

Transcript

CATHERINE: Good afternoon, everyone. Thank you for joining us today for Standardize Your Research Data with the NIH Common Data Element Repository. The purpose of today's training is to familiarize you with Common Data Elements (CDEs) and the NIH Common Data Element Repository (the CDE-R). I'm Catherine Staley from the National Library of Medicine's User Engagement Program. Here with me today are my User Engagement program colleagues, Kate Majewski, Mike Davidson, and Michael Tahmasian. We're also joined by Robin Taylor from the NLM's Health Data Standards branch, who will be helping me answer your questions. Robin has been with the NLM for six years as a Technical Information Specialist and is the Product Lead for the CDE-R.

During today's training, you'll have the opportunity to ask questions and practice searching in the repository, so let's review how we'll interact in 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, but if you have a question or a comment, please use the chat box and send it to everyone. I won't be able to answer messages sent directly to me, but Kate will be monitoring your questions throughout and I'll pause frequently to see if there's anything that I or Robin need to address. So please find that chat button now, make sure you're sending messages to Everyone, and then tell us the name of your organization and your role or title. It's always great to see who's joining us today.

Wonderful. Oh, we have a diverse group with us. I'm seeing some librarians, some NLM staff. Excellent. Hospitals. Libraries. This is great. Thank you all for contributing in the chat where you're coming from. Great, excellent. OK, you can keep introducing yourselves in the chat. I'm going to keep us moving.

So to keep us on track, we're providing a handout that you can download and follow along with. Mike is putting a link to the handout in the chat. Now, 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. And finally, during class, I will also be occasionally asking for your nonverbal 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. Some devices may not allow for that type of interaction. If you can't find the reactions button on your device, feel free to use chat to respond when I ask you to. Although I'm seeing many, many, many thumbs, so I think most of us have found it. That's great, thank you.

OK, here is our agenda for the 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.

Before we dive in, I would like to gauge your familiarity with CDEs, and we'll use a poll for this question. So Michael's going to launch the poll for us now, and we would like you to select the answer that best represents your familiarity with CDEs. You can choose from:

- Very familiar
- Moderately familiar
- Slightly familiar
- Not at all familiar

I'll give us a minute to respond. I see lots of folks already responding. Thank you for doing that. OK, great. It looks like most of us have responded. So we can end that poll, Michael, and take a look at the responses. Perfect. So it looks like we kind of have a lot of us are Not at all familiar and that is totally fine. You're going to learn all about CDEs and the Repository today. And then we have some folks in the middle and a few who are very familiar. And I think this training will offer all of you something regardless of your familiarity with the Repository. Excellent. Thank you for your responses.

We can actually use the data that you just provided in the chat and in that poll to explore some reasons that you might consider using CDEs in your projects. So a few minutes ago, you shared in the chat your organization and your role or title. Let's say I'm writing a report after this training. To summarize how it went, I want to include some data about the participants. I could start by entering each of your answers in a spreadsheet, and then I might need to standardize how your organization's name is represented. On the 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 both, with the acronym in parentheses beside the name. This is just a small subset 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'd 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 familiarity with CDEs, but they asked the question in a different way? For example, maybe they asked the audience to describe their level of familiarity in their own words. It would be really hard to compare those free-text responses to the responses that you just provided. We'd 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 asked 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 we 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 us are familiar with Common Data Elements and the Repository, and for others, it may be 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 (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. You'll see, in the following examples, that CDEs may ask questions of respondents to gauge their

opinion, collect a medical history or their experience of something. Or CDEs may collect lab data like a date or a test result.

Let's look at our first example of a CDE. On this slide is a CDE in the Repository. It's titled Usual Place of Care 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 the researcher can answer with one of these specific allowable responses:

- A doctor's office or community Health Center or hospital-based clinic.
- Walk-in clinic, urgent care or retail clinic in a pharmacy or grocery store.
- Emergency room.
- A VA Medical Center or VA outpatient clinic.
- Some other place.
- Does not go to one place most often.
- Don't know.

Here's 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. The date should be entered as a 2 digit month, 2 digit day, and 4 digit year. 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 you 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 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. The Repository can help you meet that requirement.

Let's take a closer look at how the use of CDEs supports the NIH's mission. Both the NLM and the NIH have identified Common Data Elements as a tool for facilitating the interoperability of data. Ensuring that biomedical research 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 in aligning with the FAIR principles.

Additionally, you're probably aware of the 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'd like to learn more about that plan, there is a link on the handout to it. So give me a thumbs up if you're required by organizational policy or a grant to use CDEs, or if you're helping others find CDEs to comply. I'm seeing quite a few thumbs. Great. Well, you'll be better prepared to do all of

that after this training today, because if you know how to locate CDEs in the CDE-R, you'll be better equipped to support these goals. All right.

Next, I want to review 3 examples of CDEs in use in the real world. We'll start with the NHLBI initiative called the Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies (CONNECTS) which illustrates how CDEs can be used for multiple reasons. The goal of CONNECTS is to mobilize and coordinate the US clinical trial infrastructure to find the most effective therapies for treating COVID-19. It consists of over 40 networks and cohorts with over 1,000 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 CONNECT study called the Novel Experimental COVID Therapies Affecting Host Response (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.

Now, let's take a look at a study that cites the use of CDEs. The authors compared findings from two different studies about traumatic brain injury in active military members and veterans. One study is the Defense and Veterans Brain Injury Center TBI Center of Excellence 15 Year Longitudinal Study and the other is the Long Term Impact of Military Relevant Brain Injury Consortium Chronic Effects of Neurotrauma Consortium Perspective Longitudinal Study. These two studies have a total of over 3000 participants. The authors say in the article that this comparative analysis was enabled by heavy use of NIH common data elements in both studies. For example, both studies used the CDE titled the Neurobehavioral Symptom Inventory so these authors could compare results across the individual studies.

And let's look at another example. This one is a published, systematic review from researchers at Stanford University in the journal *Frontiers of Neurology*. The focus of their review was neuromuscular deficits of children with spastic cerebral palsy. As part of their methodology for selecting articles for the review, they only accepted articles that reported using the National Institute of Neurological Disorder and Stroke CDE, titled Gross Motor Functional Classification. By only including articles that employed a specific CDE, the researchers could better synthesize findings from across multiple research studies.

The focus of our training today, the Common Data Element 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 CDEs submitted by ICs, and they work closely with the NIH CDE Repository team at the NLM. Currently, the repository has over 24,000 CDEs from across 19 organizations. In the next part of this training, I'm going to highlight three key features of the Repository that can assist with your search. These are NIH-Endorsed CDEs and forms and bundles, and this will prepare you to actually search the Repository in a few minutes. The Repository includes NIH-endorsed and non-NIH-Endorsed CDEs - 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 recognize 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 NIH-Endorsed.

On this 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 this list:

- 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.
- 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 in NIH supported and conducted research for you. This means that you can more quickly identify the CDEs that have met the set of criteria and are ready to use. Currently, there are 4 collections of NIH-endorsed CDEs. These include:

- The Project 5 collection, which was developed by an NIH-wide COVID-19 data coordination group for a wide spectrum of COVID-19 research, including translational, clinical, and applied research.
- The NHLBI CONNECTS Organ Support CDEs, which we just viewed as an example.
- Core CDEs from the Science Collaborative for Health Disparities and Artificial Intelligence Bias Reduction or the SchARE program, which was created by the National Institute on Minority Health and Health Disparities and the National Institute of Nursing Research.
- And 13 COVID-19 mitigation policy CDEs from the NIA funded Social, Behavioral and Economic COVID Coordinating Center (SBE CCC).

If a CDE is endorsed, you'll see this gold ribbon icon beside it in the Repository.

All right. The second feature I want to highlight is forms. And this is the answer to #3 on your handout. Here's our definition. a form is a group of questions or variables with specified sets of allowable responses (aka CDEs) that are used as a set for particular research or clinical reasons. In other words, forms are composed of individual CDEs. Forms can collect a lot of data about a participant all at once, like a family history, or they might assess or measure something like pain or the seriousness of an adverse effect from a medication. Some forms have instructions for the administrator that dictate how they should be completed, or for the respondent when a form is to be self-administered. These instructions also add important context for the forms use. Forms will often include a system for scoring a participant's 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 it was developed by the World Health Organization in 1998. It includes 3 CDEs. 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.

Sometimes you will see forms in the repository that are designated as bundles - and this is the answer to #4 on your handout. A bundle is a form that is indivisible and individual CDEs pulled from bundles are not considered valid. Bundles are clearly marked in the Repository with a blue banner at the top of their record that states: "This form is a bundle. Bundles are indivisible. They are not considered valid and reliable if not used in their entirety as intended." CDEs that are part of a bundle are also clearly designated in the repository with a banner that says: "This CDE is part of a bundle. All CDEs within a bundle must be used together. Go to bundle," with a link to its bundle record. We're going to see examples of all of this in the Repository today.

OK, so we're now ready to search the Repository, but before we do that, I want to review some key points that we've covered so far. Michael is going to launch a little quiz with a few questions for you to answer that will check your understanding of CDEs and the Repository so far. These are also questions 5 through 7 on your handout if you prefer to follow along there. So I'll give everyone a minute to submit their answers and then we'll review these together.

All right, I see we've got a lot of responses. Looking good so far. Michael, let's go ahead and end that quiz and we can show everybody the answers. Perfect.

All right, let's review the answers together. And remember, these are questions 5 through 7 on your handout. So if you're following along there, you can record the answers there. So for number 1, you can find CDEs, forms, and bundles in the CDE-R. So you can find all three of those things and we'll take a look at an example of all three. Next question. Endorsed CDEs have been reviewed and approved by the NIH. So Endorsed. And then finally, it is True that in the CDE-R a bundle is a form that is indivisible and individual CDEs pulled from them are not considered valid. Excellent. I think we all did a pretty good job on that. So before we open up the repository, I want to stop and see if any questions have come in. Kate, are there any questions we can answer?

KATE: Yes, indeed, there are a few questions so far. The first question is from Karen. **How does the NIH CDE Repository relate to the PhenX Toolkit? Robin, can you take that one again?**

ROBIN: Yes. So the PhenX Toolkit, if you're familiar with it, they vet instruments and protocols and assessments, and add them to the toolkit for various research purposes, like COVID or social determinants of health or so on. And they are looking at the entire instrument over at PhenX. And what we have in the CDE Repository is primarily CDEs, which are the individual variables or questions. So we have taken protocols from the PhenX Toolkit, and they have been broken down into CDEs and variables, and then we've added them to the CDE Repository. That's been sort of further in the past. We don't have any new ones endorsed yet, but there is a submission in the pipeline, so we're starting to build that relationship with them now.

KATE: Great, thank you. OK. Another question from Chris, **how well do the CDEs align with the data standards required by regulatory agencies such as FDA, PMDA, EMA and others, for submission of trials?**

ROBIN: That's a good question. I'm realizing now I don't know a lot about those particular standards. So that's something that, Chris, I'd be happy to look into offline for you. If you would like to know more about that, you can feel free to contact me.

KATE: All right. Another question from Praveena, **can CDEs be imported into data platforms like REDcap?**

ROBIN: Yes, absolutely. So there's a couple ways to do that too. We have the forms feature that Catherine just showed you in the training where you, if you're logged in, you can export forms in various formats, including into a REDcap format and then they'll be ready for importing directly into REDcap that way. There is also a feature, if you're familiar with the REDcap Online form building tool, there is a feature in there where you can actually search for CDEs in the repository directly in that tool and add them into your REDcap instrument that way.

KATE: Great. OK, I think we'll do one more question and then we'll do some more questions later. So here's a question from the FSR Research Team. **If we want to create a CDE for our disease, are there consultants or groups we could engage with that you would recommend who have experience getting CDEs endorsed?**

ROBIN: Yeah, I can't recommend any particular consultant or a group. I can say that anybody that's NIH staff, there are groups that we could connect, like working groups that I could connect you with that might be helpful to be a part of. And in addition, our team provides user support, and that includes people that are preparing CDEs to submit. So again, you can feel free to reach out to me and I can try to put you in touch with somebody that could help.

KATE: Great, thank you. So there are some other comments and questions, but I think we'll get to those in just a little bit. And let's go back to you, Catherine.

CATHERINE: Great. Thank you, Kate and Robin, and thank you all for your questions. Yeah, we'll stop a couple more times for Q&A, so feel free to drop those in. All right, we are now going to search the Repository. If you'd like, you can follow along with this demonstration on your own device. Mike is putting a link in the chat to the Repository, and I'm going to open it on my screen as well. 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 View Options at the top and adjusting the Zoom ratio. So once you have the Repository open or you can see my screen, give me a thumbs up and I'll know that we're ready to go. Excellent. I'm seeing a large cloud of thumbs, so I think we're ready. All right, let's get started.

Anyone can search the CDE-R and see results, but to access all of the Repository's features, you'll need a Unified Medical Language System Terminology Service account (UTS account). So on your screen, you'll see a Sign In button in the repository's upper right corner to either sign in to an existing account or sign up for a new 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. As we go, I'll point out any differences between what the repository looks like when you are or aren't signed in.

Next, I'm going to point out some features of the Repository's home page. The Repository's default search is for All CDEs. You can switch to searching for NIH-Endorsed CDEs or Forms by clicking on the tabs above the search box. You can also select CDEs or Forms from the menu across the top of the page to search or browse for those.

Let's start by some keywords within All CDEs. Imagine 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 with the key terms COVID-19 medications to see if any CDEs are available on that topic. So I'll type COVID-19 medications into my search box and search.

Here are my results. I want to point out some features of the results page. First, you can see that some of these CDEs have that yellow ribbon icon indicating they're NIH-Endorsed and others don't, like this one. Let's look closely at that first result. It's NIH-Endorsed. Its title is COVID-19 Specific Medication Type. Below the title is a short definition of the CDE, which is the type of medication used to treat the COVID-19 infection. Below the definition is its registration status, highlighted in green. This CDE is marked qualified. The registration status indicates how broadly a CDE is being recommended or used. We'll look at registration statuses a bit more later. Also listed is the Steward. Every CDE has a steward, which is the organization or project responsible for the CDE in the Repository. This CDE's steward is Project 5, COVID-19. If the CDE ever needs an update or changes in any way, the steward is responsible for working with the Repository to make those updates. And that's the answer to #8 on your handout. Below the Steward is Used by. This field will list the organizations that also recommend the CDE in addition to the Steward. To the right of all this information is a preview of the acceptable responses, if the CDE has one.

I'm going to click on the CDE title to see its full record. At the top of the record, we see the notice that this CDE is part of a bundle with a link to that bundle. We're going to return to that in just a minute. 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's the precisely defined question. The permissible values are the specific allowable responses to that question.

And if I scroll down a bit, I see those permissible values organized 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 Login to see values. So you can just take a look at my screen for now. The first column of the table, PV Labels, includes the allowable responses to the question. In other words, these are the responses 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 the NCI Thesaurus, or they might be what we call local codes like 123 or YN or similar. There's

also a feature that allows you to see these codes translated to other systems, including LOINC, the NCI Thesaurus, SNOMED CT US, or UMLS codes. That feature will appear above the table when it's available, which it's not for this particular CDE.

NIH-Endorsed CDEs also have concepts attached, so their meaning is clear and unambiguous in the left sided menu of the screen. I'm going to click on Concepts to jump to that section. Here, in the Concepts table you can see this CDE's concepts from the NCI Thesaurus, but the concepts may be taken from other standard terminologies or systems like the UMLS. If two CDEs share the same concept in the repository, then that means the CDEs are measuring the same thing.

OK, I'm going to scroll back up to the very top of the CDE record and I want to take a look at the notice that this CDE is part of a bundle. To see the bundle that it's part of, I'm going to click on the bundle link in that blue banner. This takes me to the record for the bundle. You'll recall that some forms are designated as bundles, which is the case here. This form is titled Project 5 - Disease Specific Medications and it's a bundle as we see from this notice at the top of in the blue banner. If I scroll below the title, I see the bundle and the CDEs that are part of it. And the second one is the CDE that we just reviewed titled COVID-19 Specific Medication Type.

OK, now it is your turn to try out searching. If you're able to be in the Repository, I would like you to search for an NIH-Endorsed CDE from the ScHARe project that collects the health insurance type of the respondent. And once you think you found it, put one of its permissible values in the chat and then we'll go over this exercise together. So try that out, put one of its permissible values in the chat, and then we'll review together.

Excellent. I'm seeing lots of our permissible values being dropped in the chat. These look good. I think folks are finding it. Excellent. I'll give us all a little bit longer to locate that. Great. All right, excellent. Looks like many of us are finding it. Perfect.

OK, I am going to demonstrate how to find that CDE in the Repository. So let me pull the Repository back up and I'm going to go back out to the home page to search. So the way that I'm going to look for this is by searching for health insurance in the search box. This gives me multiple results. The top one is from the ScHARe project as I see in the Steward field, and it's titled Health Insurance with the definition: "A type of insurance designed to cover some or all the cost of treating an insured person's illnesses or injuries." And I also see that it's NIH-Endorsed with that yellow ribbon icon. This is looking pretty good. I'm going to click on the title to see more. And when I do that and I scroll down to that permissible values table, I can view the different allowable responses listed in the PV Labels column. And these are the responses that many of you were putting into the chat. So there we go. Perfect.

All right, let's keep searching for CDEs. I mentioned earlier that the repository has over 24,000 CDEs. So when you search for certain topics, you may get an overwhelming amount of results. For example, I'm going to go back out to the Repository homepage and let's search for CDEs that measure a person's physical activity. So I'll type physical activity into the search box and click search. I get over 300 responses results. Because it's so many results, I'm going to use the filters on the left-hand side of the screen to help find CDEs that match what I need.

First, I'll try limiting to NIH-Endorsed CDEs by clicking on the box beside NIH-Endorsed. Some of you may have used that feature during our last exercise, which is great. Notice that when I click that box now underneath ACTIVE CDE FILTERS, I see my search terms and NIH-Endorsed CDEs. Using that filter gave me 11 results. But if I just take a quick look at the titles and the definitions, I don't think that any of these are measuring a person's physical activity. So I'm going to try a different filter and I'll click on the X beside NIH-Endorsed CDEs to remove that filter.

Let's look at the Collections options. These are the organizations or initiatives with CDEs related to your search. If your organization is listed here, then you might see if it has any related CDEs. So I'll click on the NHLBI to limit the CDEs from them and notice that now I have the option to filter to specific projects and topics within the NHLBI like sickle cell disease, and I can click on that to further filter it. And again, if I want to remove those filters, I'll just click on the X beside them at the top of that page.

The second piece of criteria you can filter by is registration status. Those filters are below the collection ones. You'll recall that registration status indicates how broadly a CDE is being recommended or used. If you can't find an endorsed CDE, check to see if any standard CDEs meet your criteria. A standard registration means the CDE might have been developed by a standards development organization like LOINC, or vetted by some other recognized authority. So I'm going to limit to Standard by clicking on the box beside it. And when I do that I now have 24 results. Little more manageable. And a lot of these appear to be related to my topic of physical activity.

I'm going to click on this third one titled: To what extent are you able to carry out your everyday physical activities such as walking [etc.]? I'll click on that. Notice that it does not include the blue banner indicating it's part of a bundle. If I scroll below that Permissible Values table, I see this Related Content box. If this CDE appears on any forms, they will be listed on the Linked Forms tab in this box, and I can see here that we have two. So this can be a great way to find forms related to your topic. If you find a CDE on your topic and you're interested in forms related to it, you can scroll down to this box and view any forms that may be related to that topic.

OK, this is a good time to see if we've had any more questions come in. Kate, what can Robin and I answer?

KATE: Let's start with this question from James. **Are there NIH-Endorsed forms and can a form have endorsed and non-endorsed CDEs?** Robin, would you like to take that?

ROBIN: Yes, I'll take that one. So at this time forms are not endorsed by the governance committee. That said, there are forms in the Repository that are composed entirely of NIH-Endorsed CDEs, but just there's a distinction there between the CDEs themselves and the form as a whole. So the form has not been endorsed, but the CDEs in it have been. And theoretically, yes, your organization could develop a form that incorporated some endorsed CDEs and some new or some not endorsed CDEs. It's a thing that's possible, but we would recommend developing and submitting your CDEs for endorsement if you were going to do something like that so that they could be added to the Repository for others to reuse as well.

KATE: Thank you. And from Ben, **for NIH funded studies, are we required to use the NIH-Endorsed CDEs?**

ROBIN: No. So that's, I think, a common misconception, a thing that comes up sometimes. The NIH-Endorsed CDEs are not necessarily required anywhere. So at this time, there are no NIH-wide required CDEs. There's no group of CDEs that NIH is saying all studies must include. Your individual funders at the different Institutes and Centers or your program or your research initiative, they will recommend or require CDEs for you to use. And they may include some endorsed CDEs. But just because they're endorsed doesn't mean that you are required to use them in your study.

KATE: So there's a follow up on that. I'm not sure if you saw this one yet, but **how do you find out if you do have a requirement to use a CDE?**

ROBIN: Yeah, you should talk to your Program Officer or your-- I can't remember the name. Your lead investigator, your program officer, somebody like that is going to have more information for you about what is required or recommended for your project. Also, if there's any sort of written terms of your grant, then it might be in there. I know at some institutes they write requirements for CDEs into the project milestones, if you have those. So talk to the people that are managing your project.

KATE: Great, thank you. And I think we'll just take one more before we get back to content here. And let me see, OK, from Edon, **what happens to the CDE maintenance when the Steward's project is over?** Robin?

ROBIN: That's a great question. We are concerned about sustainability of CDEs. So ideally, you know, other groups would begin using those CDEs and maybe we could find a new steward to take them over. My team at NLM is now working with the governance committee to develop some SOPs and some policies for the lifecycle of CDEs that take this question into consideration as well. So if we can't find anybody, you know, if nobody is monitoring or maintaining them and we can't confirm that they're still current and useful, then we would probably retire them from the Repository.

KATE: Thank you. And sorry, can I take it back? There is one more question I think we should get to before we get back to Catherine. This one was from Fred and it was specifically **about an example that Catherine was showing from the COVID-19 specific medication type CDE. Fred is wondering why did they choose the NCI Thesaurus in this case? Can you help us with that?**

ROBIN: I don't know, I can't say why they chose the NCI Thesaurus. We recommend that, you know, these CDEs need to be annotated with these concepts, IDs and codes as much as possible to facilitate that interoperability. But there's multiple controlled vocabularies and terminology systems that can be used for that purpose. So we most frequently see UMLS codes and NCI Thesaurus codes used for these CDEs, but I can't tell you any specific instance why they chose that particular source.

KATE: Fair enough. Thanks, Robin. All right, now we really will get back to Catherine. Thanks.

CATHERINE: Thanks, Kate and Robin. These are great questions. I always learn something from the Q&A sections of this training too, which is great.

OK. We are going to keep reviewing strategies for finding CDEs about a specific topic. So I'm going to open the Repository back up and go back out to my homepage. Let's say you want to view CDEs that measure pain or have to do with pain in some way. I'm going to search within all CDEs for pain, and when I do that, I get over 1000 results. On the one hand, that's great because it means that we have a lot of options for finding a relevant CDE. On the other hand, that is a lot of results to sort through. So I would like you to take a moment to look at the results page and consider how might you narrow down these results. Let me know in the chat one way that you would narrow these results down and think about those strategies that we just discussed when we looked for physical activity. So what is one way that you might narrow these results down? Let me know what you think in the chat.

I'm already seeing Joan and Heidi recommend the NIH-Endorsed filter. Excellent. Yes, that's a great way to maybe narrow these down. Yes, others are saying by collection or IC (Institute or Center), absolutely. We could limit to a specific collection, especially if it's a group that you work for or with great any other ways that you might narrow these results down. So you've got you were to look through that are related to your topic. Meng says registration status. Yeah, we might try limiting to Standard Registration Status to see if any CDEs match our topic that way. Excellent. All right, these are great responses.

Let's take a minute to review some of the options you provided and a couple of new options. And these are these options are also the answer to #9 on your handout: you can try to limit to NIH-Endorsed CDEs, limit to a specific Collection, or limit to Standard under Registration Type. Another strategy that we haven't discussed yet is to get more specific with your search by adding additional search terms. 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, like 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'll go back out to the Repository and I'm going to add walking to my search and I can add that to the search box at the top of my results page. I've already got pain, so I will just add walking and click Search. This retrieves 27 results, so getting more specific with my search terms really helped to focus my options. And I can tell by the titles and definitions that these look relevant to the subject of pain and walking. And keep in mind that you can use a combination of these strategies to further narrow down your results.

OK, let's do another search. Maybe this time you want to find CDEs about the flu. It's flu season. So I'm going to go back up to the top of my page where I have that search box. I'm going to delete pain walking and I'm going to type flu into the search box and search. There we go. With this search, I get 11 results. Within 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 the search a bit to make sure we're seeing as many potential CDEs as possible for our topic. So one way to broaden our search is to consider any alternative terms, abbreviations, or acronyms that could be used for your search. The Repository does some

of this work for you. It uses what we call elastic search, meaning it will add in a lot of synonyms and alternative terms to your search behind the scenes. But when you want to make sure that you're seeing as many CDEs as possible about your topic, it's a good idea to search with multiple terms. So what is another term for flu that I might search with? Let me know what you think in the chat. What's another term for flu?

Yes, influenza. I'm already getting folks saying that in the chat. Exactly. We might want to also search for the term influenza. So to do that, I can go back up to my search box and I can combine the terms influenza and flu with the word OR. So let me show you, I have flu OR influenza and this will search for influenza in addition to flu. It'll search for both and for CDEs that contain either term. So I'll click Search. I had 11 results before and I have 19 now. So by adding influenza with or increased my search results. And if you're following along on your handout OR is the answer to #10.

OK. I want to point out one more feature that will help you find more CDEs on your topic. I'm going to go to a CDE. I'll go to this first one on our list, Influenza vaccine history indicator to show you how to do this. So I'll click on that CDE and I'm going to scroll down to that Related Content box, you'll recall that we looked at Linked Forms earlier. Beside that tab is a tab that says More like this. And if I click on that, then I see a list of similar CDEs to the one that I'm viewing. So this is a great feature if you find one CDE that's on your topic, but it's maybe not quite what you need, and you want to view other options, then you can click on this more like this tab. And that's the answer to #11 on your handout, More like this.

OK, next we're going to explore searching for forms, which works very similarly to searching for CDEs. If you know the name of a specific form that you're looking for, you can search directly for it. So let's try that now. I'm going to go back out to the Repository home page and above my search box, I'll click on the Search Forms tab to switch to searching for forms. Let's say I want to find the Emotional well-being form from the Functional Assessment of Chronic Illness Therapy System (FACIT). I'll just type the title of that form into the search box. So I'll type in emotional well-being and search. Right away, at the top, I see my first result is the Emotional well being form from FACIT. So that looks like it's probably the form I want. I'm going to click on that to view its record. Normally what we would expect to see here is a form preview. Instead we see this message: "We do not have permission to display this form due to copyright or licensing restrictions. For more information, see the Copyright section below." So if we look at the Copyright section right below the general details it reads, "This form is copyright protected and cannot be displayed. More information can be found at the provided links," with links to FACIT's website and LOINC. So because of the copyright status of this form, it cannot be displayed in the Repository. All forms will have a copyright status with a link to more information if available.

So let's find a form that we can see in the Repository. I'm going to go back out to the Repository homepage and I'll click on that Search Forms tab to switch to searching within forms. This time I'm going to look for a form titled the NEI Visual Function Questionnaire. So I'll type that in to my search box, NEI Visual Function Questionnaire, and I get 2 results. This first one looks like what I'm looking for, so I'm going to click on it. And here's our form. And I mentioned earlier that forms may have instructions for their administration and context for how they should be

used. We can see that at the top here. And as I scroll down this form, you can also see that it's divided into parts like Part 1, Part 2, and so on. And if I scroll below this form, it's a pretty long form, I get to that General Details section and I see the Copyrighted field which reads, "This form is copyrighted but can be used. More information can be found at the provided link."

OK, it is your turn to try looking for a form. I would like you to locate the BRICS Social Determinants of Health form from the NINR (and that's BRICS spelled B-R-I-C-S). And once you've found it, copy and paste its copyright status into the chat, and then we'll take a look at how to do that together.

I see lots of us are finding it. Excellent. I'll give everyone a little more time here to look for it, and then we'll look together. All right, let's take a look at how to find this form and its copyright status. I'm going to open the Repository back up and I'll go back out to the home page. I'll click on the Search Forms tab above the search box and search for the title of my form, BRICS Social Determinants of Health. I get 1 result and I know it's from the NINR because that's its steward listed over here in the Steward field. When I click on the title, there's my form, and if I scroll below the form (I picked a lot of long forms for examples today), I see the copyright section that reads: "No, "this material is freely available without restriction," so I don't need to get any other permissions to use it.

All right, there's one more way to find CDEs and forms in the Repository, which is browsing, so I'm going to go back up to the top of my screen. You'll recall I mentioned earlier that you can use the CDEs and Forms button at the top of any Repository screen to browse. So I'm going to click on forms to take a look at that. And when I do on this page, I see the different collections with available forms that we can browse. I'll click into the NICHD collection to see what forms they have. So now we're seeing the 15 different forms that they have in the Repository, and I want to point out that there are filters for more specific projects within the NICHD that I can click on if I want to see forms from those particular projects, like the NBSTRN Core. So if you're looking for a specific group of forms, you can look there. If you want to browse CDEs, this process is the same except from across the top of the page I'll click on CDEs and I can browse the different collections from here.

Alright, the final feature of the Repository I'm going 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 to store CDEs and forms that you need later or that you want to export. And My Boards is the answer to #12 on your handout. In order to save and access boards, you'll need to be logged into your UTS account. Like I pointed out at the beginning of the training, you can sign up for a UTS account later if you don't already have one.

So I'm going to click into a collection. I'll go to the Project 5 collection to show you how to save CDEs to a board. Let's say I want to save this one about age. I will click on the pin icon beside the CDE name and then I see this little pop up at the bottom telling me that it's been pinned to the My CDEs board. If I want to view my boards or export from them, I'll go back up to the top of the Repository screen and click on My Boards from the menu across the top and you can see that I have two boards, my forms and my CDEs. You can't save forms and CDEs to the same board,

so you'll need to make separate boards for each type. And I'll click into the My CDEs board so you can see what that looks like. Here are all of the CDEs that I've saved, and I can click on this Export Board button to export the CDEs into a variety of formats.

OK, at this point I'm going to ask Mike to put the handout answer key in the chat. Thank you, Mike. So you can review the answers on that handout answer key and also grab any of the links that I've mentioned. And we now have time for some more questions here at the end. So Kate, what other questions have come in that we can answer?

KATE: We've got some content questions, some searching questions, and then a policy question. So let's start with the content questions. **There were a couple of questions about particular CDE collections from organizations and whether they're included in the CDE-R. So Mimi was asking about the NCI caDSR and Catherine was asking about the PDBP. (Wow, there's a lot of acronyms here, or initialisms.) So Robin, can you address that?**

ROBIN: Yeah, I can address those. So, you know, many institutes and centers across NIH are developing their own CDEs because they have their own distinct research purposes. And a lot of times they have, you know, a place where they keep those CDEs on their own platform essentially. I think the PDBP is doing that. caDSR is a little bit different. I'm pretty familiar with it. I know they have NCI CDE's, I think that started to accommodate NCI's own CDEs, but they also do accept CDE submissions from other groups at NIH. And so we're not the same as those and we're not like directly linked to them. But if those groups submit CDEs to us, then they will go through the same review process. And if they are endorsed, they get published in the CDE Repository. And I can tell you right now that we're working on the first few batches of the NCI preferred standards, which is I think the core essential CDEs in caDSR for NCI's own use. That was a lot of letters there, but I think I got that right. The PDBP, I haven't heard anything from that group, but I would encourage them to reach out to us if they have any questions about submitting to NLM and the CDE Governance Committee.

KATE: Great. Thank you. **A couple of folks were interested in how NIH-Endorsed CDEs relate to the FDA CDISC standards. Can you address that, Robin?**

ROBIN: Yeah, I did a quick check because after somebody asked about the FDA, I was thinking about CDISC. And again, I am not an expert on this on CDISC or FDA standards at all. But I do know, I did check this to confirm, that if you are logged into the Repository and you are looking at a form, you can export in various formats, which I mentioned earlier. One of them is REDcap, but there's also an option to output into a CDISC format XML file. So that's pretty much all I can tell you about that right now. It's not a question that's come up very often for us, but if that's something of interest to you, please test it out and let us know if it's something that would meet your needs or if you have any questions about it.

KATE: Great. Let's turn to a couple of searching questions. One question from Qianjin is **do Boolean operators apply in the search box?**

ROBIN: Yes. The short answer is yes. I search all the time and I use AND, OR, and NOT and I think we use the little asterisk (*) for wild card truncation as well.

KATE: Great. All right, **Fred would like to know how the system determines the More like this. Is that an algorithm or human curation?**

ROBIN: It's the search algorithm. It's the same algorithm that is being used when you enter search terms in the search box. It's through elastic search. And so it's just taking terms from the CDE you found and searching for those and finding similar things in that way.

KATE: All right. Alice has the question: **Can only a subset of questions be used from a form?**

ROBIN: Yes, with the caveat that you should make sure it's not indicated as a bundle in the Repository because, as Catherine was explaining, bundles are groups of CDEs that have to be used together as a whole or else it's not considered valid. A good example being like a standardized validated instrument, if you extract any questions, individual questions, from that instrument, they're no longer considered valid. So if you just want to check and make sure it's not a bundle. And I will also say there's another caveat, which is that some older forms have not been necessarily marked as bundles. So if you're unsure, you could maybe look at the copyright information and other details about that form to look learn more.

KATE: Perfect, thanks. And I just wanted to point out that Chris Decker has put a message in chat saying that Chris is from CDISC and would be happy to chat with anyone regarding these this connection. So great. Thank you, Chris.

All right, so Pravina has the question, **how do CDEs relate to OMOP or OHDSI?**

ROBIN: So OMOP and OHDSI (pronounced "Odyssey"), we're talking about data models now and not just and not CDEs. So I feel like the answer to this question is maybe too complicated trying to answer right now. So maybe Pravina, I would urge you, maybe reach out to me and we can maybe discuss it over e-mail or something. That might be a little easier than doing it right now live.

KATE: Thank you. All right, I'm just going to take a moment because we did get a bit of a flurry of questions recently. I want to make sure that we addressed everyone's. And in fact, if we did not address your question, it wouldn't hurt to put it back in chat and make sure that we get to it because we do have time.

CATHERINE: Thank you, Kate. I'll start our wrap-up, but I'll pause again before we end to see if any more questions have come in. So to wrap up, let's review how you can get help in the Repository and find out about updates to the Repository. First, there is an NIH CDE-R News listserv you can join to keep up with the latest updates and features, and a link to that listserv is on your handout. If you're in the Repository and you hover over the Help menu, you'll see four options. The Guides page offers step by step instructions for conducting many actions in the Repository, including those demonstrated today. New Features includes announcements and news about the Repository and the Resources page has links to additional training, information about APIs, and examples of CDEs and biomedical research. And then use the Contact Us page to send your questions directly to the NLM.

So I'll take a look. Kate, are we good on questions? I don't see any more.

KATE: I have not seen any more questions come in.

CATHERINE: OK, great. All right, Well, thank you all so much for joining us today. Thank you, Kate, Mike, and Michael for all your help. And thank you, Robin, for your expertise during our Q&A. We appreciate you all being here and we hope to see you at future NLM events. Take care. Bye bye.