TRANSCRIPT

Webinar: NIH & FDP DMS Town Hall #3

Event Date: Thursday, December 14, 2023

Event Time: 1:00 – 2:30 PM ET

Melissa Korf: All right, we're one minute after the hour, and we have a hopefully packed full of information webinar for everyone today. So, we'll go ahead and get started. Thank you so much for joining us for today's NIH Data Management and sharing Pilot Town Hall. We're very excited that you are joining us here today and for the wonderful agenda that we've got planned. My name is Melissa Korf, I'm one of the co-chairs of the Federal Demonstration Partnership, we're FDP Pilot, and I'm joined by Christi Keene and Jim Luther, also representing co-chairs of the FDP pilot.

The agenda that we've got planned for today is just to share some background information on the FDP pilot and our goals in phase one and phase two. Then we'll turn it over to our NIH Institute and Center, our IC colleagues for some presentations regarding their initial feedback on submitted data management sharing plans. We'll share some preliminary results from our submitter survey as part of the FDP pilot. And then we want to make sure that we leave a lot of time for discussion where either you, as our community members could ask to unmute and share your experience.

If we have any deaf participants that are here with us today, if you let us know in the Q&A or the chat that you'd like to share an experience, we can make that work to promote you to a presenter so you could turn on your video to be able to do that. But we're also, we're leveraging the Q&A for today. So, if you've got any questions for us throughout the session or at the end, if you want to pop those in the Q&A and we'll try to get through as many of them as we can.

So, first off, I know that we're opening this webinar up to a really broad audience. So, some of you may be thinking, what is this federal demonstration partnership, or FDP of which Melissa's speaking, and our mission can be summarized really simply to, we want to be getting back to researchers doing research, not administration. It's an organization of federal agencies, academic and nonprofit research institutions and research policy organizations that work together with our partners from our federal funding agencies to streamline the administration of federally sponsored research. We really want to be focusing on reducing administrative burden without compromising compliance.

There are about 217 institutional members right now, 48 emerging research institutions, 32 minority serving institutions. So, we really try to make sure that FDP has a representative population of member institutions. And if you're interested, I believe the slide deck has been made available on the sharing.nih.gov site, and we provided a link there to go see the list of full

membership. We have 10 federal agencies, of which our NIH colleagues are one that are also members of FDP.

So, as you might already be aware, if you're here, you might have some familiarity with the NIH data management and sharing pilot or policy, excuse me, already. But you might be aware that the policy intentionally needed to be a little light on certain details, in order so that the policy could work for the breadth of fields that are funded by NIH. Some of the details needed to be left out or left open to interpretation so that it could best meet the needs of that particular field. But we identified early on that if each IC was developing its own templates, its own requirements, this could pose quite an administrative burden for our researchers as they first had to navigate potentially various IC specific requirements and then get to the business of writing their DMS policy.

So, we developed this and a DMS pilot with our NIH colleagues with the goals of generating greater consistency in DMS plan requirements across NIH, ICs and programs, and seeking opportunities to mitigate the administrative burden for researchers associated with DMS plan development and implementation.

In phase one of the pilot, we've been testing out two pilot templates. Alpha is a very prescriptive template that's published as a very smart form fill-able PDF. As you answer a question, it may open or close other questions in the form. Bravo is more prescriptive than the sample format page you may have been familiar with that NIH published originally. It is a little bit more prescriptive than that, but it provides more free text prompts than is provided in the Alpha template. We're gathering research-- we're gathering data from the researcher perspective to help us evaluate the effectiveness and usability of those two templates. The goal here is not necessarily to say that we would wholesale adopt Alpha or wholesale adopt Bravo, but rather to gather information about what works from each template so that we can arrive at the most effective template at the end.

We're also seeking feedback from support providers, folks in the libraries or research administration or research computing that may be supporting researchers as they're completing their DMS plans. We recognize that these folks have a unique perspective on the DMS plan and the fields that it includes and might have a broader perspective actually as they're drawing from a breadth of researchers that they've supported in creating the DMS plan.

So, on this slide, we have both the link to that support provider survey as well as the link to the survey that we're collecting feedback on experiences using the pilot templates. If you are not at an institution that is a formal member of the pilot, have no fear, we would still love for you to give one of the pilot templates a try and would welcome your feedback as well using that link. So, if you're not part of an institution that is formerly part of the FDP, we would still love for you to share your feedback with us via one of these links.

In phase two, we're going to seek to develop some resources to help us address cost policies that are unique to data management and sharing, establishing some common cost principles,

identifying the types of costs that are required. How will we handle unforeseen costs or costs that may occur after the period of the grant has ended, as well as tools to help facilitate the actual development of the costs that need to be included, where we're well into the planning phase for this second phase.

I hope you all are super excited about the templates. And if you haven't given them a try already or are ready to give one of them a try, and both of them are available via the FDPs website from the page that's dedicated to our NIH pilot. This link should take you directly there, but also if you go to the fdp.org, it's available from one of the top dropdown menus demonstrations. We've also been able to work with DNP tool to be able to make the pilot templates available via DNP tool as well. They look a little bit different than the PDF or Word doc version that are on the FDP website, but the content is the same.

Some other relevant resources we wanted to share based on some of the questions that were submitted along with webinar registration. So, the FDP also does have a data transfer and use agreement template and sample project that has been going on for quite a few years. We were actually referenced in a science article, one of my most exciting moments working with the FDP. And you may find those a great resource in terms of sharing sensitive data, data that can't just be, you know, posted on a website or posted in a repository that has to be more controlled access.

And the data stewardship subcommittee with FDP is also working on a number of projects related to data management and sharing, such as a collaboration with the Subaward Subcommittee to evaluate the existing Subaward agreement templates in order to determine if any changes need to be made to accommodate the DMS policy. And now I'm so excited to turn the virtual mic, or the virtual room over to my colleague Michelle Bulls, who's the director of NIH's Office of Policy for Extramural Research Administration or OPERA, who will be introducing our IC colleagues.

Michelle Bulls: Thank you very much Melissa. Before I introduce our IC colleagues, I just want to turn your attention to a Nexus post that Dr. Michael Lauer issued on November 30th, encouraging our colleagues, you, to participate in the pilot for the data management and sharing templates. So, Kristin has thrown it in the chat, would encourage you guys to take a look at it, and please, please, please do join and participate.

With that being said, I'm going to introduce our first-- well, I'm going to introduce everyone first of all and then we'll go through the presentation. So first up is going to be our colleague Dr. Rebecca Rosen. She's the director of the Office of Data and Science and Sharing. And Valerie Cotton, the Deputy Director of ODSS, they are from the National Institutes of Children, Child Health and Human Development. We also will hear from our colleague Emily Boja who is the Scientific and Policy Program branch chief, and Heather Basehore, the Health Science Administrator from the National Cancer Institute.

And then again Greg Farber, who is the Director of the Division of Data Science and Technology. He's with our National Institutes of Mental Health, and then the National Institute of Biomedical Imaging and Bioengineering, we'll hear from Qi Duan. I mean he's the program director for the Division of Health Informatics and Technologies. So, with that being said, I'm going to turn it over to my colleagues, both Rebecca and Valerie, and they will take it away. Thank you.

Valerie Cotton: Thank you so much Michelle. I'm Valerie Cotton, and I'm from the NICHD Office of Data Science and Sharing as Michelle said. Rebecca is also here, but she's going to monitor the Q&A while I give this presentation. We're going to tell you a bit about our observations from and tips for data management sharing plans. But first, we want to start by emphasizing NICHD's commitment to data sharing. For us, data sharing is critical to our mission of accelerating research that improves the lives of pregnant people, children and people with disabilities.

In fact, it was the NICHD strategic plan that created our office to build an ecosystem that fosters responsible and innovative use of our data. And the NIH data management and sharing policy is foundational to that vision and strategy. As early as January of this year, the NICHD Data Sharing Committee, which is made up of program staff from across the institute and representatives from our office, looked at a subset of data management and sharing plans and thoroughly assessed them against the policies expectations.

This was before the FDP pilot templates went out, so we don't have feedback on the templates yet. But this is where we started seeing the key issues that I will present today. Over the summer, we came together, and we deliberated on those issues and agreed on what was acceptable or not, and then we updated our public and internal resources based on those discussions. And we've held office hours and NICHD wide training for all program staff.

This is a summary of the key issues that we keep seeing, and I think you'll see that many of these could be addressed by a better template. Overall, we want to highlight that the elements of the format page are almost never used the same way. So, we had to train program staff to look at the whole plan to determine whether something was acceptable. There's a lot of contradictory information across the plan elements, and other overall issues include a lot of vague reasons for not sharing and only sharing in the form of publications or presentations at conferences.

We identified a variety of issues related to which data are shared, when it's shared, where it's shared, and how it's shared. And we also found that the data management and sharing budget justifications are often missing or not clear, or it's hard for the program officer to tell how the budget is going to support the data management and sharing activities described in the plan. I'm going to dig into some of these a bit more in the next slides in the form of tips for writing a data management sharing plan.

Here are some of our overall tips that we think would be useful to keep in mind. Remember that data management and sharing are inherent parts of the research process, and good data management and sharing is critical to good research. Remember that the goal of data sharing is to do what we believe in, accelerating scientific progress, that's why we're doing this. So, your approach should also support that goal. When in doubt, refer to the definitions in the policy materials because that's what we are relying on to determine whether or not a plan is acceptable.

For example, data sharing means making data accessible to the public or the larger research community. And the definition of scientific data is tied to principles of data quality and reuse regardless of publication. Be clear about which data, when, where, and how, not only for NIH's sake, but also for the public to understand your plan, because the intent is to eventually make these public. And also, so that you know exactly what you need to do to implement your plan. Finally, for anything that deviates from what is expected in the policy, justify it. So, think of the guidance and expectations in the policy as do this unless you have a reason not to, and then NIH can review whether that justification is acceptable.

When describing which data will be generated and shared, think about including details like species format and amount. You'd be amazed at how many plans we've reviewed where we could not figure out whether the data had come from a human or not. And all of this information matters for us to consider whether other aspects of the plan are appropriate. Should there be privacy protections? Probably not if this came from a fish. Is this the right repository? Is special software needed to access these files? And does the NIH genomic data sharing policy even apply?

It's important to clarify details associated with each data type. Sometimes it was hard for us to tell which data will be shared compared to all the data being generated, or which standards apply to which data type. Many plans still say that they will only share publication related data, but the policy expects sharing of data not associated with publication as well. So, provide all the same details for non-publication associated data, as you would for data underlying publications. This is something that was built into the policy and its definition of scientific data from the start but has since been further emphasized by the White House's call to ensure free, immediate and equitable access to federally funded research.

Identifying and committing to data repositories early on will actually help you prepare and manage your data for sharing throughout the life of the project. Once you know which repository, you can start preparing your documentation and your metadata and your formats and your de-identification strategy, all according to the repositories guidelines, even as soon as your data collection begins. And remember, if something comes up and you have to change your repository, you can do that with-- you can make that kind of change with your program officer's approval.

Make sure that you're using data repositories that make data accessible to the public or the larger research community and not just a limited group of investigators. Again, in alignment

with that definition of data sharing. Sharing data in journals is not the same as sharing data in a data repository, and it's generally not sufficient for making that data accessible and useful to the research community.

The policy guidance emphasizes that domain or data type specific data repositories are always preferred over generalist or, especially over generalist repositories. And this is because domain specific data repositories often specialized in curation services and serving the needs of that specific research community. Generalist repositories are great, and they should be used, but only if there's not an appropriate domain specific data repository available.

Now, the biggest issue that we are seeing are data management and sharing plans that propose sharing data by request. So, a researcher would have to contact that PI to get access to the data. And not only is there a growing body of literature demonstrating that this approach is not effective, having one person or a small team serve as a gatekeeper inhibits the ability to make the data accessible to the larger research community. So, it's not aligned to that definition of data sharing. And we realized that there was actually some potential confusion about this in the community because there's a difference between by request and what NIH means by controlled access, which is a privacy protection described in the policy materials.

So, when you're concerned about privacy, you should consider things like de-identification and using controlled access repositories. In this case, it's the repository that has established processes for verifying appropriate use of data, including things like centralized data use agreements. So, to Melissa's point, it doesn't have to be the researcher on their own who's making these data use agreements, if you're leveraging an existing repository, you're leveraging the established processes for data sharing agreements.

For any approach that deviates from what the policy expects, provide a justification. Tell us why it's not possible or reasonable to share certain data or to share by the end of the award period. And we have not seen a good reason for not using a data repository, but if you think you have one, you should state it. When citing laws, regulations, or policies as a reason to limit data sharing, cite the specific law or policy. In some cases, you might even have to invoke a specific protection. For example, some of these rules mean that you can delay data sharing, but you don't necessarily have to. So, if you choose to, you have to make it clear that that's what you're invoking. When citing human subjects issues as a reason to limit data sharing, it helps to explain your IRBs role in that determination, so that we understand the expertise that went into the decision.

We've had a lot of missing or vague data management and sharing budget justifications. But program officers are looking. They want to make sure that researchers are requesting sufficient funds to support the data management sharing activities described in the plan throughout the life of the project, and not just at the very end. We suggest you think about the time and effort that your staff and personnel commit to preparing data for submission to data repositories. It does take a lot of cleaning and curating and preparing that documentation.

Maybe some of you don't need as much support if you are using your local librarians or data coordinating center to help, but it might actually help if you tell us that, so that we understand why the cost might look low. And we've observed that it's actually more expensive for researchers who are proposing to share by hosting data on their own server or their own website, rather than using an established repository. So, this is yet another reason to use an established repository where the costs shift from the researcher to the repository. And this is especially true since many repositories do not charge fees for data submissions, or they charge a minimal one-time fee, so you don't have to worry about recurring costs beyond the award period.

And finally, NICHD does not have its own version of the data management sharing policy. Everything I showed you today is based on the NIH data management sharing policy materials. But we have created NICHD specific resources to address feedback from and considerations for our specific research communities. And those are all available on our website. And with that, I will be handing it over to our NCI colleagues. Thank you.

Heather Basehore: Okay, I think you'll probably find similar lessons learned from all of the ICs that are on today. So just a quick introduction to the National Cancer Institute and our look on the data sharing landscape. So, we really look at this kind of arc or trajectory of how policies can influence data sharing within an institute like NCI. So, we know that there can be very project specific data sharing requirements, and that's kind of where we've been, right? Where it's up to the individual applicant or individual grant to tell us how they're going to share their data. And what we anticipate is that we're moving toward a state where there are more uniform expectations across everything that's funded by the NIH. Right now, we find ourselves somewhere in the middle of this arc. But we do have experience within the NIH and especially within the NCI with our experience in the Cancer Moonshot Initiative, which had very specific data sharing requirements for Moonshot funded projects.

Now within the National Cancer Institute, at any given time, we're administering funding for approximately 10,000 grants. We don't have a centralized mechanism for reviewing the DMS plans from every grant. We want to make sure, we felt that it was important for our program staff to have that expertise to be reviewing the DMS plans along with the grant applications. But we do have a lot of resources in place through our office of data sharing for program staff to get help when they need it and to collaborate and make sure that we're all learning together. We also have a large number of branches and intramural labs that are generating scientific data.

So, when the plan was implemented in January, our office put together a strategy of reviewing a really small, relatively small random sampling of plans that were submitted. And I'd like to share with you some of those results. We decided to go through the policy element by element. So, if you're familiar with the data management and sharing policy, you know that there are six major elements that are discussed in terms of data sharing requirements. So, we randomly selected around a hundred DMS plans that had gotten pretty good scores. So, you know, we

felt that they were pretty likely to be in the fundable range. And I know that there have already been questions in the chat about this, so just to clarify, we looked at the plans that were submitted, not anything that was corrected on just in time. We wanted to know what are the areas within the DMS policy where people might need some additional help. Okay, so our systematic evaluation was guided by elements of the DMS policy.

Just for a quick look at what we evaluated. We chose 101 at random. Not surprisingly, you can see the breakdown by grant type. Most of, or the highest number of those was R ones. But you can see we also reviewed K grants, U grants, P grants, and other RT awards.

So, if we look element by element, we start with the first where applicants are asked to summarize the data types and the amounts of data that will be generated in the project. And you can see a quick breakdown of the types of science that were submitted in the plans that we reviewed, largely preclinical but also some genomic and imaging data and other clinical data as well, and then some other categories. And we found that about half of the plans that were submitted did list the data type and the amount of data that they expected to generate. The rest of them only listed the data type without listing the amount. And most of those that did list data amount listed it in terms of number of samples, number of human subjects participants, or something along those lines, not in terms of computer space that would be taken by the data.

We also found that only 25 of the 101 plans specifically talked about data that would not be shared. So, we thought that that was pretty interesting. As far as our tips based on element one, because grant applications typically are generating more than one type of data, and we'll be answering a lot of questions about each data type within their DMS plan, we found that those plans that were submitted using some kind of a table, bulleted list, or a way to organize that information were a lot easier to follow and a lot less likely to miss further elements.

The second element asks the applicant to talk about specialized tools, software, or code that a secondary user would need to have in order to access or use the data. And we found that about 79 plans listed software or code, where eight said that there were no specialized tools that would be needed, and 14 of the applications just didn't list anything. We did also find though, that only about half of the applicants described how those tools could be accessed.

Our office was also really interested in knowing how a secondary user would be able to get to that data though. And so, we wanted to know, are the tools that are being proposed free of charge for a secondary user or something that would require a fear of subscription? And we were a little surprised to find out that even though many of the plans did list free tools, most of those also listed some tools that had a fee associated with them. So, we feel that that's an area that we may need to help our investigators, or our applicants find out more about how they can make their data more accessible, in a more broad way.

Our third element asks the applicant to talk about the common data elements or the standards that will be applied to the data collection, providing the names of the data standards. And if there are none to indicate that no standards exist. We found that this was an area where there

was a lot of misunderstanding and definitely someplace where we are going to need to work with our applicants and investigators to have a better knowledge about what actually we're looking for, and the importance of common data elements and standards.

We had about two thirds of our responses met the policy by talking about common standards, about one in 10 Applicants said that there were no standards that exist, although we weren't able to go and independently verify that, and the remainder didn't list any standards. So again, we want to be sure that we are pointing our applicants to the resources like the NCI thesaurus or the NHS Common Data Elements repository, so that people know how to find their standards and how to list them.

The fourth element is really where we find a lot of detail. So not surprisingly, this was the area where we found the most room for improvement. It asks to name the repository, which 91% of our applicants did. It asks to describe the methods that make the data findable, which only 77% of the applications did. And it also asks to describe the timelines when and for how long the data will be available, and 85% of them have that. Importantly, though, the policy expects that data are going to be made available at the time of publication or by the end of the award, whichever comes first. And this was definitely the area that needed the most improvement or the most-- probably would require the most second look by a program staff because only 51 of the 101 applicants listed both of those criteria. The rest listed maybe one or the other or didn't address when data would be made available.

Importantly, 68% of the plans that we reviewed met all of the sub elements in this element number four. And elements five and six were both pretty well answered or answered very adequately. So, for element five, which asks the applicant to talk about limitations or considerations for reuse, most of the plans that were submitted answered it appropriately. We did find anecdotally that human subjects researchers were very good at answering this question. They were-- and we assumed that that might be because they are used to working with IRBs and familiar with the protections that are required for human subjects research. We just offer as a tip that there are lots of factors that could affect secondary access, including human subjects protection, legal or policy considerations, IP or licensing agreements, and so on. But as Valerie mentioned that there are ways to look for maximizing data sharing even within these limitations.

And Element six talks about the oversight, and so around 88% of our reviewed applications did a good job answering that, and I have lost my slides, but I think I know what I'm talking about. Can anybody else still see the slides? Nope, okay, slides are gone. All right. So, while that's coming back up, we were also interested to know who is being proposed as the oversight monitor for our DMS plans. We found that by and large, when somebody was named, it was the principal investigator, which wasn't surprising, but we were also encouraged to find that about 35% of our applications listed some type of administrative office or institutional resource that also bears some responsibility for oversight of data management and sharing. So, our tip there is that we want to make sure people are checking with their institution, that slide is perfect,

thank you. Check with your institution because there may be resources or even policies or recommendations from your institution on how to answer that question six.

Okay, so we've skipped to the overall summary, which is fine. We found that out of 101 randomly selected plans, 60 of them were likely to be acceptable as written. And additionally, 31 of them were pretty good. They would probably require some minor corrections, but we're definitely on the right track. And then 10 of them would've required some major revisions to meet the policy recommendations. We weren't able to go back in and look at the just in time corrections to those plans at this point, but that's definitely something that our office is keeping an eye on.

Our general observations were that when a plan used some template, any template, they were much more likely to hit all six elements and to do a good job answering each of the sub elements. We also found that those plans that were submitted using tables or lists that organized their responses were much more likely to meet the policy requirements as our, you know, kind of highest-level general observations. We wanted to point out that it's important that all of the data that are described in the research plan of the grant application should be discussed in the DMS plan. So that was an area where we found, you know, some data types discussed either in the grant or even in that first element, and then there was no further information on that data type throughout the DMS plan. And also, as I mentioned, that data should be shared at the time of publication or at the end of the grant period, whichever comes first as an area to highlight.

So, I would just like to point out that the National Cancer Institute's Office of Data Sharing is here to help. And I would like to definitely highlight the work of Sophia Coco who was an intern with us over the summer and did a lot of the framework and research on this particular review. So, with that, I'm going to hand it over to my colleague Greg Farber.

Melissa Korf: Heather, can I ask one quick question that I think is directly relevant to some of your comments before we move on to Greg? There was just a question for how you might characterize the difference between a minor and a major correction.

Emily Boja: So, this is Emily Boja from and NCI, I can answer that. So minor, we, in our assessment, we characterize them as having to correct or revise one to two element and or sub element. And anything above three we call it major corrections.

Melissa Korf: Thank you, Emily. Now to you, Greg, sorry.

Greg Farber: No worries, Melissa, Thanks. So, I'm Greg Farber from NIMH. And as many people have discovered, I'm sure writing a good DMS plan really does require some knowledge of informatics, some knowledge of data archives, data standards, how software is being made available, things that all of-- that both of my colleagues have talked about previously. A number of investigators are having trouble with that, they're just not familiar. Many program staff at NIH are equally unfamiliar with these areas. And because of that, at NIMH we've created a centralized infrastructure in the institute to evaluate all DMS plans that are being submitted.

The assigned program officer still makes the determination about whether the plan is acceptable, but that determination is aided by the evaluation, and the evaluation is done by a team of folks in my division that have informatics expertise. So, the real goal here is to help program staff understand what's really been submitted, so that they in turn, can help the PIs make better plans if that's needed. They do bring us in as needed, but the program officer is really the, you know, the key decider, as is almost always the case at NIH.

So, my team has assigned overall scores to each data management plan. And we do this after, you know, a rigorous evaluation of the plan. So, we have a set of questions that we fill out for each plan. They go through all of the criteria that are in the guide notice that are, you know, related to the DMS plans. Applications that get a score of A are great. Generally, in our-- what we've discovered is that people who have these really good DMS plans have downloaded one of our sample plans, and many other institutes also have sample plans, they're all on the sharing.nih.gov website.

They've downloaded the plans; they've made some edits, and they've submitted them. I would say in response to one of the questions, plans don't need to be lengthy, but they do need to be complete. Applications that we gave a score of B to, have submitted a pretty good plan. The team believes that the plan could be accepted as submitted, but the program officer may want to discuss a few issues with the PI and the two of them can then make a decision or program officer makes a decision about whether a new plan should be submitted or just some amendments done in some other way.

Applications that have a score of C, have submitted an insufficient or non-compliant plan. Unless my team has missed something, the DMS plan needs to be revised before an award is made. And applications that get a score of D are just unacceptable. These are plans that really did not follow the guidelines at all. So, with that as a background of our scoring, in the October council round at NIMH, we evaluated 291 DMS plans. Those were all plans from applications that got a pretty good score that might be funded. So, you see that roughly a quarter of the plans got an A, roughly a quarter got a B, roughly a quarter got a C, and roughly a quarter got a D. It was interesting that the applications that come from human subjects, that involve human subjects, scored better than the applications that don't have human subjects.

There could be a variety of reasons for that, but we suspect that the real reason is that all applicants to NIMH who are doing work with human subjects have to deposit their data in the NIMH data archive. And so we think that what this is really revealing is that if you're going to deposit data to a well-established archive that has longstanding clear rules about how to deposit data and timelines and things like this, that the DMS plan is pretty easy to write because you just have to say, "Well, I'm going to do what the archive requires, I'm going to use these data dictionaries, you know, I'm going to do this, that, and the other thing." My guess is that's going to end up being true, that good archives make it much easier to write good DMS plans, and that that has more to do with the archive than it does with the template or the format of the DMS plan itself.

We've asked, you know, we've started to do since all of the data is in a database, you can do analysis. So, we wondered whether better funded institutions did, you know, submitted DMS plans that were better than on average than less well-funded institutions. And we couldn't see any trends here at all. We wondered if there was a correlation between score and DMS plan quality. So, you know, we divided this into applications that scored 10th percentile or better, which at NIMH are very likely to be funded. And those that scored between 10th and 23rd percentile, which are of course less likely to be funded, and we didn't really see any correlation here. Well, like we'll continue to watch this in additional council rounds, but I'll be surprised if a correlation does show up.

One place where we really saw a quite believable correlation was with activity codes. So, we did not have enough center grants P codes or cooperative agreements, U codes to aggregate the information from those, but we had plenty of R activity codes, and we had a number of Ks. So, it is really clear that the DMS plans for the K applicants were much better than for the R applicants. Could be a couple of reasons for that, most of our Ks in this council round were mentored Ks.

So it could be that having two sets of eyes on it really made a difference. It could be that K applicants pay a little more attention to the instructions than our R applicants do. And so, we'll, we'll continue to watch that as, to try to see if we can figure out why that is. What of course we would like to do is figure out why we're being more successful with the K applicants than we are with the others, and then see if we can improve our instructions to everyone.

In terms of data archives, here was the listing of archives. These, you know, most of the data is going to the NIMH data archive because we require that. But you can see the other archives that our applicants are planning to submit data to or proposing to submit data to. We manage a variety of brain initiative data archives, and they were also fairly frequently listed the dandy archive, much more so than the rest. And we also separated out the NIH supported generalist research repository ecosystem archives, the gray archives. And I was a little surprised, I thought by where data are ending up, there were three or four archives that were very popular and three that were far less popular.

Here are the data types, and as others have said, for each of these data types, we, you know, we evaluated what archive the data were going to, what standard they were using. It's very important to think about your DMS plan in terms of data type, and then answering all of the questions that need to be answered. And the last thing we looked at was software availability. And we weren't terribly worried about whether the software was coming from a commercial or an open-source supplier, because both of those are readily available, even if one costs, you know, something and the other is free. What we were much more concerned about is custom built tools, and if a researcher is proposing to do that, is there an adequate plan to share the tools and software?

So, we found that, you know 80% of our-- 80% of our applicants are proposing to use commercial or open-source tools. But for the 20% who aren't, more than half of them really

didn't have a good plan, a good way to share that software. And we were very liberal in terms of good plan. If they mentioned GitHub, we kind of said, "Sure, fine, that'll work." You know? So, a lot of groups that have custom software just don't seem to be sending us plans that you know, that describe how they're going to make that software available. So, I believe that's my last slide, and unless there were pressing questions, I will pass things over to Qi.

Melissa Korf: Greg, I think there was one question specific to your content that really, I think relevant to the mention of GitHub not being a good repository for the data or, you know, that kind of consulting is bad grade and. Is it just that they weren't using NIMH as specified repository, or is it when someone was saying they would use GitHub to share data and not their software?

Greg Farber: Yes. So, I should-- I need to make clear, GitHub is a fine place to store software workflows, tools like that. GitHub is not a fine place to share data because it's very hard to find there. And especially since NIH has spent significant resources setting up these generalist repositories where data can be shared, people should send data to the gray repositories, to the generalist repositories and not propose to put it in GitHub.

Melissa Korf: Thanks, Greg.

Qi Duan: Hi good morning or good afternoon. My name is Qi, I'm a program director at NIBIB. Today I'm going to present some analysis and feedbacks we gathered as the NIBIB Data Management and Sharing working group. So, I'll first introduce or explain the uniqueness of technology and engineering focused research, which might be different from what you have seen so far. And then I'll go through the results for the evaluation analysis on the October Council, or the first round of the application that's subject to the policy in our institute.

And so generally speaking, there's a majority of the research might be considered as a hypothesis driven research, where you start with the hypothesis, then you go through a study design, and you need a lot of data to test the hypothesis, and which lead to your research finding. In contrast NIH also supports technology development or engineering focused research. And depending on the year and the keyword that you are using in mind ends up about 10 to 15% of NIH budget are invested in this type of research. They do have a broad spectrum, but they start from first principles or engineering design. They involve a lot of preliminary steps and iterations in the research and development step before they can reach to a deliverable technology which might enable another iteration. And this technology eventually can lead to data generation that support hypothesis driven research such as MR imaging or coil or sequence and so on. And they have a different relationship when comes to data manager sharing.

We actually present this relationship in detail in September Council 2022. And here is the link that you can see the full presentation. And so, in terms of the data management plan analysis, the working group for data administrative by analysis, which I'm going to present in detail, and we also start gathering the PO feedback during the JIT process. And just one caveat is that the

data they probably can provide some insight on the first round of the application, but just keep in mind that given the uniqueness of the NIBIB applications, the data might not be representative or is relatable to other ICs or some other type of research. And for the, I'm sure the analysis, we focused on the first round of the applications that are subject to the new policy.

And we further limited the scope by exuding the activity codes that are exempted, or any withdrawn applications or application without a score. And that's come down to the scope of the application we look at. Then for each of them, we had to manually go through the budget and DMS plan because currently there's not a good way to automate the process, even the diverse type of the application we got. And we look at the templates used or not, and what is the DMS cost and also whether the budget justification is provided, which is by the way, is required by I should know this. Due to the limited resource, we didn't have the capability to fully better each of the plans for their justification, whether the cost is just well or not, that's have to lead to the PO during the funding process.

And all the data I'm sharing here today has been validated by the NIBIB data team. So, in terms of template use, about 87% of the application used the OER template. The rest of them with one inception come up their own format. They're not using any of those three companies available. And there's no significant variation across activity codes. And it seems like the ratios are pretty similar in different type of applications.

And in terms of DMS cost request is similar to other ICs funding that there's a majority of the grant they requested at zero cost, and there are about 10% of grant they didn't address the cause at all. And there's a clear trend that the smaller grants like R21, R03s, and small business grants, they tend to request at zero cost, and just look at it as some of these costs may be just declined to be zero because they're only considering the repository cost.

And given the uniqueness of our application, usually generalist repository or even sharing data through publication may be sufficient to satisfy the DMS policy requirement. But there are other costs like my other IC colleagues already mentioned that it has to be considered during the planning phase. And the justification, there's another area that we see like there are only 42% of the application that has provided us some meaningful DMS justification, the budget justification section.

And the 28% of the application only just mentioned the DMS cost again, and there's 30% of them they didn't have any budget justification regarding data manage sharing. And the major, the R01, the R21, and R03s, they appear to have more policy compliance. And we also did some association studies to look at other factors associated with the factors we're looking into. In summary, there's a clear trend to show that using template definitely helps and application without template tend to have some issues like missing the cost or have a justification issues.

And for grants, smaller grants are using modern budget they tends to have a zero or missing DMS costs. We didn't find any significant association between the modular budget and the DMS

plan template use, and we didn't find any significant association with the-- between the PI career status to any of the factors we looked at.

We start to gather some feedback from POs. As mentioned earlier, NIBIB does not have a centralized you know, a DMS plan review office or group like my other three colleague mentioned. The review is really relies on individual POs, and they're voluntarily providing their feedback to us. So far, the most commonly reported issue are, there's a frequent confusion between, you know, confusing the MS plan with resource sharing plan, and the review panel actually indicating the resource sharing plan is now missing, and there's some missing information from plan that's usually happen when they're not using template. And occasionally we do see lack of sufficient justification for the repository that are being used or being chosen, or the missing justification for limited sharing.

And so, in summary, it seems like the investigator education need to be further, you know further needed for the culture change that the policy has to be part of the research plan development and using template definitely helped. And there are some issues on the justification to be provided that can be further improved. With that I handed to my FDP colleague, named Christi Keene.

Christi Keene: Thank you, Qi. So, we've heard a lot of great insights from our IC colleagues, and we deeply appreciate that, them just sharing with us today. And so those insights are really important. And now we also want to look at the insights into what researchers have experienced. In your experience using templates, submitting plans. And as you might be aware, Melissa referenced earlier, we are collecting feedback from those who have submitted plans. And so, let's look at that data.

So, we asked which template you're using, and of the four, the 104 that have completed the submitter survey thus far, 46 have used Alpha, which again is the more prescriptive template, that smart form, and 33 have used Bravo and about 15 that have used sample format or other formats.

We also wanted to find out how much time you spent completing the plan. That is an important measure of the process. And over half have spent two hours or less on their plans. There might be some variance in the data, you could see there's 40 hours referenced, but it may not be 100% accurate. We also wanted to find out how much time others have assisted you or who else has assisted you.

Because we know that this often doesn't happen on its own and you know, many of you are getting assistance most people are getting assistance of some variety. And where is that coming from? Primarily from research administrators. So, we're hearing that research administrators are assisting librarians especially are assisting, and then a few others along the way. Again, also important is how much time are those folks spending to assist you?

And in most cases, it's an hour or less of assistance. So, you know, this could be a function of what resources are available on your campus or just how connected your offices are and how

well they work together. We also know that this policy, you know, it adds additional requirements, but is it changing your practices in your lab? So, to be successful with the policy, are you doing anything differently? And about half of you will need to make changes in your lab in some way. Some will be substantial changes, others not so much. And I think in some cases, you know, maybe TBD on that, how much change is needed. And finally, we also-- we wanted to partner with DMPTool and they've been great to help us to roll out the pilot templates. That was feedback we heard early on that they really need to be in DMPTool and as expected, many people are using them.

So, we're really happy to have that partnership and keep using those templates and complete the survey if you do. But an important part of our submitter survey is that qualitative feedback. So we asked a number of qualitative questions and we just summarized a few of the themes we've heard so far. As we mentioned, Alpha is a smart form. It's a little bit more technical in nature, and in some cases folks are having issues using it. That could be dependent on which version of Adobe that you have. But definitely hear you that there are some technical challenges there. We've also heard that you just need more guidance on the level of detail required, instructions, maybe like an example of a good plan using those templates. So definitely hear you there. And more guidance specifically when using secondary data would be helpful.

We've also heard that maybe the pilot templates are more targeted towards human subjects research. Again, we're hoping that these templates will kind of stretch across the wide variety of research that's happening. And you know, maybe one template does work better for your research than another. You've also said that it's very difficult to know some of the requested information at the proposal stage. I think, you know, hearing the ICs feedback, they're seeing that it is difficult for you to know that information and provide it at proposal stage. And the templates, the Alpha and Bravo definitely make those requirements more prescriptive at the beginning. And in some cases, folks have felt that the plan seems long or request redundant information.

And while we're talking about the pilot specifically, and we're very grateful to hear the feedback from the ICs, if you're still interested in participating in the pilot, Melissa, Jim, Michelle, Kristin, and I, would be very happy to have you join. We appreciate the participation of our pilot members. They are doing a lot of work to work with their faculty and gather this feedback that is so critical to making this successful. This is truly a partnership from day one. It is a partnership between NIH and FDP to find what works for the ICs and to find what works for the researchers and to make all of that compliant to the policy. So, we'd love to have more participation and you can reach out to us at the email shown here. And you do not have to be an FDP member to participate.

Okay, so I think at this point we will work through some of the questions in the Q&A, and work through those with our IC colleagues. I think many of them are a little bit specific. So again, thank you to our IC colleagues and thank you to Michelle and, and Kristin for the partnership.

Melissa Korf: So, two separate questions that I think there are on the same theme that got a lot of up votes, is that a lot of us research administrators or, you know, data librarians, we're trying to assist our PIs in completing these plans. And the tips that have been shared here today are so wonderfully helpful. Is there any opportunity where we might be able to create some sort of a rubric or, you know, a checklist to support PIs and the research administrators that are working with them, too, you know, hit the mark a little bit better the first time or, you know, provide some additional guidance. I don't know if there are any plans on the NIH end, or if there might be resources that are available already that we could point folks to. Valerie, go ahead.

Valerie Cotton: Yes. I'll just say one of our resources that I shared for our website is called Tips for Writing a Data Management and Sharing Plan. And that is specifically structured to explain element by element the policy relevant pieces. So, the program officers are trained to analyze your plan against the policy's expectations. And so, though that tips document talks through you know, here's how you get to that, here's how, you know, what we're expecting to see here, timeline, repository, et cetera, et cetera. We would be happy to collaborate across, you know, with our IC colleagues and OER to generate something central. But that is something that we made with that into it.

Jim Luther: Melissa, obviously stating the obvious, the templates, right? The templates drive that process. I know you were maybe setting up that discussion point there. And Michelle, I know you had a partial conflict, but I think you're still on, maybe you can join us on camera, right? I mean, we've been meeting with you and your team as co-chairs for nine months now. And those, the people on this screen, the presenters today were so instrumental in creating the thought behind that, and again, this is so generous with the time to communicate on that. But I know the prevailing themes in many of these presentations is the more formatted and structured the plan, the more likely you are to do it well. That's said in a layperson's term. I'm sure many of you, Michelle, you and others can say it more articulate than I can.

Michelle Bulls: No, I actually think that all of our colleagues within the ICs have said it nicely just in different ways, right? They're saying that, you know, you've used one template, and I think Greg said it very nicely. You know, you use a certain template and you're not getting enough information, so you quickly realize that you might need to spin off to another. I just think that it's really important for us during the pilot, that we begin to continue to promote the use of all of the various templates so that we can really see which one works best, because the goal is success. But just having the colleagues here today and hearing the feedback has been tremendously helpful.

And I'm just grateful, I know Kristen is as well. And I hope you guys have really enjoyed it and learned and are able to take this and move forward in a very positive way. It also all shows just the difference too in how NIH ICs are able to provide a really good perspective on what that success looks like. And sure enough, it seems like there was a common thread all across and all through the IC. So, this is really good.

Melissa Korf: Along that same theme there was one question regarding I think the sample format page not seeming detailed enough. And again, I think that the two pilot templates are an effort to provide information on the level of detail that is hoped for by our IC colleagues in the submitted DMS plans. And then another very similar question regarding how do ICs view the acceptability of the FDP templates, and kind of going back to some of Jim's comments, I think it's important to note that those templates were created by our IC colleagues very carefully to make sure that those templates were providing the information that they need and would accept DMS plans prepared on those templates as part of an application.

You know, I can't promise that, you know, depending on how you complete that template, you know, whether or not you know, all of the information would be found acceptable, you know, upon JIT stage review. But the thought is that if all of that information is provided, it's an indication of the level of detail. And I know that a lot of our IC colleagues that are or with us today on a town hall participated in developing those templates. I don't know if there's anything further you would want to say related to the process you went through to create them?

Michelle Bulls: Valerie?

Valerie Cotton: I have nothing to say, that's great.

Emily Boja: This is Emily from NCI. So, the template that we created was with the goal in mind to maximize structured data, yes or no, you know genomic data, other data, even though there is no data type ontology out there. We're trying to structure it in a way that, you know, people don't have to write a whole bunch of free text, but with the option to free text it in other line items. So, structure in a way that it's in a table and line up all the plan elements expected from the policy so that if people are writing several different data types generated from the same project or research proposal, they're not mixed and matching information, they're not missing information with regard to each data type.

Melissa Korf: Thanks Emily. And Heather did-- was there something that you'd wanted to add to that or?

Heather Basehore: Yes Melissa, you had just asked about the development of the plans, and I mean, that really was a very collaborative process, you know, not just from even one IC, it was collaborative across several institutes and centers you know, with the idea of trying to come up with a template that's going to be universal enough to address everybody's needs, but to really, you know, have kind of a guided method for filling out these DMS plans.

Valerie Cotton: And I'm sorry, Melissa, I didn't realize you wanted us to address the rationale behind the template structure? Okay, I'm trying to say neutral here because, you know, we have problems that we're seeing regardless of structure, you know. And unfortunately, we haven't seen enough use of the template from our researchers to analyze how it's going. But we went through so much internal user testing and we in fact used our Intramural researchers

before Intramural was required to use the format page. So, there are things I would change still, you know, this was before we realized what the biggest problems were going to be.

I'll tell you; I thought the biggest problem was going to be timeline, but most researchers are actually telling us the timeline, you know, they're going to do it, they're going to do the timeline of the policy, and we weren't able to analyze what the issues were until the plans actually came in. I still feel confident in the structure of Alpha in terms of addressing some of these pain points about, you know, which data type's going to which repository, which data type, which data standard applies to each data type. Like that has been such a challenge with the format page structure. But, you know, I'm kind of ready to move on to the next phase and make another template based on everything else. So based on everything, you know, that we've talked about today.

Michelle Bulls: Valerie, I thought I was losing it for a minute there. Wait a minute, Valerie didn't you participate really heavily?

Valerie Cotton: I'm trying to hide on that today. I'm trying to be neutral.

Michelle Bulls: I understand, understand. But you still did a yeoman's job on that.

Jim Luther: I was just going to comment that Valerie in your closing comment you talked about iterating to the next version. That's always been the vision, right? You know, we didn't think we were going to get it right right out of the gate? And it's wonderful to hear that you got the Intramural perspective, but we knew this was going to be an iterative process. So that's why we need the format. That's why we need the input.

Melissa Korf: And there are a couple questions I think maybe related to the data that our IC colleagues have shared in terms of their reviews. And one question related to if there's any sense of the percentage of plans that required modification at just in time, and then about how much time on the IC program side does it take to review each plan? I don't know if each one of our colleagues has any intel to share on that, Valerie?

Valerie Cotton: Yes. I started to answer this question and I don't have a percentage to give you, but how many plans require revision? The answer is most. You know, we're learning together. We trained our program officers pretty hard, so they're looking out for everything we talked about today, and the plans are not quite there. We are hearing from them, it's taking them on average one to two hours to review the plan, but then it spikes if there are just a lot of those issues we talked about, you know, particularly the PI control. Why do you need to control this access? Just put it in data repository. And, you know, once you pick your repository, you can complete the rest of your data management sharing plan. And some, and we do have a couple NICHD designated repositories, like NIMH, but we also recognize, you know, many of them need to be used.

And it's the plans that are not doing what's expected that take the most, right? Like if you're sharing data in a repository according to the timelines, you are maximizing the sharing of scientific data and being clear about, you know, the file formats that you're generating. And of

those, which ones are the scientific data that will be shared, right? We don't necessarily need your fast cues for genomic data anymore, but we do need your crams or bands. And that's genomic data sharing policy, but it's a really good example of like the type of details. So, the point is, you know, if you're clear and we can see that, that plan could be approved very quickly. If you're trying to tell us you can't share for some reason, that's where you need to do more work to actually justify to us why.

Greg Farber: Yes, and Melissa, I would say that from our experience, all of the application, all of the scores that, what that got all of the DMS plans that got a C or a D, which was half of them, we need to need to be revised, and some additional fraction of those that got Bs will also be revised. I think since the evaluation team at NIMH we're seeing a lot of these. We've gotten pretty fast at reading through the DMS plans and finding what we need to find. So, I think it probably takes us 20 minutes to half an hour per plan. But relaying that information to the program officers so that they can in turn relay it to the PI, that can take a while and we do, we have been having consultations with, you know, program staff and occasionally taking part in calls with the PI to really explain what needs to get fixed. That takes a lot more time, honestly.

Heather Basehore: Yes. I guess for the National Cancer Institute similar answers. I don't have part in fast numbers on, you know, how many are returned for just in time, things like that. But you know, we can definitely see a difference in those who are used to data sharing. You know, their plans are pretty good off the bat because they've been doing this for a long time, versus those who are working towards it, and I think our program staff are doing a great job, you know, kind of working with our PIs and applicants to get the plans you know, even better.

Emily Boja: So just to add to what Heather was saying, so we do have a sort of a customized evaluation support tool that we've developed, and it's obviously iterative as well as we learn to assess these plans, you know, that to include tips on like how to evaluate, what to look for, you know, those kinds of things. And it gets, you know, constantly updated. So that helps a lot in terms of sort of grading those, right? Like level one, two, and three, sort of in a very similar fashion as what NIMH is doing. Sorry Qi.

Qi Duan: No, no, no. I just want to add the NIBIB perspective as well, and just similar to my analysis colleagues and we don't have a, you know, statistically significant number to provide you. And just based on my interaction with older POs who asked me to chime in, I think like probably a significant portion, if not majority of the untouched plan that does need some additional iteration during the JIT process.

But the majority of the plan, as I showed, that they were using some template which apparently helped to organize the information and-- so most of the cases there are really focused on the question, like the additional justification for the choice of the repository or the justification for giving the assess. One thing that NIBIB is facing the challenges that our applications has a very broad spectrum in terms of technology, in terms of data type, in terms of their face in their field. Some are really pioneer and probably the first few applications in the entire field, there's no data standard yet. And that might come like five to 10 years later.

So, there are unique challenges. But I just want to kind of emphasize the previous question that using the template it does help to organize, to sort, to help to develop the plan. And I would recommend to use it in the early phase when you're developing the research strategy and make it a culture or a habit that, you know, think about how the data is managed and shared throughout the research. And then that you will have a better starting point and then that really will in turn benefit your research and the entire field. Thank you.

Melissa Korf: Thanks all. And Michelle and Kristin, there's sort of one more question about that JIT phase that you might be able to best help us out with. And that is, does this feedback go just to the PI or is there an expectation it would go to the authorized official? And I believe it's just like other JIT information where the authorized official would need to sign off on any revised plan that's submitted. Is that right?

Michelle Bulls: That is correct. Yes. We have not modified our JIT process for this. We should follow the existing process. It's very important for the authorized organization official to know what's going back to the NIH. So, they're the responsible entity at the PI.

Melissa Korf: And there was one question that got a lot of up votes, just sort of asking about this sort of you know, then what happens after award of some of those logistics. And there was one sort of very technical question about information that's requested about the grant number at the top of, I believe it's Alpha, which I think lends itself well to this topic.

Because I think it's well recognized that a data management and sharing plan in so much as the research changes and adapts to the findings that a data management sharing plan may also need to adapt. And so, in the templates trying to kind of account for the fact that we may need to be keeping track of version control, you know, that there may be the version submitted with the proposal and then the just in time version and trying to keep track of those and keep them associated with the grant.

So, I think if it's a competing renewal or a resubmission and, you know that sort of core grant number already would be fine to include that in alpha. But just generally, you know, some of the maybe more system comments, common aspects to it, you know, what might we be expecting to see if a new plan is approved at JIT? Would that sort of be uploaded also as a document in the commons? And if there are any kind of thoughts as we try to maintain these plans through the award lifecycle, post award.

Michelle Bulls: Kristin you're going to take that? Because I think they're asking about where they upload revisions.

Kristin Ta: Sure. So, I think I did see that question. When you submit your plan with your application, it will show up in commons as the DMS plan that's there. If there are changes to that down the line that are submitted through a prior approval request or anything like that, that comments version is still going to stay the original, but you will see updates to your notice of award saying that this award is being revised to, you know, to update the DMS plan

submitted on X date. So, there will be a trail of when changes are approved, it just won't update that original version in your application.

Michelle Bulls: And that's really important for audit purposes.

Melissa Korf: And so, Kristen, that might happen whether you submit via the comments module for prior approval, or if you include an updated plan as part of the RPPR, right? We might expect in both of those cases to see a revised NOA providing the approval?

Kristin Ta: Yes. And for your RPPR, it would be the type five, you know, your next year that you get, that would approve. It wouldn't be something separate.

Michelle Bulls: That's why I started shaking my head just because it would be, yes.

Melissa Korf: However, it comes, we would expect to see some documentation approval.

Michelle Bulls: Absolutely. Yes. Because once the PO reviews it and accepts it, you have to revise the notes of award because it then becomes a term and condition of your award that you have to abide by.

Melissa Korf: I mean, we have received a lot of great questions, but we're closing in on our two-minute warning here. And Jim, I think you would agree to bring us to completion with some closeout comments. So, if you want to hand it off to you for that.

Jim Luther: Great, thanks Melissa. And Michelle, I'll turn it over to you for the final thanks. But I wanted to thank everybody here, the leadership group again. This takes a village, right? Between OPERA and the ICs and FDP and the institutions. I really want to encourage everybody. Melissa and Christi showed the slides of how to participate, become part of the pilot or give us your input. Phase two about costing is going to be concurrent with this.

We're going to continue to get this input, but as we go into the costing, I forget who presented, but one of the presentations showed that 60% of the plans had a zero budget. We don't think that's the case. We do think there's more cost. We really need to spend some time getting into that. And again, I really wanted to thank everybody. Thank you all for taking your time, but really thank the NIH team. Michelle, do you have a final closing comment?

Michelle Bulls: I really just want to echo what you said Jim. We are so grateful for everybody's participation. This is truly a different type of a partnership I always say. We always have the research administrators with the grants administrators at NIH, but to engage and involve, you know, all of NIH. OSP, OER, OER OPERA of course, because we are part of OER. And the institutes and centers and program officials, it makes me really, really happy. This is a really great thing. I'm so honored that the ICs that, you know, decided to participate are with us in this partnership. So, thanks everybody.