Welcome, everyone. Hi, my name is Kate Majewski. Welcome to the National Library of Medicine or NLM office hours. NLM office hours are an opportunity for people who use NLM products to Interact with NLM staff. We are recording and posting these sessions so that those who aren't available to join us can watch them later. Each office hours session begins with a brief presentation by our product experts, and the remainder of the session will be devoted to answering Your questions. Now you are muted by default, but you will be invited to raise your hand to Speak, or you can enter your questions at any time into chat. Remember to send your chat messages to everyone. We appreciate that.

Now note that this is our regular office hours session, but if you are interested in hearing more about and commenting on MeSH vocabulary related to population groups, we are holding a special listening session at 2:00 p.m. Eastern time today. That's just 90 minutes from now. And there's no restriction on registration, everyone is welcome, and I'll put a link to that registration in chat right now. So, if you're interested in that, you can come join us, please do.

So, today's special NLM office hours are co-hosted by the Massachusetts Health Sciences Library Network or MAHSLIN, so welcome to those participants and a Special welcome to our guest moderator, Elaine Alligood. Elaine recently retired from the Boston VA healthcare system. She was a co-leader of the 2011 new England VA library redesigned task force, where she developed a network-wide virtual knowledge library That won the 2015 federal library of the year award. Elaine was the manager for North America in the 80s, she was at Countway medical library for five years and she joined VA Boston in 1995, where she spent her first 16 years as a clinical informationist. As one of the co-developers of the MLA CE program on diagnostic care -- Elaine believes that medical librarians transforming roles perfectly align us with the clinical and research staff ever-changing information needs. And that adjacency, moving librarians out of the library to where Information needs to occur at the point of care in research benefits, everyone. Clinicians, researchers, and patients. Welcome, Elaine.

>> thank you. So, one of the things that was a big surprise to me is when I took the job at the VA, I'd been at the VA in Baltimore way earlier just for a couple of years, was this whole new thing of systematic review writing. In those days, we called it healthcare technology assessment. One of the things I loved about it was all about searching for me, and really that's the heart of what we do as medical librarians. We know where to find it when no one else on the clinical team does, and I Love a good search challenge. I'm sure you all do too. And what's better than developing a search strategy that you believe, Hope, and pray will capture everything you need, and reveal other things about the topic that you didn't know, that you can then follow up on. It's a really iterative process. So, today's session is really all about MeSH, and our guru on MeSH, Dan Cho, is going to talk about how it's developed and what's going on with it Right now, and then the rest of the presentation is really all about you and asking questions and getting guidance or, you know, whatever you Need to know about MeSH. It has changed over the years. Back when I learned MeSH in the 70s, it was supposedly limited to no more than 16,000 terms but we busted through that marker a long time ago.

>> long time ago!

>> so, I want to say just a few words about Dan. Dan is a senior MeSH analyst, MeSH content lead. He's project coordinator and point of contact for MeSH on demand. Dan has a Ph.D. in genetics from the lab of Charles livings at Carolina State university, he was a postdoc and research faculty member at Penn med Working on the single-cell microarray projects related to human mental Disorders. Sounds utterly fascinating to me. He has a special interest in projects that help MeSH users with systematic Reviews and topical domains that are underrepresented in MeSH such as social determinants of health. So, there's always some new and hot topic to construct MeSH terms around and that's not ever going to change, I suspect. So today, Dan is going to talk about the vocabulary, how it's developed, and after it, it's all open for questions and discussion from the 88 Participants that we're lucky to have. So, I leave for Dan right now.

>> thank you for the lovely introduction. I was actually listening to what I used to do, and I can't even fathom that -- Single cell microarray, I don't know if I could do that again. My name is Dan Cho, I am the content lead and project coordinator in MeSH. Today we're going to share my desktop for the PowerPoint.

>> yes, if you'd go ahead and share your -- yes.

>> I will do that. Okay. S so meanwhile, I'll share Screen. So, can you see it?

>> we see your PowerPoint but you're not in the presentation view. There we go! Thanks.

>> so, do you see the presentation mode?

>> yes.

>> okay, great. Thank you. Appreciate it. So, thanks for the introduction. Today what I'm going to do is I'm going to talk about MeSH and how we handle Our user request, and then really talk about in the future how we're going to Do what we do, but for the better, we've switched to machine indexing and That has ramifications on how MeSH is going to develop. So first the history. MeSH is a control of vocabulary we are using here at NLM, and our bottom line Is for us to help others to find biomedical literatures and serials and Whatnot, and we have been doing this for 60-plus years and we've done it so Good that most of actually our users don't even know that MeSH is behind the scenes helping you getting your product, and most typically in the PubMed search.

So, MeSH or medical subject heading is part of a control vocabulary. Medical vocabulary has a long history. Our first edition was in 1960. When it was introduced, it actually revolutionized a couple control Vocabulary aspects. First it was the first time that a single authoritative source, MeSH, was Used for both cataloging and indexing, and it was also the first large scale Experiment on using the combination of subheadings with descriptors so we can Limit the number of MeSH terms we can actually have. But it is actually in 1963 when we first printed our second edition. We oftentimes here in MeSH consider that to be the birth year of MeSH. We've

gone through many iterations of making it better and helping the Users, and one of the biggest changes that we probably made was in 2001, When we went from term Centric to concept Centric, so in a single MeSH Record, we actually have a lot of different -- potentially have a lot of Terms, synonyms, as well as concepts. And these concepts are so closely related, we felt like that it would be Helpful for our users to have them together. So, for example, if we introduce a new descriptor that says food security, then we'll have a later one that says it's food insecurity because lack of Food security would be food insecurity. Another good example of that is alien hand syndrome. So, a lot of times a pathological entity would be depending on the penetration and where it actually appears, so alien hand syndrome is grouped together with Alien leg syndrome, and even broader concept which is alien limb syndrome.

We're thinking about -- one of the terms we're actually thinking about is Data Bias, and I'm pretty sure we'll grouped together with that other Concept that's data fairness. So those are the kinds of things that we actually think about when we have Switched to concept Centric format. Since approximately 2015, we've actually went out and systemically Asked our users and stakeholders outside of NLM, how do you use our MeSH data and what kind of product do you build around it? And we've learned quite a bit, and truly learning about how we actually Use our MeSH data and structure outside of NLM is going to also tell us a lot About what kind of content we need to actually develop here in MeSH in the Future. I'm going to talk about components very briefly and the reason that I'm Actually doing it is because they have different lifecycle. We refresh them at a different rate. So, as you all know, we have descriptors, these are like what is the article about, right? Covid-19 is a good example. And then we have a supplementary concept record or what we call SCR. You like to sort of remember it as a junior descriptor. Then of course we have subheadings. Which limit and further define what the topic is all about.

As I said earlier, the lifecycle refreshment -- descriptor and Subheading is updated on a yearly basis, so today, 3:00 in the Afternoon, we're going to have our last batch review meetings, and then when All said and done during the summer, we're going to make it perfect and We're going to release it come November. So that's how we do it. Once a year in November is when we release them. SCR, supplementary concept, is being adjusted and created and updated and Deleted on a daily basis, and we release it on a daily basis, nightly Basis from Monday through Thursday. And this is important to remember because if you send us a request and then you don't see a change, don't get frustrated because we might be Actually considering to be as part of the descriptor then, which means it's on an annual recycle base. Whenever we make changes, we release it free of charge to our users, which you Can download it and NLM data distribution program releases in Multiple formats, the latest of which is MeSH RDF, it's become a darling in the data science community. And first off, descriptors, we have about 30,000 of these guys.

About 29,000 of them is topical. We have also type 2, which is the publication type, so these would be Your, what, abstract, clinical trial papers and whatnot. We have check tags, officially we have two of them, male and female, but unofficially we have many others. We also have geographic terms. Everything from Africa to Antarctica and all the countries in between, and

Including territories, Palestinian territory has made MeSH because U.S. State department recognized it for the first time. So, we include it. We have all the countries that you can imagine in MeSH descriptors. SCR or supplementary concept record, we have more than 300,000 of these. Of course, most familiar is chemicals. We have 240,000 of these. But we also have protocols.

These are oncological protocols that hospitals use a lot, and we have them. We also have rare diseases, about 10 years ago now, we worked with obssr -- I'm sorry, rcdc and obssr and rare disease communities to introduce all The rare diseases into MeSH as SCR, and as you can imagine, these are rare Diseases so you're not going to have like a lot of articles about it, but They clamor for it and we have it, and from what I understand, we have more Rare diseases than any other controlled vocabulary. About two years ago, we further extended our SCR reach to include Organisms. So, we work with the NCBI taxonomy, which has emerged as the authoritative source of the taxonomical terminology. We introduced 65,000 of hand-picked these because they appear in PubMed Which is very important for MeSH, so they are either medicinal plants or Causes disease in humans. We like SCR because they're easy to make.

We can update it on a daily basis, but it also retains all the important MeSH Fields for our searchers to effectively use to search, let's say, PubMed, such as synonyms, entry terms, all the related terms, and it's also treed in A way because it's heading map to descriptors, and we can also use the SCR to find out how often these SCRs are being used to index. So, for example, this SCR called dostarlimab was in the news quite a Bit these days lately, but if you search in PubMed, you can quickly find Out that it's been around for a while and in the PubMed, there are about 19 Articles that are about it. About the monoclonal antibody. And you can actually see the latest one that everybody heard about is about These particular articles.

So, SCR is really important, we love it, but we also have subheadings. So instead of just saying Covid-19, we can talk about the history of covid-19 Using history subheadings, transmission of Covid-19 is another good future if not already search term. Now the most important part. Users and stakeholders. Indexers, catalogers, HMD, UMLS and whatnot users, PubMed, of course, is -- uses MeSH indexed with NIH CDE repository, clinical trials, MedlinePlus, pub chem, they all use it. But outside of NLM, who are our stakeholders? We've gone and systematically asked who they are and then what they use it For. And here's just a good sampling. You know, it includes Wikipedia, wikidata, hl7, many NIH offices, U.S. Government tall agencies, EPA. Many publishing companies use our data and further refine to their fee-based Customers. And when we did -- in 2020, and after five, four years of releasing MeSH Rdf, we actually did a survey with our MeSH RDF users, and then 25% of them Says that not only do they actually use MeSH data, but they also make Applications, tools and systems around MeSH and using MeSH and PubMed data to Cater to their customers. And that's something that you need to keep in mind because whatever change That we actually make, all these people, you know, 164 products Outside of NLM need to retool it and just have it ready, right? So, it's important to keep in mind, for me to keep in mind, at least.

In the survey, we also heard that we actually need to do better training and better documentation, and then we also need to develop more content. Because we're missing a lot of it. And that's true. Here's what we do when we receive a request from you. So NLM follows a consistent and thoughtful and deliberate process to Informed by stakeholders. Typically, on an annual basis, we receive about 700 requests, which is Typically divided into seven review cycles. And these processes always include your request, and then we go much further Than that. So, a typical request will require, depending on the scope, a day or two to just gather all the information to propose changes. When we do need outside help, like we will primarily rely on PubMed, but Whether we do need outside help, we reach out to subject matter experts Across nih, we're lucky, most areas, we can find world renowned scientists Within the NIH community. But we also reach out to other federal agencies, particularly about Education, psychology, whatever, it's slightly outside of the biomedical Domain, and we'll reach out to other federal agencies who specialize on those, and we at times will actually ask outside of NLM and outside of federal Government for help as well. Most of the time, though, we rely on PubMed and then -- since that's our Number one customer in a way, we always look for literary work. So, if somebody says, hey, can you add, I don't know, almond milk, and then -- On the milk, we first look for how many times almond milk appears in PubMed, and then we look at the FDA library to see, is almond milk considered to be a Type of milk? Those are the kind of questions that we do. But whatever the decisions that we make, we make sure that it's a Consensus-based and not skewed on somebody's -- including the users, right? I mean, you may think almond milk is a type of milk, but it may not be, depending on what our sources are, right? But whatever the changes are consistency based, we always announce It either through the MeSH webpage or through NLM technical book.

When we receive a request from you to change or ask something, here are the Three questions that we always ask. So, what do you really mean, right? So, if the scope is not within biomedical library content, then we'll Refuse it outright. So, if you request us to add broccoli, we'll probably say no thank you. Go to USDA library. But if you say, you know, could you just add broccoli because broccoli Contains a component that's medicinally important, then we consider it and look for postings and how it's been indexed and all that. You'll be surprised that a lot of your requests already are in MeSH, because They're in SCR or entry term. And then if it is not truly in MeSH already, then we need to find out what the posting is, right? Literary warrant trumps everything else. If it's not there in PubMed, we'll just simply not make it. Or we'll make it as SCR, if it's a disease, protocol, organism or Chemical, right? If it's other than those four topical areas, then unfortunately we cannot Make it.

This is a really busy slide, I'm not really quite sure if I want to go into Much more details than to emphasize that we really, really take your requests seriously and then go through various systematic review of what it is exactly that You're requesting, whether or not it should be added. But the most important thing that we actually keep in mind going through All these different types of process is, we always ask, am I going to help You find more articles related to the topic that you just requested. And most of that decision is based on primary source. So, we will do PubMed search day and night to see, like what the definition Is, and what the synonyms are. UMLS helps quite a bit.

We actually grab all the synonyms related to the one request that you Have made, and then we search all of those terms to see what the posting is. We rely on a lot of other governmental sources, so other NIH IC terms, we Rely on them, NCBI Taxonomy, we rely on that a lot. But if -- let's say if we truly don't have the term or the concept that you Requested in MeSH, then we go through step three, and this is where we Determine, hey, does this appear more than 100 times in MEDLINE-indexed Articles of title? So that's the question that I ask. Whatever analyst comes back to me and says, hey, should we make it, this is What I said. Look for it in title. And see if it appears more than 100 times. If it does, most likely we need it. If it doesn't, and you cannot make an SCR, then, you know, that's what it is. So, we also ask, is this a fad? A lot of dietary diet is a fad. And first of all, it's not going to have high enough postings to make us Make the particular descriptors, but also look at it, whether or not it's A fad or it's a legitimate increase in number of interest articles.

Another way we actually add terminology into MeSH is we systemically go out and search for, you know, concepts and terms that are missing in MeSH. And we do that oftentimes by harmonizing data and then looking for Underserved domains. So here are some of the recent projects. I already talked about ORDR or rare disease. Medicinal plants from NCBI Taxonomy. We worked a couple of years ago, two years now, SDOH, worked with the NIMHD to Identify all the social determinants of health. Talking about EPA, we've worked with the EPA with identifying all the PFAS, Or -- these are the forever chemicals that are highly fluorinated. We've identified all of them, and we feel very comfortable that we have Almost all of if not all that appear in publications. In MeSH. And we're currently working on ethnic minorities. So, we got a master list from various governmental sources supplemented with Other primary sources, and we think we're working with approximately 1200 Ethnic minorities around the world, but mostly in the U.S., and then we have All the American native tribal names sanctioned by the U.S. Federal Government being reviewed by our sibs.

So, here's the new trends, I think as we move toward machine indexing and when we switch to machine Indexing, we kind of realized that you know, QA is going to continue to be Done so that the quality of MEDLINE indexing is going to be there. But now we don't have to worry about overwhelming our indexes with too many Terms. But the machines don't complain as much. So we are going to make sure that MeSH is going to be ai-ready, it's going To be all the fields are going to be machine-readable, and we're going to Simplify our MeSH, getting rid of some of the subheading combinations, so for Example, last two years, we made elbow injury, shoulder injury, knee injuries, and we only did that because, in a previously indexing was not all over, it was perfect. But we were missing some of the very specific eponym shoulder injury being Indexed properly that way.

So, by making it a descriptor on its own, we were able to add a lot of Eponym and synonyms into these terms. We're including a lot of natural language-friendly entry terms, heart Attack, bowel movement, and whatever the nurses say, and the hospitals say, we try to add them. We continue to interact with other international users because we think This is important. We are a U.S. government agency, obviously, but it is good to have a Broad perspective for our

international users to ward off really Machine-indexed issues and make sure that we are being -- our data is being fair. Internationally. We're adopting -- at times we're adopting, I'm sorry, some of the Formats, such as MeSH RDF, which was a big hit thanks to a lot of work that people have done in our MeSH RDF, and this is, you know, making a lot of People in data science and other systems who already had their systems Set up to leverage into MeSH and PubMed, and they absolutely love it. We also are integrating and harmonizing a lot of other databases, and the Bottom line is, we're trying to increase the coverage and close the gap, and this is really important to keep in mind because as we go toward Machine indexing, we need to be cognizant of the fact that we need to Include prototypes as well as a typical form. So, we need to include all the minorities into our vocabulary to make sure that we're not having an issue with machine bias. That's really all I have for now. So, if you have any questions.

>> thanks so much, Dan. That was a great overview of the MeSH process, we greatly appreciate that. We have questions. We have a number of questions in chat so what I'm going to do is I'm going to read them to you. They span the gamut. So, I actually -- I'm going to come back to Christine's first question in just A moment because the second question actually relates to the process of Submitting suggestions. You covered that a little bit, but I did want to point out that you have the MeSH has a guidance document up on the web that helps walk you through what You should be submit when you're making MeSH suggestions. So, I thought I'd put that link into chat and then, Dan, if you want to say Anything more about what information you would like when people make Suggestions to MeSH.

>> I think the most important thing is, tell us exactly what you want done. And tell us your contact information. I don't know why people don't want to give the contact information because oftentimes what's so obvious to you is not so obvious to us. So, when you request something, please put your contact information so that if we do have questions, we'll come back to you. I think there's no such thing as bad requests. If we don't make descriptors or SCR this year, we keep those requests Forever. We have databases to keep those. And then we go back to them at times. If you do put your contact information, you'll get a letter, nowadays an Email, you used to get mail, but you get an email reply as to what happened to your request. We'll say, hey, I know you requested it, it's been almost a year, we're Sorry we're very late, but we made it, you made it. And so, you'll get a reply from us.

>> Elaine, I think you're talking about you're muted.

>> thank you. So, the second question was from marina to everyone. What is the protocol for dealing with automated indexing having difficulty with indexing a work, what ai recognizes that -- when ai recognizes It cannot index a work properly in the first place? And then what action do you take on your end? So, she gave an example, and it's the pmid number 358-64-1623. Shall I repeat it? 35641423. 35641623.
>> right.

>> marina has a sense of humor.

>> right. I mean, I don't even have a -- I'm sorry, I don't even have a full text on this, so you know, this is -- I'm sorry, there's a full-text link. The bottom line is, that the more terminology covers all the Disciplines, the better it's going to be. As far as working into -- I see where this is going.

>> can I jump in here because I do want to -- just to explain, we don't have indexing staff with us today, and your question really is about Indexing. But I can give a general answer which I think is perfectly correct, which is If we don't have enough data to provide indexing, it would be flagged, and our Indexers would review it. This particular citation, I believe, is a letter so I don't think it gets Indexed, but just in general, we have Q&A processes so that if there are issues with a particular record, we're going to review it. So, I think that's the answer to your question.

Could we jump back? Because I only skipped Christine's question momentarily, I really meant to go back to it. This is for Dan. The question is, how do you search to see how many times a word or phrase Appears in PubMed since PubMed is phrase-indexed? So, she's talking about the limitations of PubMed in terms of phrase indexing. As we all know, PubMed doesn't index absolutely every phrase. PubMed phrase index is based on a noun analyzer, and so there's a minimum number of times that it needs to appear for to be indexed for PubMed. So, I guess the question is to you, Dan, what's your procedure for when you're Searching PubMed to see how often it is -- a particular word or phrase exists in the literature?

>> I think we're stuck just as the user is. I mean, phrase -- if you put more than three words the number of times that You're going to hit, especially if you put it in parentheses. So that's typically what we do. We put the two words, combination of two words in parenthesis and then we Search in the title aspect. Then we go in the UMLS browser and find all the relevant, what we think is Relevant synonyms, and we'll search in addition to what you have just Requested. So, in general, that's what we do, but searching phrase in quotations in PubMed is problematic. So, anything more than three words in phrase is going to -- unless you put It in quotation, it's going to pick up a lot of things that you may not want.

>> thank you, Dan. I think the next question was from Courtney. How many people are part of the decision process for a new MeSH term Request?

>> I'm sorry, so how many people are within MeSH division I guess is the Question?

>> I think she means when a new request comes in, how many sets of eyes are Looking at it?

>> well, so MeSH staff, we have five full time members, so that's how many People are actually handling it. We have 15 member reviewers associated with it. If it's a part of a project, then we have an unlimited number of people who Wants to volunteer within nih. So let me give you an example. So, I'm actually working on social terms, and we have monthly meetings on This, and there are typically about 30 people within NIH who's talking. And then I will bring that in as a liaison, and then within MeSH, there's Five of us who's reviewing my work which I brought in from NIH community. And then when I present to external review, there's 15 members outside

of MeSH who is also reviewing it. So, by the time it actually is done, the entire process, a lot of sets of eyes Have seen it and reviewed it.

>> so, Tracy shields has a question. What determine when is a supplemental concept gets moved up to a MeSH term?

>> so, we have a very systematic way of doing it. Any time it actually is above 200 times, in what we call frequency, so 20 times an article is indexed with it, then it flagged and then it should be Promoted into descriptors. Except on Certain things, we just cannot do it, but in general, that's What we do. 200-plus SCR should be promoted to descriptors.

>> okay. Fair enough. So, another question from Carol Mita is, how is it determined whether a phrase Is included in the phrase index? Is it the three-word or -- that you mentioned just a minute ago, or Something else?

>> I mean, we do a lot of searches in the PubMed, so we put into quotation The phrase that you're looking for. We put in quotation and then we search either the title or title of the Abstract. That's really what we do. And then we look for any permutations of the phrase that you have in mind, and then we look for other synonyms and antonyms and quasi synonyms and what Not to try to build up our phrases that we want to search for. That's really the process that we go through, which I think should be part of any systematic searcher's process.

>> and then the Galter library wants to know are there any plans for making the review process more transparent? Like for example, sharing the proposed terms that are not yet accepted?

>> you know, we talk about this quite a bit within ourselves. And I think there are pros and cons on this, but at the moment, what happens, once we receive it, once NLM receives it, you will get a quick turnaround, say yes, we received it, thank you very much, you know, it's been given to MeSH, they will work on it. Once we receive it, then what I do is, I assign it to an analyst, and the Analysts have a Certain number of weeks that he or she can actually work on It, and if it doesn't get worked on it, then she needs to send out a letter During the summer. So, within a year, you should receive something. Now, should we tell outsiders about all the processes? I think there are pros and cons to that. At the moment, we're not doing that. But if you sent back to us and said, hey, I sent you a request, you know, three months ago, what happened to it, we'll happily reply to you. But we get a lot of requests, to be honest. 700. If we keep track of the 700 and let every one of you guys know, that's not the process that we're going through at the moment.

>> okay. And then another question from Diana Lowdone to everyone, one of the most Useful subjects -- capturing concepts that can be expressed in many different ways, like integrated healthcare. Will ai algorithms look at more than the MeSH entry terms or title word Terms to assign MeSH terms? Will the algorithms make use of search filters that have been developed by an outside expert, for example?

>> I certainly don't have an answer for it.

>> that's a great question. That's not a question for MeSH, but it is a question for our PubMed staff, which makes me think, gosh, many of you might really enjoy our next office Hours with the PubMed staff. So, we'll announce those probably in the next month or so, we're going to have Another office hours with PubMed. A lot of the MeSH questions are over the PubMed questions, so I think we'll Have lots more to talk about. But right now, I believe we are pretty much out of time for this session. I'll remind you that we're having another session at 2:00 p.m. Eastern Time, that's just 45 minutes from now, if you want to talk more about MeSH, you're welcome to join us. It's open to the public, just let me get the registration link back up in Chat. And meanwhile, I hope I'm sharing on -- in my Screen hopefully a few links to Learn more about MeSH, and I'm going to ask Brittney -- oh, thank you, Brittney. She's already got it. To put those links. She's got them in chat, so to learn more about MeSH, please explore the MeSH home page. We have a number of training materials on MeSH. Also, the FAQ, and of course if you don't already subscribe to the NLM Technical Bulletin, please do so, so you can find out what's going on at NLM Specifically with our products and services. You can find out what's going on through the NLM technical bulletin. So, if you have additional questions, if we didn't get to your question today, please use the link to the NLM Support Center on the NLM webpage to get assistance. And when you leave the session today, you will get a pop-up for an Evaluation form. Please fill out our evaluation form. NLM office hours is still a new kind of thing. Please tell us what we could be doing better. Also fill out that evaluation if you want CE credit, you can follow the instructions on the screen and you'll get your MLA CE. Thank you all for coming. Thank you to Elaine and to Dan and to everyone. Hopefully, we will see you all soon. Maybe we'll see you at 2:00. Have a wonderful rest of your day.

>> thank you all for coming.