Live Captioning by Ai-Media

JAMIE KOENIG:
OK, hello, everyone. Welcome to this webinar on the opportunities with the National COVID Cohort Collaborative. My name is Jamie Koenig, and I am a specialist at AUCD. Thank you for joining us today.

I just wanted to address a few logistical details before we begin.

So, we ask that you mute yourself during the meeting. If you have any questions, you can put them in the chat box. Or you are welcome to just shout them out. Leslie and Sharon said this could stay a pretty informal conversation. So if you have a question, feel free to shout it out. If not, put it in the chat and it will be got to buy the end.

We do have captioning available, you can click the CC button at the bottom to view subtitles. If you do speak, we do ask that you begin by saying your name to help the captioner and everyone else. If you're comfortable providing a short image description or providing your pronouns – which I now realize I did not do. Whatever you are comfortable is good.

This meeting is being recorded, and will be available a few days after the event, as well as the written transcript. Thank you for coming and please join me in welcoming Sharon Patrick and Lesley Cottrell. From the West Virginia clinical and translational

--
West Virginia Clinical and Translational Science Institute.

LESLEY COTTRELL:
Thank you, and welcome everyone. I will just add a bit to Jamie's fabulous introduction. Thank you, Jamie. Sharon is my fabulous partner here in Morgantown on the WV campus.

She is the coordinator that brought this opportunity to me to consider, and as my colleague for bringing to you in the AUCD network. We really appreciate this opportunity.

So, the overall purpose, the reason why we wanted to touch base with you is, we have had this opportunity to talk about and review the N3C database. We will talk about what that is, and how it might be used.

Or how it has been used in the past, how it relates to COVID. And many elements of our mission as in a UD -- AUCD network. So, we will really dive into the database, what it can do, also some of the variables.

We will not go into the complete list exhaustively, but we will have a variable list available to you, for you to take a look at. And show you where to find additional resources if you are interested.

And then really start to focus on the connections to our network, and review next steps. If you are interested in this, what we are ultimately looking at is a collaborative that is focused on using this database, to really examine COVID, and its effects on individuals with disabilities. And their families.

So with that being said, let's get into a bit more detail, and I will turn it over to Sharon.

SHARON PATRICK:
Hello, so thank you all for being here today. And allowing us to have this opportunity to speak with you about N3C. I'm sure you are all wondering what is N3C? That is the National COVID Cohort Collaborative. The N3C is a secure cloud-based data analysis resource, for COVID related research.

The data enclave is a secure platform through which the harmonized COVID related electronic health records data provided by the contributing members, is stored. The data itself can only be accessed through a secure cloud portal.

Let me advance my slides. So, as you can see, the map here. The N3C enclave is the largest limited electronic health record dataset in US history. There are billions of rows of data. And millions of patient records. Procedures, visits, medications, that go back two years prior to the COVID testing.

So that takes us back to January 1 of 2018. So, the requirement for the patient record to be in the enclave is that the COVID test has been done on the patient.

So then the enclave combines the data from the sites across the United States, and as you can see the dark purple are sites that have ingested data. Into the enclave.

So, I this slide, because it shows the trajectory of N3C. I'm sorry.

As you can see, N3C was created in a response to the pandemic, this kicked up in April 2020. I like to look at November, because that is when I started as a coordinator with N3C. In November, we can see that there are 73 data transfer agreements on file. And there are 103 Data Use Request. The data transfer agreements are the sites that have agreed to upload data.

And then this slide is just slightly updated, from May 2021. Just last month, but as you can see, we are on the track to 5.9 million patient records. And of those 5.9, there are 1.6 million COVID positive patients.

(Multiple speakers)

SHARON PATRICK:
Sorry, this has not happened to me before. Let me put myself on mute for just a second. Sorry.

OK, so sorry about that. So, let me start over by saying that we are well on our way to having many patient records for you all to use, for a data resource.

And I want to also say, when I say a site has contributed data, that is not necessarily one healthcare institution. For example, here at WVU, that includes a 12 or maybe 14 hospitals within our WV U system. So there is a lot of information available.

I do not know if Dr. Cottrell, you would like to take the next slide?

LESLEY COTTRELL:
When Sharon brought the opportunity to me I thought, how are they using this? What types of

question are they asking, or could they ask this large database. Just to get an idea, wrap my head around it. So, in general N3C literature, I provided just a few here. But these are general areas, several articles within.

They talked a lot about the establishment, NIH setting up this database, and what is involved. Like Sharon started to describe to us.

Then we went into mechanisms related to differences in COVID-19 symptom severity. We started to look at that. In general samples.

Early severity prediction, trying to look at what factors could we use to predict certain progressions with certain types of treatment? Health disparities in terms of gender, race, and some of those key elements. Our in the literature right now.

Phenotyping, the COVID-19, and then getting the lay of the land on treatments. There is not a deep dive. Beyond that, there is not deeper dives based on other questions.

And so, we wanted to provide that to you, so that when we start talking about domains, and types of research questions we could ask, it seems pretty wide open. Particularly for our network, and the types of questions we might have, to use this database.

As we all know, you know, the ACD network, has representatives that do research on the whole spectrum. The translational research spectrum. So, our research questions, other than purely basic science – but the translational piece, could be in this database. So, in terms of who we would want to invite to our collaborative for the domain on individuals with disabilities...

Their background and expertise could be very wide, as long as we are talking about translational. The more variety of our experiences, the better discussion.

I think it is back to you Sharon, for the structure of this all.

SHARON PATRICK:
Sorry, I could not find my unmute again. Alright, now we are going to talk about the N3C data, and some structure/formatting, to get where we are today.

So, not every site has the same electronic health records, and not every site uses a common data model,. So what the NCATS has done, is that institutions have a data transfer agreement, like I mentioned.

They submit that to the enclave, and then it is harmonized. This is the important part that I think is helpful to our researchers. Because then you can use the (unknown term) data dictionary, to find your variables that you are looking for to identify your cohort.

And to also identify the variables which you are looking out for your research. So, based on that there are three different levels of data.

The first one is the synthetic data which is computationally derived, that is level I. Then the de-identified data is level II, and that has 17 of the 18 HIPAA direct identifiers removed. That is level II. Then the limited data set has the 16 of the 18 HIPAA direct identifiers removed. Level 3 data will require institutional IRB approval to get access to that level of data.

I should also say, for the limited data set, it is important to know that this keep the true dates of service. Otherwise, the dates of service have been -- (unknown term) by approximately 180 days...

Level II has the dates of service shifted to protect patient privacy. Patient ZIP Codes for level II are truncated to just the first three digits. Then the synthetic data consists of data that is computationally derived from the limited data set, and it resembles patient information statistically, but are not really actual patient data.

This slide also talks about the registration, which we will talk a little more about in a few slides.

Then, how you have to have the Data Use Agreement on file, and then we will register for an enclave account, and then we will go to the Data Use Request.

So, also, I think it is important for folks new to the N3C to know that there are also external databases. And datasets that are available within the enclave.

An example of one that is used is public information. Such as the Royal urban continuing codes. So if you're looking for (Indiscernible) that our senses track based classification are available within the enclave. That number continues to grow as well. Also the link here, which I will have the slides available this will allow you to request an external database to be in the N3C enclave.

Next we will talk about the domains and the resources. So, with N3C, there are a lot of resources available. We want you to be able to use those for your research. So the N3C aims to improve the efficiency of an accessibility using a very large patient level COVID-XIX clinical data sate -- said,

So, the whole purpose of this is the collaboration. As we said, in your network, your collaboration then we will bring it to the N3C, so you have the data available. And then we can all work together on projects. This is a valued organization, because access to a large-scale COVID-19 data across the nation. This is good for data, for many things. Access to Domain Teams, which we will talk about more. Statistics, machine learning approaches, informatics expertise, and training on machine landing analytics, tools software, and additional datasets.

There are several ways to become engaged with N3C. One item noted on this slide is to join a Domain Team. A Domain Team brings everyone together, that has a similar interest.

And they can work together on a project, and then they also get help from N3C with use of a data analyst as well. Which is suitable -- super helpful, a data liaison person.

Also to get access to the N3C data enclave. This N3C landing pages very helpful. It gives you quick links to the Domain Teams, frequently asked questions, and other training. As you can see, other items listed as well. Very helpful.

So, the N3C Domain Teams enable researchers with shared interests, as I said, to analyse data within the N3C data enclave and collaborate more efficiently in team science environment. These teams provide an opportunity to collect pilot data for grant submissions, train algorithms on larger datasets, informed clinical trial design, learn how to use tools for large-scale COVID-19 data, and validate results.

And so, also part of getting an N3C enclave account, is also signing up for the slack channels.

That is another means of communication, also for the Google groups.

There are some examples of the Domain Teams, are the cardiovascular disease, there is perioperative team, acute kidney injury, really, anything tailored to your clinical interests. It is out there. If you find something that is not, she identified there was – there had not been any Domain Team started for disabilities, or lifespan disabilities as she has titled it, she has formed that team.

So we have a meeting, I do not know, Dr. Cottrell if we want to talk about that at the end. I do not know, Dr. Cottrell if you wanted to talk more about this initiative?

LESLEY COTTRELL:
OK. So, in terms of the lifespan disability, and what I am going to do is in the... In the chat is the list, or the link to the different domains that exist. As well as ours.

So when we submitted the request for a new domain, we needed to list a mission and a purpose. By design, it was generally worded so that it could represent the wide variety of contributions and characteristics of our usage, within the network, as well as other partners.

Within the network. So, I will not read it verbatim, but, basically we are – we want to focus on all aspects that we can possibly get from this database. To understand the effect of COVID-19 on individuals. So, it might be with regards to treatment access, continuation of treatment, severity of symptoms, progression over time.

In terms of the COVID symptoms themselves. But also, there might be other variables within and I will attach the variable list as well in a few minutes. On subsequent effects. Emotional, developmental, structural, cultural effects that might be in there. There is not a lot of those, because these are medical records by nature.

But there are some captured in there. So, we could start to take a look at that. Overall, the mission is to characterize that and document any disparities and patterns that we see, within this larger database.

And then the purpose is very similar. The wide variety of disability types, we are not confined to a particular definition of disability. A lifespan approach, so pediatric, youth-adult transition, as well as adults, and geriatrics. So again, very generally worded, and we can go in a number of different directions.

The link, also on the slide is two hours specific domain. And so, we also had to list an initial research question. When you register, if you are interested in using the larger database, and then the domain specifically, in terms of a collaborative, you are asked, what particular research questions are you interested in?

They may not be like this, this is what I included, and I was developing the main, so it was very general. You might have a specific one. These get the light -- collected. As we have those Domain Team meetings, and having the assistance, both locally if you have that available in terms of biostatistics, epi, things of that sort, those skill sets, you can do that locally.

Then bring it to the group, as particular analysis that support your research question, but as Sharon said, we also have support from the N3C, and NIH that as a whole, we can start really producing some of these analyses, and answer multiple questions. At the same time.

Next slide, please. OK, skip that one. (Laughs)

Hopefully we can send you these, we were hoping is that other domains have been going, for various lengths of time. And give you an idea of what a domain meeting looks like. So, we will get that to you. We also might have some individuals join us on the call today, to share some of their experiences. Depending on the questions.

SHARON PATRICK:
I think it is important to note as well, that collaboration is very neat. Everyone comes with an idea, and you know, you do not really know what you're getting into, and then you join the team.

There is a bunch of support, and the help. So, we will talk more about now, then the N3C registration. And training tutorials that are available. There are so many resources out there. I really recommend you use them. They have live office hours on Tuesdays and Thursdays at 1:00 PM Eastern time.

And have life-support with a helpdesk representative. It is very helpful, either to have your concerns addressed, or even just to listen to others as well.

There is also, as you get more comfortable within using the N3C enclave, there is an enclave users group that meets on Fridays at 1:00 PM.

So, there is a lot of information, I encourage you once you create your account, and you feel comfortable getting in there, and navigating around, to join these different tutorials that they have, and different groups.

Another resource that is available that I also highly recommend is the N3C enclave orientation sessions. Session A is more for those that are interested in learning about N3C, as well to how to engage with project teams, and access the data. They do recommend that you attend session A prior to session B. However, it is not a requirement.

Session B is set up for analysts, statisticians, scientists, anyone who is looking at using the tools within the data, within the enclave.

Now we will get to the good part. Which is what you will need to create your enclave account.

So, the first time user prerequisites are your institution needs a signed Data Use Agreement on file. And in the slides, there is a link that shows you the Data Use Agreement's that are on file. So you can check that, and confirm your institution has a DUA on file.

You will also need to have human subjects training, like city training, also security training from the NIH, that is actually a prompt while you are creating your account. Lastly of course is to register at N3C.

The Data Use Agreement is required, and it outlines the terms of the use for the enclave portal, as well as the governed policies around the data within it. The Data Use Agreement can access an umbrella. So anyone associated with your institution can create an account, if your institution has a DUA on file.

I also recommend, please sign up with your institutional email. If I sign up with my Google

account, they do not know which institution I am working with. So we need that checkpoint, to see that your institution has a DUA, along with your email.

So, that is a really important step. And then, the goal of the Data Use Agreement is privacy protection to promote broad access. So, this is related to COVID research only. There will be no re-identification of individuals or data sources. You cannot download or capture the raw data. And it is an open platform for all researchers.

Also the security aspects are, that the activities within the N3C data enclave are recorded and can be audited.

Now we will create our N3C account. So, as I mentioned earlier, when you sign up for your account, you not only are requesting access to the enclave, which houses the data, but also to the N3C team drive, which include the Google groups, and slack groups as well.

So we are going to go to the website, we will click here, and we are going to go to the 'access the enclave' button.

Then we are a first-time user, so we will go here, for first-time users. And then, I said, you have to have the DUA on file, you have to have NIH security training, you have to have human subjects training, and then you can proceed with your request.

So, there are two different routes to create an account. Some folks have the income and account here, we have the federated version, so we are able to login through the InCommon method. The second method is the login.gov.

They really do a great job at having step-by-step directions. Without you create your account, with the InCommon account, there is also a link that will allow you to know if your institution has an InCommon account, or if you need to go through the login.gov.

I will tell you, you need to know that it takes about five days to create your account. This is not like a Google email, where bang, you have access. This is vested with humans, so it does take a little bit longer. And they will cross reference that your institution has a DUI on file, to protect our data.

So, as Dr. Cottrell talked about the ideas that you're going to bring to the N3C Enclave... So you need to start with the Data Use Request. The Data Use Request has to be approved by the data access committee.

And after the Data Use Request is approved, that is when you will have a workspace provision, and that is when you can start to navigate around in the data. Until then, you might be able to get into the enclave and see some things, but you will not really get true access to the data, and the concept sets, and the data store, until you have your Data Use Request on file.

That takes approximately 21 days. So, I have my idea, I submit it, and it takes a bit of time to be reviewed.

They are not like the data police or the science police, they are really checking to make sure if anyone has done this type of research. And if they have, maybe suggest you join that project. Or also, just to confirm that it is COVID-19 related.

So, here is the inside the enclave, what you will see when you are submitting your Data Use Request. So, at this point is where you will have your collaborator information. So, if I am coming and I know I have a team I am working with, I will need everyone's email address, everyone's name.

And there are pieces of this that are private. That only the data access committee can review. That is the project rationale. The other items are public, and those will be shown on the public dashboard, which we will get to in the next slide.

At this point is also where you can make the decision if you're going to allow others to join your project. You do not need to allow others, it is just that a lot of people do because of the nature of the team science approach. But this is the step in which you will enter, if you do not allow researchers to join the project, or to allow them.

Of course, it is going to ask your title, and the abstract information. Then here is the public view. So everything you just typed in on the Data Use Request, within the enclave, and it has been approved through the data access committee. It will then be available on the public version of that.

As you can see, I see a short abstract, I see who the lead investigator is, and I see a title. Also the search button is important. This is where you can search by a provider's name, or a subject, to find other projects to collaborate on.

So, it is a little small, but this shows us within the enclave, where you can request to join a project. So I have searched, and I find something that looks exciting, that is an interest of mine, and I can request to join the project.

Once I request to join the project, the request goes to the PI, and the PI will approve you, and then it will go through the data access committee again, just to review.

It does take a little bit of time, approximately two weeks for the approval process. Again, once you have that approval, that is when you have your workspace provision. Again, I do not think I can stress it enough, that you do not get access to any of the data until you have the Data Use Request on file, and that has been confirmed.

And then, that is when your workspace is provisioned. As I mentioned earlier, there is a lot of assistance, tutorials, videos etc. that are available. So this is the landing page for the support desk.

It is very helpful, I mentioned about the live office hours as well. All very valuable information. And they are pretty prompt, and your support ticket questions. If you have questions, you can put in a ticket. They are very quick to respond to those.

Lastly, I just want to give some attribution to the N3C administration and leadership team here. That's our – that are beneficial to the N3C, keeping it moving, and keeping all the great information safe. And keeping publications coming, and just are very helpful to the slides today.

And that is really, all I have for today. I thank you all, I am so sorry, my son... (Laughs) Had to come and visit me several time. I do apologize for that. If you have any questions, I will provide the slides, to Jamie, and then also if you need my email, I am happy to help with anything N3C.

Whether you are creating an enclave account, if you need me to look anything up, I am happy to help and enjoy the opportunity to speak with you all today.

LESLEY COTTRELL:
Thank you for giving those details, Sharon. When she said two weeks, and gave certain timelines, I can only say, in my situation it was pretty fast. But even with that, Sharon, you had mentioned...

I was going to say, as you are going through this process, please join us for the first Domain Team. And as you continue, if there is a second Domain Team that comes up – I think the biggest, or most important issue here, initially, is that we can come together and see what would those research questions B.

After we have had a chance to look at the variable list, and go through them as a group. But what would those research questions B? And really start to vet it. Locally, for you, but also larger group, and get that going.

I know I had a chance to sit in on the rural health. In the first length, by the way, where we sent the domain names, they have a schedule at the bottom. And you can join those. The rural health one, it was so smooth.

They have been meeting, I do not know, how long Sharon?

SHARON PATRICK:
Since early February, I think.

LESLEY COTTRELL:
So even from that period of time. They had already asked a question, they had some analysis back, they were asking additional questions. The analyst was – I mean, it was just perfect. As a researcher, just this brainstorming session, I was just loving it.

So, getting those discussions started I think is really important. For the moment. I am here, and Sharon is fantastic at troubleshooting anything on the document side. The registration side. All of that piece. But it was pretty smooth.

SHARON PATRICK:
I also wanted to let everyone know, we are going to kick off the lifespan disabilities meeting, on June 30 at 2:30 PM Eastern time. Very happy to have all of you there.

We are going to work with analysts from our sides, and Dr. Cottrell has her Data Use Request out there. What we have done in other teams, and how we got so smooth with the rural health team, is we have one common project, and really just broke that up into many different small groups.

We actually had two submissions into a journal already. We are really picking up, help also with the cardiovascular team. We are really moving along there. The N3C analyst, and data liaison, and logic liaisons have been very valuable to our teams.

So you do get that support from N3C, and it is great to have local analysts, and statisticians as well. Because then you really get things really hashed out. There is a lot of information, it can be slightly overwhelming. But I really think that it is... It is this huge, incredible database that is out

there, to help answer your research questions.

LESLEY COTTRELL:
And so, invite – if you're interested, and we will open it up for questions. But if you're interested, and you think others are interested but they were not able to join us on this call, invite. Feel free to share.

It is open to anyone who would have a shared mission and focus. So, what questions do people have initially? To get familiar with it? Or, previous experience with N3C, maybe?

It is at the end of the day... People are tired. OK. Great, thank you, Sharon.

SHARON PATRICK:
I do not know if anyone has ever worked with a database this large. You know, there are 7 million, I think it is 7 million, or billion, I would have to look it up... Rows of data.

Right now I think there is 2 million positive patients. I should also say, there is also a matching, or a 2:1 matching as I indicated. For your information to be distributed to the enclave. You would have to have a COVID test that is negative, or positive.

So, for every two negatives there is one record, basically. So there is that, you know, more of a control as well. If you are interested in that. There is that information as well.

JAMIE KOENIG:
Well, thank you so much for taking the time today to talk to us about this opportunity. As they said, they are free to get questions emailed anytime. When I send out slides, and a recording, I will make sure to include their contact information as well.

I am putting a survey link in the chat. If you could let us know how this webinar went, and if there are other topics you are interested in. That would be great. Otherwise, have a great rest of your day.

LESLEY COTTRELL:
Thank you Jamie, thank you everyone.

SHARON PATRICK:
Thank you!

SPEAKER:
Thank you all, this was a lot, but it was great to hear from you all.

Live Captioning by Ai-Media