

Transcript

August 22, 2024, 4:59PM



Alvarez, Marta 4:12

Enter Jared.

I have a slide to share.

When you guys are ready to start, just watching to see who's on, if there's anybody new.



Williams, Andrew E 4:21

Yeah.

What kind of a slide is it, Marty?



Alvarez, Marta 4:27

It's a pretty slide.

It's just this.



Williams, Andrew E 4:29

It's a pretty slight.



Alvarez, Marta 4:30

No, no.

It's just so funny I put together for the announcements.



Williams, Andrew E 4:36

Fantastic.

Why don't you put it on it?



Alvarez, Marta 4:41

There's two I will just.

Do that.



Williams, Andrew E 4:50

And tastic.

Great way to begin.



Alvarez, Marta 4:52

This is not showing up, are there?

It is OK.



Williams, Andrew E 4:59

Nice.

So indeed, thank you all who did that.

We kind of suggested if you could please please do that.

And if you did it last time, no need to now if you haven't, just love it.

If we get as many many sites to update that data acquisition Google form so we can umm, and thanks for putting that up there Marty and for for everybody who did that right away so great.



Alvarez, Marta 5:38

Am I still sharing?



Williams, Andrew E 5:39

You are.



Alvarez, Marta 5:40

OK, great.

So that's your announcements.

I'm going to put the link in the chat right now.



Williams, Andrew E 5:46

The link for the up Google form that people might update.



Alvarez, Marta 5:49

For the Google form, yeah.



Williams, Andrew E 5:51

Great.

Do you want to go over any of these?

Smarter.

Do you want me to?



Alvarez, Marta 6:08

I love you too. Umm.



Williams, Andrew E 6:11

Umm, a while back we.

Began a process that we want to continue to help sites who have to manage the fact that they themselves don't have direct access to the HR data that they're putting in up form, but get it from another group and their institution.

Several sites are in that situation and we had a meeting where what the situation is like at each site was described as well as some possible avenues, including, you know, asking other people to.

Help escalate any kinds of requests that might.

Be made on behalf of the project to relevant people who could help expedite things as well as kind of making the case for the value added at the institution.

So if there's a group that's already.

Uh providing.

Almost data for either centralized resource or for particular projects.

Saying this additional work that you're doing for bridge to AI is actually going to add value to increase confidence and comprehensiveness, and the data that they're providing, and rather than a favor to you for the purposes of the project, it might be feather in their cap to show how great their work is and that that kind of a strategy and helping to facilitate any communications around that or descriptions around that.

And then other sites have other descriptions but that and other stuff has been summarized and put into a discussion thread.

We're going to uh, I think provide a a way to who sign up so that when sites other than your own who are in that same situation add to the discussion thread, you get a notification or you could unsubscribe to that.

I think that's a feature that's in the discussion thread.

There's a stub there.

And so that's just what that's about.

I think before we have another meeting, we kind of want to see if we can.

This isn't something that the Central Group of standards or data acquisition kind of has a a unique and universal answer to that we can do.

We can mostly play this kind of facilitative role to help sites help each other, and convene conversations and help centralized and disseminate useful strategies.

And that's what this is an effort to do.

We might have other meetings in the future.

In the meantime, we're going, we've consolidated some of what was discussed previously in the beginning of that discussion thread and urge you all to participate.

There.

I don't know, Jared.

You everything else to say about the discussion thread for that?



Houghtaling, Jared 8:53

No, I think it's just kind of this idea of community engagements.

It's named at sites who are working with a third party and coordinating the ETL's and trying to kind of relay the information that you're getting here to somebody else.

That's not necessarily involved in the calls and really with a focus on like ways that you've maybe addressed challenges or things that have come up.

Maybe you didn't expect or anything that might be relevant for somebody who's earlier on in the process than you are just as kind of a place to share this and discuss solutions to problems that have come up along the way so.



Williams, Andrew E 9:30

Yep.



Houghtaling, Jared 9:33

Awesome.



Williams, Andrew E 9:35

And I'll just note, I think we generally tend to think of some of the kinds of discussions that come up as being things that we might be being kind of folks who are deeply enmeshed in some of the main kinds of institutions and practices for defining, you know, the best way to do something like some of the Etling work into OMAP or other kinds of things that we're we might be able to provide.

And we probably can provide definitive solutions and kind of asking us in some cases might be the best way to do it.

Uh to get to a a A conclusion and and to have a discussion that gets recorded go and other cases we don't have that we're all kind of starting from a similar spot and it's going to be unique to each individual situation we lot of the community engagement activity is going to be required because we don't have there isn't such a thing as a central group of experts who might know what the what the right answer is sort of a thing and so we'll try to make that clear where discussion that's you know using the same.

Format we kind of regarded as like here.

These are sort of the authoritative resolution of of a question or discussion that's been opposed, or one that's just sort of an ongoing thing where best practices are being shared and and people are thinking, so they'll help guides like what you're looking for or how you might think about participating in the discussion. UM.

Not that we won't always be learning from each other.

We always will, but it's there's a bit of a bit of a difference there and probably topical areas that how we think about the discussions functioning E.

So volunteers are needed to present ETL challenges or other aspects of data that may help sites with similar concerns thought there would be a lot of value to kind of going through some.

Particular experiences that individual sites have had to understand what's been most challenging and difficult, how you've resolved issues, and to do it in a way that's, you know, facilitated by these calls and these meetings probably with fairly brief overview. And so we're really we think that's going to be useful.

I'll just seeing what others have done, how kind of challenges they've they've met and overcome umm lot of expertise and the community that we were forming here and sharing what you've learned discussing how it does or doesn't relate to other sites.

We think is going to be very broadly useful.

So we're looking for volunteers to help with that and to be scheduled on upcoming

calls.

Umm.

So please reach out to Marty if you are willing to to do that.

Anything more to be said about that, Jared?



Houghtaling, Jared 12:38

I just So what I kind of going back to your community engagement versus standards question.

I do have now labels that I'm starting to label the threads with to hopefully identify which kind of the intention of the thread based on the topic that's discussed.



Williams, Andrew E 12:48

Umm.



Houghtaling, Jared 12:55

So I can, if you go on there, you'll see that there's like a green label that says community engagement where we're looking for people that like kind of chime in and give feedback.

And then there's an orange label, which is the standards question, which is more like.

Defining a particular convention, and there's gonna be a response, probably from the standards team, which is less oriented, maybe toward like just a general discussion.

So those are hopefully clear.

I'll try to make a brief announcement or like put in the guidelines essentially that summarized, but hopefully it's fairly self explanatory in that regard. So.

OK.



Williams, Andrew E 13:34

OK, take it away.



Houghtaling, Jared 13:34

Umm yeah.

Great.

Alright, so today.

We're gonna kind of continue on the path that we had started on the last few weeks, which is being able to actually implement all of this lovely validation work that clinicians have done, umm, with a focus on what you do for those things that don't necessarily match exactly.

So before I even dive into that, I want to just give a shout out to both.

Well, actually three sites, but specifically Seattle and UCSF, who have met with Nur last week and had very successful, uh kind of integrations are implementations of the of the ontology.

So these delta tables that we've talked about, both sites have been able to successfully

match more than 200 distinct flow sheet descriptions like in total about 500 between the two of them, which is fantastic and that makes it a pretty big chunk of the distinct flowsheets in general at those sites.

So that's already it's an encouraging, encouraging results and then a bit of a shadow to nationwide as well who we've started our kind of 1 on one feedback sessions where we go through these reports that have been developed and kind of assessing the OMOP data.

So for those of you who haven't heard about this happening or or haven't seen the emails we're trying to schedule these kind of 1 on one sessions with those of you who have delivered data to kind of walk through what the data looked like where the gaps might be.

Some of our recommendations for fixes and things like that.

So but be on the lookout for those being scheduled if you've already uploaded data to the to the portal.

Umm wanted to give another shout out quickly to Tony and will.

So Tony has been doing a ton of work actually on the data extract, upload.

Umm package.

I don't think Tony's on the call today.

Unfortunately, they don't see him.

Will is here however, so I'll put the link in the chat to that.

I don't know how many of you are following the development on this, or even know about it, but I think it is maybe more relevant for data acquisition context.

But here we're also looking at how we can deal with home updated with what that means in terms of getting it ready for submission.

So I'm gonna put that here just a second because I just wanted to give a shout out to Tony and will for their efforts on that.

So awesome job guys.

They don't know if it's been widely broadcasted enough yet.

It's like to that repo.

If you wanna check out what's happening, I think most of the work is actually being done on the packaging Branch if you're interested.

Umm, that's another thing.

And what other announcement again before I get into the technical bit of today is after the the discussion with UCSF today and specifically with Kasturi, it's clear that that we should look at making a an SOP for this integration process.

So all of these office hours that we've had, I think it's now three or four that talk about, OK, now we have these like flow sheets mapped, but what does that actually mean?

How do I actually get those things?

How do I use that to transform my data and so I started to put together a proposal for a SOP which is effectively integrating validated mappings into ETL processes.

We'll share that just a second.

What I would like is for this also to start out as a bit of of a almost a forum type thing.

So we have a general structure of how we expect you to do this, but what's become clear in already 2 calls is that depending on how you're OMOP database is set up and what your system looks like, you might have pretty different approaches to actually going through these processes.

And so I would like to be able to compile as many different examples as I can of this kind of walk through.

And so you've seen me do all of this in data bricks.

I've done it now in Postgres and I've also sorry in big query and I've done it in SQL Server and so we're able.

I think if we can start to compile the ways that this is being done, we should pretty quickly get to a full coverage of of what's represented across the consortium.

And I think it will be a valuable resource.

So all that to say, I will create another thread that I'll I'll link to you, umm and as either we have one on ones or as as you want to contribute to this, we'll start to like make make headway on what this will actually look like with the idea that here's the here's the workflow for doing so.

And then here are within each step examples depending on how you might have to handle CSV files or might have to work with your vocabulary tables.

So I'll show just quickly what that looks like as I share my screen here in a second.

Umm yeah, a lot of announcements I guess today, but hopefully that's I think that that's it from from my sides before the main main topic, umm and maybe I'll just pause now after throwing a lot at you.

Are there any pressing questions or topics that you would like to cover that we can already cover now before we inevitably run out of time at the end of the call?

I see it's a smaller group today, so maybe it's less daunting to to to share or or ask things if you would like to do so.

OK.

All right.

And now I've let that pause happen.

I thought of one other announcement, which is.

Coordinating an Azure data share for those sites who are on Azure.

That's also an SOP that needs to go out sooner rather than later.

Alex Ruiz from the Microsoft team I believe has documentation already on this.

He's assigned to that SOP, so be on the lookout for how to do that.

I think that's directly relevant right now to Duke or no.

Sorry, maybe to Colombia, I believe I don't know if Dan's on the call, but I think it's Columbia, yes, he, Dan.

So we will have more documentation for you there.

I think it's not super well, umm, well documented at the moment on how to actually practically share a set of the data share.

So more coming your way there.

Umm, OK, so now I'll dive in into the work here.

Let me share my screen.

Unless there are any questions.

OK.

So first, this is the this is the with the SOP kind of will be structured like we'll have like main main steps I was actually should be numbered.

Probably which more or less described the work that I've shown in office hours?

Uh, so this idea of you go and get the ontology or some odontology, but the CSV files from this from this nice repository.

Umm you create delta tables that are empty in your OMOP instance you load those CSV files into your OMOP instance, you create views or tables that combine those delta tables with your standard vocabularies.

Uh.

And then you actually implement the ETL logic.

Create kind of a a mapping source to concept map table ish.

Thing that you can then reference in all of your flowsheets logic downstream, and so I think especially for these first three pieces, maybe even this middle two, I think it's going to be useful to have different database flavors represented.

So I'll add the ones I know, but as you are all going out and doing this in your various database flavors and the jungle of of OMOP databases, umm, feel free to comment on this thread and add exec other examples or ways that you actually approached getting these delta tables in or if you've encountered any problems.

Umm.

So in any case, that's there.

That's that's the plan for now.

Umm.

And in terms of that process where we ended up last week was with a view on exact matches, UM, and we had this kind of ETL that I showed where he created exact matches with those exact matches, we were able to actually kind of create a OMOP representation of a flow sheet where we looked at not only the description as a standard concept ID, but also the value as a standard concept ID, which is great because then that's more or less a way that we can view that flow sheet as an OMOP entity.

But at least with the MIMIC data, a lot of our descriptions are not exact matches.

So like only 6% were distinct matches and the bulk or even the bulk.

But like another a bigger chunk, I should say 1/4 have.

I would like bag of words similarity where they have terms in common with some of the the terms that I've have validated mappings but not all of them.

And so.

How do you approach this right?

I think here there's then another part of the SOP that I will will create is like in this non exact case.

How and that's what we're gonna talk about today.

What are the strategies to try to make and say fuzzy linkages or non exact linkages to existing mappings?
Validate those linkages and then be able to use them in your exact matching paradigm.
So I think for people who are thinking about this, a lot of hopefully that's clear.
I wanna just pause there though.
Just does that make sense that we're now talking about the things that are not exact in the synonym table?
That we're now like flowsheet descriptions that you can't.
You can't join on, but there are maybe close to something else.
How do we handle those?
Is that is that kind of set the stage appropriately?
Eric is here.
Off off 8.



Rosenthal, Eric S.,MD 24:31

I joined late, so I'm sorry I can I can hide my face.



Houghtaling, Jared 24:32

Yeah.

OK, no worries.

No, no, no worries.

I didn't know.

OK.

So that's that's at least some context.

For what we're gonna show and what I showed a few weeks ago, is is this concept of a bag of words match where in this case we're splitting our.

Are descriptions into arrays, and we're looking at effectively the the distinct overlap of those arrays, so Hello Beta I demo goodbye B2 AI demo and our overlap is just this two term set for our second.

And the ordering in this case it's not.

It's not respected.

It can.

These terms can fall anywhere in the array, and in this case this would be a two term match and so that's what I'm doing.

At a vectorized level in this, in this particular SQL SQL Code here for the MIMIC data and I can I think just run this out of the box.

So that's double check and make sure that it's.

It's, umm.

And as you might expect, some of these look pretty good.

I mean, one of the things I can also do is before we even go down this road, let me.

Let me already create a a table here.

So subjects.

Alright, so we're going to just create a table with our exact matches here and we're going to exclude those from this list because we are, again, we're not talking, we don't care about the exact ones we already got, we have them. Uh.

Cool.

Let me just double check some stuff.

Cool.

And so these are our exact matches and need and there's we're actually gonna do.

We'll do select distinct.

And this is actually what we want excluded.

OK, so this is then what we're going to exclude in.

Perfect.

And just gonna be.

Alright, so now we're filtering on uh, we're preventing any exact matches from ending up in our bag of words.

Output so I can then take a look at what that looks like again.

We should see now all of these descriptions are.

Maybe not exact or they're they should.

All not be non exact, but some of them we actually don't want to to include, so nine is an awesome example of something I think that is a very relevant to chorus.

It's a non exact match, but we want to say for me I would look at this and say this seems like a legit match that we wanna make a synonym, right?

We say this is Glasgow Coma scale GCS eye opening because it's abbreviated.

That's not in here.

I don't know if any clinicians on the call have a idea what the CPN is, if that's gonna be relevant or not.

If that needs to be considered in this, but.

If not, let's say we want to.

We want to include this and So what would we actually do here?

So at a high level, I would take this table and I would send it probably or work together with somebody locally with clinical expertise to go through and evaluate these matches, and ideally we'll set up some sort of thresholds.

Right now I'm only looking at the overlap terms, but there's ways that you can calculate a match percentage based on the length of the terms that match and the length of the overall descriptions to kind of give some sort of similarity context.

And you can also do a Levenshtein distance and things like this.

There's like more lots of different algorithms to calculate ohm A particular like quantifiable value out of these matches, but for now crudely, we're looking at overlap terms and we've actually only limited our results to everything that has at least two terms in common.

Uh, and I would go through this list and effectively try to identify those ones that that are are sure thing matches that I think OK, I see something as just essentially abbreviated, but everything else looks good.

And I would then already put those into one bin.

And then I would have a second validation piece where there's maybe things that are close and I don't necessarily know the meaning of things and those that I would send off for additional clinical review locally, umm.

And then a third bit, which would probably like discard things that actually don't make sense.

And so then again, this is a bit of a manual exercise to do some of this.

In my case I have 200 or 375, so there's the nontrivial amount of things.

But through this process, the idea would be to develop a more automated threshold that you could start to sort these matches into appropriate bins and speed up kind of subsequent fuzzy fuzzy matching.

But once you have this, so let's maybe let me just choose a few of these.

So if I, I'll just.

Maybe I'll say that this is a good one.

Squares in there, right?

And I'll just grab a couple of these quickly.

This isn't.

Graphic.

And what you can see, I mean some of these are more, uh, can be more difficult than others to identify.

Uh, some of them are are close.

I don't know.

I think I don't think Paulina is on the call today to talk a bit too to that. Uh.

Reasonable.

I sort by label here.

I mean, you see a lot of these labels are actually duplicated for these are all different quadrants.

Of course, for the bowel sounds, I would probably say in this case we actually even have.

Yeah, there's there's two concepts there.

If it's not specified, I would imagine that it's going to be a general one, but maybe you even have a value that's associated with that.

OK.

So let's just start here with he is use three UM.

To that.

We're just gonna already filter that down.

You actually have detail on what I've.

It's OK.

And so effectively then and again, there's more ways than one to do this, but you would go through, identify the the sure matches and then those ones you can add to the synonym table.

So if we go back to this idea of our concept synonym delta, this is really where you're

trying to do your matching.

You'll see again, these look like your flow sheets.

This is in the logic of both the exact matches and the uh and the bag of words were actually looking at that synonym.

So in this case, everything's coming out of the concept synonym table and we're joining that onto the concept table.

And so we're here.

We're focused on the synonyms and what you can then do is to say, all right.

And maybe I should have done it this way in the 1st place, but if I go let's order here by label O, let's just say that we wanted to do the bowel.

Sounds all four quadrants.

We see that there's a standard mapping, so we're just going to already say this is my combination.

So bowel sounds in this case.

And then I have another one where.

Let's see people response.

Let's see, it's people responsive pupil reaction.

So I'll do this one again.

And in this case I see it also has a standard mapping, so we're going to grab that one.

And the third here, we're going to do people response rights.

And so then these are effectively going to make up the tuples that we need in our synonym table so I can show you that quickly.

And what else do we have in our synonyms?

Here we just have a concept of the the name and our language concept that he saw.

All of these will actually be in English language.

That's that's so these are then entries that could effectively be filled into the concept synonym table.

Students.

So now we've updated our synonym table.

So now if we were to run the exact matches again up here, what we would then see is those exact matches should now be captured.

In.

Sorry, sorry.

Deed.

So if we look let me Scroll down.

Also.

Just a little check on it.

OK.

Alright, I just need to double check on this, but let me see if we've added these.

We should be able to see.

Uh, yes.

OK.

Umm so sorry and I this is actually good that we're going through this.

So I what I ended up doing in this case was these are.

See that shouldn't be in there.

Those are in and then if we if we took this.

I'm just going.

Let's.

That's because we're matching now in the concept synonym name.

Project.

Not exactly seeing why this is not this joint is not capturing it, but.

You know what, it's because of this.

So if I let me, if I take the the substring out.

And there they are.

OK, so.

All that to say.

So what we what we have done or what I had done before and this is a point that also kind of came up in our previous call is it can be the case that in Epic you'll have or in any I guess source system you'll have a set of common descriptions uh for flow sheets and then you'll have a site specific set that might have a prefix something like UCSF or Tufts or anything like that necklace of MIMIC this are prefix is typically not in the umm is not in the.

And the description itself.

And So what I did crudely was just to remove the 1st 2 characters of the concept synonym to match only on the ladder.

Uh, assuming that my stuff had did not have R and most of the concepts synonyms do and that was able to get me some matches, but because we've added these directly where we know they need to match exactly that logic conflicts, so we would need to come up with in this case more sophisticated logic to ensure that that kind of trimming or sub stringing match plus the non substring match could coincide in the same.

In the same logic or in the same let's say mapping table creation? Umm.

But even just stepping back from that, which is, uh and all these specific example here, I think the whole point of this is to try to say I have this flow sheet description and I have this set of 1600 validated mappings that are high quality and core of specific.

I want to first do an exact match on those synonyms and I think what I've seen now from 2 sites already is that already can can make pretty big gains in being able to map flow sheets.

Another interesting thing from.

From the call with UCSF was that very few of their internally mapped flow sheets were represented in this exact match set.

So they went through this process, they matched something like 250 descriptions from the Delphi process and only like 10 of those were mapped elsewhere in their code.

So they had.

That's that's a lot of additional data that they're pulling in from like an hour of work,

effectively at least on the site side.

It's a lot of work from the clinicians to actually do the validation, of course, but with minimal effort you can get a pretty big gain as a site to implement this and what I am hopefully showing a little bit here is that for those things that don't don't match exactly what an approach might be to try to capture just a bit more.

Umm.

And so I think there's going to be an element of, let's take a look at bag of words.

We could do Levenshtein distance.

We could do other fuzzy string matching.

There's lots also in Python there's lots of packages to look at.

Fuzzy string matching?

Umm, you could open it up and look at also like the synonym table at large and try to match on that.

Umm, but in general, if you don't have a match within the bridge to AI set how you actually do that, and then the nice thing with the concept synonym table is you're not having to mint any new concept ID's.

All you're saying in your own local logic is that, hey, my flowsheet description matches with some other concept ID description.

So what?

What's not being shown is actually the true description for this concept ID, which I can show you here.

It's gonna be some like 3 tiered.

But description.

Let's take a look and see what.

Gastrointestinal, bowel sounds, general balance sounds.

So this is the overarching kind of multi tiered concept that got created as a standard.

It's an observation.

The relevant right now for this discussion, but.

This is then what you're linking this kind of broad description.

Let's say to this very specific bowel sounds flow sheet description in MIMIC in this case and in the concept synonym table all you need to have is just whatever concept ID was already provided by the the the those who validated and and you can then link to that.

So this is then what gets represented in that table and then this is your source description in the concept synonym name, and that's actually the only two columns you care about, because everything in this project will likely be in in English.

So that's a lot of information that you need to pause there and maybe see also which direction we wanna take this or if there are questions and and if I can elaborate on any of this stuff I just showed.



Williams, Andrew E 46:36

That was really great, Jared. As always.

I just want to sort of repeat some of the things that you said, high level descriptions of

what's, what's motivating, you're demonstrating this and what people can expect. So as a result of all the great validation work, we have a set of things, some of which are going to provide exact matches.

You just going to point your data in the right way and you're going to see an exact match that's going to be straightforward incrementally.

We're going to go from that set to as complete as possible in a way that's fully automated and is a variety of things that are falling into this fuzzy matching category.

Bag of words is one way to get the probability that things that are similar are the same. It's often used.

That's what Jared's been showing.

But the O that you'll eventually be following will kind of layout in steps what you do to get gradually more and more of the things that you actually see at your site mapped to this stuff that we've validated and to contribute things that can't be.

Uh confidently done and automated way to a centralized resource for us to process there.

So to minimize the amount of effort on your part and the minimize the amount of work, that's just a high level stuff.

So the expectation on your end should be in addition to kind of 1 on one sessions with us sort of a more formal less on how you how you deploy that kind of stuff that Jared has been showing as a group reply going to have to come to some sort of consensus about let's say we particular pick a particular bag of words approach to quantify the similarity of strings and say this is high enough threshold we think it's safe to assume in an automated way without any additional review of this is good enough umm maybe that.

Isn't feasible to do at all, but maybe there's a threshold where the probability from a version of a bag of words based algorithm says that's we're confident this is the right thing.

We can just do that at at whole batch.

Nobody else has to look at it. Sara.

We're gonna be trying to come to a consensus about that.

That's probably going to require some input from you.

Uh, or for others.

You know, there's a broader conversation where do we feel like that threshold is for the probability of that, and it'll get more formalized.

But this is a great walk through a really fantastic amount of yeah, expertise thrown at a challenging problem.

I think it's going to really lead to a great incrementally more and more of the data being mappable at each site in a way that we're confident with.

So just reemphasizing the stuff you've already said, Jared, really.

Thank you for doing that.



Houghtaling, Jared 49:27

Thanks.

Maybe just a question for those of you still in the call or any of you using the concept synonym table or have you used it in the past or is this kind of a new a new piece of of the vocabularies for you?

And I would be curious if any of you have used it, but.

We cannot, OK.

Uh.

Cool.

So I would say if you are, if you are looking to try to to map things, especially if you're trying to do it in some sort of algorithmic way or programmatic way, I should say, uh, it's a, it's a very useful table.

Like let's say you know, I don't know, you know, very small subset of concepts that your source descriptions should fit into.

It could be super valuable to expand that small subset using the synonym table, so you're matching on a larger space, so you have essentially multiple representations of the same same, uh, standard concept ID.

And if you're trying to create an algorithm that will link source descriptions with standards, umm, expanding descriptions and being able to look over a broader description space is a super valuable and will really improve performance.

So again, we're gonna getting kind of into the nitty gritty of not on the ETL thing, but like doing mappings in a I don't know, reproducible and programmatic way, but umm definitely worth worth knowing about this table and kind of the power of it.

And I think maybe even to to step back from this and get a bit of perspective on where this is headed, what the vision would be, is that sites will go through this exercise and we will kind of the that the vote have in the IDs that we've minted will be somewhat static.

I mean, they're there may be will be updates as we find, hey, this should be mapped this way and we review things but we'll end up with a set of kind of distinct terms that will be slowly changing or growing.

But then we'll have a more rapidly growing set of synonyms as sites begin to map their somewhat distinct descriptions into that kind of standard space.

And so it's and and having them kind of semantically, I guess, what is it semantically similar, similar but syntactically different?

Did ways of representing these concepts will essentially give us a better performance when we do these exact matching exercises?

Effectively, we're trying to increase our ability to just simply join on a synonym table and be able to find the descriptions that we see in our data sets and link those with things that have been that have been validated.

So again, this is now all far off in the distance, but this whole process of creating synonyms and tying those synonyms in with existing validated mappings is something that we will leverage and try to grow up or kind of build upon as we as the the the set of terms becomes more mature throughout the project so.

And I see both Satish and and Dan and actually and Javier also have all this is all new to use a synonym table.

So I'm happy that you at least had a few few weeks worth of introduction into that table.

Umm, I don't know if any of the three of you have questions specifically about what I've shown or how you might implement it.

Umm you can also discuss this offline if you would prefer, but I think definitely for Dan and Javier like we will have meetings in the not so distant future where we can go one on one and and implement this at your sites so.

Yeah.

All right, I'm gonna stop sharing.

Any other questions not related to what I showed today or or ontology?

Just general Standards questions that that we can field during this time.

Alright, well if not, I don't know Andrew Marty.

If either of you have anything else but give a few minutes back.



Williams, Andrew E 53:58

No, I think this is really exciting progress.

I think we're entering a new phase.

We're going to be getting more feedback and standardized form and more tools to help increase progress towards getting everything completely, getting it in the standard form we want and we'll try and make that as painless as possible through these tools and other kinds of things.



Houghtaling, Jared 54:19

Yeah.



Williams, Andrew E 54:20

And look forward to more of that.

And thanks again, Jared, if you want to hang for a second after folks leave, I want to just quickly chat about something, but that's it for me.

-Marty anything else we should be he has to for folks.

Go people are leaving.

See you all next week, everybody.



Houghtaling, Jared 54:39

See all.



Houghtaling, Jared stopped transcription