Transcript

>>Welcome everybody to Standardize Your Research Data with the NIH Common Data Elements Repository. The purpose of today's training is to familiarize you with common data elements, or CDEs, and the NIH Common Data Element Repository or the CDE-R. I'm Catherine Staley from the National Library of Medicine's Office of Engagement and Training. Here with me today are my colleagues Kate Majewski, Mike Davidson, and Michael Tahmasian also from the Office of Engagement and Training. We're joined today by Robin Taylor from the NLM's division of Library Operations, who will be helping me answer your questions. Robin has been with the NLM for over 4 years as a Technical Information Specialist and is the product lead for the CDE-R.

During today's training, you will have the opportunity to ask questions and practice searching in the repository. So let's review how we'll interact. And Zoom captions are available for this class. Click the Show Captions button at the bottom of Zoom to see those. We do have everyone muted to cut down on background noise. If you have a question or a comment, please use the chat box and send it to everyone. I will not be able to answer messages that are sent directly to me, but Kate will be monitoring your chat questions throughout, and I'll pause frequently to see if there's anything that I or Robin need to address. To keep us on track, we're providing a handout that you can download and follow along with, and Kate has put a link to that in the handout or in the chat now. Thank you Kate. If you're working on a small screen and it's not feasible for you to work on a handout during the presentation, feel free to just listen along and we'll share a link to the answer key at the end of class, which you can use to review. During class I will also be occasionally asking for your non-verbal feedback using the Reactions feature. The Reactions feature is accessible from the Reactions button at the bottom of your screen, so please find that button now and give me a thumbs up to indicate that you found it and I already see some thumbs, thank you. Some devices may not allow for that type of interaction, so if you can't find the reactions button on your device, feel free to use chat to respond when I ask you to. And thank you very much, getting lots of thumbs, which is fun to see. All right, I'm going to turn off my camera and we will get started.

So today we'll explore a variety of ways that CDEs and the repository can support you in your work. But to get us started, I want to put them in the context of the NIH's mission. Both the NLM and NIH have identified common data elements as a tool for facilitating the interoperability of data. Ensuring that biomedical research data is interoperable is a step towards making that data FAIR. The FAIR principles recommend that data be findable, accessible, interoperable, and reusable. The repository is one way that the NIH supports researchers and aligning with the fair principles. Additionally, in the NLM Strategic Plan for 2017 to 2027, CDEs support the objective to connect the resources of a digital research enterprise. In the NIH Strategic Plan for Data Science, the use of CDEs is encouraged to improve accuracy, consistency and interoperability among data sets within various health and disease research. Finally, you're probably aware of the new NIH data management and sharing policy that went into effect in January 2023. Common data elements can be part of a compliant plan. If you

would like to learn more about that plan, there is a link on the handout to it. If you know how to locate CDEs in the CDE-R, you'll be better equipped to support these goals.

All of that being said, today's training is going to specifically focus on how to navigate the repository and find CDEs that will assist you with your data collection process, and this is the agenda for our session. First, we'll review what you need to know before using the repository, like what a CDE is and special repository features that will aid in your searching. Then I'll demonstrate the CDE-R live for you and you'll have the opportunity to practice on your own. And finally, we'll review how to save CDEs for later and how to get help if you're stuck.

But before all of that, I would like to get a sense of who you are and your work. So please take a moment to tell us in the chat your title or your role in the name of your organization. Oh, wonderful. See lots of responses in the chat. So many. I can't keep up with them all, but I can tell we're all coming from different places and have a variety of roles, which is great to see. Great. Thank you very much. Feel free to continue to put that in chat.

But next I would like to know your comfort level with using CDEs and we'll use a poll for this question. So Michael has launched our poll. Now please select the answer that best represents your comfort level with CDEs. I already see the answers coming in. All right. Thank you. Michael, let's go ahead and close that poll and show the results. Thank you very much. OK, let's review these responses. So it looks like most of us know what CDEs are, but we've never used them. And some of us don't know what CDEs are, which is perfectly fine. We're going to go over that together. And then we have some participants who are pretty comfortable with using CDEs. So no matter what your comfort level is, I think you're going to walk away from today's class knowing a lot more about the repository and CDEs than you did before. All right, so I will close out the poll. Wonderful.

OK, so thank you for sharing those responses. We can actually use the data that you just provided to explore some reasons that you might consider using CDEs in your projects. Let's say I'm writing a report after this training to summarize how it went, and I want to include some data about the participants. I could start by entering each of your answers in a spreadsheet. Then I might have to standardize how your organization's name is represented. On this slide I've used the acronym like NLM to standardize the results, but I could also write out the name as in National Library of Medicine, or maybe both with the acronym and parentheses beside the name. This is just a small set of data, so it wouldn't take me too long to standardize. But what if I wanted to compare the responses from all NLM trainings, not just this one? With more data, it could take longer to standardize, and I would have to ask other trainers to standardize their data too, so that we could combine our results.

Plus, what if another trainer asked about their audience's comfort level with CDE, but they asked the question in a different way? For example, maybe they asked the audience to describe their level of comfort in their own words. It would be really hard to compare those free text responses to the responses you just provided. We would have two very different sets of data,

like the ones on this slide. If my colleagues and I came together and created a survey that asks all of our participants to answer questions using the same format, we'd be able to combine and compare our data quickly and efficiently. We'd have results sooner and could adapt our trainings a lot faster based on who participates. You probably want to compare your research results to those of your colleagues who study similar topics. CDEs can help you do that.

As we just saw, some of you are very familiar with common data elements and the repository, and for others it's maybe a brand new topic. So let's review a definition of CDE so that we're all on the same page, and this definition is the answer to #1 on your handout. Common data elements, or CDEs, are standardized, precisely defined questions paired with a set of specific allowable responses used systematically across different sites, studies, or clinical trials to ensure consistent data collection. In other words, it's a standardized way of collecting a particular piece of information from a respondent. No matter who is asking the question, it's asked in the same way, and respondents have the exact same answers to choose from. This applies to when respondents are filling out their own answers, or if the researcher is recording the answers on their behalf.

Let's look at an example of a CDE together. On this slide is a CDE from the repository. It is titled Usual Place of Health Care Type and it asks, "Is there a place that you usually go to when you are sick or need advice about your health. Select all that apply. That is the standardized, precisely defined question. Respondents or researchers can answer with one of these specific allowable options: a doctor's office or community health center; walk-in clinic, urgent care center or retail clinic, and a pharmacy or grocery store; emergency room; VA Medical Center or VA outpatient clinic; Some other place; There is NO place; Don't know; or, Refused/Not reported; or, Unknown.

Here is another example of a CDE. This one is titled Date of Chest Examination. It asks, "What was the date of the imaging examination?" The specific allowable response for this one is the date, and you'll see other CDEs in the repository that allow responses like the time, a number, or a date.

After viewing those two examples, you might be thinking about how CDEs can fit into your workflow. First, CDEs can save resources, including your time and your labor. Rather than writing your own instrument or data collection tool, you can find one in the repository. Using CDEs also advances science because it allows researchers to compare their data and results across studies. Imagine if you could seamlessly compare your results with researchers across other institutions without having to coordinate your efforts because you're both already using CDEs. And finally, you might be required to use CDEs to comply with organizational policies, or you may assist researchers at your institution with writing a data management and sharing plan for a grant. Many grants awarded by the NIH now require or strongly encourage the use of CDEs so the repository can help you meet that requirement.

Take the case of the NHLBI initiative, called the Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies, or CONNECTS, which illustrates how CDEs can be used for all of these reasons. The goal of CONNECTS is to mobilize and coordinate the US clinical trial infrastructure in order to find the most effective therapies for treating COVID-19. It consists of over 40 networks and cohorts with over 1000 participant sites across 6 studies. To coordinate and connect these studies, the NHLBI created a set of common data elements that trials funded through this program will implement. With these CDEs already available, this gets the research off the ground faster.

A recent CONNECTS study called the Novel Experimental COVID Therapies Affecting Host Response, or NECTAR, has multiple sites across the country, as illustrated by this map. Because they're using CDEs, researchers at these sites can combine their data easily, regardless of where it was collected from.

The focus of our training today, the common data elements repository, was developed to help you find CDEs. It's a free, collaborative platform designed to provide researchers with CDEs that have been recommended or required by NIH institutes and centers or other organizations for use in research and for other purposes. The NIH CDE Governance Committee reviews and assesses CDE submitted by ICs, and they work closely with the NIH CDE Repository Team at the NLM. Currently, the repository has over 23,000 CD's from across 18 organizations. In the next part of this training, I'm going to highlight two key features of the repository that can assist you with your search NIH-endorsed CDEs and forms. This will prepare you to actually search on the repository in a few minutes. The repository includes NIH-endorsed and non NIH-endorsed CDE. NIH endorsed CDEs have been reviewed and approved by an expert panel and meet established criteria, and this is the answer to #2 on your handout. NIH-endorsed is the designation in the repository. That means CDEs were reviewed and approved by an expert panel and meet established criteria. This feature is intended to help you identify CDEs that are ready to use. NIH recognized bodies like institutes or research initiatives may submit CDEs to the NIH CDE Governance Committee for consideration for endorsement. When the Governance Committee deems the CDEs as endorsed, the NLM will publish the CDEs in the repository and designate them as endorsed. On the slide is the criteria for endorsed CDEs. These criteria were developed by the NIH Scientific Data Council. To be endorsed, a CDE must adhere to these criteria: Clear definition of variable and measure with prompt and response. Documented evidence of reliability and validity. Human- and machine-readable format preferred. Recommended or designated by a recognized NIH body and licensing and IP status clear. The expectation is that the designation of CDEs as being endorsed by the NIH and making those CDEs discoverable and accessible through the repository will advance the use of CDEs and NIH supported and conducted research. For you, this means that you can more quickly identify the CDEs that have met this set of criteria and are ready to use.

If a CDE is endorsed, you'll see this gold ribbon icon in the repository. You can search directly in the NIH-endorsed CDE collection or limit your search results to endorse CDEs, which we'll see in

a few minutes. Currently there are two collections of NIH-endorsed CDEs. One set is from the Project 5 collection, which were developed for a wide spectrum of COVID-19 research, including translational, clinical and applied research. The other set is the NHLBI connects organ support CDEs.

The second feature I want to highlight is forms. Sometimes CDEs are combined to create forms in the repository. Forms are a group of questions or variables with a specified set of allowable responses that are used as a set for particular research or clinical reasons. And this is the answer to #3 on your handout, a group of questions or variables with specified sets of allowable responses that are used as a set for particular research or clinical reasons are called forms. Forms can collect a lot of data about a participant all at once, like a family history. Or they might assess something like pain or the seriousness of an adverse effect from a medication. Forms will often include a system for scoring participants' responses. Once they've completed filling out or responding to a form's questions, the administrator will tally up their score. Different scores will lead to different conclusions or interventions.

Some repository forms recreate validated instruments that were created outside of the repository. For example, this Audit-C questionnaire shown here is intended to identify alcohol use disorders and was developed by the World Health Organization in 1998. It includes 3 CDEs.

Other forms were created in the repository. This includes a form like the BRICS Social Determinants of Health from the NINR. This form is a collection of 63 CDEs related to social determinants of health. It actually includes the Audit-C questionnaire that we just viewed, plus CDEs about access to care, food security, depression, housing and more. The key thing to remember about forms is that they are composed of individual CDEs, and those CDEs come together to measure or assess something.

All right. So now you're ready to search the repository. But before we do that, I want to review what we've covered so far. Michael is going to launch a poll with a few questions for you to answer that will check your understanding of CDEs so far. And these are also Questions #4 through #6 on your handout, if you'd like to follow along there. So I'll give you a moment to answer those questions and then we will review the answers together.

OK. It looks like most of us have responded to the poll, so I will ask Michael to share those results and we can review the answers together. Great. Thank you. All right, so the answer to the first question, which is #4 on your handout, "What can you find in the CDE-R?" is CDE and forms. We'll look at both today. The answer to the second question and #5 on the handout, "Which CDEs were reviewed and approved by the NIH?" is endorsed. And finally, the answer for the last question, and #6 on your handout, to identify an endorsed CDE, look for the gold ribbon and we will see that a lot today, so you'll definitely remember it by the end of class. Alright, wonderful. Thank you. OK. OK. So Kate, have any questions come in that Robin and I can answer before we dive into the repository?

>>Yes, we have a number of questions in chat. We're going to start with a question from Maria and she asks, "Are there CDEs for preclinical studies?"

>>I can take that one, Kate. Thank you.

>>This is Robin and Kate, I'm sorry, I forgot the name of the asker already, but yes, the answer is yes. There's not a lot in the repository, but there are some in the CDE repository. Now there are CDEs for preclinical traumatic brain injury, TBI from NINDS. And I have also heard of groups at NIH that are working to develop and they're more preclinical CDEs. So this is a question that I hear from time to time. So I know the interest is out there and we have a little bit now and I think that will be growing.

>>Great. Thanks, Robin.

>>Next we had a few different questions about specific CD's that were used as examples. So for example, Catherine had shown a CDE on the usual place of health care type early on and there was some questions about the specific elements and how they were defined. And I think there was another question about date fields. I think if you wouldn't mind, I'd like to defer those questions until after we actually get into this CDE-R, so you can look at specific examples and see the kinds of information that are on the records. So if we don't answer your question by the end of class, please repeat your question in chat so that we can be sure to address the questions that you have. So moving to another question Steven asks, "Are endorsed CDEs redundant with other CDEs?" Robin, would you like to take that one?

>>Sure, I'll take that. I think Steven was asking in response to the criteria for endorsed CDEs that were shared, and so those criteria were sort of headed down from the NIH Scientific Data Committee and, you know, not being redundant is actually technically not one of the criteria, but of course it's a goal of CDEs because to make them more common and shared across studies, then we need to reduce redundancy. So that is-- while it's not one of the specific criteria is absolutely something we strive for.

>>Thanks very much, Robin. So I think we'll get back to the content with Catherine and we'll address some more questions in just a bit. Thank you.

>>Great. Thank you, Kate and Robin. OK, so we are ready to search the repository. If you'd like, you can follow along with my demonstration on your own device, and Kate is going to put a link in the chat to the repository. I will try to make the screen bigger when I can, but you can also adjust the size of the demonstration on your own screen by going to the view options at the top and adjusting the Zoom ratio. So once you have the repository open and you're ready to go, give me a thumbs up and I'll know that we can get started. OK, I'm seeing lots of thumbs so I think we are ready.

All right, anyone can search the CDE-R and see results for CDEs and forms, but to access all of the repositories features you'll need a unified medical language system, terminology service account or UTS account. You will see the Sign In button and the repositories upper right corner

to either sign in to an existing account or sign up for one. I'm already signed in, so my profile name appears there. If you don't have an account, I encourage you to wait until after this training to sign up. You can still follow along on your own device and see almost everything. And as we go, I'll point out differences between what the repository looks like when you are or aren't signed in.

Next I'm going to point out some features on the repository's home page. The repository's default search is for NIH-endorsed CDEs. You can switch to searching for all CDEs or forms by clicking on the tabs above the search box. You can search with a topic, keyword, or organization name. You can also select CDEs or forms from the menu across the top of the page to search or browse within either type and we'll look at how to browse later. Let's start with searching within the NIH-endorsed CDEs group using keywords. Imagine that you're designing a research study and you want to collect data about the medications a patient being treated for COVID-19 receives. We can search within the endorsed group with the keyword medications to see if any CDE's are available on that topic. So I will type medications into the search box. And search. I'm going to adjust my screen just a bit. There we go.

So here are my results. I'm going to point out some features of the results page here. Let's look specifically at the 2nd result titled COVID-19 specific medication type. Notice the yellow ribbon that indicates a CDE is NIH-endorsed. Beside that is the CDE title which in this case is "COVID-19 Specific Medication Type." Below the title is a short definition of the CDE which can be helpful for determining if it fits your topic or your need. Below the definition is the registration status, highlighted in green. This CDE is marked qualified. The registration status indicates how broadly CDE is being recommended or used, and we'll look at registration statuses more later. Also listed is the steward. Every CDE has a steward, which is the organization or the project responsible for the CDE and the repository the CDE stored is the Project 5 COVID-19. If this CDE ever needs an update or it changes in any way, the steward is responsible for working with the repository to make those updates. This is the answer to #7 on your handout. If the CDE ever needs an update or changes in any way, the CDE's steward is responsible for working with the repository to make those updates.

Below the steward is Used by and typically the steward is also who the CDE is used by, like our example here, although anyone can use it. To the right of this information is a preview of the values list, if the CDE has one.

I'll click on this CDE title to see its full record. If you'll recall our definition of CDE earlier, it should have a precisely defined question and a set of specific allowable responses. This CDE's question text, listed right below the title, is "What medications did the patient take to treat COVID-19?" That is the precisely defined question. The permissible values are the specific allowable responses to that question, and if I scroll down a little bit I see those arranged in a table. You should be aware that if you're not logged into a UTS account, then the permissible value section could be missing some information. You might see the message, "Log in to see values."

The first column of the table PV Labels includes the allowable responses to the question. In other words, these are the answers that someone answering the question can choose from. The next column PV Definitions includes how those values are defined or described. The other columns relate to concepts and codes. Concepts describe the meaning of a permissible value. Codes are how the responses would be recorded during data collection for interoperability between studies. Codes might come from a system like LOINC or NCI Thesaurus, or they might be what we call local codes 123 or YN or similar. There is also a feature that allows you to see these codes translated in other systems like LOINC, NCI Thesaurus, SNOMED CT US, or UMLS codes, and that feature will appear right above this table when it's available.

If you keep scrolling past that table, you'll see a lot more details about the CDE. We'll cover some of these features today, but I'll refer you to the repository help guide for more information about them. OK. Now it is your turn to try out searching, so please search within the NIH-endorsed collection for a CDE that collects the employment status of the respondent. And once you found it, put one of its permissible values in the chat.

Alright, I'm seeing lots of questions. Employed full-time, employed, part time, working without pay, retired. Yes, and we will take a look at how to find this together in just a minute. Wonderful. It looks like many of us were able to find that. Let's review this exercise together. I'll go back to the repository and I'm going to go back out to the home page. And I'll search for employment status in the NIH-endorsed collection. That gives me 4 results, and based on the definition of the first one, which is a textual description of a person's employment status, that looks like a good fit, so I'll click on the title. And then I can scroll down to see the permissible values table to view the different allowable responses that are listed in the PV Labels column. And many of these are the responses that you provided in chat. Great.

OK, so next we're going to explore searching within all CDEs in the repository, not just the NIH-endorsed ones. You'll recall that the repository has over 23,000 CDE's and 137 of them are NIH-endorsed for now, and that number will grow. So you might want to search outside of the endorsed group for certain topics. I'm going to go back out to the repository home page and I can do that anytime by clicking on the logo in the upper left corner on the home page. I'll click the search all CDEs tab above the search box. This time I want to find CDEs that measure a person's physical activity, so I'll search for physical activity. I get almost 300 results. Because it's so many results, I can use the filters on the left hand side of the screen to help me find CDEs that match what I need. So first I will try limiting to NIH-endorsed CDEs by clicking on the box beside NIH-endorsed. Notice that now under active CDE filters I see my search terms and I see the NIH-endorsed filter. Using that filter gave me 4 results, but I don't think that any of these four are related to measuring a person's physical activity, just glancing at them. So I'm going to click the X beside the NIH-endorsed CDEs to remove that filter.

If you can't find an endorsed CDE that matches what you need, there's a few other criteria that you can use. First, look at the collections options. These are the organizations or initiatives with CDE-R related to your search. If your organization is listed here then you might see if it has any

related CDEs. I'll click on the NHLBI to limit to their CDEs. Notice that now I have the option to filter to specific projects and topics within NHLBI like sickle cell disease. And again, I can just click on the x beside NHLBI to remove that filter from my search.

The second piece of criteria that you can filter by is registration status. I'll scroll down. Those filters are located right below the collection ones. You'll recall that the registration status indicates how broadly as CDE is being recommended or used. If you can't find an endorsed CDE, check to see if any standard CDE meet your criteria. A standard registration means that the CDE might have been developed by a standards development organization like LOINC or vetted by some other recognized authority and I will limit to standard CDE's by clicking the box beside Standard. Now I have 24 results and they appear to be related to my topic of physical activity.

There's one more way to filter your results that I want to show you. You can check to see if the CDE is used with any forms that you're familiar with or from a widely recognized survey or instrument. And remember that forms are sets of CDEs. I'm going to show you an easy way to see if the CDE is part of any forms. So first, click on the CDE that you're interested in. I'm going to choose this third one titled to what extent are you able to carry out your everyday physical activities such as walking, etcetera. And if I scroll below the permissible values table. I see the related content section with a tab that says linked forms. If a CDE appears on any forms, they'll be listed on this tab. And here we have two. So if one of these forms is one that you're familiar with, then this might be a good CDE option, and we're going to search specifically for forms in a few minutes too. OK. This is a good time to pause for questions. So Kate, have additional questions come in that we can answer.

>>Yes, quite a few. So let's start with some questions related to what you just showed and then I want to get back to a question that we sort of missed earlier. So Stephen asks how is the used by field populated and Steven would like to know if I download a record, would that used by field be included? Robin, do you want to take that?

>>Yeah, I can take that one. So Used by actually refers to organizations, NIH organizations that have recommended that exact same CDE and that's what we call reusing the CDE in the repository. And so we hope to see more and more reuse of common data elements in the repository as the Governance Committee continues to endorse more CDEs and we publish them there. There are older CDEs in the repository that are not yet endorsed, but you-- for examples of seeing reuse, you might see in that field, you might see like NINDS and NHLBI and NCI. You know there might be multiple organizations that are all recommending the same CD. So it's at the organization level. It's not going to not going to add your name there.

>>Great. Thank you. OK. And Thomas asks how are CDEs updated?

>>Well Catherine mentioned this a little bit. She touched on this briefly already, that it's up to the steward of the CDE. That would be the group that is submitting the CDE to the repository is also responsible, continues to be responsible for maintaining it and so we rely on them to bring changes to our attention and the attention of the Governance Committee and we actually are

developing now some SOP's for and processes for handling sort of regular reviews and updates of the CDEs. But the CDE-- at the CDE Repository, our team, we are not subject matter experts, so it's really up to the stewards to update and maintain their own CDEs.

>>Very good. Thank you.

>>We had at least one sort of back-end question Marie is asking does the system use MeSH on the back-end to help find similar terms? Robin.

>>Marie, that's interesting question. Not at this time, we are-- we currently use elastic search for our search and elastic search has its own synonyms and things. So you will you will notice some synonyms being used in your search term from time to time if you look at what you've entered and what's being retrieved. I will say that we are exploring ways of using terminology, not necessarily MeSH, but some kinds of terminology to help categorize and make CDEs more discoverable. Similar ones. I hope that helps.

>>Thanks. OK. All right, let's go to a sort of bigger picture question. Fred had asked earlier can a CDE be considered validated if the field is removed from its original context?

>>Yes, that's a great question and the answer is that it depends. So, Fred, there's a type of form that's known as a bundle, and a bundle is a group of CDES that's considered almost always to be indivisible. And bundles are usually-- often they're standardized instruments, but there might be other groups of CDEs that can't be separated because they'll be invalidated. And if you extract a CDE from a bundle, then yes, that would invalidate it. You're not going to see this feature in the repository today, but it's coming soon and you will see we're going to be adding a feature where if you are looking at a form that is-- I see somebody asking how would you know if a CDE is part of a bundle, soon! In the repository, if you were looking at a form that's a bundle, or if you are looking at a CDE that is part of a form that's a bundle, there will be an alert, like a very clear alert at the top of the screen letting you know and explaining that a bundle is recommended not to be divided. And then you can sort of, you know, review that and decide what you want to do. So that is coming soon. It's not going to be in there today if you go looking. But then aside from bundles, we have just what we call individual CDEs. And those can be, you know, mixed and matched like as you see fit. And there's no-- they're not going to be invalidated by their context changing. I hope that helps.

>>Thanks very much. OK. Sorry, I'm not sure if we already answered this one, so I will repeat it just in case. So Bobby asks, can you remind us who endorses a standard registration status? So Bobby, the registration status has to do with-- It's not about the endorsement by the NIH CDE Governance Committee. I know it's a little confusing, that sort of one thing, which is NIH CDE Governance Committee endorsement, that's the gold ribbon. That's the NIH group that's reviewing and endorsing CDEs. The registration statuses they existed before the Governance Committee did, and what they refer to is how widely a CDE is recommended to be used. And so most CDEs that you see in the repository will have that qualified status, which means they're recommended for use by, you know, in an IC at NIH or individual project. The standard ones are

usually marked standard because they come from a standards development organization or they come from LOINC for instance. You know, larger, I would say larger standards development organization for broader use. And there are some definitions of those in our the guide in the repository. I think at the end Catherine will tell you where to find the guide.

>>Great. OK. So there are a few additional questions, but a couple of them I think we're going to answer in the next section and one, Becky's question in particular, I'd like to just hold off until closer to the end because kind of a bigger question. So I think we can move on and we'll get to some more questions in a bit. Thanks.

>>Great. Thank you, Kate and Robin, and thank you everyone for those wonderful questions. OK, so next we're going to review some strategies for when you're searching for CDEs but you're struggling to find one that's related to your topic. So let's say you want to view CDEs that measure pain or have to do with pain in some way. I'm going to go back to the repository home page. And I'll switch to searching all CDEs and search for pain. With this search, we get over 1000 results. On the one hand, this is great because it means that we have a lot of options for finding relevant CDEs. On the other hand, that is a lot of results to sort through. So take a moment to review the search results page and think about how you might narrow down these results. Let me know in the chat one way that you would narrow these results down. And think about the strategies that we just discussed when we searched for physical activity. How might you narrow these down?

Jaylene says start with endorse. Yes, I'm seeing lots of folks recommend that. Wonderful. Marie says, "Pain severity." And I think that Marie is referring to maybe adding some additional terms. I like that idea. Saraswati says collections. Yes, we can limit by collections. Great. Standard registration status, yes. Thomas says acute versus chronic, again, maybe adding some additional search terms. Good. OK, that's those are great suggestions.

I actually want to review some of the options that you provided and that we covered. So these are the options that we would recommend. And also they're the answer to #8 on your handout. So you can try to limit to NIH endorsed CDE's, limit to a specific collection, or limit to standard under registration type. Another strategy that we haven't discussed yet, but many of you recommended, is to get more specific with your search by adding additional terms. So for example, is there a type of pain or pain in a specific part of the body that you're concerned with, like a limb? Is there a condition or disease that the pain is associated with, such as a fracture? Is there an activity or movement that you're trying to gauge pain during, like walking? You can add these terms to your search to see if a more specific CDE is available. So I will go back to my search and I'm going to add walking and I can add that to the search box at the top of my results page. So I'll add walking. And hit search. This retrieves 24 results, so getting more specific with my search terms really help to focus my options. And keep in mind that you can use a combination of these strategies to further narrow down your results.

Let's do another search. Maybe this time you want to find CDEs to record body mass index, so I'll go back up to the search box at the top of the page. This time I'm going to type in body mass index. Notice that some results do start to auto populate, so if I see a CDE that I want, I can click on it from there. For this example, I'm going to run my search and see all of the results that I get. So with this search I get 22 results. In those results, I can probably find what I need. However, unlike our previous example with pain, when we got over 1000 results and we wanted to narrow our focus, we might want to broaden this search a bit to make sure that we're seeing as many potential CDEs as possible for our topic. One way to broaden our search is to consider any alternative terms or abbreviations or acronyms that could be used for your term. Some CDE's might use a different word to describe the same topic that you're searching for. So what is another term, abbreviation, or acronym that someone could use instead of body mass index? Tell me what you think we could try in the chat.

Yes, I'm already seeing folks saying BMI because that is an acronym for Body Mass Index. I'm also seeing other alternative search terms like weight, which you could also try. Yes. Wonderful. So some of you did say that we could try searching for BMI because that is a common acronym for Body Mass Index, and maybe there are CDEs that use that term instead. So one way to search for BMI in addition to body mass index is to use the word OR to connect the words in the search box. This will search the repository for CDE's that contain either term. OR is the answer to #9 on your handout to expand your search, add similar terms or acronyms to your search and connect each term with or so I can go back to my search box at the top of the page and I already have body mass index, I'll add or BMI. And click search. Now I have 48 results by adding BMI to my search which is more than when I just searched for body mass index.

I want to point out one more feature that can help you find more CDEs on your topic. I'm going to click on the first CDE in my results list to show you this. I'll scroll down to the Related Content box. And beside the Linked Forms tab that we looked at earlier is the More Like This tab. And when I click on that, I see a list of CDEs that are similar to the one that I'm viewing. This is a great feature if you find a CDE that's on your topic, but it's not quite what you need and you want to view other similar options, and this is #10 on your handout. The More Like This button will display similar CDEs or forms to the ones that you're viewing.

All right, we're going to shift gears and explore searching for forms, which is very similar to searching for CDEs. I will go back out to the repository home page. And this time I'm going to click the Search Forms tab above the search box. Let's say you want to study participants for depression and you specifically want to use the NIDA's self-administered version of the Prime MD tool. You can use this information to find that tool in the repository. So I'll type NIDA Prime MD into the search box and search. We get one result with that search. Let's click on the title so that we can see its details and confirm that it is the form we want. Right away we see our form at the top of the record. Below that, I want to draw your attention to the description section of the form. The first sentence of the description reads the Patient Health Questionnaire or PHQ, is a self-administered version of the Prime MD tool for common mental health disorders

administered by health professionals. If I keep reading, I also see that this particular form is the PHQ-2, which includes the first two items of the PHQ 9. Maybe after reading this description you decide that you want to see the version with all nine items. What would you search for? Tell me in the chat what you think we could search to find the nine item version. Yes, I am seeing lots of folks say PHQ-9 exactly. We see in the description that the nine item version is called PHQ-9, so we can try searching for that. My point here is that the descriptions can be really useful for not only making sure that you found the CDE or form that you're looking for, but also if there are similar ones or different versions that you can use for your data collection.

OK, so let's try searching for a form together. This time I would like you to locate the Clinical Decision Support for Substance Abuse form from the NIDA. And once you found it, tell me in the chat how many questions it has.

Already seeing lots of answers, I'll give everyone a little bit more time to look for that. Great. Thank you so much for participating. I'm seeing 28 across the board, 28 questions, that is correct. So let's take a look at how to find that together. I'll go back to the repository. Back out to the home page and this time I'm going to search forms for clinical decision support for substance abuse. I get 2 results and I want the one from the NIDA which I see is the first one and now I can see that that form has 28 questions. There's one more way to find CDEs and forms in the repository, which is browsing. You can use the CDEs and forms buttons at the top of the home page screen to browse.

So let's look at browsing forms. On this page I see the different collections with available forms that we can browse. So I'll click on the NICHD to see what forms they have. And now we see the forms from the NICHD. I want to point out that there are filters for more specific projects within this collection, like the MBSTRN core. So if you're looking for a specific group of forms, you can look there. If you want to browse CDEs, the process is the same except that I will click on CDEs from the menu at the top. And then I can browse collections from there.

So let's try out browsing. This time I would like you to tell me the name of one NEI CDE collection that you can browse in the repository. So this time we're browsing CDEs and tell me in the chat the name of one NEI CDE collection that we can browse.

I'm seeing lots of different answers, which is great. OK, this one's a little trickier. Some of us are getting it, others aren't. So I'm going to demonstrate this and we'll go over it slowly together because it can be a little bit harder to find. So I will go back to the repository. And again to browse I will click on CDEs from the menu across the top of the page. Then I'll locate NEI. Click on that. And then as you can see over here the NEI has two collections of CDEs in the repository that we can browse the LASIK Quality of Life Collaboration Project and eyeGENE. And then you can click on those to view just the CDEs in those collections.

OK, the final feature I want to demonstrate is how to save the CDEs and forms that you find so you can come back to them later. The repository has a feature called My Boards where you can create a board which is kind of like a folder to store CDEs and forms that you need later or that

you want to export my boards is the answer to #11 on your handout. In order to save and access boards, you'll need to be logged into your UTS account. And like I pointed out at the beginning of the training, you can sign up for a UTS account if you don't already have one. So let's try adding a CDE from this page to a board. To do that, I'm going to click on the pin icon beside the CDE name. Let me try another one. Think I may have timed out on being logged in, so I'm just going to pause my share and log back in quickly. OK, now I'm logged back in, so now I can do this. Alright, so let's try adding a CDE. I'm just going to click on the little pin icon beside its title. When I do that, I get this pop up window asking me if I want to add the CDE to an existing board, of which I have two or I can use the Add Board feature to make a new one. I'm going to select an existing board, the NEI CDEs board to add it to that one. And then I get this message that says it was pinned to my NEI CDEs board. It's kind of a small message at the bottom. To view all of your boards, scroll to the top of the page and click on My Boards from the menu across the top. Now we see my boards that I've created. I have three. Notice that one is for forms and the other two are for CDEs. You can't save them to the same board, so you'll need to make separate boards for each type. I'll select my NEI CDEs board so you can see what that looks like. I can see all of the CDEs that I've saved to this board, and I can also click the Export Board button to export this board in different formats.

All right. At this point, I will ask Kate to put the handout answer key in the chat. So if you missed any answers and you want to review that handout, you'll have that answer key. And now we will answer any remaining questions.

>>Alright, fantastic. So we have a lot of questions that have come since you've been talking, Catherine. So what I'm going to do is I'm trying to sort of group the questions by theme. And so I see some questions about searching, some about display, some about linking, and some about the content itself. So I'm going to start with the questions about searching because we were talking about searching body mass index or BMI and one question we had is, "Do you use Boolean in searching the CDE-R?" And I think that's for you, Catherine.

>>Yes, I can answer that, thank you. Yes, you can use Boolean. You can use OR to combine search terms if you want to look for you know, similar terms, synonyms, acronyms. You can also use AND. And there are more details on the help guide about using Boolean to search and I'll show you how to get to that help guide here in a couple minutes.

>>Fantastic. And so a related question was something to the effect of is the BMI search mapping to anything or is it really just a keyword search?

>>Robin, you can correct me if I'm wrong or I leave anything out, but it is searching specific fields. It's searching for keywords and specific fields in the CDE, so it's not necessarily mapping to anything. And Robin, feel free to jump in if there's more to that.

>>No, that's correct. Each search that you run is a keyword search, but it's searching almost every text field in the database.

>>Great, wonderful. Thank you. OK, there was a question about the registration status. And the question was, "Why would you limit this to standard under registration status instead of qualified?" Not sure who should address that question.

>>I mean I can take that. Simply put, yeah. Simply put if it's a standard CDE it's a more broadly recognized standard like it might be used outside of-- might be recognized outside of NIH. I know I used the example of LOINC before. I think most are all of our LOINC based CDEs are considered standard because they are developed by that organization and used more broadly than just an NIH biomedical research. So it's just if you have too many results and you're trying to narrow it might help you find some more widely recognized standards.

>>Great. Thank you. OK, I think there's another question for Robin. This question is how the data types field works. The mention was that in some CD's the value lists are just numbers. Robin, are you able to address that? Thanks.

>>Yeah, we have a limited number of data types and of course, I'm not going to be able to list them all at the top of my head, but there's text, number, date, time, value list which is when you have like a list of choices. I know there are others and I'm forgetting them. Those are the most common ones and so when we talk about CDEs having a specified set of answers. Sometimes it's a specific answer list, but sometimes it's just a format, like it needs to be a date. And so that is set by the submitter and the steward of the CDE determines what the what format it is, and that's the data type. I hope that answers the question.

>>Thanks. And yeah, Armory, if we didn't quite answer your question, please ask it again in chat. Thank you. So from Carrie, just sort of a display and functionality question. In the future will you have hyperlinks to other forms in the form descriptions? Robin.

>>Yeah, we don't have that now, but that's a great feature to request. And so I've made a note of it. So thank you. Wonderful. Thanks for that suggestion, Carrie. So then there were a few questions about relationship between the CD-ER and other databases and other systems. The first one was will the CDE-R be linked to PubMed?

>> Not right now. We are-- This is something we've been exploring for a while. What would be useful things to link to and what would be possible, and you know, everyone is interested in being able to cite CDEs and track them and then link to where they've been used. This is definitely something there's a lot of interest in, but the practicalities of how that's actually achieved where there's still a lot to work on there, so not yet.

>>Thank you very much. And I think you've addressed Gazal's question as well related to citing. Another question Gazal had was, is there a software or a tool that could import CDEs and compile them?

>>I am not totally sure if Gazal wants to unmute or put more in the chat that would help me understand the question. I'm not sure I understand what they want to do. Yeah, thanks. If we could get a little bit more feedback on that Gazal let us know what you want, what you're trying

to do, and then we'll try to address that. Thank you. OK. Lots of questions still coming in, but I'd like to address some of these content questions. Some of them were asked a little bit earlier, but I wanted to save them to kind of address them together. Becky was asking. It seems like most of the CDEs are for projects that are developing. Are we expected to develop a CDE if they don't already exist?

>>I think no, not at the individual level. The idea with CDEs is that your organization, your funding organization or even higher up than that is going to be developing CDEs for use across multiple studies. And so you as in like individuals are not expected to be coming up with CDEs. They should be getting guidance from their organizations on where to find CDEs to use and whether they are required or recommended or so on.

>>OK. Thank you. I may just need a moment to catch up with all these questions coming in. Actually, let me pose one more question to you, Robin, while I look for more.

>>Sure that's fine. I'll answer very slowly, give you some.

>>Becky's commenting that it seems like the NIH endorsement or the NIH-endorsed CDEs are very focused on COVID-19. And are there plans to expand the NIH endorsement to other CDEs?

>>Yes, I'm really glad that you asked that question because I sometimes forget to emphasize this. The endorsement process, the Governance Committee and the whole NIH endorsement process are a relatively new. And so the very first group of CDEs that were endorsed were from Project 5, which was the COVID-19 [inaudible] NIH Initiative. And then the second group comes from the NHLBI CONNECTS, also a COVID initiative. NIH, the Governance Committee was actually given guidance to prioritize any kind of COVID projects in their endorsement process. So those are the first two groups that have gotten endorsed, but this is an ongoing process. We have-- I think we have half a dozen different collections in the pipeline right now and hopefully something else will be endorsed in public soon. And this, you know, we definitely want to keep growing the collection of endorsed CDEs there and the focus will, you know, and of course it will broaden and not be so COVID specific and offer more and more choices for other domains in the future. So yeah, do check back.

>>Thanks so much. Here's a question about coding. Is there any standardized coding advice for endorsed CDEs?

>>So we at the repository and also at the Governance Committee don't advise any particular coding. Like we're sort of agnostic on what you might use for coding. We do believe that, you know, coding definitely facilitates interoperability. So we encourage but don't require submitters to code their CDEs. So some of them I know if you look at I'm not sure about all the Project 5 ones, but I think if you look at the NHLBI CONNECTS CDEs and the repository-- there's only a few there now. They're going to be adding to their collection later. But if you look at the ones there now that are the value list data type, I think almost all of those have coding for the responses,

and the coding comes from, you know, LOINC and SNOMED and the NCBI thesaurus and so on. So yeah, we do definitely encourage that, but it's not required.

>>Thanks so much, Robin. Anthony asks, is there an API for the search?

>>Yes, there is. And maybe while you're finding the next question, I'll find a link and share it with you. There is indeed.

>>All right, I found the link and I'm putting it in the chat to everyone.

>>Fantastic. Thank you. All right. So there were some other questions about development of CDEs, sort of the process of development which I think you addressed, but if you wanted to say something more about development of very common, common data elements like [inaudible]

>>No, it's funny to see these questions, actually very validating for me to see these questions. These are all like-- almost everything that people are asking about is are things that we are talking about and exploring in the Governance Committee. And other groups outside of that have definitely been talking about how do we achieve like yeah, common common, the most common, universal data elements that cover that would be applicable to almost any kind of research at NIH. There is definitely a strong interest in that it also seems again not, you know, not easy to accomplish something like that and have it work across all of NIH. So that is a conversation that's ongoing. But we, you know, someday we hope to get there.

>>Fantastic. And I think you addressed this, but I think it's probably worth repeating. We had another question about, you know, who can create forms and who can share them.

>>At the moment the only people that can create forms are the stewards of the CDEs, you have to have certain permissions to do that. And so probably, unless you are somebody who's representing, you know, a group that's going to submit, an NIH group that is going to submit CDEs for endorsement, then you probably would not be allowed to create forms. But then I might encourage you to look at the boards feature that Catherine highlighted at the end. That's one way to sort of, you know, group together CDEs that you're interested in. And store them in one place. And of course you can export them in multiple formats and then you know, do what you want with them in other applications.

>>Great. I'm answering a question in the chat that I can answer right now. Very quickly.

>>Go for it. I want to because it's an important one. Do the CDEs have identifiers or DOI? They don't have DOI, but they do have identifiers. Every CDE in their repository has what we call an NLM identifier. It's actually part of the URL, but it's also stored in the page. And like in the CDE details page that Catherine showed with all the metadata about the CDE, there's a section called identifiers and you'll find the NLM ID there. So everything in the repository has an NLM ID. It doesn't change. It may have a different identifier as well. And there is a related question, are they versioned? Yes, CDEs are versioned, and if you scroll all the way to the bottom, if there are earlier versions, you can see the history as well.

>>Right. Karen asked a related question. I think it's related and that's whether the concept IDs are used outside of the CDE repository? Not sure if we can answer that.

>>I don't-- I can't-- The concept ID's come from other systems and they-- so far we've had concept ideas come from UMLS and from the NCI thesaurus. So those are the only two sources I've seen so far for concept IDs. But those are very, very widely used sources, so I know that they-- those concept ideas get used elsewhere. I don't know a lot about [inaudible]. I know of it, but I don't know what kinds of I think there. I think they're working on-- They might be working on UMLS concepts, but I might not be correct about that, so please don't quote me. But they are from recognized systems that are widely used.

>>Thank you. I see at least one question here about requirements for use of CDEs and I guess in the generic question would be, you know, to whom should one refer about what your requirements are for your grant? And that would of course be your-- the granting organization. So you probably should talk to your project scientist. OK. And oh, here's a fun question. How do you cite a CDE?

>>That's a good question. We have not established the standard format for citing a CDE yet. As I said, we do have these permanent identifiers, but we have yet to establish the citing format. So I'm sorry, I can't help you more than that, Leslie.

>>Yeah, so stay tuned. That's a great question. Thank you for that question. OK. So I think actually at this point we need to wrap up the questions so we can wrap up the session. So oh, Kim is requesting--

>>I can briefly address that, yes. So we do know that there is-- There's duplication in the repository if you just a couple of things briefly. One is that sometimes something that looks like duplication might not be. There might be a slight difference in the type of the question, or there might be different allowable responses and so they're not actually duplicate. We're very literal about duplication, like if it's not exactly the same, then it's not a duplicate. So you might have two CDEs, for instance, that are measuring something about employment. But if they one only has 3 answers and one has like 10, then those are not, in our view, the same CDE, even though they're measuring the same thing. So that's one thing to know. The other thing that I'll acknowledge is yes, there is duplication in the repository and we are aware of that. And so on the one hand we have the new endorsement process which we hope is going to start to reduce a lot of the duplication of things, new things coming in and at the same time we're aware of it, the older CDEs in there that are duplicative and we're going to, we're working on cleaning those up as well. So you can use the strategies Catherine taught you to narrow down if you're getting too many responses and you can always reach out to us. I think Catherine is also going to cover how to get in touch at the end here. So thank you.

>>Great. Thanks so much. And if we didn't get to your question, we apologize, but Catherine is going to explain how to reach us. So take it away, Catherine. Thank you.

>>Yes, thank you. OK, so to wrap up, let's talk about how to get help and to find updates to the repository. First, there is a new NIH CDER news listserv that you can join to keep up with the latest updates and features. And Kate is going to put a sign up link in the chat for that. And that link is also on your handout, so you can get it there. When you're in the repository, if you hover over the help menu, you'll see these four options. Guides has step by step directions for a lot of what I demonstrated today and other actions that you may take in the repository. The New Features page has announcements and news. The Resources page has links to additional trainings and information about API's and examples of CDEs being used in biomedical research. And then you can use that contact us page to send your questions to the NLM.