

## Transcript

**MIKE:** Thank you all for joining us for another NLM Office Hours. As I mentioned before we started during our sound checks, my name is Mike Davidson from the Office of Engagement and Training here at the National Library of Medicine. My pronouns are he/him/his. The goal behind these Office Hour sessions is to give you a chance to learn more about NLM's products and to get your questions answered by our trainers and members of our product teams.

Today's focus is on PubMed and we've got a great roster of folks with us who know a lot about PubMed and all different aspects of PubMed. We're going to kick things off with a brief presentation from Amanda Sawyer, who's part of the PubMed team for NLM's National Center for Biotechnology Information, or NCBI. Amanda's team are the folks who develop and maintain PubMed, and she'll be sharing some updates about recent improvements to PubMed, plus a little bit about what you might expect to see from PubMed in the near future.

We're then going to use the rest of the session to have our panelists answer your questions. And in addition to Amanda, we have a great panel of NLM experts, including Alex Sticco and Melanie Huston from NLM Index Section, Dan Cho from the NLM MeSH team, Jessica Chan of the NCBI PubMed team, and Kate Majewski, my colleague from the Office of Engagement and Training. And we also have Michael Tahmasian here helping us out with the tech side of things. Between all of these incredibly knowledgeable folks, we should hopefully be able to answer any PubMed questions that you might have.

A few quick logistical notes before we get started properly. We are recording today's session to share with those who are unable to attend. The recording will be posted shortly following the office hours and everyone who registered should receive a link to that recording. We have muted all attendees for today's session, but you can put questions in chat at any time. Please make sure you send those questions to everyone so that we can make sure our whole panel can see them and answer them. Over the course of the presentation and throughout the Q&A period, I'll be taking note of the questions that come in and then we'll try to direct them to the right expert. But before we get to the questions, I'm going to hand things over to Amanda Sawyer to bring us up to speed on what's new with PubMed. Amanda, take it away.

**AMANDA:** Thanks, Mike. I am Amanda Sawyer. My pronouns are she/her/hers and I'm very happy to be here today with you. We're going to talk a bit about how PubMed has grown over the last year and then hopefully everybody will walk away with something new. We do have some updates from just the last two months.

So like I said, we're going to talk about PubMed's growth over the past year. First, I'm going to give you the same update that was given at the Medical Library Association Conference in May, plus I'll be mentioning those two updates that we've made in these last few months and as always, I will close my presentation with some information about training and support resources and we'll be leaving probably most of the time today for your questions at the end.

We always start with PubMed by the numbers, PubMed now includes more than 36 million citations, over 2 million of which link directly to the full text. Nearly 2 million citations were

added to PubMed in the last year since MLA of May 2022. PubMed's user base is growing too. Our site averages over 3.5 million users on a typical weekday who are coming from all over the world, from every continent, including Antarctica. Those users are conducting around 5.5 million searches every day in the PubMed web interface alone. That doesn't even account for all of the searches done in the API.

So now let's take a look at some of the new PubMed features and improvements that we implemented over the last year. In November 2022, we released the updated PubMed E-Utilities API. The API now uses the same technology as the PubMed website, keeping the results from the website and the API consistent and in sync. This update was of course good news for our E-Utilities users, but also for other products and services that use E-Utilities API to access PubMed data like DOCLINE. The API now has the most up to date information instead of lagging behind the website. Since this update that was made last November, you should no longer run into the issue where PMID was available in PubMed but it was not available in DOCLINE yet. Following this release, we also updated the PubMed User Guide with more information for users who are navigating large result sets with more than 10,000 citations, whether they're accessing those citations in the web interface or through the API. And to get the latest news and announcements about public data, E-Utilities API, and year end processing, please subscribe to the E-Utilities mailing list which I have linked here, and these slides will be shared when the recording is posted as well.

Our next update was to the journal translation table. Last summer, we updated this translation table that is used by automatic term mapping to make journal searching more flexible. The PubMed Journal Translation Table was updated to include journal names without initial articles, such as "the," "a," "and," etcetera. This means that such journal titles can now be searched with or without the initial articles and they will still map to the appropriate journal. So the example that I have here on the screen is for the New England Journal of Medicine. Previously you would have had to include that leading "the" in front of New England Journal of Medicine, but now you can search either New England Journal of Medicine with the journal tag or The New England Journal of Medicine with the journal tag and get the same results. And if you want more information about this update, I really recommend checking out the technical bulletin which is here on the slide, and I think someone's going to put that link in the chat as well. There's a lot of information and some other examples to make it clear exactly what we did here.

Our next update was to the Collections button. Last fall again, the Favorites button on the Abstract page was updated to use the same functionality that the Send to Collections button that you see elsewhere in PubMed uses. We made this change from a Favorites button to a Collections button so that we could provide the same functionality and user experience in both PubMed and PubMed Central. So now when you're on the Abstract page of a citation you'd like to add to a Collection, you can click that Collections button and a menu will pop up. And your default is Favorites, so if that is your workflow, you can click Add and continue on. But you can also create a new Collection from this menu or choose from the drop-down menu from one of your other Collections that you may have created in the past. And I'll just step back for one second to also mention that the little flag icon you see next to Collections is empty here, but once

you've added a citation to your Collection, any of your Collections, when you come back that icon will then be filled in so you can know if you've already added something to one of your Collections.

Earlier this year, we began looking at cleaning up the additional filters interface to ensure a consistent user experience. We started by renaming the Journal filter category to Other and then the MEDLINE filter was moved to this new category. This filter category also includes a new Exclude Preprints filter that was created in response to feedback received on the NIH Preprint Pilot. Lastly, we also updated the underlying structure and tagging of the Filters menu to make it easier for people who are using a screen reader to navigate this part of the website. Our next feature improvement was made thanks to feedback we received from PubMed users who are receiving citations from their colleagues and needed to know who was sending them the e-mail. Users now have the option to include both a To: e-mail and a From: e-mail. The newly added From: field is optional, so if you're emailing citations to yourself, this will not add any extra steps to your workflow. If you're signed into your My NCBI account, both the To: and From: emails will be populated with the e-mail that's associated with your account, and then you can change those to whomever you need to direct those citations to. When you receive the e-mail in your inbox, or the person you're sending it to receives the e-mail in your inbox, the sender still appears as NLM NCBI nobody. The emails are coming from NCBI that hasn't changed, it's information from that from field will then be included in the text section at the top of your message if it was supplied by the sender. If they didn't put anything in that field, you just won't see it in the e-mail.

Another highly requested update was made in March of this year to streamline the display of link the author lists and search results. When viewing search results in the Summary Display format, author lists are now truncated after 1200 characters, followed by an ellipses and a link to See abstract for full author list. In practice, this change only applies to citations with approximately 100 or more authors, or around .01% of the citations in the PubMed database. Though the change only applies to a small percentage of citations, we're really hopeful that this is an improvement that will make a big difference for users who previously were having to scroll through extremely lengthy citations in their search results.

Next up, our Phrase Index Warning. As you may know, PubMed maintains a phrase index in order to support phrase searching. The phrase index is currently the most efficient way that we are able to provide free searching while still maintaining system speed and performance for all three and a half million daily users that we were talking about earlier. In the past year, we updated the warning, that yellow box you see at the top here. This is the warning message that appears when you search for a phrase that's not found in PubMed's phrase index. We also added a link to that warning that will take you directly to the PubMed User Guide section with more information about what you can do when your phrase is not found, including instructions to write to the Help Desk to request that new phrases be added.

That said, we think our biggest and most exciting update from the past year is the addition of a powerful new search tool, which also offers a way to search for phrases that are not found in the phrase index. In November of 2022, we were excited to announce that proximity search is now

available in PubMed. You can search for multiple terms appearing in any order within a specified distance of one another in the title, title/abstract, or affiliation fields. There are three main components to crafting your proximity search in PubMed. First, you'll need your search terms and closed in double quotes. Remember that you must use two or more terms because proximity searches are based on the distance between those terms. Secondly, you'll need the field you want to search in PubMed. Remember that proximity searching is only available in the title, title/abstract, and affiliation fields. You can use the spelled out field tag or the field tag abbreviation, which you can find in the PubMed User Guide. Finally, you'll choose your N value. N is the maximum number of words you want to appear between your search terms. A higher N value will give you broader results, and a smaller end value can help you narrow down your results.

So let's take a look at a quick example by returning to that screenshot from the previous slide. To search for citations addressing treatment options for chronic migraine in the title field, we might consider a search strategy like the one you see here on this slide. Here we have a search for the terms treatment, chronic and migraine and the title field with a maximum of four words appearing between them. This search returns results that include phrases like treatment options in chronic migraine, treatment of resistant chronic migraine, chronic migraine pathophysiology and treatment, chronic migraine long term treatment, and more.

When it was first introduced last November, proximity was only available in the title and title/abstract fields. But just last week we released an update to expand proximity searching to the affiliation field in addition to the title and title/abstract fields. This update is based on feedback we received in our last PubMed Office Hour session earlier this year. With this update, you can now use the proximity search to look for multiple terms appearing in the same affiliation without requiring an exact phrase match, since the same institution can often be represented in a lot of different ways, and affiliation data using a proximity search can be very helpful. For proximity searches in the affiliation field, an N value of 1000 or less will search for the double quoted terms together within the same affiliation rather than spread across any affiliations on the record. If you have questions about proximity searching in PubMed, please see the PubMed User Guide which has been updated with details about proximity searching. The training team also produced a quick interactive tutorial to help introduce proximity searching, and you can find even more examples and FAQs about PubMed's proximity searching in the Technical Bulletin announcements from last November, as well as the one from last week. We would also love it if you would share these resources with your colleagues and help us get the word out about this new search feature in PubMed.

Our second update we made this summer was made last month. This update was made to make it easier and faster to change the sort order of your results. The sort drop-down menu used to change your sort order has been moved out of the Display Options menu and now appears as a standalone feature at the top of the search results page.

So what's coming up next in PubMed development? We continue to evaluate new updates based on user research, including the feedback we received from you, as well as usability testing. Updates to the PubMed website include regular ongoing maintenance work. This can mean bug

fixes, some of which we learn about from folks like you. So thank you for writing to the Help Desk when you notice something acting strangely in PubMed. A lot of our maintenance work is behind the scenes and it's not something that most people will notice when they're visiting the website, but it's essential to keeping Pub Med running smoothly. These updates can include things like security updates and upgrading software to the latest versions. For example, in the future you may notice some minor changes coming to the styling of alert banners in PubMed. We are currently working to standardize these banners to ensure consistent appearance and to follow the most recent version of the United States Web Design System, USWDS, which is the standard for federal government websites. Additionally, in the coming months we intend to take a closer look at the Filter menu interface on the search results page. This review is prompted by feedback we've received via the Help Desk, as well as comments collected at MLA in May.

We have a thorough process for evaluating comments, testing potential new designs, including hands on testing with volunteers, and communicating and releasing any major updates to the web interface. This process starts with collecting user feedback about a specific topic or feature, so please write to us at the Help Desk with any feedback you have about the Filters menu in PubMed. Letting us know how you use a feature, what you like about it, and if there are any places where you get stuck or you wish something was different helps us a lot when we are looking at making changes. Updates to major features like filters take time and we want to make sure we do our best to get things right. It may be some time before we have further news to share on this topic. As always, we will communicate any changes to PubMed through the Technical Bulletin as well as through the New and Noteworthy PubMed RSS feed.

And speaking of those resources, we encourage you to subscribe so you can stay up to date with PubMed announcements. The NLM Technical Bulletin provides regular updates on new features, often with detailed descriptions on how those features work. The NCBI Insights blog also posted about features and planned updates for NCBI products, and the PubMed New and Noteworthy RSS feed is a great way to stay up to date, specifically about PubMed.

I also want to direct your attention to where you and your patrons can find more information, training materials, and opportunities to stay informed about the latest updates to come in. If you're looking for more information about how PubMed works or if you are a trainer directing others to this information, I encourage you to check out the PubMed User Guide and FAQ which is linked here on the slide, but it can also be found on every page in PubMed. We also offer a variety of tutorials as well as a PubMed Trainers Toolkit that conveniently collates handout slides and other resources that you may find useful for your work. And we hope that you'll continue to join us for future Office Hours events.

If you don't see an answer to your question in the user guide, or if you notice a bug or that PubMed doesn't seem to be performing correctly, or if you have feedback for us on PubMed, we encourage you to reach out to the Help Desk. A lot of the features added or updated in PubMed this year were based on user feedback. When you write to us and provide us with your use cases and suggestions for improvements, we use that information in our planning and developing process. So really the more specific you're able to be when you're providing feedback on a feature or asking for something to change, that's very helpful for us to understand what the

specific use cases and then to go from there and see how it might impact other users as well. You can reach the PubMed Customer Service Team and that includes me using the Help link that you will find at the bottom of every page in PubMed or by using the link you see on the slide here. I want to close by saying thank you for joining us today to talk about PubMed. I'm happy to answer any questions you may have and for now I'm going to pass it back to Mike.

**MIKE:** Alright, thank you very much, Amanda. So we're going to now spend the rest of this session answering as many of your questions as possible. If you've already started typing your questions into the chat, great, keep going, keep typing those. Make sure you send those questions to everyone, not just the host, so that we can make sure that all of our panelists can see them. We also have a few questions that were submitted ahead of time when we announced this Office Hours, and when we announce every Office Hours, we put out a call for your questions and we're going to mix in some of those questions with the questions that you're asking live today in the session.

So I think we'll actually start with one of the pre-submitted questions. This is a question that Tom submitted asking **how can I check if a word or phrase is included in pub meds phrase index? Is there a way to tell when a phrase was first added to the phrase index?** And I think that's a question for Amanda.

**AMANDA:** Sure, I'm actually going to share my screen while I explain how you can do this. You can check if a word is included in PubMed's phrase index, and the way to do that is to go to the Advanced search page and you can type in a phrase like-- oh my keyboard is not on, one second there we go-- heart attack and then click the Show Index button that appears under Add and we can see that obviously is in PubMed's phrase index and you can scroll and see what appears near it as well. If you don't know what the phrase is you want to look for, you can always click Show Index and scroll through them. But I will say that PubMed's phrase index is millions of phrases large so that's a lot of phrases to scroll through. We don't keep track of when a phrase is added to the phrase index. We don't have that data. But you can always search for a term or a phrase and then sort your results by publication date and go to the end of those to see when the term first appeared in PubMed.

**MIKE:** Excellent. Thank you for that explanation. I see a question in chat from Linda about allowing the chat to be saved so we folks can get access to the links. We will post all of these links when we post the recording so that those links will be right there, accessible to anyone who comes across the recording or wants to refer back to that recording later. Alright, while y'all are still entering your questions for our PubMed experts, let's see what else we have that was pre-submitted. Ah, here we go. Here's another one. **How can I tell which PubMed submission citations were indexed by humans versus indexed by the MTIA algorithm?** And this is a question I think referring to our somewhat recent transition to an automated MEDLINE indexing process. And Amanda, I think you might be able to address this.

**AMANDA:** Absolutely. I'm going to share my screen again. So yeah, you can tell how a citation, a Medline citation was indexed. We have these search strategies that you can use to do that, and there are two of them. It's indexingmethod, one word, underscore, and curated

(indexingmethod\_curated) is for citations that were indexed by MTIA, which is the algorithm but a human then later went back and reviewed them so we could search for those. I will note that PubMed spell checker tool will give you a suggestion, like maybe you meant to put a space between indexingmethod. We did not mean that this is intentional and I don't recommend clicking on that. It will--It's kind of a meaningless search at that point. So just ignore the spell checker suggestion.

The other search term that you can look at-- Let me move my Zoom bar here for the next one. Well the other option is for automated citations, so fully indexed by the algorithm, you just change that curated to automated. You can do the same thing. And then if you want to look for what citations were indexed only by humans, you can create that search strategy. It would look something like this. This would be a search for citations from this year, that were added to PubMed this year, that are in the MEDLINE subset because only MEDLINE citations are indexed with MeSH terms. And then we're going to NOT out the curated citations as well as the automated citations.

And I'll tell you right now that that search gives you 0 results because we moved to automated indexing before 2023. But when you get 0 results, PubMed tries to come in and save the day again by suggesting a different option. But we can see here we got the warning message that my search returns 0 results because again, there weren't any human index citations this year. But if we went back a few years to when there were still citations being indexed by humans this search we still get the spell checker recommendation, but now we have 661,000. Almost 662,000 citations from the year 2019. And I'll also mention you can tell if you are someone who uses the E-Utilities to look at the XML of the articles, these values are also stored in the XML so you can find them there and then. These search strategies are in the PubMed User Guide, so don't worry about trying to write them down. As I read this out, you can go to the user guide and find these there along with some more information about it.

**MIKE:** Excellent. Thank you very much for that, Amanda. We have another question coming in via chat from Andrea who asks, **is there a way to limit your PubMed search results to only preprints?** And Kate, who in addition to being the head of our training team here, has done a bunch of work with preprints. I think you might have a good answer for that.

**KATE:** Thanks for that question. Yes, you can search for preprints using the Publication Type of preprint and there are two ways you can do that. One is by adding preprint[pt] to your search. I hope I'm sharing my screen.

**MIKE:** Yep, you're good.

**KATE:** So that's how you would do that. You can add it that way or because preprint is a publication type, if you use the article type filter, you'll need to go to your additional filters and go to your article type. You can find preprint on the list there. And add that as a filter to your search. All right. Hopefully that helps.

**MIKE:** Excellent. Thank you very much for that. Going to go back to another question that was previously submitted. I think this is going to be for one of our indexing experts on the call, either

Alex or Melanie, if you want to take this asking about **now with the advent of automated indexing why MEDLINE automated indexing does not index based on the full text of the article and only the title and abstract?** Alex, do you want to take a crack at that one?

**ALEX:** Sure. Yeah. So there's two reasons that we don't use the full text. So the first is a practical licensing reason. So we don't actually have permission from the publishers. We do not have licenses from the publishers to do what they consider data mining on the full text of the articles that we indexed. So we can only do-- we can only run our algorithm over the titles and abstracts. We are negotiating for full text access, but it's really not known whether we will ever get full text access to the entire sort of corpus underlying PubMed, so we can't sort of count on it or promise that we will be using full text at any point in the future.

The other reason that we aren't using full text, for example for the articles in that are in PubMed Central, is that the algorithm has not been trained to use full text. And when we have experimented with it on full text we find that its performance drops significantly because there is a lot of noise compared to signal in a full text article. So it would require some additional research and development training the algorithm to understand what is important when it's looking at a lot more text. The abstract and titles are distilled down for it, to what is important so it knows what to pay attention to. So there would have to be additional R&D for us to be able to use the full text in meaningful way.

**MIKE:** Excellent. Thank you very much for that explanation. I know that was something that I was curious about when we were first moving, when NLM was first moving into automated indexing, whether that was possible and that explanation makes a lot of sense. All right, let's see-- Again and if the floor is pretty clear, so if you have any questions about PubMed, feel free to drop those in. Oh, here we go. We have a question from-- here we go, **in the past there used to be sometimes significant lag time for indexing articles, for MEDLINE indexing articles. What is the lag time now?** Melanie or Alex, do you want to take that one?

**ALEX:** Sure, yes. Essentially, when an article arrives from the publisher with us and is put up on PubMed, it's usually run within a day through our automated algorithm. So the average lag time is about 24 hours.

**MIKE:** Which is substantially better than it used to be.

**ALEX:** It is much better than it used to be.

**MIKE:** Awesome. Thank you. Thank you for that question and thank you for that answer. Let's see what else we have here. Ohh, here's a question. **This is a question that was submitted earlier and this is a question, Amanda, for you about Best Match. The Best Match algorithm which for folks who don't know Best Match is the sort order that intends to elevate the articles that are most likely to match your search to the top ranked ranking of your search results. And folks have some questions about bias present in the Best Match algorithm. Any thoughts on that?**

**AMANDA:** Absolutely, yeah. I know that this topic has come up a few times and you may have heard that last year we had a codeathon to look at this. And then this year this is something that



Library Operations and NCBI, which are two divisions at NLM, we've been working with the National Institutes of Standards and Technology to apply what was recently published an AI Risk Management framework that's coming out of the National Institute of Standards and Technology. So that's applied to Best Match search result ordering and also being worked into our framework for future development opportunities as well. So it is something we're looking at and addressing. And that's in progress as we speak.

**MIKE:** Excellent. Thank you. Thank you for that update there. Let's just see what are the questions we have either previously submitted or if folks want to keep submitting their questions into chat-- Let's see. All right, here's a question. This question I think Melanie or Alex you might want to address. It's from Carol and it's another automated indexing related question. **Is the automated indexing based on requiring entry terms from the mesh to be present in the title or abstract? For example, for a record to have the MeSH young adult, does the phrase "young adult" or "young adults" need to be in the title or abstract?** And I know that's a question sort of about the workings of the automated indexing algorithm. So if you don't have the expertise on that, we can see if we can find another answer.

**ALEX:** No, we have it. So I can answer that question. So the current algorithm is based on text matching, but it does go beyond the actual MeSH terms and entry terms. So for example, there are certain sort of things that we could recognize as number ranges that might indicate something like young adult that would tell us it's looking at ages of cohorts or something like that. So there actually is a very, very extensive list of additional sort of triggers that are not within the MeSH entry terms that are used to find those phrases.

Now that is of course for the current algorithm MTIA, we are planning to launch our machine learning algorithm MTIX this November at which time the workings would change. And MTIX is a more sort of advanced machine learning neural network based algorithm, so it kind of creates its own rules for how it extrapolates what the indexing should be. So it's able to use a lot more context from the article besides just those exact kind of trigger phrases or text strings. So we're expecting to see actually a significant improvement in performance when we do launch MTIX because it's able to identify concepts in more cases with broader language.

**MIKE:** Excellent. Thank you for that for that answer. And while we're still sort of in the same realm of indexing actually, sort of taking a step back to look at more at the development of MeSH vocabulary itself. **Another question that we've had sort of in reference to some recent conversation around the development of the MeSH vocabulary and sort of calls for additional transparency in that development process. Dan is on the MeSH team and I think you might have some information for us about that.**

**DAN:** Yeah. So I mean I think it's obviously very important to maintain the clear and transparent communication with our users. MeSH has implemented a number of other things as with. So we had two listening sessions last year or so and I think that we have some recordings that's available if you missed that.

We also have made a new page, I think it's called What's New. And in it we are having a lot of reports such as the Upcoming Match 2024 that you can browse through in certain formats and we

also have a lot of reports such as how we are going to or how we did the year end processing. So we're trying to reach out and just work with our users and developing MeSH.

**MIKE:** Thank you very much for that update. Then we have another question that just came in. I think this one's going to be for Amanda. And this user asks, **I use PubMed periodically all day and every time I want to save a citation I have to log in. Is there a way to stay logged in without keeping your browser open all day?**

**AMANDA:** This is a great question and, without knowing specifically what's going on in your browser, I would guess it's related to your settings about cookies. That's how PubMed stores a lot of that information. So I would check there first and if you're still getting signed out, send us a case to the Help Desk and include as much detail as you can, like what browser you're using when you're closing out of it, that sort of thing, and then we can look into it to help you further. But start with the cookie settings and see if that helps anything.

**MIKE:** Yeah. And the second part of that is also a good lesson to reinforce, which is if something is behaving the way you don't expect, let us know because you know it either we can help you make it work or it's something that's not working right and we can fix it. I also noticed that Michael dropped in the chat the new technical bulletin article regarding the sort of MeSH 2024 preview information that Dan was talking about.

Got another question for you, Amanda. This is from a different Michael. A third Michael. **How do articles get seen by the search subset hasplainlanguagesummary? Is this collection of plain language summaries manually curated, or is that tagged from the publisher? How does that information get on those records?**

**AMANDA:** That information is tagged by the publisher when they submit it to us, so they're supposed to submit it with specific tags in the XML, and then a flag gets added to the citation on our end when that's tagged correctly, and you can use the hasplainlanguagesummary search to find them. This is something, that search strategy is newer and plain language summaries are newer. So this is something that I think is starting off a little slowly. But we are encouraging publishers to supply us this information whenever they have it available. And if you are publishing, encourage you to talk to your publisher about that as well.

**MIKE:** Yeah, I feel like this is something that we always see when we add new-- add the capacity to store new types of information to PubMed records. Like many years ago, at this point, with adding ORCIDs to help identify authors. You know we at a certain point are depending on the that information coming from the publisher. But we can also depend on you to encourage-- to submit that information to your publishers when you're publishing and to encourage more of that information to be sent to us. So that's definitely great.

All right, let's take a look and see if I missed any questions or if there's other questions that have been submitted. Let's see. Addressed that. We talked about staying logged in. Alright, well let me go back and see if there's any other pre-submitted questions that we've missed. Let's see. Oh well, here.

This is a question, I guess that ties into some of the indexing stuff we were talking about before, especially sort of in light of the backlog, right and the lack of backlog now in indexing and how fast things get indexed. I know that some folks have asked either in previous office hours and then also at MLA and at Customer Service if there's any movement to adding MeSH to non-MEDLINE journal articles, so to citations that are in PubMed because their articles are in PMC but they're not part of the MEDLINE subset. Melanie or Alex, one of you want to sort of **address that question as to whether non-MEDLINE journal articles will be getting MeSH attached?**

**MELANIE:** Sure, I can answer that. We don't plan to index non-Medline articles at this point. We are still focused on developing our algorithms for automatically indexing just MEDLINE journals.

**MIKE:** Yeah, I know that with a lot of these developing of these new systems and improving these new systems, there's still a lot even if the backlog is down, there's still a lot of work to be done to make sure that that things are working as they should be. All right, let's see what other questions we had submitted in advance. We did that one. We looked at that one already. Amanda, I'll give you a couple of sort of rapid-fire proximity search questions, because I know that these are some questions that folks have had about how proximity search works. **Can I combine PubMed's proximity searching with truncation, with the asterisk truncation operator?**

**AMANDA:** No, at this time proximity searching is not compatible with truncation. So if you do include that wild card which is the asterisk symbol within the double quotes when you do a proximity search, the proximity operators will be ignored and it'll be run like a regular double quoted phrase search with approximate with a truncation operator in it.

**MIKE:** Next one, how many terms can I include in a proximity search?

**AMANDA:** There is no limit to the number of terms that you can search together with one proximity operator, but the more words you include, the narrower your search becomes. So I think at some point it's maybe not worth putting all of them into a proximity search, and you might consider a different type of searching, but we're not imposing a limit on how many words you can put between your double quotes.

**MIKE:** Alright, and one last one, at least for now about proximity search. This is one that confused me when I was doing some initial testing of it, so I feel like this is a good one to answer. **I see the results of my proximity search with the terms highlighted in the results, but the terms are not near each other. What is going on there?**

**AMANDA:** So PubMed has a highlighting tool in the search results that is completely separate from your search syntax other than seeing your terms. PubMed's highlighting tool does not incorporate your query syntax. It just uses simple term matching to show you words from your query highlighted, wherever those words are appearing in your search results. And that includes any terms from the automatic term mapping translation, if automatic term mapping was applied. So you're only seeing a snippet of the abstract when you see the search results. So even though you may see maybe just one term highlighted in that snippet, if you click on the abstract and look

at the thing as a whole, you'll probably encounter where your terms appear closer in proximity as you get further into the abstract.

**MIKE:** Excellent explanation. All right. I have another question. And again, we do have a bit more time left. So if you do have more questions about PubMed, about MeSH, about indexing, about PubMed training, any of those things we can address your questions on. So please feel free to just put those questions in chat so we can answer them. I lost my place-- there it is. A question that was submitted ahead of time. **Sometimes I find articles that are not systematic reviews with the systematic review publication type applied. Is there a better way to find real systematic reviews?** And I think we'll go to Kate on this one as it's sort of a training related type thing.

**KATE:** Sure. Thanks for the question. So NLM relies on authors and publishers to identify the research methodology that they're using, including systematic reviews. So NLM doesn't verify the research design manually, but we use the description and the title and the abstract to apply a publication type term, or we use the publication type as applied by the publisher, and you can find those using the PT or the filter. So yes, occasionally authors or publishers confuse methodological terms. But it's not possible for us at NLM to evaluate the research methodology on an article-by-article basis. And so we recommend that if you do find an article to be mislabeled, please please contact the publisher and request a correction. And once published, a correction can be noted in PubMed. Thanks.

**MIKE:** And I think Sharon has **a sort of follow up question that you sort of addressed, which is are scoping reviews included in the systematic review filter?**

And I think that the sort of answer to that is you know if that systematic review filter is attached-- Ohh wait hang on one second, I'm getting some more information in here from a colleague here about the systematic review filter. One moment. Let me see. Yeah, I'm actually just going to-- We'll just put this in chat. So this is the search strategy for the systematic reviews filter. Put that in there. And that's sort of what it's searching.

So if scoping reviews have that-- would be caught in that search filter, then they would show up or if they are labeled with that publication type.

Carol asks **why do I find the same citation listed more than once in a search results list, neither of which is an epub citation?** Amanda, any thoughts on why this might happen?

**AMANDA:** Yes, occasionally we do end up with duplicate citations in PubMed. We have checks in place, automated processes to check that publishers aren't resubmitting content. But there's no perfect way to do that without, like, putting a lot of restrictions on content we do want in PubMed. So occasionally we end up with duplicates. If you find them, send them to the Help Desk. Send us both of the PMIDs and we will look into it. We'll determine which one is supposed to be kept. Usually this means we'll keep the one that was submitted earliest. And we'll remove the other one and then the link to the other end will redirect to the citation that remains. And even if you're not sure if they're actually duplicates, you can send them to us. We will investigate and we'll let you know what the outcome is either way.

I will also note that sometimes publishers publish an erratum or something with the same title, and they don't include the word erratum in the title or it may look like a duplicate. So sometimes it's not. But most of the time I would say that's probably what's happening and please let us know.

**MIKE:** Yeah, I know Amanda and her colleagues have seen a lot of interesting data come through from the publisher sometimes. I mean, the amount of data, the sheer amount of data that goes into PubMed is massive. So the fact that there would be issues from that from time to time is unsurprising. But they do a great job at making sure that everything is cleaned up. Carol asked **a follow-up question, what if the PMIDs are identical?**

**AMANDA:** That would be very unusual and so we would need an example to look at that. So in that case, send us your search so that we know how you got to this issue as well as the PMID that you're seeing duplicated in your search results and then we can take a closer look.

**MIKE:** Getting into the really weird stuff late in the Office Hours. Alright, let's see if I come through any other questions that were submitted in advance. Again, if you have any questions for our panel please feel free to put them in there. We still have a little bit of time left. Here's a question that came in earlier. **Somebody's asking about embargoes and embargoes on journals, on journal articles appearing in PMC. And sort of figuring out which journals are embargoed and when things would get into PubMed faster than the full text would get into to PMC.**

**AMANDA:** Is that question for me? Sorry.

**MIKE:** Yeah, sorry. You looked like you were confidently nodding. So I was like, oh Amanda will take that, but—

**AMANDA:** I was flipping between screens. So I'll answer about embargoes and then if I miss a part of the question, Mike please remind me. So PubMed does include citations to articles that are under embargo that are implement central like you noted. There are some links that we can add to the chat. But it will give you more information about embargoed articles. If you want to know the complete list of embargoes and what the default is for each journal, you can find that information on PMC's journal list page, which I think is one of the links we're putting into the chat, so you can take a look there.

I'm also going to note that the NIH Public Access Policy is changing. That was in response to an August 2022 OSTP memo, which is doing away with the embargo and making results of taxpayer supported research more immediately available to the public at no cost. That new policy is going to go in effect no later than December 31st, 2025. So there may be some differences with embargoes due to that in the future. Did that answer all the questions, Mike?

**MIKE:** I think you got it all. Great. Thank you. All right. Well, we'll start to wind down here. We have a few minutes left. So if you have any last minute questions and want to get those in now, I'll give you a few more minutes to type some stuff in, while I just go through and make sure that all of our pre-submitted questions were addressed. Let me see if there's anything that I'm missing

here. We already addressed those. We talked about preprints. Bear with me just a moment. Looks like most of what we had submitted ahead of time we've gotten to.

I'll actually throw to Kate one more time to sort of **talk about what might be things, other things in the realm of PubMed training that you should keep your eye out for, new things, new developments, things of that nature.**

**KATE:** Definitely. So first I'd like to invite everyone here to join us for our next session of our popular class, How PubMed Works in November. How PubMed Works is a four part online interactive class and if you watch the NLM Technical Bulletin you will see an announcement for the next session when it's available. And if you just can't wait, we have a wide variety of Just in Time training resources on our PubMed online training page. I'll put the link to that in chat as well. So a couple of recent additions to our online training is a long form tutorial called Topic Searching in PubMed and a quick tour on proximity searching and you'll see those linked from the top of the page, so check those out. Thanks.

**MIKE:** Thank you very much Kate. So as we wait a few more moments for last call for questions, I want to sort of get ahead of this and start thanking our panelists for being here to help answer our questions today. As you can sort of see by who's represented here, a lot goes into making PubMed work the way it works. We have, you know we have the folks on Amanda and Jessica's team who are sort of building the application and making it, making sure that it works. We have sort of at the other end of things we have Dan developing the vocabulary, Dan's team, developing the MeSH vocabulary. Alex and Melanie's team making sure that MeSH terms are attached. And then folks like me and Kate who are, you know, helping get the word out and helping making sure people know how to use things. So I encourage you, when you're sort of thinking about PubMed and using PubMed, remember that it's sort of there's a lot that goes into it and a lot of different people doing a lot of things that go into it. And, you know, sometimes things go wrong. I think we can acknowledge that. But as Amanda said, we're always eager for you to let us know if something's going wrong and we can help fix it or if there is something that you'd like to like improved, you know, we're always eager to get that feedback. That's what drives a lot of the changes that we make in PubMed, is actual user feedback. I also, of course, want to thank all of you for attending and asking questions.