



This also, and I'm just gonna kind of add notes as I see them, but this also raises the potential data quality issue of having the overlapping visits. If you have multiple visits in your visit occurrence table and they're not kind of in series like a train but you have like any number that can participate this type of matching because then I can be not anchor which which row you should refer to. Like I had a second or a third visit here that would come fall in the same range, but maybe be a shorter period of it. Which one did I associate a visit?

In theory you should not have that at the visit occurrence level, or even at the visit detail level.

But yeah, sorry go for it Andrew.

Williams, Andrew E 21:13

Just for the sake of simplicity, Jared, I wonder if we are should maybe just state like a small set of assumptions. That we'll take for granted is obvious. So I think that might include just as you were saying that any workflow have taken care of, that the mapping is being done locally and that there's no need. Even necessarily be understood.

Kind of the crosswalks that might need to be figured for.

Things centrally.

I think as long as the mapping is done locally in the way that we want, we'll handle any crosswalks centrally.

So I think it might be nice to know about that in some theoretical way, but mostly we're not asking you to kind of really even be up to speed on that.

We want to just.

Assume that your data have been transformed to valid already locally that they're available in a file structure where there are no overlaps amongst.

Visit start and stop times so that you can get sort of a clean environment.

So I think those complexities all are things that have to be dealt with, but we're not really asking people here to deal with them.

And I think it's really just really the expectation. If we just sort of say assume you don't, we've got sort of a clean start and stop time for your visits.

It's already in valid format.

You don't have to worry about other things beyond what we've gonna cover here is that, is that all right?

Just to just to keep things a little bit clear.

Hughlaugh, Jared 22:50

That's awesome.

Top. Clear perfect.

Williams, Andrew E 22:51

Alright, great.

Hughlaugh, Jared 22:52

That's great.

Thank you so much.

OK, so instead assuming single visit, we're not, we don't have overlapping visits and so this query, what I'm I'm sending here returns a single visit. That's great.

And what I can do is confirm in my case the visit occurrence ID 54347 is the one that I've already used to organize.

This is this visit structure as everything checks out so far for that being great.

Next guess is we're going to try to try to define the visit detail that's associated. And so here I'm going to not only filter on the person that I know and the visit occurrence ID that I know right because visit occurrence ID is a required field and visit detail.

Every visit detail has to be associated with kind of an encounter, visit, occurrence or encounter encounter.

And then now I'm instead of looking just within the single day as I did here, I'm I'm now looking.

Specifically at the times that are.

That are specified in that header file.

And we'll do this.

I can submit this and I find again a single visit detail ID, so this 555352.

All right, I'm my visit detail ID that's returned and now we're getting to the point where we're gonna start to put this in a format that could be for the registry itself.

So.

I am gonna create a new table called waveform registry and currently this has the columns that are associated with the multimodal data linkage that that we've talked about with this.

This structure.

And we're adding to that.

What we talked about last week, the visit detail ID.

So we have now visit occurrence ID and visit detail ID. But as Andrew mentioned at the start of this call we have two additional fields as well.

So we're gonna have a field called linkage concept ID.

And we're going to relate to the manner.

Linked the visit file.

And then we're gonna have linkage.

Service. Yeah.

CH-3 facts.

OK.

And so this these last two columns are going to give us that insight.

Did you need?

Where there certain valid files in this case that you needed to match based on timestamp?

Or were did you have AET information or otherwise this linkage concept ID will be associated concept and we will have in a subsequent call we'll go through what the possible linkages are.

I think we're talked.

Maybe there's gonna be like for so discover.

Categories of ways that you could link this.

And we will define those in in two billion.

For you and source value like I said, it's a free text.

You can include any sort of comments you want concisely on how you have you link to particular file or set of files, let's say.

And that's gonna be Temporal. I think that that should be good.

Let's make some everything looks good.

Let's do this and we'll create that.

OK.

So now I have my table and now we're gonna insert into that table.

The values associated with with what we just the files that we're just looking at now insert into.

Public data waveform.

Waveform registry.

And then I'm not gonna specify the columns because I'm gonna go through 1 by 1.

OK.

So our first one is file ID.

Here I have Tony and kind of kind of and others who worked on the data upload update upload had talked about maybe doing a hash value of the file name or an actual hash of the file itself.

I'm not exactly sure where that landed.

Where what this needs to be in general is a unique string referencing that file.

If you make a hash, you could actually just use an MD5 sum of the file like I can go like this here. Maybe just for demonstration sake I will do that.

So I'll do sum 244.

On the header. We're gonna I'm gonna work just with the header for now.

OK.

So this is the kind of the MD5 sum of that file, and I'm going to use this as my file ID for now.

So let's copy that.

And we're going to say.

Value.

So we have our our MD5 sum.

OK. Next next one is procedure occurrence ID.

I think I might have missed one here. I'll sorry.

I'll need to quickly drop this.

And then I'm gonna add procedure concept ID real quickly.

Alright. And we're gonna recreate this table. Sorry about that.

Script file.

Perfect. So we've created that.

So we have our file ID which is our hash we have for procedure occurrence ID.

This is just the primary key where you will insert this into the procedure occurrence table.

The only restriction on this is that I don't want conflict with what else is in your procedure.

Current table.

So what you can do is something like.

If I'll do something simple. So we'll do select bounce.

From.

Session.

Right?

And so what this is going to do is it's going to take the count of that table and it's going to increment it by 1.

And so just because I'm just meeting a single row as a table.

When you can also do I'm gonna go to multiply the form as from a select statement for instance, you could do plus row number.

Over something like that and that would then give you this offset of your what's in the procedure occurrence table and then it would each row of your insert would have.

Automatically incremented number.

There's other ways to do this in different database management systems, but all that to say you just want have a unique procedure occurrence with you.

Or sorry. Yeah. procedure occurrence ID.

Quick note on that we had said here to do offset this by something like 1 billion.

It's also possible, you see in this example, you can offset it if you want, you can just have it directly adjacent to yours.

It doesn't.

I don't think we we have an explicit or need to have an explicit convention for that.

The only restriction is that I just cannot conflict with any other procedure occurrence ID in that table.

Full stop.

And I needs to be unique.

So that's the primary key.

OK, I'm running out of time.

I'm gonna try to wrap this up in the next 5 minutes procedure. Concept ID is where Paula comes in and this is the mapping of the type of.

Waveform file.

To this set of potential mappings that that we have defined in this document, right?

So we have all these different mappings that we have defined in this document, right?

And so I'm gonna use this 418-187 as my mapping.

And so I'm gonna use this 418-187 as my mapping.

There's ways to join content of your header file or other things like that to actually get these mappings.

Doing it manually for now, we're inserting a single row.

Person ID we have.

It's a match.

And I think actually I can't do so in this case since values.

Sorry for that.

And then we have our start date and stop date or start time and stop time of our.

OK.

That's our person ID.

Group ID we had said to test it.

In this case, we're going to just use the file name.

Without an extension as our group name, so it's this.

It's the general WAVE file the file name of the header without extension in my case will be the group ID and that's how all of these files are.

Are associated in my case, that might change for you.

You can assign other group IDs, but just some way to group the session or the set of files.

That are associated is what we're looking for here.

So I'll do that as a group ID.

Visit occurrence ID is.

Here which is 54347.

Visit detail ID make a comment 555352.

Waveform start date again, that's coming out of this guy.

So I can grab this.

I think that's going to be improperly formatted as well.

If I try it.

Stop time again.

This should be straightforward here and then.

What are the last ones here?

And we have source file and target file. In my case those are going to be the path to where I found this and I didn't change any sort of naming conventions.

So what I'll do is actually take the path and use it for both of those.

Both of those entities.

And then it's the header file that I want that I'm actually talking about right now.

You would do the same thing for the data file.

And you would think we talked about something.

So repeating all files in this case, so each file is going to have a procedure row as to just not get confusing about how we want to group things and that's going to be a decision for later down the line.

So that's our source file.

One I'll do the same thing with.

Target file.

And then we have the last two are the way in which we link things.

So I'm going to say 2.

Something like this, which is gonna be.

In my case.

Same time.

And then.

Something like that.

And that then should do it.

So that's there a single row.

That would then fit into your waveform registry.

And I'm going to then provide, like I said, this table here and going from that into procedure occurrence.

It's almost a one to one mapping, right?

I think there is a couple things that you would not fill in. Group ID, for instance, would not make it.

Or maybe it gets concatenated?

But that's the general idea and I don't know if this is going to run off the hat, maybe we'll see, but.

And it does.

Cool. And so I've select all.

You can now see that we have our first row in our waveform registry that has the contents from that diagram.

And this is the person that you would need to report for all of it, in the case where file is your file.

So with that, I do have to go and hand it over to Paula to take a little bit about this mapping step that I did very quickly where I took the mapping for EEC and I inserted it here.

But thank you all. Hopefully this was useful and.

Williams, Andrew E 30:27

Great. Thank you so much.

Hughlaugh, Jared 30:22

Top.

Williams, Andrew E 30:23

Yeah. And I think before Paula we. And so Jared was to go.

Go to where?

Where you got to go.

Hughlaugh, Jared 30:29

Thanks.

Williams, Andrew E 30:29

Well, I won't talk to you next week.

So I'm apologizing to everyone for creating this confusion.

The sign I have defined for a valid file in the waveform format.

Clearly sign it can be submitted in other valid or other formats with, I think, WFF 3 being the prominent or the named alternative example I guess.

We want to spend the rest of the mapping of the rest of the call.

With Paula talking about the concepts that are gonna be used in this you know the other the complement to it. Jared was just showing.

I do want to give them a chance to kind of clarify anything that we need to communicate now about what we know right is different.

I have to be done locally to perform this kind of linkage.

I think the SOP instruction says do the analogous thing essentially to what you would be doing here, but with your.

Your files.

In a different format.

Is there anything more to be said? I guess before Paula goes in there and again I apologize for having forgotten that this is now being done centrally and not locally, so anything else, Brian?

Gene, Brian 37:35

Well, that's OK. But I think for everybody, the SOP we're looking at, Andrew, in the previous one, there is a if you go to the SOP repository, there's a pull request for an updated version.

This is what we want people to follow and it does change things so that it's explicitly done not recommended calls.

We want to now format standardized centrally as possible so that we can.

Handle things like ages and data.

In a programmatic way across sites.





A recommendation

So I I guess I'm implicitly saying we're gonna do this. What you just described, but probably in a smaller group and then come back rather than asking people to kind of dig into these sources of alternatives and and so on.

So I guess, would you agree with that generally. OK, yeah.

Polina, Polina 10:50

Yeah, I do. I do agree.

And this is introduction, the place where everyone just can.

Can be aware that something is happening right now and just in case maybe some person would like to join.

What can we say?

We can invite this person.

But yeah, it's just introduction.

No solid decision yet.

Williams, Andrew E 10:50

Thanks. Thanks so much.

Any questions for Polina?

Polina, Polina 10:50

I see the hand raised.

Alajum, Tannam Wad A 10:42

Yeah, I think my question might be a bit more general and to relate to what Brian is saying. I think in our site, Colombia, we, our file our way from files are in PDF format.

So, for example, ECC and PDF and many more.

Heart rate and much more labels are all found under one PDF file.

Have created a workflow registry currently.

You need to fill the procedure concept ID empty and I'm unclear about how to fill it.

Should I?

Should I basically create duplicate so that one HOF file would have many different concept IDs pertaining to the labels that are in them?

Or should I just have one?

How and within that procedure concept ID fill it up with a telemetry.

Concept ID.

Polina, Polina 10:46

Good question.

Andrew, what do you think?

I think many.

Williams, Andrew E 10:50

I think as many in the general guiding principle is get as granular as you can.

That's just to get as granular as you can in the calculation.

But just to get as granular as you can in the calculation. One of the points of having a really well-organized standardized vocabulary like the ones that Polina creates is if you need to say I'm gonna grab all kinds of telemetry you can, as long as you have the more granular information available, because the Hierarchies and other.

Relationships are that.

Who can have a classification that says, you know, classifiers telemetry? It subsumes these other things, etcetera.

And to what we need for everybody to get the most granular information that you can, ensuring it's in a hierarchy to a, in a vocabulary that does that subsumption work. Then you can always roll it up.

But if you start off and that's all you get, then that's all you ever get, right?

And in the hierarchy, like it's get as granular.

As you can.

Does that make sense?

Alajum, Tannam Wad A 10:50

OK, I don't.

And then another question. I think when I was seeing that's filling out the wave form registry and the daytime columns, he ended up using both start time and end time and it was unclear if he was measuring the entire period of that file of that ECC ID.

For if he was just using.

Or adding the start and end time. I think I did it with the start.

Start time of the first waveform label.

And wanted to know if that needs to be changed.

So.

Williams, Andrew E 10:50

I think for the purposes of the linking only.

You just need the procedure that's associated with each of those most granular representations. In his case, it was different.

You know ECC leads.

And you could need separate concept IDs for each of those things. I think when we get in a later call to the tabular representation of things, we'll have things like start, stop times and offset times for each of those different leads. If like if they vary in terms.

Of their start and stop.

Times like the interpretation is we're going to have a tabular representation that has each of those distinctly. For each of those leaves, if, for whatever reason like one of them stopped sooner than the others, or whatever.

And you know where? You've not just got a ECC, but other kinds of information collected at the same time. You'd want to have all of that relevant information in a tabular form, extracted probably from.

HOF or you know with headers and population in that way.

I think that's but again for right now, I think all you need is the linking information that specifies a different.

You know the specific procedures for generating that waveform or other file for.

For right now.

Gow, Brian J 10:49

I guess I'll just add that I think it would be possible to do this procedure linkage centrally as part of the conversion if desired.

We ask for a mapping files that standardize.

What types of signal a site is submitting?

And we will.

We will know what what signal types are in each file as we convert them. Having said that, I don't think it doesn't hurt.

As well.

Williams, Andrew E 10:50

Whenever we can safe.

Gow, Brian J 10:50

It doesn't hurt to have them do it.

As well.

Williams, Andrew E 10:50

No, it doesn't.

It's a time people need to do other things. I think where we see know we have a reliable centralized strategy that takes that off of people's plates.

The only reason to do it then locally is you want to use it locally and not just.

Stop to off to the British in project and not be able to query you locally.

I think there's a throughout the whole project there's this.

Options sites have we want to ship this. On it right, do it in standard and we're good versus we're building out a centralized infrastructure for multimodal analytics that's leveraging this project to do that in a way that brings all of the value for querying and working with the.

Data there that needs to be done locally in order to get that at your institution, so it, you know, participating institutions in the project vary.

In terms of how they use this for they want to do it, everybody wants to do it well, but.

On they really kind of just want to do it and then move on to the next thing or in this part of an ongoing infrastructure they're building out for multimodal analytics that requires this being done locally in a particular way? And I think that's where.

We want to be making that available for sites who have that second option in mind, so I think it's great to be able to do things locally.

I think making it so that sites default to something that really reduces the overall value they get out of it.

In the end of the day, that's where we want to be careful. I think. Does that make sense?

Gow, Brian J 10:49

Yeah, Yeah. Makes sense. Yeah.

Williams, Andrew E 10:55

Right, we're a couple minutes over.

Thanks everybody.

Talk to you next week.

Polina, Polina 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50

OK.

Alajum, Tannam Wad A 10:50