

CHoRUS B2AI Standards Module Office Hours Transcript

November 02, 2023

0:0:0.0 --> 0:0:0.480

Talapova, Polina

I want.

0:0:5.700 --> 0:0:5.940

Guan,Ziyuan

Hello.

0:1:0.270 --> 0:1:2.810

Talapova, Polina

I hope you're about to start in several minutes.

0:2:3.970 --> 0:2:4.290

Alvarez, Marta

Paulina.

0:2:6.690 --> 0:2:7.110

Alvarez, Marta

How are you?

0:2:8.740 --> 0:2:10.870

Talapova, Polina

Hi, Martin, I'm fine, thanks.

0:2:10.880 --> 0:2:11.240

Talapova, Polina

How are you?

0:2:12.100 --> 0:2:12.320

Alvarez, Marta

Good.

0:2:16.740 --> 0:2:17.520

Alvarez, Marta

Welcome, everyone.

0:2:18.950 --> 0:2:20.60

Alvarez, Marta

Well, wait a little bit.

0:2:20.530 --> 0:2:25.80

Alvarez, Marta

The last hops on maybe starting two minutes.

0:3:25.390 --> 0:3:26.80

Alvarez, Marta
Happier.

0:3:26.90 --> 0:3:27.180

Alvarez, Marta
I know I think.

0:3:27.190 --> 0:3:28.940

Alvarez, Marta
I'm not sure if you've come to the meeting before.

0:3:28.950 --> 0:3:34.430

Alvarez, Marta
It's possible you came once and I missed it, but we'd love for you to introduce yourself if you're up for it.

0:3:38.960 --> 0:3:39.560

Alvarez, Marta
You're on mute.

0:3:42.260 --> 0:3:42.670

Sanz, Javier
See.

0:3:43.400 --> 0:3:43.690

Sanz, Javier
OK.

0:3:43.730 --> 0:3:44.260

Sanz, Javier
Can you hear me now?

0:3:43.330 --> 0:3:44.990

Alvarez, Marta
There you go. Yes.

0:3:45.760 --> 0:3:46.90

Sanz, Javier
Alright.

0:3:46.100 --> 0:3:55.330

Sanz, Javier
Yes, my name is Javier Dissave and I'm with UC with the UCLA group and I recently started working on this, so I'm trying to catch up.

0:3:55.860 --> 0:4:5.100

Sanz, Javier
I'm working on a 25 sample patient cohort and they started to gather some of the old map data using some of the resources that I've gathered from you guys.

0:4:7.910 --> 0:4:8.290

Alvarez, Marta

Awesome.

0:4:8.300 --> 0:4:8.750

Alvarez, Marta

Welcome.

0:4:8.760 --> 0:4:12.640

Alvarez, Marta

Thank you so much for joining Umm, I also see.

0:4:14.40 --> 0:4:19.740

Alvarez, Marta

Umm, I want to is it Kim Long?

0:4:20.840 --> 0:4:28.640

Alvarez, Marta

I think you also maybe join last time, but I wanted to give you an opportunity to introduce yourself as well if you're available.

0:4:31.930 --> 0:4:33.160

Vong, Kin

Ohh yeah hi everyone.

0:4:33.210 --> 0:4:36.150

Vong, Kin

My name is Ken Wong and I'm a data analyst at Cell Children's.

0:4:38.900 --> 0:4:39.290

Alvarez, Marta

Awesome.

0:4:39.460 --> 0:4:40.200

Alvarez, Marta

Thank you so much.

0:4:41.880 --> 0:4:45.420

Alvarez, Marta

OK, I think that captures everybody.

0:4:46.510 --> 0:4:48.630

Alvarez, Marta

Everybody else, familiar names, I believe.

0:4:50.60 --> 0:4:50.280

Alvarez, Marta

Yes.

0:4:51.790 --> 0:4:52.880

Alvarez, Marta

So yeah.

0:4:52.890 --> 0:4:55.50

Alvarez, Marta

So thank you everyone for joining.

0:4:55.60 --> 0:4:57.900

Alvarez, Marta

And Paulina, I'll send it over to you.

0:4:57.910 --> 0:5:2.470

Alvarez, Marta

Unless Andrew would like to add anything before we get started, OK?

0:5:2.90 --> 0:5:11.450

Williams, Andrew E

No other than my great gratitude for pulling and doing this, given the extensive travels and everything, she's still recovering from, I'm sure.

0:5:11.460 --> 0:5:12.660

Williams, Andrew E

So thank you for being here.

0:5:15.410 --> 0:5:15.800

Talapova, Polina

Thank you.

0:5:15.810 --> 0:5:16.100

Talapova, Polina

Andrew.

0:5:17.800 --> 0:5:22.930

Talapova, Polina

Umm, thank you everyone for coming and I hope that we can start.

0:5:23.110 --> 0:5:26.280

Talapova, Polina

I will try to share my screen in a sack.

0:5:37.840 --> 0:5:39.210

Talapova, Polina

And his see my screen.

0:5:42.330 --> 0:5:42.850

Williams, Andrew E

Yes, we can.

0:5:39.360 --> 0:5:42.990

Talapova, Polina

Sorry for this annoying question, but I have to ask.

0:5:43.80 --> 0:5:43.660

Alvarez, Marta

Like, yeah.

0:5:43.900 --> 0:5:45.30

Talapova, Polina

Super cool.

0:5:45.580 --> 0:5:49.170

Talapova, Polina

So let's start.

0:5:48.920 --> 0:5:49.270

Alvarez, Marta

There you go.

0:5:49.500 --> 0:5:56.340

Talapova, Polina

Today's topic is principles of mapping in general and vocabulary gaps identification.

0:5:57.710 --> 0:6:3.300

Talapova, Polina

So let me just tell you about content that we are going to cover today.

0:6:3.310 --> 0:6:10.780

Talapova, Polina

First of all, recap, we will talk about mapping, what is it that and what the problem space that we have there.

0:6:11.110 --> 0:6:16.680

Talapova, Polina

Then we will go to Maplin principles and at the end vocabulary gap identification.

0:6:17.330 --> 0:6:32.730

Talapova, Polina

I'm not sure that I will cover the whole hour, but if we will be able to get some conversation after that, it will be great and maybe you will be able to share your experience that during the map and it's also very important for us.

0:6:33.480 --> 0:6:45.340

Talapova, Polina

So first of all, umm, we need to recollect the recall that the mapping where map it's association between two terms.

0:6:45.850 --> 0:6:53.160

Talapova, Polina

First term, it's a source term, usually from some coding system or data set, and another term.

0:6:53.170 --> 0:7:23.910

Talapova, Polina

It's usually some standard or the term which we are considered to be standard, and this association in the ideal world should be direct one to one where first concept is equal to the second concept and particularly in Omloop mapping it's also the process of harmonization of the data or transformation of source terms to standard terms from the home of vocabularies and why it's important.

0:7:23.920 --> 0:7:27.530

Talapova, Polina

Because for our ETL we have to populate standardized.

0:7:27.700 --> 0:7:34.920

Talapova, Polina

Nickel data tables, also called event tables, and these tables allow us to store only standard concepts.

0:7:35.130 --> 0:7:41.310

Talapova, Polina

So it means that all source storms should be converted into these standard concepts.

0:7:43.690 --> 0:7:44.210

Talapova, Polina

Umm.

0:7:44.450 --> 0:7:55.690

Talapova, Polina

Also, it is important that it's a more about technical aspect of the process, but mapping usually is performed you through the records in the concept relationship table.

0:7:55.700 --> 0:8:16.520

Talapova, Polina

We covered this in previous office hours and if you want to know a bit more you can go to our recordings and look at the presentation and video record about this about the structure of standardized vocabularies, and the idea is in the concept relationship table.

0:8:16.530 --> 0:8:17.650

Talapova, Polina

We have concept idea.

0:8:17.660 --> 0:8:19.350

Talapova, Polina

One uh, which is?

0:8:20.900 --> 0:8:29.570

Talapova, Polina

Connected with our source term concept ID too, which is about standard or target term and something between them.

0:8:29.580 --> 0:8:36.390

Talapova, Polina

It's relationship ID and in almost there are not a lot of relationship ideas which indicate mappings.

0:8:36.400 --> 0:8:58.200

Talapova, Polina

In fact, we use only two maps to and maps to value, and this is a disadvantage of the system because unfortunately we cannot distinguish when mapping is full equivalence or for instance with some lack of details or additional details that we don't want to see there.

0:8:58.210 --> 0:9:17.620

Talapova, Polina

But in fact we have only that and these records in the concept relationship table allow us to populate even tables as you see here, and particular feels like condition concept, adding condition, occurrence, table observation concept, ad measurement concept ID and etc.

0:9:17.890 --> 0:9:19.990

Talapova, Polina

And also value various relevant.

0:9:20.0 --> 0:9:25.670

Talapova, Polina

Usually we use it for measurement and observations, the next one.

0:9:25.920 --> 0:9:55.790

Talapova, Polina

So here you can see the AH concept of ideal mapping and ideal mapping in ideal world is a fully equivalent mapping where we have source term which is equal to the target standard term and this target standard term fully reflects the meaning of the source and the quality of terms is determined by the extent to which a target term semantically covers a source term.

0:9:56.30 --> 0:10:3.660

Talapova, Polina

So the more details from the source are captured, the more precise the mapping becomes.

0:10:3.730 --> 0:10:7.220

Talapova, Polina

This is a rule and also here I want to show you.

0:10:8.0 --> 0:10:13.90

Talapova, Polina

Hmm, maybe I need to switch, maybe cannot see the another window?

0:10:13.640 --> 0:10:14.170

Talapova, Polina

Or can you?

0:10:17.250 --> 0:10:18.920

Talapova, Polina

You see presentation maybe? Yeah.

0:10:19.660 --> 0:10:21.210

Alvarez, Marta

No, actually we see.

0:10:21.30 --> 0:10:24.580

Williams, Andrew E

We we see the CDM documentation, how the vocabularies built.

0:10:23.440 --> 0:10:25.30

Talapova, Polina

Oh, perfect, this is.

0:10:25.110 --> 0:10:35.120

Talapova, Polina

This is great, so here I want to show you examples of mappings which we have in the OMOP for instance full equivalent mapping.

0:10:35.130 --> 0:10:37.450

Talapova, Polina

The best of the best ideal scenario.

0:10:37.660 --> 0:10:46.530

Talapova, Polina

For instance, here we have source term trigger finger maps to relationship ID and the target term trigger finger.

0:10:46.830 --> 0:10:56.240

Talapova, Polina

So the same target term fully covers the semantics of the source, but unfortunately there are other scenarios.

0:10:56.250 --> 0:11:0.260

Talapova, Polina

For instance, refractory anemia with multilineage dysplasia.

0:11:0.490 --> 0:11:17.260

Talapova, Polina

Here we have also mops too, but the target term is only about refractory anemia and there is no information about multilineage dysplasia and there is no information that this data is missing in the mapping.

0:11:18.140 --> 0:11:21.610

Talapova, Polina

Another example, idiopathic osteoporosis.

0:11:21.620 --> 0:11:31.400

Talapova, Polina

Multiple sites also maps too, but idiopathic osteoporosis without mentioning of multiple sites and involvement of the in this process.

0:11:33.140 --> 0:11:33.520

Talapova, Polina

Umm.

0:11:34.370 --> 0:11:42.620

Talapova, Polina

Also, there you can see one too many mappings when the source term covered by several target terms.

0:11:42.910 --> 0:11:59.420

Talapova, Polina

For instance, open wound or fingers with damage to nail here mapped 22 concepts open wonder finger, which covers the first part of the source term and another one injury of nail damage the nail.

0:11:59.430 --> 0:12:4.800

Talapova, Polina

Yes, covered to this this term, it means that we have two records for one patient.

0:12:6.200 --> 0:12:22.360

Talapova, Polina

Both of them are conditions and the only way to know this to check how many maps to relationships we have in concept relationship table for one concept source term and this scenario is not good.

0:12:22.370 --> 0:12:30.900

Talapova, Polina

It means that we we already have the gap in the OMOP vocabulary and this gap potentially should be covered.

0:12:31.10 --> 0:12:41.170

Talapova, Polina

For instance, we can create additional new concept standard concept which cover both parts of this source terms so stormed.

0:12:41.840 --> 0:12:44.50

Talapova, Polina

So the same story for them.

0:12:45.170 --> 0:12:47.590

Talapova, Polina

So let me come back to the presentation again.

0:12:49.660 --> 0:12:56.130

Talapova, Polina

And the next slide, now we have to discuss uh mapping principle #2.

0:12:56.140 --> 0:13:11.630

Talapova, Polina

The first principle is 1 to one, mapping is the most uh now preferable way to map if we have one too many mapping or mapping to one but with the loss of granularity it means that this is a gap for sure.

0:13:12.20 --> 0:13:24.10

Talapova, Polina

So the second principle is to follow to triad of importance and first point of this triad is standardness of the target term.

0:13:24.60 --> 0:13:26.640

Talapova, Polina

It means the target term should be standard.

0:13:27.460 --> 0:13:37.870

Talapova, Polina

The second point, it means that you have to pay attention to the domain with which you work and the domain which you will select to.

0:13:38.190 --> 0:13:42.750

Talapova, Polina

For instance, find the target term and the Third Point is vocabulary.

0:13:42.760 --> 0:13:56.20

Talapova, Polina

It means that vocabulary also matters and for different domains we use some particular vocabularies, mostly some we don't use or avoid they usage and they will show you a bit later.

0:13:56.30 --> 0:13:58.320

Talapova, Polina

What is the idea of that?

0:13:59.710 --> 0:14:4.480

Talapova, Polina

So first of all standardness, it means that we are interested in standard concepts.

0:14:4.970 --> 0:14:7.460

Talapova, Polina

There are two ways how to select them.

0:14:7.690 --> 0:14:17.220

Talapova, Polina

If you do it manually, first one in Athena, you can just uh click on standard concepts and search only through this list.

0:14:17.400 --> 0:14:22.910

Talapova, Polina

Another option is to work with the database and just add the close.

0:14:22.920 --> 0:14:41.80

Talapova, Polina

That standard concept should be as and this also can help you to find only standard concepts if you use for instance Usagi, you also can enable the checkbox for standard concepts only and get only standard concepts.

0:14:41.190 --> 0:14:43.180

Talapova, Polina

So this is the most important.

0:14:43.470 --> 0:14:46.460

Talapova, Polina

Don't look to non standard concepts.

0:14:46.790 --> 0:14:48.380

Talapova, Polina

Use standard one.

0:14:48.650 --> 0:14:57.700

Talapova, Polina

But there is also interesting idea that sometimes you can find standard concepts using non standard concept, but it's another story.

0:14:57.710 --> 0:15:1.870

Talapova, Polina

But it's also possible, but for mapping we use only standard.

0:15:1.940 --> 0:15:3.120

Talapova, Polina

This is the gold rule.

0:15:4.60 --> 0:15:22.680

Talapova, Polina

Next point point, don't domain selection uh to be able to find correct domain for our terms and to limit the space for the searching for standard term we have to know what is the idea of each domain within the OMOP.

0:15:22.690 --> 0:15:37.560

Talapova, Polina

So it's a theory part and I recommend you to remember this information if you for instance not sure, please just read this in the documentation or you can reference to this presentation as well.

0:15:37.570 --> 0:15:41.290

Talapova, Polina

So a let's talk about this a bit.

0:15:41.380 --> 0:15:42.760

Talapova, Polina

The measurement domain.

0:15:44.310 --> 0:15:53.990

Talapova, Polina

This domain covers standardized diagnostic and evaluative tests or assessments conducted on individuals or their biological samples.

0:15:54.490 --> 0:16:16.70

Talapova, Polina

What is important these tests implies structured data represented in and not, not just data, but values represented in multiple formats, and it can be numerical values, ordinal interval ratio, nominal texture, binary time based, image based, composite, and scaled values.

0:16:16.450 --> 0:16:28.210

Talapova, Polina

So all these tests, these values should be in the measurement domain when you look at this terms in Athena, you usually don't see values.

0:16:28.260 --> 0:16:32.610

Talapova, Polina

Values usually are at your part at the data source data part.

0:16:32.860 --> 0:16:43.70

Talapova, Polina

It means that you first map your source term to the standard measurement and then populate measurement table and add values to the respective fields.

0:16:44.950 --> 0:16:49.160

Talapova, Polina

So next one close to the measurement, it's a procedure domain.

0:16:49.170 --> 0:17:3.720

Talapova, Polina

This one covers activities or processes ordered carried out by a healthcare provider on the patient with a diagnostic or therapeutic purposes, and usually procedures do not imply values.

0:17:3.790 --> 0:17:15.800

Talapova, Polina

This is important, but again in some cases it can be controversial because some procedures can have values in your data, but this is another scenario.

0:17:16.70 --> 0:17:19.50

Talapova, Polina

In the ideal world, procedures don't have values.

0:17:20.130 --> 0:17:27.540

Talapova, Polina

Another domain condition, simple one covers diseases or medical conditions stated as diagnosis signs.

0:17:27.550 --> 0:17:32.810

Talapova, Polina

Symptoms, which is absorbed by a provider or reported by person.

0:17:34.120 --> 0:17:42.410

Talapova, Polina

In fact, in the home of vocabularies, we have just names or terms indicating these medical conditions.

0:17:43.870 --> 0:18:7.480

Talapova, Polina

They were reported by someone we don't know who it was, but in fact it's usually standard coding system systems and in OMOP we use SNOMED for this purpose, but also in the world in the real world data you can see ICD's like ICD 10, ICD 11 accords which also about the condition domain.

0:18:8.130 --> 0:18:10.640

Talapova, Polina

Next one the observation domain.

0:18:10.650 --> 0:18:22.620

Talapova, Polina

This domain captures clinical facts about a person obtaining the context of examination, questioning procedure as well as social, lifestyle and historical facts.

0:18:22.710 --> 0:18:27.840

Talapova, Polina

So mostly in observation you can see terms which are not covered by all other domains.

0:18:28.960 --> 0:18:37.270

Talapova, Polina

So some people call these domain like garbage domain, but I cannot say that it's fully useless.

0:18:37.280 --> 0:18:44.950

Talapova, Polina

It's not, uh, so you can actually find their some quite useful terms for sure, so don't avoid it.

0:18:45.560 --> 0:18:58.240

Talapova, Polina

The next domain it's a drug domain and this one includes formulations of biochemical substances that were administered to person and what is important. This substances?

0:18:59.500 --> 0:19:5.320

Talapova, Polina

Uh influences on the metabolism inside the person body.

0:19:5.930 --> 0:19:11.620

Talapova, Polina

This is important because, for instance, device domain is a bit different one.

0:19:11.630 --> 0:19:21.250

Talapova, Polina

It's also can have substances, but these substances usually don't affect metabolism, so talking about devices this one captures.

0:19:21.340 --> 0:19:34.490

Talapova, Polina

It's a exposure to foreign physical objects, substance or instruments used for diagnostic or therapeutic purposes, but again, through a mechanism beyond chemical action, usually.

0:19:35.990 --> 0:19:38.750

Talapova, Polina

So let's keep this information in mind.

0:19:39.870 --> 0:19:43.380

Talapova, Polina

Next one what is important for domain selection?

0:19:43.660 --> 0:20:2.380

Talapova, Polina

Of course it's sources we have to listen to our sources, to trust our sources, and first of all, we pay attention the source tables from which we get the source terms and this will be helpful for choosing the preferred target domains during the mapping process.

0:20:3.390 --> 0:20:16.800

Talapova, Polina

If you have table for instance, call diagnosis, you expect to see or to find condition domain concepts or observation domain concepts for some input events or medications.

0:20:16.870 --> 0:20:28.920

Talapova, Polina

They can be drugs and devices also possible if you have source term source table called lab events you expect to see their measurement domain terms.

0:20:29.830 --> 0:20:42.610

Talapova, Polina

Like you said, the next one like micro events as Simon and this one should have specimens only micro events test indicates measurement and we can expect this.

0:20:43.100 --> 0:20:46.250

Talapova, Polina

But again, you have to be familiar with the source data.

0:20:46.260 --> 0:20:51.380

Talapova, Polina

If you not, you have to define the preferred domain by yourself.

0:20:54.40 --> 0:20:56.840

Talapova, Polina

And we will try to do this.

0:20:57.40 --> 0:21:1.210

Talapova, Polina

So in examples source term X ray clavicle.

0:21:1.220 --> 0:21:12.430

Talapova, Polina

So we we hope that we're not, we don't hope, we just expect that this term refers to procedure domain because it's a actual procedure.

0:21:14.370 --> 0:21:15.220

Talapova, Polina

Umm.

0:21:15.850 --> 0:21:33.380

Talapova, Polina

Which just usually performed with patient and the next one the same MRI of thoracic spine also procedure and usually such source terms don't have values as I said before, the next one hallux flow and literal per minute.

0:21:33.910 --> 0:21:42.470

Talapova, Polina

We expect that this storm contains values, so it means that this one should be bought measurement another one.

0:21:42.520 --> 0:21:46.930

Talapova, Polina

It's a assessment scale score also values there.

0:21:46.940 --> 0:21:53.830

Talapova, Polina

So we expect to see measurement and we will look through the concepts inside this domain for sure.

0:21:54.340 --> 0:21:57.930

Talapova, Polina

Then gender, gender is not a measurement.

0:21:57.980 --> 0:22:9.170

Talapova, Polina

It's not a procedure, and it's not a condition, not a drug, so it's observation domain should be then lower, not not lower left upper extremity sensation.

0:22:9.300 --> 0:22:15.530

Talapova, Polina

So we perform a test to understand the level of sensation, the extremities.

0:22:16.10 --> 0:22:24.430

Talapova, Polina

It's about measurement because we expect to have the value of this examination and the same with tidal volume.

0:22:25.560 --> 0:22:30.400

Talapova, Polina

Even it's a setting for ventilator doesn't matter, it's still measurement.

0:22:31.770 --> 0:22:33.800

Talapova, Polina

Then let's try to do it together.

0:22:34.110 --> 0:22:34.980

Talapova, Polina

How do you think?

0:22:34.990 --> 0:22:37.180

Talapova, Polina

What's about urine color?

0:22:37.450 --> 0:22:39.370

Talapova, Polina

What is the source domain of this term?

0:22:45.210 --> 0:22:45.880

Talapova, Polina

Urine color.

0:22:56.120 --> 0:22:56.610

Sanz, Javier

Measurement.

0:22:54.470 --> 0:23:4.990

Talapova, Polina

I'm going to wait, but yes, yes measurement because color can be different and this is a value then absolute count buzzes, buzzes.

0:23:5.0 --> 0:23:7.700

Talapova, Polina

Basil fills, so this is also.

0:23:12.780 --> 0:23:13.210

Sanz, Javier

Measurement.

0:23:13.40 --> 0:23:16.860

Talapova, Polina

Also measurement great, the third one, headache.

0:23:17.830 --> 0:23:18.300

Talapova, Polina

What did that?

0:23:23.50 --> 0:23:23.730

Talapova, Polina

It's a symptom.

0:23:24.920 --> 0:23:25.430

Sanz, Javier

Condition.

0:23:26.20 --> 0:23:26.890

Talapova, Polina

Condition.

0:23:26.900 --> 0:23:27.390

Talapova, Polina

Great.

0:23:27.720 --> 0:23:29.390

Talapova, Polina

Then we have pain intensity.

0:23:32.80 --> 0:23:32.720

Talapova, Polina

Not so easy.

0:23:37.530 --> 0:23:38.330

Yaxing Liu

That was a vacation.

0:23:38.960 --> 0:23:40.800

Talapova, Polina

Yes, it's observation.

0:23:41.650 --> 0:23:47.40

Talapova, Polina

Uh, but also some people can say ohh, but also they can give some values.

0:23:47.50 --> 0:24:2.220

Talapova, Polina

Why it's not measurement we cannot answer this question why it's not measurement, but in normal it's observation and in source data it's usually also observation this something that we can observe on the patient's next one.

0:24:2.830 --> 0:24:4.770

Talapova, Polina

Coma scale, eye opening.

0:24:11.50 --> 0:24:11.760

Talapova, Polina

How do you think?

0:24:11.770 --> 0:24:12.320

Talapova, Polina

What is it that?

0:24:13.390 --> 0:24:13.990

Sanz, Javier

Naturally.

0:24:14.410 --> 0:24:17.260

Talapova, Polina

Measurement correct, then chest tube.

0:24:23.480 --> 0:24:24.680

Talapova, Polina

Without any details.

0:24:29.800 --> 0:24:30.330

Sanz, Javier

Procedure.

0:24:32.680 --> 0:24:33.20

Yaxing Liu

Vice.

0:24:31.150 --> 0:24:33.300

Talapova, Polina

No, it's a device.

0:24:34.0 --> 0:24:35.150

Sanz, Javier

Otherwise yes.

0:24:33.490 --> 0:24:35.630

Talapova, Polina

Just you yeah.

0:24:35.640 --> 0:24:49.520

Talapova, Polina

Device because if you don't have any additional information in the source, don't try to find this information in the target and avoid to use concepts with additional abundant information and the last one was oppressing.

0:24:54.670 --> 0:24:55.210

Guan,Ziyuan

Medication.

0:24:56.60 --> 0:25:0.230

Talapova, Polina

Medication or drug domain in normal PS correct.

0:25:0.240 --> 0:25:10.800

Talapova, Polina

So it sometimes it can be challenging, more challenging or less challenging and but in some cases we even don't know what to choose.

0:25:10.810 --> 0:25:19.280

Talapova, Polina

But in the next slide, so they should be the slide about what to do when you're not sure what the, what is the source domain.

0:25:19.730 --> 0:25:24.40

Talapova, Polina

But before this, let's discuss vocabulary selection principle.

0:25:25.90 --> 0:25:46.370

Talapova, Polina

Umm so once you know domain were you are a bit sure about this domain then you can filter vocabularies, standard vocabularies with storms to limit the space for searching for standard term for instance measurement domain number one priority.

0:25:46.420 --> 0:25:48.330

Talapova, Polina

It's allowing vocabulary.

0:25:48.640 --> 0:26:3.510

Talapova, Polina

It means that if you have several terms indicating almost the same source term but in different vocabularies at the same time in low income SNOMED or in low income normal extension, we recommend you to use low ink at first.

0:26:4.60 --> 0:26:6.530

Talapova, Polina

If you don't have low ink, you go to SNOMED.

0:26:6.540 --> 0:26:10.690

Talapova, Polina

If you don't have SNOMED, you use OMOP extension and all others.

0:26:10.750 --> 0:26:25.900

Talapova, Polina

If you don't have any of them, but alright, it had to say, I expect that if I don't have standard term in LOINC, SNOMED or OMOP extension, it means that this is a gap in OMOP vocabulary.

0:26:25.910 --> 0:26:32.500

Talapova, Polina

Because these three guys are the most important, standard vocabularies, they should have this.

0:26:33.860 --> 0:26:41.460

Talapova, Polina

But anyway, procedure domain a bit different scenario number one priority is to SNOMED for procedure.

0:26:41.780 --> 0:26:43.570

Talapova, Polina

Then we have OMOP extension.

0:26:43.580 --> 0:26:46.360

Talapova, Polina

Homophobic tension always goes after SNOMED.

0:26:47.560 --> 0:26:50.60

Talapova, Polina

Then numbers sort priority.

0:26:50.70 --> 0:26:51.690

Talapova, Polina

Vocabulary is allowing.

0:26:52.250 --> 0:27:0.500

Talapova, Polina

Thank if you don't have any strong three, we can use ICD 10 PCs or ICD 9 procedure.

0:27:0.770 --> 0:27:5.210

Talapova, Polina

They also contain standard terms and we can use them.

0:27:5.660 --> 0:27:11.910

Talapova, Polina

Except this you may see standard procedures in opcs 4IN, CPT.

0:27:11.920 --> 0:27:18.910

Talapova, Polina

Four and he picks also possible to use, but for instance he picks or CPT four.

0:27:18.920 --> 0:27:29.550

Talapova, Polina

They have too much granular standard terms and there are a lot of details and for large scale analytics they are not so nice to use then.

0:27:30.80 --> 0:27:36.640

Talapova, Polina

Sort one sort domain or a certain force condition or observation doesn't matter for them.

0:27:38.10 --> 0:27:40.380

Talapova, Polina

High priority vocabulary is a SNOMED.

0:27:40.490 --> 0:27:43.980

Talapova, Polina

If we don't have SNOMED equivalent, we use OMOP box tension.

0:27:44.390 --> 0:27:51.210

Talapova, Polina

If there there are no equivalents, it means that we have gap devices for devices.

0:27:51.220 --> 0:27:56.350

Talapova, Polina

We use SNOMED at 1st and all others at the next, and the drugs for drugs.

0:27:57.380 --> 0:28:0.310

Talapova, Polina

Umm, very different vocabularies we use.

0:28:0.320 --> 0:28:3.690

Talapova, Polina

So first place goes through RX norm for U.S.

0:28:3.700 --> 0:28:4.20

Talapova, Polina

Market.

0:28:5.220 --> 0:28:5.530

Talapova, Polina

Uh.

0:28:5.540 --> 0:28:9.550

Talapova, Polina

2nd place is for our economy extension for non U.S.

0:28:9.560 --> 0:28:16.30

Talapova, Polina

Market like Europe, Asia, whatever and sort one vocabulary that we use is CVX.

0:28:16.240 --> 0:28:26.730

Talapova, Polina

This one used for vaccines mapping, usually because analytics norm there is a chaos for vaccines, so they can be too much granular with a lot of details that we don't have in the source.

0:28:27.40 --> 0:28:30.430

Talapova, Polina

That's why CVX can be quite useful for these cases.

0:28:32.360 --> 0:28:34.650

Talapova, Polina

And now how to define a source domain?

0:28:34.700 --> 0:28:40.630

Talapova, Polina

If it is unclear, first of all, read and understand the arm of definition of domains.

0:28:40.640 --> 0:28:45.240

Talapova, Polina

It's very important to remember what are their names and what are they about.

0:28:45.620 --> 0:28:49.430

Talapova, Polina

Then comprehend the concept of a source term.

0:28:49.440 --> 0:28:53.190

Talapova, Polina

It means just understand deeply the meaning of the source term.

0:28:53.200 --> 0:28:59.570

Talapova, Polina

If it's possible, sometimes source terms are represented by some strange short.

0:29:0.640 --> 0:29:17.50

Talapova, Polina

I don't know pieces of words and it's impossible to guess what is meaning, but in the ideal scenario you have to understand the idea, then recommendation to recode abbreviations and acronyms, if any, in the source.

0:29:17.60 --> 0:29:18.710

Talapova, Polina

Those can be very helpful.

0:29:18.840 --> 0:29:30.510

Talapova, Polina

So you can use synonyms for the search and maybe you know that Athena sometimes doesn't work properly, especially in terms of synonyms.

0:29:30.580 --> 0:29:31.290

Talapova, Polina

Search.

0:29:31.420 --> 0:29:37.830

Talapova, Polina

You can try one second, third combination of word and only the first one will give you the result.

0:29:37.900 --> 0:29:46.10

Talapova, Polina

That's why be creative and this also search for relevant scientific literature can be articles or some guidelines.

0:29:46.20 --> 0:29:46.930

Talapova, Polina

But what is important?

0:29:46.940 --> 0:29:47.870

Talapova, Polina

Scientific.

0:29:47.880 --> 0:29:51.640

Talapova, Polina

Scientific not for just civilians.

0:29:52.210 --> 0:29:53.760

Talapova, Polina

UM and another way.

0:29:53.770 --> 0:30:8.100

Talapova, Polina

It's also consult Chad GPT, but the best idea to provide chat GPT with some context like additional information which can help chat GPT to understand what is the source term about.

0:30:8.470 --> 0:30:12.970

Talapova, Polina

But again, this is useful only for manual mapping, not for automated one.

0:30:14.780 --> 0:30:16.890

Talapova, Polina

And now all vocabulary gaps.

0:30:16.900 --> 0:30:17.990

Talapova, Polina

Identification.

0:30:18.320 --> 0:30:19.680

Talapova, Polina

Let's start from the types.

0:30:20.810 --> 0:30:30.780

Talapova, Polina

Umm, there can be different types of gaps in the vocabulary, but first of all I want to distinguish two of them, quantum gaps and hierarchy gaps.

0:30:31.190 --> 0:30:38.620

Talapova, Polina

Quantum gaps means that we can see absence standard equivalent term for our source term.

0:30:38.850 --> 0:30:40.920

Talapova, Polina

We can see missing synonyms.

0:30:41.230 --> 0:30:46.660

Talapova, Polina

For instance, we're trying to find some abbreviation, but it doesn't work with.

0:30:46.700 --> 0:31:5.650

Talapova, Polina

Then we try different variations of name and then we luckily find the term and when we check synonyms in Athena, we cannot see the synonym which was required to find this concept in this simple way, it means that it would be great to add such synonym to the vocabulary.

0:31:6.100 --> 0:31:10.430

Talapova, Polina

The third content gap option is absent vocabulary.

0:31:10.440 --> 0:31:16.930

Talapova, Polina

In general, we had such situation with GIS vocabularies.

0:31:16.940 --> 0:31:22.960

Talapova, Polina

For instance, social determinant of health such space was not covered totally in OMOP.

0:31:22.970 --> 0:31:28.240

Talapova, Polina

That's why it's also possible just not to see the vocabulary you need in their OMOP.

0:31:28.250 --> 0:31:35.290

Talapova, Polina

It's a more complex problem than just steps and term, but also possible to cover missing attribute.

0:31:35.660 --> 0:31:44.330

Talapova, Polina

Missing attribute can be important for those who are working usually with database with concept relationship table.

0:31:44.340 --> 0:31:56.130

Talapova, Polina

Maybe with lowering sources with noise sources and want to filter the concepts using some attributes or constitutional parts of the concept.

0:31:56.260 --> 0:32:7.210

Talapova, Polina

For instance, you want to find all concepts that have some particular law Inc method or some all concepts which relate to particular specimen.

0:32:7.580 --> 0:32:15.470

Talapova, Polina

That's why you can use attributes as well and some of them can be missing, especially in SNOMED.

0:32:15.480 --> 0:32:35.210

Talapova, Polina

They have a lot of not fully defined terms with the not a full set of attributes attached to the term and according to hierarchy gaps it means that you can find the target term which covers fully your source term.

0:32:35.260 --> 0:32:43.30

Talapova, Polina

But this target term for some reason is not connected with some branch hierarchical branch in this moment.

0:32:43.40 --> 0:32:58.910

Talapova, Polina

For instance, like finding related to pregnancy and some complication of pregnancy and when you use Atlas and when you want to pick the highest hierarchy term to cover all descendants, you can lose some.

0:32:59.730 --> 0:33:4.60

Talapova, Polina

Some of these storms, that's why hierarchy gaps also important.

0:33:4.650 --> 0:33:11.60

Talapova, Polina

But in this presentation, we're not going to cover hierarchy gaps and normal for sure, but just to let you know about that.

0:33:12.470 --> 0:33:14.400

Talapova, Polina

Umm, what about principles?

0:33:14.410 --> 0:33:22.700

Talapova, Polina

So for Gabriel gaps identification, the first one is that these gaps can be found just during mapping and after the mapping.

0:33:22.710 --> 0:33:27.550

Talapova, Polina

It's impossible to predict what will be missing always.

0:33:27.610 --> 0:33:38.850

Talapova, Polina

You can find the gaps once mapping is done and what is another principle that is very important according to the standard module team.

0:33:39.220 --> 0:33:46.170

Talapova, Polina

We recommend you to use Cesar mapping metadata to preserve information and mapping inconsistencies.

0:33:46.700 --> 0:33:59.130

Talapova, Polina

Whether it's abundant granularity or lack of granularity or absence of mapping at all, so you can use this metadata and let's look at the examples.

0:33:59.140 --> 0:34:0.730

Talapova, Polina

For instance, object label.

0:34:0.740 --> 0:34:2.900

Talapova, Polina

It's a source storm blood glucose.

0:34:3.640 --> 0:34:10.840

Talapova, Polina

And when we look at Athena, we will find at least 2 standard concepts which cover blood glucose.

0:34:12.10 --> 0:34:17.560

Talapova, Polina

First one, it's allowing glucose must volume and blood and glucose measurement blood.

0:34:18.130 --> 0:34:28.140

Talapova, Polina

And if you look at our vocabulary priority, let me just look here like measurement first priority, allowing second ones SNOMED.

0:34:28.150 --> 0:34:39.90

Talapova, Polina

Let's come back and what we have, we have low import these confidence, 0.9 and SNOMED code, second row 0.8.

0:34:39.250 --> 0:34:44.410

Talapova, Polina

It means that the highest confidence goes to the law in vocabulary and lowering term.

0:34:44.780 --> 0:34:57.340

Talapova, Polina

It means that during ETL we can filter out the highest confidence the terms with highest confidence, but we will preserve this information and we can submit this.

0:34:57.760 --> 0:34:59.850

Talapova, Polina

For instance, SNOMED glucose measurement.

0:35:0.10 --> 0:35:3.520

Talapova, Polina

What for the duplication to the vocabulary team.

0:35:3.950 --> 0:35:14.920

Talapova, Polina

That's why this approach can be helpful, especially when you reuse your mappings over and over again and then see and the eye movement facial.

0:35:15.450 --> 0:35:19.260

Talapova, Polina

And if we look, for instance, what does it mean this acronym?

0:35:19.270 --> 0:35:28.970

Talapova, Polina

We know that this is a checklist of nonverbal pain indicators, and this panel contains different items.

0:35:29.40 --> 0:35:36.530

Talapova, Polina

This one is about facial expression, but we don't have such a normal, but we have some panel at least.

0:35:37.720 --> 0:35:38.710

Talapova, Polina

And what does it mean?

0:35:38.720 --> 0:35:46.120

Talapova, Polina

It means that we can use broad match as a predicate ID and confidence one.

0:35:46.130 --> 0:35:51.900

Talapova, Polina

It means that we are very sure that this is a broad match, but we don't have respective item.

0:35:51.950 --> 0:36:5.640

Talapova, Polina

It means that this is a gap where we don't have full equivalent, but we have just some equivalent and there's also can be useful for the gap detection and coverage of these gaps in the future steps.

0:36:6.90 --> 0:36:17.860

Talapova, Polina

And the last one and I'm calling center graph in nuclear medicine means and unfortunately there is no standard equivalent in the OMOP for such a term for such procedure.

0:36:17.950 --> 0:36:22.640

Talapova, Polina

So we can put scores no match or just leave it empty.

0:36:22.710 --> 0:36:32.20

Talapova, Polina

But object ID is null and then we can just obtain all object ID new concepts to cover these gaps.

0:36:32.670 --> 0:36:36.120

Talapova, Polina

How and how exactly we will cover them?

0:36:36.250 --> 0:36:37.600

Talapova, Polina

I will tell you a bit later.

0:36:39.810 --> 0:36:46.110

Talapova, Polina

Next one uh, this slides is about slide is about vocabulary, term gap identification.

0:36:46.440 --> 0:36:56.360

Talapova, Polina

So it's a types of term gaps I already told you about them a bit, but let's just focus on them, particularly here.

0:36:56.440 --> 0:37:3.570

Talapova, Polina

So first type it's a mapping results in a loss of source term detail or granularity.

0:37:3.580 --> 0:37:11.710

Talapova, Polina

It means that source is and umm how to say more granular than target term.

0:37:11.920 --> 0:37:23.280

Talapova, Polina

In this case, we use broad match and let's come back like here sinpi movement, facial broad match, more granular cmpi concept.

0:37:24.310 --> 0:37:30.140

Talapova, Polina

Then second type mapping brings extra details which are not mentioned in the source.

0:37:30.330 --> 0:37:33.540

Talapova, Polina

In this case, we use schools, narrow match.

0:37:33.850 --> 0:37:35.280

Talapova, Polina

Let's come back and look at that.

0:37:36.700 --> 0:37:59.190

Talapova, Polina

Uh, there is no such example unfortunately, but anyway it means that, for instance, in source we have

blood glucose and the target will be glucose measurement in blood using some specific method and this specific method will be abundant information which we don't want to have in our mapping.

0:38:0.290 --> 0:38:12.230

Talapova, Polina

This sort one sort type umm the case when we have several similar standard concepts for one case this one is represented here by blood glucose.

0:38:12.410 --> 0:38:18.320

Talapova, Polina

So we use narrow match to identify what's of them and different confidence.

0:38:18.370 --> 0:38:37.500

Talapova, Polina

So the higher confidence, the more preferred term we have and this is the rule score predicate that is caused their own match in different values of the confidence field and the first type of the gap term gap is no equivalent at all.

0:38:37.770 --> 0:38:45.200

Talapova, Polina

In this case we can use predicate ID's cause no match or empty predicate ID, but object ID is null.

0:38:48.310 --> 0:38:53.40

Talapova, Polina

OK, So what to do if we already have the list of gaps?

0:38:53.170 --> 0:38:54.160

Talapova, Polina

Different types of them.

0:38:55.260 --> 0:39:4.710

Talapova, Polina

So once the mapping is ready, we can create concepts with nonequivalent mappings or and or result mappings as standard.

0:39:5.120 --> 0:39:10.170

Talapova, Polina

This is a how to say useful scenario and a lot of people use it.

0:39:10.180 --> 0:39:20.930

Talapova, Polina

So we just create custom plus 2 billion concept IDs and make this terms standard and once the standard we can see them in Atlas.

0:39:20.940 --> 0:39:34.140

Talapova, Polina

If we use Atlas for our analytics, then second step we can submit missing content as a vocabulary contribution and we can click here to look at this page.

0:39:34.490 --> 0:39:39.480

Talapova, Polina

So this is a detailed explanation how to contribute to the omit vocabularies.

0:39:39.490 --> 0:40:3.390

Talapova, Polina

If you find some gap, especially if it's non drug vocabularies, what you should do or what you have to populate there is Google Drive with different templates according according to the type of content that you want to modify and you can click on this Google Drive and find the template that you need etc.

0:40:3.560 --> 0:40:5.180

Talapova, Polina

But this requires some reading.

0:40:6.730 --> 0:40:12.940

Talapova, Polina

Let's come back and the third step is to share missing concepts with the standard module team.

0:40:12.950 --> 0:40:25.600

Talapova, Polina

So we will be happy to have the list of unmapped concepts or concepts with not equivalent mapping that you have to be able to help you with the submission process.

0:40:25.610 --> 0:40:36.870

Talapova, Polina

Creation of the additional new terms, which can be grouped as a one bridge to AI vocabulary which can help us to cover terms which are not covered yet.

0:40:37.500 --> 0:40:38.750

Talapova, Polina

So this is the idea.

0:40:41.20 --> 0:40:43.250

Talapova, Polina

Questions I will be help.

0:40:43.620 --> 0:40:45.870

Talapova, Polina

Happy to answer if you have any of them.

0:40:47.810 --> 0:40:49.150

Williams, Andrew E

I yeah.

0:40:49.230 --> 0:40:50.630

Williams, Andrew E

Thank you so much, Pauline.

0:40:50.640 --> 0:40:51.390

Williams, Andrew E

That was great.

0:40:52.790 --> 0:40:53.50

Talapova, Polina

Thanks.

0:40:51.670 --> 0:40:59.160

Williams, Andrew E

Umm so I I wanna make sure people are clear on what's being asked and the importance of it.

0:40:59.170 --> 0:41:7.140

Williams, Andrew E

So the OMOP vocabularies were not initially developed with sort of critical care in mind.

0:41:7.150 --> 0:41:27.470

Williams, Andrew E

There's all kinds of things that are going to be important for bridge to AI that may not be covered by the standard vocabulary, so I want to make sure everybody's firmly aware of what the importance of having as complete coverage as possible in the standard vocabulary space of the things that are needed.

0:41:28.820 --> 0:41:30.150

Williams, Andrew E

Why that's important.

0:41:30.160 --> 0:41:43.750

Williams, Andrew E

I think that all the analytics that the Odyssey community has developed for developing and validating computable phenotype definitions, it allow you to say who's in a study or what the features and a prediction model are.

0:41:43.760 --> 0:41:58.820

Williams, Andrew E

All of those things require things be mapped to a standard, and so if there are gaps in the coverage of those standards, you just that's a fundamental limiting step for the value of the product that we're all working on.

0:41:58.830 --> 0:42:1.840

Williams, Andrew E

And it's a really great opportunity.

0:42:1.850 --> 0:42:20.470

Williams, Andrew E

This project is a really great opportunity, not only for us and for bridge to AI, but for anybody who comes afterwards who wants to use this whole very well developed infrastructure to do this work, to have a complete way of representing all the important data in and these standard terms.

0:42:20.480 --> 0:42:46.600

Williams, Andrew E

And so your identification of gaps in the way that Polina has described and your contributions and working with us to fill those gaps is going to really, it's not only just going to allow you to kind of meet the goals and so on of the of the project, it's really facilitating this kind of work on a very broad way because the obviously community in OMOP using Community has this global reach beyond what we're doing and and bridge to AI.

0:42:46.610 --> 0:42:48.310

Williams, Andrew E

And so it's just really important.

0:42:49.270 --> 0:42:51.870

Williams, Andrew E

And then Pauline, I did have a question.

0:42:55.900 --> 0:43:0.570

Williams, Andrew E

What do you see as the expectation in terms of developing a mapping?

0:43:0.580 --> 0:43:7.980

Williams, Andrew E

I think is I think you know one thing that we're going to do in the standards module is help develop.

0:43:10.540 --> 0:43:20.110

Williams, Andrew E

At a process as you've outlined for for filling these gaps in some cases, that process is going to include minting a new concept, right.

0:43:20.910 --> 0:43:21.140

Talapova, Polina

Umm.

0:43:20.280 --> 0:43:37.300

Williams, Andrew E

And so I think of this as something where people might try to do an ETL and say, hey, I don't have a thing to map this to and note that and report it to us or report it as a as something to the vocab team or both.

0:43:37.390 --> 0:43:40.560

Williams, Andrew E

I guess that's what we're recommending to the Odyssey vocabulary team.

0:43:41.470 --> 0:43:48.260

Williams, Andrew E

UM, But the ability to actually generate a a a new term?

0:43:48.270 --> 0:44:0.280

Williams, Andrew E

A candidate term that could be added to the vocabulary is a little bit higher level of understanding of like

what that requires and so I think we should help people understand what the expectations are with respect to them.

0:44:0.290 --> 0:44:23.580

Williams, Andrew E

I think for the most part, we're really interested in helping people get an understanding of the gaps where they want to map something and can't and to the extent that they feel able to work in the way that you've described to generate a new term that's suitable for inclusion, we could help with that.

0:44:23.590 --> 0:44:28.840

Williams, Andrew E

But I don't think we're expecting everybody who's doing this mapping to have that capacity, is that right?

0:44:28.850 --> 0:44:32.800

Williams, Andrew E

Would you agree overall with that? Yeah.

0:44:31.310 --> 0:44:33.920

Talapova, Polina

Yeah, yeah, for sure, I agree.

0:44:41.410 --> 0:44:41.610

Williams, Andrew E

Yeah.

0:44:34.750 --> 0:44:44.680

Talapova, Polina

But new term or vocabulary generation process requires maybe separate office hours because it's a bit more technical aspect of this process.

0:44:45.670 --> 0:44:46.460

Williams, Andrew E

Yeah, I was.

0:44:46.550 --> 0:44:50.400

Williams, Andrew E

Maybe I was misreading a little bit of of last couple of slides.

0:44:50.410 --> 0:45:14.880

Williams, Andrew E

It sounded almost as if we were expecting people to once they identify these gaps to kind of make contributions that fill them in, assess some format and and to do all of that on their own, whereas I think we see one of the roles we have in the in the standards core is really being able to either facilitate that or do that for them once adapt gaps or identified and we'd like people to take a stab at mapping them.

0:45:15.620 --> 0:45:22.250

Williams, Andrew E

But we're not expecting that they do that for a living and and know exactly how to do it perfectly right off the bat.

0:45:22.260 --> 0:45:27.800

Williams, Andrew E

So I just wanted to clarify that expectation, but other questions for Polina.

0:45:32.780 --> 0:45:35.230

Dandayudhapani, Satish

Uh, yeah, this is Satish from Seattle children.

0:45:36.120 --> 0:45:51.880

Dandayudhapani, Satish

So my understanding is like this mapping from source to concept, not the gas part of it, but it's just a general when we are starting from scratch where we have identified all the source values and we want to and look for the target concept ID.

0:45:51.970 --> 0:45:53.360

Dandayudhapani, Satish

Is this looks like.

0:45:53.370 --> 0:45:54.760

Dandayudhapani, Satish

Is this a manual process?

0:45:54.970 --> 0:45:58.500

Dandayudhapani, Satish

Is it we have to go one after the other or is it a?

0:45:59.290 --> 0:46:6.340

Dandayudhapani, Satish

Is there a tool or something where we can import the source values and it does match some of the target concepts?

0:46:7.590 --> 0:46:8.530

Talapova, Polina

Great question.

0:46:8.540 --> 0:46:9.520

Talapova, Polina

Thank you very much.

0:46:9.610 --> 0:46:15.860

Talapova, Polina

Usually mapping process includes both automatical or automated part and manual part.

0:46:16.140 --> 0:46:27.860

Talapova, Polina

So we do automatically as many concepts as we can using Asagi or some string match lexical match similarity of the strings, whatever we use.

0:46:27.970 --> 0:46:33.440

Talapova, Polina

But usually we have some percentage of source terms which are not covered automatically.

0:46:33.630 --> 0:46:39.180

Talapova, Polina

That's why this portion of terms have to be to be mobbed manually.

0:46:39.640 --> 0:46:45.630

Talapova, Polina

So it depends on the source data and results of automated part of the mapping.

0:46:47.710 --> 0:46:48.750

Talapova, Polina

This is the answer.

0:46:48.310 --> 0:46:53.990

Williams, Andrew E

And then pair presentations went over some of those other tools for the automated part.

0:46:54.80 --> 0:47:0.170

Williams, Andrew E

I think they're some of the particularly time intensive aspects of doing an ETL.

0:47:2.160 --> 0:47:2.330

Talapova, Polina

Yeah.

0:47:0.180 --> 0:47:15.630

Williams, Andrew E

Are the manual part in making sure that that's done correctly and so you can find and Marty can put in a chat some of the links to the prior presentations on some of the automated tools that help take care of the automated part that the Odyssey community has built?

0:47:15.980 --> 0:47:27.530

Williams, Andrew E

So that's definitely the place to begin the bridge to AI because of the need to get certain things that we know are really important for defining populations, exposures outcomes.

0:47:27.540 --> 0:47:34.610

Williams, Andrew E

And you know, developing features and prediction models, we know that some of the things that you might get are more important than others.

0:47:34.620 --> 0:47:55.710

Williams, Andrew E

And so there's been a very uh in depth process for getting at what are the priority things and that's where some of the decision making about the labor invested to do the manual part, that's that's a little more time consuming is going to be based on those priorities, things that have been already marked out as being of particularly high value.

0:47:55.720 --> 0:47:59.800

Williams, Andrew E

And so that's going to be shared in an official capacity, if it hasn't already.

0:48:1.630 --> 0:48:1.900

Dandayudhapani, Satish

Yeah.

0:48:2.140 --> 0:48:8.870

Dandayudhapani, Satish

OK, so the way I understand is like run through the automated process and then whatever doesn't match.

0:48:8.920 --> 0:48:11.430

Dandayudhapani, Satish

Then you go do the manual fill.

0:48:11.440 --> 0:48:12.10

Dandayudhapani, Satish

The gaps.

0:48:12.680 --> 0:48:14.100

Talapova, Polina

Yeah, correct.

0:48:12.480 --> 0:48:14.270

Dandayudhapani, Satish

Is that right? OK.

0:48:14.110 --> 0:48:21.260

Talapova, Polina

And also I'm sorry after automated part usually review is required, if it's not full string match.

0:48:22.90 --> 0:48:25.150

Talapova, Polina

Yeah, like subject domain expert house.

0:48:25.160 --> 0:48:41.510

Talapova, Polina

To look at this results of mapping and validate them and to validate them correctly, these people have to know the mapping principles, so that concept should be standard which we capillary to use which domains to use etc.

0:48:43.970 --> 0:48:44.180

Dandayudhapani, Satish
OK.

0:48:44.950 --> 0:48:45.300

Dandayudhapani, Satish
Thank you.

0:48:46.180 --> 0:48:48.310

Talapova, Polina
Come thanks to you.

0:48:55.880 --> 0:49:5.510

Schmidt, Heidi E.

You may have already covered this, but I'm gonna ask, umm, the source to concept MAP OMOP table.

0:49:7.220 --> 0:49:12.750

Schmidt, Heidi E.

It comes empty at in default when you deploy the DDL.

0:49:14.280 --> 0:49:17.590

Schmidt, Heidi E.

Is that something that has to be populated?

0:49:18.800 --> 0:49:19.0

Talapova, Polina
Yeah.

0:49:19.820 --> 0:49:27.660

Schmidt, Heidi E.

And is that something that you do with your own SQL or is there a specific tool that you should be pointing?

0:49:33.660 --> 0:49:34.140

Talapova, Polina
OK.

0:49:34.150 --> 0:49:35.430

Talapova, Polina

So thank you for your question.

0:49:35.770 --> 0:49:36.280

Talapova, Polina
Uh.

0:49:36.360 --> 0:49:45.280

Talapova, Polina

Source the concept map is not populated by the default, so you have to populate it by yourself using SQL queries or manual mapping like.

0:49:45.290 --> 0:49:49.50

Talapova, Polina

Just upload the table with mappings and populate source the concept map.

0:49:49.820 --> 0:50:3.330

Talapova, Polina

Umm, another way to get source to concept MAP table is from Usagi Usagi allows to upload the like output of those saggy one of the option is to store it as a source to concept MAP table.

0:50:3.700 --> 0:50:12.510

Talapova, Polina

These advantage of this table is that you can store only maps to maps to value relationship ID without any additional mapping metadata.

0:50:12.520 --> 0:50:23.310

Talapova, Polina

So it's impossible to identify whether it's full equivalent mapping or partial equivalent mapping, or just like too much narrow mapping or whatever mapping.

0:50:23.990 --> 0:50:30.300

Talapova, Polina

So there is no additional fields which can help to sort this information and reuse it in the future.

0:50:30.370 --> 0:50:40.410

Talapova, Polina

So you you have to do this again and again every time you have to deal with the duplicate standard duplicates again and again and it can be painful.

0:50:41.700 --> 0:50:48.250

Schmidt, Heidi E.

Have have we had a session on you? Yeah.

0:50:42.230 --> 0:50:48.460

Williams, Andrew E

It's important, Heidi, maybe to sorry to add quickly a little bit to what Polina said.

0:50:48.470 --> 0:50:52.700

Williams, Andrew E

I I think you know it's optional.

0:50:52.810 --> 0:50:53.620

Williams, Andrew E

There's just a start.

0:50:53.630 --> 0:50:58.480

Williams, Andrew E

What you asked, I think, do you have to do that and you don't have to.

0:50:58.550 --> 0:51:11.60

Williams, Andrew E

It's there as a way to record and preserve what you've done in a way that needs some standard location tools don't run off of it.

0:51:11.310 --> 0:51:12.220

Williams, Andrew E

It's not.

0:51:12.650 --> 0:51:13.770

Williams, Andrew E

It's not meant for.

0:51:16.880 --> 0:51:35.90

Williams, Andrew E

Anything other than kind of a library of of decisions you've made about what to map to what I think it can be the source of other things, but it is not like all the machinery of using an OMOP CDM work just fine with an empty source to concept MAP table.

0:51:35.180 --> 0:51:37.930

Williams, Andrew E

It's there to kind of say these are the decisions we made.

0:51:37.940 --> 0:51:49.530

Williams, Andrew E

And uh, there's a lot of ways it can be useful in, particularly when you're doing an ETL or you have a lot of non standard concepts that you you want to make sure you're preserving in a single place.

0:51:49.540 --> 0:51:56.100

Williams, Andrew E

So just to be clear about what that's for and and how how it gets used, it's so I want to make sure that was clear.

0:52:1.680 --> 0:52:2.800

Talapova, Polina

Source the costs have moved.

0:52:2.810 --> 0:52:11.590

Talapova, Polina

Table can be used for the population of event tables and the way how you will populate the source to concept MAP are different.

0:52:11.600 --> 0:52:30.90

Talapova, Polina

So as I said before, queries manual mapping or for instance session mapping table converted into the source to concept MAP table and in source to concept MAP you can remain only high confidence mappings with exact matches already.

0:52:30.100 --> 0:52:35.670

Talapova, Polina

And be sure that all your maps do are full equivalent maps too, but not something else.

0:52:37.990 --> 0:53:5.90

Schmidt, Heidi E.

OK, so it could be essentially a dimensional lookup table just to know the heuristics of what you've been using inside your database, and then when you have to reconcile future Athena vocab versions, you're going to be looking to that is your main source, but it's up to you to keep it updated and In Sync with your data as you're populating it.

0:53:6.320 --> 0:53:6.700

Talapova, Polina

Correct.

0:53:8.440 --> 0:53:9.30

Schmidt, Heidi E.

Thank you.

0:53:9.80 --> 0:53:10.90

Schmidt, Heidi E.

It has there been?

0:53:11.60 --> 0:53:14.190

Schmidt, Heidi E.

I'm not always able to make all of the wonderful demos.

0:53:14.360 --> 0:53:17.0

Schmidt, Heidi E.

Has there been a you soggy demo in the past?

0:53:18.100 --> 0:53:18.670

Schmidt, Heidi E.

Uh.

0:53:18.680 --> 0:53:21.400

Schmidt, Heidi E.

Since like May or.

0:53:23.980 --> 0:53:29.220

Talapova, Polina

I have to ask Jared because I'm not sure I didn't for sure, but maybe Jared did.

0:53:31.990 --> 0:53:32.940

Alvarez, Marta

Don't think so.

0:53:33.840 --> 0:53:34.400

Alvarez, Marta

Excuse me.

0:53:34.480 --> 0:53:39.450

Alvarez, Marta

I'll take a quick peek, but he may have mentioned it in passing, but I don't think it was specific about you soggy.

0:53:40.320 --> 0:53:40.560

Schmidt, Heidi E.

OK.

0:53:43.390 --> 0:53:43.750

Schmidt, Heidi E.

Thank you.

0:53:44.290 --> 0:53:44.840

Alvarez, Marta

But thank you.

0:53:44.850 --> 0:53:47.170

Alvarez, Marta

That might be a good option to add.

0:53:48.700 --> 0:53:49.750

Talapova, Polina

Yeah, for sure.

0:53:49.840 --> 0:53:51.800

Talapova, Polina

We can demand this one as well.

0:53:54.680 --> 0:54:5.990

Williams, Andrew E

It was covered, I think, as one of the tools that are bundled up in Perseus, which were no longer suggesting people use, but it's one of the one of the things that's a part of that.

0:54:6.0 --> 0:54:8.250

Williams, Andrew E

So I think it's it was just very, very briefly describe.

0:54:8.260 --> 0:54:10.760

Williams, Andrew E

It wasn't like a demo of how to use it.

0:54:11.240 --> 0:54:18.400

Williams, Andrew E

I think a demo of how we are using source to concept MAP table and how some people use it to drive loading of event tables.

0:54:18.410 --> 0:54:23.550

Williams, Andrew E

As with pulling the describing is, it's also maybe a useful thing.

0:54:24.0 --> 0:54:24.280

Williams, Andrew E

You're.

0:54:24.650 --> 0:54:26.380

Williams, Andrew E

It was a good question, Heidi.

0:54:26.530 --> 0:54:31.360

Williams, Andrew E

It's people differ on that, so there isn't an exact standard way.

0:54:31.370 --> 0:54:37.480

Williams, Andrew E

It's a useful thing, and maybe a demo of how it's useful is there's a good thing to think about covering also.

0:54:38.340 --> 0:54:49.670

Schmidt, Heidi E.

Yeah, I kind of have to think about it from like the what do I need to do and then what would be useful to have around if other people are putting everything into a centralized place.

0:54:57.430 --> 0:55:2.560

Talapova, Polina

Thank you, Heidi, for your active participation and questions and brilliant ideas.

0:55:3.530 --> 0:55:4.660

Talapova, Polina

We really appreciate it.

0:55:9.370 --> 0:55:10.10

Schmidt, Heidi E.

Thanks for your help.

0:55:29.280 --> 0:55:29.520

Alvarez, Marta

OK.

0:55:30.870 --> 0:55:33.410

Alvarez, Marta

Any other questions before we wrap up for the day?

0:55:37.20 --> 0:55:40.80

Williams, Andrew E

There's a question, but I'll say that we are using uh.

0:55:40.90 --> 0:55:43.900

Williams, Andrew E

Thank you very much for your wonder who's been updating your status.

0:55:43.910 --> 0:55:54.420

Williams, Andrew E

We are looking at that in our weekly meetings and hope that people will start using discussions a bit more for the questions that come up.

0:55:55.970 --> 0:56:8.780

Williams, Andrew E

I think we are, uh, starting to kind of get the hang of understanding how people are updating things and what it implies in terms of what we should do, but it's still an evolving thing.

0:56:8.790 --> 0:56:21.140

Williams, Andrew E

So we're I hope that that's been useful for you and that you feel free to ask questions in that uh, in that setting where you're adding your updates to it.

0:56:21.670 --> 0:56:28.110

Williams, Andrew E

And so that work is is very valuable and thank you very much for for continuing to do that on a regular basis.

0:56:45.760 --> 0:56:46.730

Williams, Andrew E

That sounds like it's it.

0:56:46.740 --> 0:56:48.190

Williams, Andrew E

That's a deafening silence there.

0:56:48.200 --> 0:56:48.830

Williams, Andrew E

We want to end.

0:56:48.840 --> 0:56:53.0

Williams, Andrew E

We should like pop a quirk or do something to like end on a high note there.

0:56:53.10 --> 0:56:54.910

Williams, Andrew E

That was it was a great presentation.

0:56:55.0 --> 0:56:56.330

Williams, Andrew E

Polina has always thank you so much.

0:56:56.340 --> 0:56:57.530

Williams, Andrew E

And thanks everybody for coming.

0:56:57.800 --> 0:57:0.70

Williams, Andrew E

And we'll see you next week like.

0:57:1.50 --> 0:57:1.210

Alvarez, Marta

You.

0:57:0.470 --> 0:57:1.740

Talapova, Polina

You take care.

0:57:2.10 --> 0:57:2.260

Talapova, Polina

Bye bye.

0:57:2.840 --> 0:57:3.140

Dandayudhapani, Satish

Thank you.

0:57:6.10 --> 0:57:6.370

OZRAZGAT BASLANTI,Tezcan

Thank you.