Robin Taylor
Behind the Mask
June 11, 2021

Barr: Good morning. Today is June 11, 2021. My name is Gabrielle Barr, and I am the archivist with the Office of NIH History and Stetten Museum, and today I have the pleasure of speaking with Ms. Robin Taylor. Ms. Taylor is the lead for the NIH Common Data Element Repository at the National Library of Medicine [NLM], and today she is going to speak about her work on COVID-19 common data elements. So we are really excited to speak with you.

Taylor: Thanks. I am really honored to be part of this project.

Barr: Data is very important and helps with research, as we will learn more. What are common data elements, also known as CDEs, and why are they so important?

Taylor: Common data elements, which we will probably just call CDE from now on, are a type of health data standard, to be specific. Data standards are important because they support interoperability and reuse of data. There are many kinds of health data standards, but common data elements, in particular, are defined as a precisely defined question, which might be like birth weight, or what is your race. It could be a question or a field on a data collection form. And the other part of it, the other key part of it, is the set of specific allowable responses. For instance, the answer might have been a number, or it might be a multiple-choice list to choose from, or it might need to be a date or a time. What makes them common is that these data elements get used across different sites, studies, and clinical trials, and that ensures consistent data collection which facilitates that interoperability and reuse, as I mentioned.

Barr: What is your role as the director of the NIH Common Data Elements Repository?

Taylor: I will just say that I do not call myself the director of it. You will usually say that I am the lead. I think the term we use internally at NLM has been Product Owner. There are many aspects of that. I do a lot of planning and thinking about strategy, and I am writing proposals, and getting funding for us, because we have been getting some sort of special funding from NIH OD [Office of the Director], and some COVID-related funds. I also write our FTS and award contracts to support our project, and I am overseeing daily operations including any enhancements, production, quality assurance (or QA as we call it), and user support. I am not doing all that myself, but I am overseeing all of that. I am also setting the priorities, and because I am the lead for the repository, I am sort of the point of contact at NLM and across NIH for the repository. So I participate pretty actively in the NIH CDE task force, which has been hosted by NLM for the last five or six years. I am also an ex-official member of the new NIH CDE government committee that Dr. [Patricia] Brennan convened last year.

Barr: How many people do you work with directly on your team?

Taylor: It varies a lot. Our federal staff team is pretty small. At times, it has been only me and maybe my supervisors, obviously. I now have a more recent hire who has joined the team and has been giving me a lot of support so that is great. She came in through the Pathways [intern] program last year. We had three developers, but now we have four, who work in the ONC, the MLB, the medical language branch, and they are all contractors, but they have been pretty consistently on our team since the beginning. And now we also have a handful of contractors that are helping us with QA and curation, and just getting our processes improved and documented, and that kind of thing.

Barr: Are there representatives from all the institutes on the task force?

Taylor: Yes, in theory every institute is represented on the task force. I say in theory, because I think some institutes do not regularly send representatives to the meetings, but they are on the roster. There is somebody from every IC, and I will say, most ICs do have somebody. It has been growing. For a while, it was sort of a community of practice, where people would share what they were doing with CDE, and just have presentations, and now it is kind of under recent direction. Lisa Federer in the Office of Strategic Initiatives (I think that is where she is) has split this off with Mike Huerta who is the chair. They have decided to split it off and make half the meetings just presentations, and the other half more active working groups where they talk about communications and outreach and other ways to encourage people to use CDE across NIH.

Barr: Has this resource increased since you came to NLM, and what kinds of individuals use this resource?

Taylor: The short answer is yes; the usage has increased since I came on board. When I looked at the number of users over the last few years, it has increased by a little over 50%. Actually, it is a government website so of course, we have limited data about our users, we do not know who they are exactly, but we do know that there are two kinds of users that we envision for the repository. There is what we call the end user, which would be an NIH-funded researcher who needs to use CDE for his project and he would come looking for data from CDE that he could use. And on the other side we have the NIH groups that own and maintain CDE for their projects or their institutes and they are giving us content so they are also visiting the repository. We do not know who these people are and who is coming to the repository in greater numbers now. We just know that CDE is just a growing conversation across NIH, and there is definitely more and more to talk about them.

Barr: Can you define the metadata under a CDE in the NIH Common Data Repository? What makes it qualified or not? What do the value in codetags mean?

Taylor: This is a pretty technical question. We do collect a lot of metadata about CDE, but I am going to try to keep it focused on some key points. As I mentioned earlier, the two most crucial pieces of metadata are the question—which may not actually be a literal question but a field on a form—and the response. The response format has to be specified also, like a list, or a date, or a number. We also encourage people to link their CDE to standard terminology as much as possible because that is going to facilitate interoperability as well.

There are many other pieces of metadata that we collect. A key piece is what we call the steward and that is who is in charge of maintaining that CDE, and this would be an organization, not an individual, so who does it belong to, and who is responsible for it is very important, obviously. You also asked about the registration status so, without going into too much in the definitions of those, they indicate levels of completeness and of quality, and also of broad adoption, like whether or not they are recommended. To meet certain minimal status, they have to just be complete, all the required pieces have to be there, and if they are more fully specified than that, or have more documented validity, then they would get a higher quality status. If they are recommended for use very broadly—a good example of that would be the OMB race and ethnicity CDE, which are recommended for use across many agencies— then that would attain a higher status.

I just want to mention here something that does not yet appear in the repository, but it will soon. I mentioned the NIH CDE government committee that was convened by Dr. Brennan last year. They are charged with reviewing CDEs that are submitted to them to see if they meet certain criteria and those that meet the criteria merit what we are calling endorsement. NIH-endorsed CDE are going to be promoted above all others in the repository. The committee has been working on its processes and we are now starting to review the first couple of batches of CDE. We are focusing on COVID CDE because we want to get those out as quickly as possible so that is something that is not visible in the repository yet, but it will be very soon. It is going to be very important.

Barr: That is exciting. Will you, please, discuss what COVID-19 specific CDE initiatives transpired at NIH and how you and others at NLM and NIH have guided the various discussions?

Taylor: Sure. Last spring, multiple projects sprung up very quickly across NIH to address different aspects of COVID research. I think that the first group I became part of was called Project Five and was about COVID data coordination, and there were many aspects even within Project Five, but there was a specific working group for common data elements and a common data model, and that was a pretty fascinating group to be part of because it had representatives from all across NIH, and it was not just theoretical. The representatives were talking about what their researchers were interested in collecting and what data they wanted to collect for their COVID research. That group has been working on developing a sort of core set of COVID CDE that could be applicable across domains and across NIH. Different programmers will have their own specific focus, but this should be a core that maybe all groups would collect or are recommended to collect.

Since then, I have been participating in some more specialized groups, such as a pediatrics group. There is the PAS group for the Post-Acute Sequelae of COVID, that is a very new group. The RADx initiative which has underneath it multiple programs, but they are all focused on diagnostics. And every one of those groups, they are all funding research, they all have multiple studies going on, they all have data coordination centers where the research data is going to be coordinated and harmonized, and I will say, the NLM's role is coming from Dr. Brennan. She has been making the case for CDE with these programs. I know she serves as chair and co-chair of several of these groups, and she has a vision for what the CDE can do for them. She has been a real champion for common data elements and at the same time, she has been very active in our government committee meetings. She is not afraid to get down in the weeds and go through a spreadsheet with us at the same time. I think NLM has just been trying to champion common data elements and show how if you get into the workflow early, they can benefit downstream.

Barr: Has that been a problem then, being introduced way too late in the research process?

Taylor: Yes. These data coordination centers are trying to harmonize data post hoc—after the fact—and that is difficult and expensive, and sometimes impossible. So of course, if you can get people to use the same data elements from the beginning, it makes that part of their job a lot easier, and a lot more is possible. There is still, I think, a lot of outreach and education to be done around common data elements and to explain why it is important and we are definitely focusing on that.

Barr: Well, that is really great. Who is a part of the COVID CDE committee, and what expertise do each of these members add?

Taylor: NLM again, has stepped up to convene this COVID CDE coordination committee, or as we call it 4C, and that is due to Dr. Brennan and Dr. Lisa Federer's efforts. They have convened this group to connect bi-weekly so everybody can communicate with each other, and I believe that with all the efforts I just mentioned, there are representatives from all these groups on that committee; there are people there that are from those data coordination centers, there are also some people there who are IC directors, there are program directors, program officers, there are data scientists, there are all kinds of people. There is a real variety of research fields which is helpful because we are trying to arrive at some commonality among all these participants who have very different levels of technical and policy experience, which enriches that conversation, and of course, we all have the same shared goal of finding CDE that are useful for the entire NIH community. When the projects get more specific and focused on their own domain, we will be providing guidance for them on how, with their own group of researchers, they can encourage them to use certainties.

Barr: Do you all look at other CDE repositories as a guide and if so, which ones have you consulted for ideas?

Taylor: We definitely do. I mentioned the CDE task force. That has been a great way to learn what other

efforts are going on across NIH. In addition to that, we have done our own periodic environmental scans to see what other groups are doing. Some of the larger and more well-known CDE repositories at NIH are NCI has one called caDSR that has been around for decades; NINDS has a large repository of CDEs that they require for their researchers; NIH CIT has a program called BRICS, where CIT does not have their own researchers, but they are supporting CDE efforts at other ICs (like they support NINR and the certain initiatives in INDs). We, definitely, looked at all those, and representatives from all those. I am working very closely with them right now on the government committee and so we are learning a lot from them. And I would say, in addition to that, when we need additional guidance, we look at standards, like there is the International Standards Organization (ISO) that has a standard for metadata registries which has also been helpful. We all face similar issues and so we can learn a lot from each other.

Barr: Have you ever shared CDE in development before and if not, why did you choose to do that with COVID?

Taylor: That is a really good question. We have always tried to find a balance, and I am speaking of things that even predate my time on the repository, but I know that we have always tried to strike a balance. On the one hand, we have to help people coordinate their CDEs because that is what helps make them common, but on the other hand, NLM's role is to disseminate only the highest quality information and so we have to be careful. However, if each group is developing their own CDE and they are not working together, then the likelihood of them actually being common and standardized is much reduced. Until not long ago, anybody publicly could visit the repository, and if they knew where to change their settings, they could see CDE of all kinds of statuses, and all levels of quality. The default was to show only the highest quality and that is something that we always stood by, but it was possible for them to go in and see things that were basically in draft form. We have made some changes very recently, and now users from outside NIH are only going to see the highest quality on the published CDE. NIH users will have a special setting, and I think this is what you are asking about. They do have it now. We have created a special setting for NIH users so if they log in with their NIH credentials, they can change their settings and they can see what is in draft format. And that is to support what I was just mentioning, coordination between these groups that are covering similar domains, or maybe covering these more universal things, like demographics, and vital signs, things that they can coordinate on. We just have to make it very clear that what they are looking at is not a finished product, but I think we have done a decent job with that.

Barr: What are some of the obstacles that you and others on the COVID CDE committee have encountered?

Taylor: I have mentioned harmonization, already. Arriving at a common set of CDEs for COVID researchers across NIH was really a huge task. Even when it was things that were not real domain specific, or disease specific event, that was a big challenge across ICs because everybody agrees in theory that standards are great, but nobody wants to change what they are doing. Also, the event within the projects themselves, like I know in certain projects, they will tell their researchers that a certain set

of CDEs are required or recommended, and the researchers will say: “Okay, that is great. I am just going to change some of the answers that are in this one.” They do not understand that actually that is changing the CDE, and it is reducing the usefulness of the CDE. I guess, with COVID urgency, researchers were getting started immediately, and we were not getting out in front of them as much as we could, but on the bright side, we have learned a whole lot in this process.

Barr: What has been the most rewarding aspect for you?

Taylor: Getting to work with colleagues all across NIH, which I really like. I love NLM, and I love my NLM colleagues, but I also like getting outside and seeing things from other points of view; there are clinical researchers, data scientists, and people with very different specialties from mine, and I have learned so much from all of them. Also, it felt good to be doing something to support COVID research, even if it was many degrees removed from that. I felt like I was sort of using my powers for good.

Barr: Do you think that COVID has brought a heightened awareness to CDE?

Taylor: Absolutely. Yes. It is interesting because before COVID, we were already talking about raising awareness of CDE and finding some way to encourage using them across NIH, and really getting out there, and evangelizing for them basically, and then, COVID came along, and I will say the silver lining of it is, it made it very clear why we want to use them, and how it could benefit research.

Barr: We are going to transition to you as a person who has been living through the pandemic. What have just been some personal challenges and opportunities that have arisen for you due to COVID-19?

Taylor: I have to say, I just want to be clear that I feel very privileged. I have definitely seen my privilege during this pandemic. I was able to just bring my laptop home and start doing my job from home without any issues. Basically, I was, I would say 99% or 100% able to do my job from home from day one because NLM provided the hardware that I needed and then any kind of support that I needed so that, for me, has not been challenging, not technically anyway.

Barr: It is great!

Taylor: Yes, it is really great, and I really appreciate it. I will say also that I am looking forward to being in the office, maybe not every day, I do not have a very long commute but a little bit long, and so I am looking forward to making more use of the telework in the future now that I have seen how much I can do from home. I think the biggest challenge was the fact that I have two kids who have been in virtual school since last March and working from home while they are also in school from home, that has probably been my biggest challenge. I know it is a very common one.

Barr: Is there something that you have enjoyed that has made the pandemic more manageable for you?

Robin: I was thinking about this, and I was thinking back to last year. I do not know if you were aware of it, but NIH sponsored a fitness challenge. It was called Coast to Coast, and you could sign up and form a team, and it was basically like counting steps, and you could have this friendly competition. In our section, maybe eight or ten people signed up, and we formed two teams to compete, and then, we wound up making our own little chat room for friendly sort of talk, egging each other on, and cheering each other on.

Barr: Did you share pictures of you exercising with each other?

Taylor: I would not say there were so many pictures. I am trying to think, I mean maybe some, but not so much. The friendly competition aspect really got a lot of us very motivated, and just that sense of camaraderie, when it was over, we really missed it, and in fact, we set up our own little challenge, sort of a basic challenge, just to keep it going because it was very motivating, and it helped us feel like a team again even though we were all at home.

Barr: This is one of my last questions. It's a thought-provoking question: How did COVID-19 inform how you could or should approach CDE for other conditions and diseases?

Taylor: That is a good question. I have already mentioned that getting in early is one of the most important things you can do, and I think we already, even before the pandemic, had our work cut out for us in communications, outreach, and promotion, so, getting in early, we had a lot to overcome to get there, but the researchers want to move fast. They just want to get out there and do their research. They do not want to slow down and learn about data standards when there is a pandemic happening, right? So, I think, maybe a more important thing is to have as much input in place as possible, before a crisis hits, and I think that is sort of the flip side of what we are doing now. Even though maybe, some things did not move as quickly as we would have liked during COVID, what we are doing now is laying groundwork for next time, whether it is educating people on the importance of CDE or setting up these communication channels between groups and the NIH CDE government committee is a huge part of that. They are going to be endorsing CDE and people will be able to come to the NIH CDE repository and find that gold standard CDE for whatever their research is quickly. I think that is the biggest part of it.

Barr: Is there anything else that you would like to add that you have not mentioned?

Taylor: No, I think your questions were very thorough, and we talked a lot about the big things going on with CDE at NIH so, thank you.

Barr: Well, I wish you and your team continued success and your family continued safety.

Taylor: All right, thank you very much.