --> 0:53:7.620

Houghtaling, Jared

And you can then give it any number.

0:53:13.60 --> 0:53:13.380

Bold, Delgersuren

Uh-huh.

0:53:7.630 --> 0:53:15.780

Houghtaling, Jared

If you do set the constraints to where that cohort definition ID has to be a foreign key to this primary key, then you would need to have this table populated as well.

0:53:16.110 --> 0:53:17.960

Houghtaling, Jared

If you don't constrain, it's not a problem.

0:53:16.210 --> 0:53:22.460

Bold, Delgersuren

Well, what what would our data our schema would require?

0:53:22.670 --> 0:53:23.40

Bold, Delgersuren

Do you?

0:53:23.50 --> 0:53:25.830

Bold, Delgersuren

Do you want that to be with constraint or no?

0:53:27.940 --> 0:53:28.850

Bold, Delgersuren

So like for what?

0:53:27.970 --> 0:53:28.970

Houghtaling, Jared

What do you mean with what?

0:53:28.860 --> 0:53:29.90

Bold, Delgersuren

What?

0:53:29.100 --> 0:53:30.730

Bold, Delgersuren

What would chorus required?

0:53:33.890 --> 0:53:34.460

Bold, Delgersuren
Constraint.

0:53:31.870 --> 0:53:37.930

Houghtaling, Jared
So Chorus will not require either of these tables actually in the delivery.

0:53:39.40 --> 0:53:47.590

Houghtaling, Jared
You can use them because they are like referenced by audyssey packages, but these are more for let's say down the road.

0:53:47.600 --> 0:53:57.290

Houghtaling, Jared
If you're creating models to predict a certain outcome of interest and you wanna compare 2 cohorts of patients and use those as part of your model, then this all becomes very relevant.

0:53:57.860 --> 0:53:58.780

Bold, Delgersuren
Ah.

0:53:57.340 --> 0:54:9.640

Houghtaling, Jared
But just for the purposes of constructing an extract, it can be useful like with Dan was saying to isolate a small subset of patients and run only your checks on those 200 patients.

0:54:9.650 --> 0:54:17.610

Houghtaling, Jared
And you can do that simply by populating this cohort table that looks like this, and in that case you don't need.

0:54:17.620 --> 0:54:18.950

Houghtaling, Jared
This doesn't need to be constrained.

0:54:28.110 --> 0:54:28.860

Bold, Delgersuren
OK.

0:54:18.960 --> 0:54:29.170

Houghtaling, Jared
This can be any ID you want, you just need to make sure that whatever ID you place here is the same one that you specify when you're trying to filter your checks later on.

0:54:29.170 --> 0:54:36.90

Bold, Delgersuren
But you you the the application you are deploying the RC start that includes Atlas, right?

0:54:39.230 --> 0:54:39.820

Bold, Delgersuren

Yes.

0:54:39.910 --> 0:54:39.990

Bold, Delgersuren

Yes.

0:54:37.930 --> 0:54:40.740

Houghtaling, Jared

On the central look and the central server, eventually, yes.

0:54:40.810 --> 0:54:41.850

Houghtaling, Jared

Yeah, it will include Alice.

0:54:41.250 --> 0:54:44.960

Bold, Delgersuren

Then can I assume eventually it will have a constraint?

0:54:52.0 --> 0:54:55.200

Bold, Delgersuren

Ah OK got it got it.

0:54:46.60 --> 0:54:55.810

Houghtaling, Jared

So yes, but the sites are not expected to create, they're populate these tables so that that then will happen once all the data is in a central location.

0:54:55.820 --> 0:55:5.580

Houghtaling, Jared

Then we'll have people accessing that data in Atlas, and then there's gonna be kind of a running constrained cohort table depending on what those people are doing and how they're interacting with the data and constructing models.

0:55:5.140 --> 0:55:6.500

Bold, Delgersuren

OK, got it.

0:55:7.330 --> 0:55:7.670

Bold, Delgersuren

Thank you.

0:55:6.620 --> 0:55:8.370

Houghtaling, Jared

Uh, but the sites don't have to worry about that.

0:55:10.70 --> 0:55:10.260

Bold, Delgersuren

OK.

0:55:12.830 --> 0:55:13.410

Houghtaling, Jared

Great questions.

0:55:18.760 --> 0:55:19.110

Houghtaling, Jared

All right.

0:55:19.120 --> 0:55:20.50

Houghtaling, Jared

Anything else?

0:55:20.190 --> 0:55:21.710

Houghtaling, Jared

Uh on either.

0:55:22.200 --> 0:55:23.40

Houghtaling, Jared

Yep. Yep.

0:55:20.970 --> 0:55:25.390

Dandayudhapani, Satish

Of yeah, this Satish, I have a good question on a different topic.

0:55:25.570 --> 0:55:26.550

Dandayudhapani, Satish

It's an image.

0:55:26.560 --> 0:55:30.360

Dandayudhapani, Satish

The identification I don't know if this is the right forum to ask this.

0:55:30.990 --> 0:55:32.160

Dandayudhapani, Satish

I would like to know.

0:55:32.670 --> 0:55:37.290

Dandayudhapani, Satish

I know there are somebody working centrally on the denotification of the images.

0:55:38.200 --> 0:55:42.0

Dandayudhapani, Satish

Umm I'm I'm doing some work on that.

0:55:42.10 --> 0:56:2.360

Dandayudhapani, Satish

I just wanted to touch base with him to see what kind of because I wanted to see what is that algorithm is covering different types of like we do have OCR's in our images right there is metadata and how that has been all cleared so that we wanted just to make sure that it's all being done as well or is it something we have to do on our side.

0:56:4.700 --> 0:56:5.890

Houghtaling, Jared

So we just, I gotta.

0:56:8.210 --> 0:56:8.530

Bold, Delgersuren

Umm.

0:56:5.900 --> 0:56:8.960

Houghtaling, Jared

There's a long discussion, I think on Tuesday about this exact topic.

0:56:10.510 --> 0:56:10.740

Houghtaling, Jared

It's.

0:56:11.640 --> 0:56:24.220

Houghtaling, Jared

I don't have a super satisfactory answer for you now, other than the fact that there's now a push and I think that we're converging on a set of processes to kind of standardize that, then sites can deploy.

0:56:24.680 --> 0:56:31.860

Houghtaling, Jared

I think in general though, the DEIDENTIFICATION itself will have to happen at the sites, at least all of it that's relevant for the.

0:56:35.660 --> 0:56:38.490

Houghtaling, Jared

Yeah, the the data used agreements that have been put in place.

0:56:39.500 --> 0:56:44.70

Houghtaling, Jared

So it will be the case that you'll probably need to process your images on your sides.

0:56:44.340 --> 0:56:57.470

Houghtaling, Jared

Remove whatever offending data there might be, depending on those standard conventions, and then the transfer, but I don't have a ton of specifics for you yet, because I think they're still being ironed out by the the general course team.

0:56:57.670 --> 0:56:58.780

Dandayudhapani, Satish

OK. Thanks.

0:56:59.540 --> 0:57:2.50

Bold, Delgersuren

He didn't ask when you do it on the fly.

0:57:2.310 --> 0:57:4.410

Bold, Delgersuren

Uh, do you did identify?

0:57:4.820 --> 0:57:10.210

Bold, Delgersuren

Do you remove the private tags and things like that or just the subject ID's and they?

0:57:13.90 --> 0:57:13.790

Houghtaling, Jared

Any question for me?

0:57:15.770 --> 0:57:16.730

Bold, Delgersuren

Of our satish.

0:57:24.280 --> 0:57:24.580

Bold, Delgersuren

OK.

0:57:17.970 --> 0:57:33.700

Dandayudhapani, Satish

Ohh yeah, so right now we are having going up to the metadata of the DICOM images and yeah we are we are seeing from our side what are that metadata fields we need to dentify that's what we are taking care of.

0:57:33.710 --> 0:57:36.140

Dandayudhapani, Satish

But there are some there are.

0:57:36.560 --> 0:57:37.220

Dandayudhapani, Satish

We are.

0:57:41.660 --> 0:57:43.220

Bold, Delgersuren

Yes, yes, yes.

0:57:37.230 --> 0:57:46.470

Dandayudhapani, Satish

We are having different kind of screen captures which are coming as part of the DICOM images as well as there is OCR's inside the DICOM images.

0:57:47.940 --> 0:57:50.820

Dandayudhapani, Satish

So yeah, that's that's where we are having some challenges.

0:57:51.950 --> 0:57:56.170

Bold, Delgersuren

Yeah, I think I think that's what they're trying to solve.

0:57:59.390 --> 0:58:1.200

Dandayudhapani, Satish

Yeah, that's yeah, that's where I am.

0:57:56.640 --> 0:58:4.430

Bold, Delgersuren

The Pixel reduction or anything PII PII screening on all the tags and all that.

0:58:4.900 --> 0:58:32.460

Bold, Delgersuren

So we had a good discussion on Tuesday, I think more discussion will happen and I'm in the also what I understood was that also wanna construct OMOP image occurrence table which will have certain features, modalities, anatomic uh, you know anatomic off that image and things like that needs to be preserved before you deidentify them.

0:58:33.580 --> 0:58:33.800

Dandayudhapani, Satish

OK.

0:58:33.190 --> 0:58:35.400

Bold, Delgersuren

So I think you know.

0:58:37.550 --> 0:58:44.660

Bold, Delgersuren

Just uh, wait for them to give us, like A and no good information that how we perceive.

0:58:46.290 --> 0:58:48.300

Bold, Delgersuren

I think it's a good idea, yeah.

0:58:49.140 --> 0:58:54.90

Dandayudhapani, Satish

Yeah, I was looking into the the Microsoft Presidio ohm, a library.

0:58:52.980 --> 0:58:54.750

Bold, Delgersuren

Ah, yes. Uh-huh.

0:58:54.780 --> 0:58:59.510

Dandayudhapani, Satish

That's what I was looking into, but that can only for the OCR images.

0:59:1.930 --> 0:59:2.110

Bold, Delgersuren

Umm.

0:58:59.520 --> 0:59:4.230

Dandayudhapani, Satish

It only looks for certain area to the corners where it can pixelate.

0:59:4.240 --> 0:59:12.810

Dandayudhapani, Satish

I mean like the identify, but if it's the the Phi data anywhere around within the image, like any part of the image doesn't recognize it.

0:59:14.10 --> 0:59:18.160

Bold, Delgersuren

Ohh OK yeah, we did try it and we and also MGH try it.

0:59:18.170 --> 0:59:23.970

Bold, Delgersuren

I think MGH was successful in, you know reducing their images.

0:59:24.190 --> 0:59:35.380

Bold, Delgersuren

But for us, some of the very bright images for me able to pick up the HCI so and I don't know if they're gonna pursue and try to love, you know.

0:59:37.510 --> 0:59:38.640

Bold, Delgersuren

It's open source project.

0:59:38.650 --> 0:59:45.960

Bold, Delgersuren

So, uh, try to, you know, change the code and you know, change the saturation or things like that.

0:59:46.30 --> 0:59:51.660

Bold, Delgersuren

I don't know if they're gonna do that or whether they're gonna get a a private product.

0:59:52.480 --> 0:59:53.840

Bold, Delgersuren

That's what they're discussing right now.

0:59:55.180 --> 0:59:55.590

Dandayudhapani, Satish

OK.

0:59:55.760 --> 0:59:56.70

Dandayudhapani, Satish
Yeah.

0:59:56.80 --> 0:59:56.710

Dandayudhapani, Satish
Thanks for the update.

1:0:6.760 --> 1:0:7.880

Houghtaling, Jared
Alright, great. Yeah.

1:0:6.610 --> 1:0:8.60

Bold, Delgersuren
Up, Jared, I have a question.

1:0:8.100 --> 1:0:22.330

Bold, Delgersuren
Just last question, so on on my note table, I think that on the last couple weeks ago, Jill said we he wanted us to construct the types note types in concepts.

1:0:24.420 --> 1:0:31.210

Bold, Delgersuren
Will we get a a assistant from the standards or is it something we have to consider for ourselves?

1:0:32.330 --> 1:0:47.550

Houghtaling, Jared
Umm great question I I I believe and polina feel free to step in here, but I think that there's gonna there's been some work to create some guidelines for note types and note type mappings when those will become available where they will be available.

1:0:47.560 --> 1:0:51.440

Houghtaling, Jared
I guess we will announce at some point on one of these calls, but.

1:0:53.570 --> 1:0:56.720

Houghtaling, Jared
Yeah, I I think that that is something that that you should expect.

1:0:56.730 --> 1:0:58.910

Houghtaling, Jared
There's there to be some guidance or assistance on.

1:1:0.900 --> 1:1:1.160

Bold, Delgersuren
OK.

1:1:1.210 --> 1:1:1.370

Houghtaling, Jared

Yeah.

1:1:4.560 --> 1:1:4.710

Bold, Delgersuren

But.

1:1:11.450 --> 1:1:13.900

Bold, Delgersuren

No, it did also said on the standard.

1:1:13.910 --> 1:1:14.970

Bold, Delgersuren

Uh, the Edu call.

1:1:14.980 --> 1:1:23.950

Bold, Delgersuren

Uh, he wants the device information from the waveforms that we produce are what do we put that?

1:1:23.960 --> 1:1:28.910

Bold, Delgersuren

Do we give you device table or a how does that work?

1:1:30.860 --> 1:1:31.190

Houghtaling, Jared

Yeah.

1:1:31.200 --> 1:1:32.810

Houghtaling, Jared

So that's also a great question.

1:1:35.300 --> 1:1:37.750

Houghtaling, Jared

I think there's multiple ways, of course, of capturing that.

1:1:37.760 --> 1:1:43.560

Houghtaling, Jared

I think one of the things that we have talked about with with Eddie in particular is.

1:1:45.850 --> 1:1:52.780

Houghtaling, Jared

This idea of a registry table where you you you kind of capture some of the metadata about your images or your waveforms.

1:1:55.50 --> 1:1:58.350

Houghtaling, Jared

In kind of 1 location and say hey, here's my source waveform file.

1:1:58.360 --> 1:2:6.780

Houghtaling, Jared

Here's what I ended up translated in the name to here are the various metadata that are relevant, like the device and the anatomic site and things like that.

1:2:6.790 --> 1:2:18.840

Houghtaling, Jared

So for you even go about this whole process of mapping this into OMOP, which is its own challenge, just to be able to catalog what types of information you have in that like registry format.

1:2:18.850 --> 1:2:24.270

Houghtaling, Jared

And I think we've talked about that and then some extent on these calls with regards to the MIMIC data.

1:2:25.80 --> 1:2:36.980

Houghtaling, Jared

Umm but yeah, I would say it would be good for us at least in the as kind of to provide guidance to have a list of what types of of metadata need to be captured.

1:2:37.910 --> 1:2:47.30

Houghtaling, Jared

But like a in this type of registry format and then the idea would be that you would submit that registry together with your data exports.

1:2:53.700 --> 1:2:55.110

Bold, Delgersuren

And when when do we?

1:2:55.160 --> 1:3:1.30

Bold, Delgersuren

When will we have that the waveform registry table and the schema?

1:3:1.710 --> 1:3:7.570

Houghtaling, Jared

So we have a we have a meeting today, this afternoon, a wave form meeting and I think that that's one of the agenda items.

1:3:13.650 --> 1:3:13.930

Bold, Delgersuren

OK.

1:3:7.580 --> 1:3:17.80

Houghtaling, Jared

So it could be soon, I don't know, and depend on how quickly these decisions get made and how quickly we're able to propagate them across site.

1:3:16.930 --> 1:3:17.290

Bold, Delgersuren

Got it.

1:3:22.320 --> 1:3:22.630

Houghtaling, Jared
Great.

1:3:23.310 --> 1:3:23.930

Houghtaling, Jared
Yeah, I see that.

1:3:26.220 --> 1:3:26.500

Houghtaling, Jared
Alright.

1:3:22.210 --> 1:3:26.990

Alvarez, Marta
We're over, so we should probably wrap no, alright.

1:3:26.510 --> 1:3:28.130

Houghtaling, Jared
Well, thank you all so much for your patience.

1:3:28.140 --> 1:3:35.250

Houghtaling, Jared
Hopefully this is helpful and like I said, take a look at the discussions and some of the resources there and then I'm always.

1:3:35.260 --> 1:3:42.540

Houghtaling, Jared
I'll try to be active in terms of responding to questions and then things that come up there and feel free to reach out by email or otherwise.

1:3:43.60 --> 1:3:45.540

Houghtaling, Jared
If you have any problems, so thanks again.

1:3:46.100 --> 1:3:46.430

Bold, Delgersuren
Bye.

1:3:47.580 --> 1:3:48.300

Houghtaling, Jared
Talk to you next week.

1:3:49.0 --> 1:3:49.870

Talapova, Polina
Thank you everyone.

1:3:49.970 --> 1:3:50.690

Talapova, Polina
Thank you, Jared.

1:3:51.90 --> 1:3:51.330

Talapova, Polina

Bye bye.