The NIH Final Policy for Data Management and Sharing is in Effect

Session Transcript: 2022-2023 NIH Grants Conference

Elyse Sullivan: Thank you for joining today's presentation, "The NIH Final Policy for Data Management and Sharing is in Effect: Planning for Success." My name is Elyse Sullivan, and I am your moderator today. And presenting, we have Dr. Cindy Danielson and Dr. Julia Slutsman, both from the NIH Office of Extramural Research. Cindy and Julia have led the implementation efforts for the data management and sharing policy, and we're thrilled to have them here to answer your questions. And with that, Julia, I'm going to turn this over to you.

Julia Slutsman: Thank you so much, Elyse, and welcome to everyone, and thank you for joining us for the session. We want to start by getting a bit of sense of who's in the audience, so we're going to call up a poll question for you to respond to asking whether you're a researcher, a research administrator or serving in another capacity. Just give a minute for the poll results to come in. So just looking for ... As the results are coming in, it looks like we have a really nice mix of groups today for this presentation, and we'll be working to make sure to address areas relevant to all of you. While the NIH Data Management and Sharing Policy is new, NIH's commitment and investment in data sharing is a long-standing one, and this policy is the most recent one that promotes the NIH mission to improve public health through research and a legislative mandate to make publicly available results of research activities that are NIH-supported or conducted. The data policy went into effect last Wednesday, January 25th, but has been in the making for over 6 years. The policy has really benefited from and evolved in response to community consultations and a consistent, iterative process of community engagement that's really informed the final policy that went into effect last week. I'll just say a word about the high-level goals of data sharing before diving into the specifics and the scope of the policy. The first main goal of data sharing is to advance rigorous and reproducible research. And this is done through enabling validation of research results, making high-value data sets accessible, to accelerate future research directions and increasing opportunities for citation and collaboration. The second broad goal of data sharing is to promote public trust in research, to foster transparency and accountability, to demonstrate stewardship over taxpayer funds and to maximize research participants' contributions while supporting protections for participants' data and privacy. So let me talk a little bit more specifically about the requirements of the policy, and we'll be talking for the remainder of the session about how the policy is implemented and how it meets the goals that I've shared. In terms of policy requirements, there's two requirements that the policy puts in place. The first is that applicants submit a Data Management and Sharing Plan as part of their application. The second requirement is that awardees comply with the NIH-approved Data Management and Sharing Plan, and noncompliance might affect future funding. So in terms of what's in the scope of the policy, the policy applies to all NIH-supported research generating scientific data. And we have a definition of scientific data in the policy, and what we're referring to is data of sufficient quality to validate and replicate research findings. We also define what's not considered scientific data, so some examples of materials that do not meet this definition are lab notebooks, preliminary analyses, physical objects, peer reviews. With regard to when data should be shared, data should be shared no later than at the time of publication for data that underlies a publication or manuscript or at the end of the award for unpublished data, so whichever of these milestones comes first. And the policy gives a good deal of flexibility with regard for how long data should be shared. It really encourages applicants to consider relevant requirements and expectations from their discipline, from journal or repository policy, so there's a great deal of flexibility in how long data can be shared. On our scientific data-sharing website, we list out comprehensively all activities that are subject to the DMS policy. Generally, as I said, the research that generates scientific data is within the scope of the policy. And so examples of activities that generate scientific data include research projects, some career development awards, small business SBIR/STTR awards and research center awards. We also have carved out some activities that do not generate research as well as activities, as well as codes that do not apply, and we've enumerated these, as well, so these include training awards, fellowships, construction awards, research-related infrastructure programs and others. And I encourage you to look at our resources on the scientific data-sharing website to look for information about particular activity codes. In terms of expectations for Data Management and Sharing Plans, our expectation is that sharing becomes the default practice, and that means that we encourage applicants to think about how you can maximize appropriate data sharing. We understand that there are justifiable reasons that may limit the scope of data sharing, and in a minute I'll say more about ethical, legal and technical factors that might limit the scope of data sharing. I also want to stress that while not all scientific data must be shared, all scientific data should be managed. In terms of responsibly implementing data management and sharing, Data Management and Sharing Plans should outline how privacy rights of participants will be protected and how confidentiality will be maintained. Plans could also indicate what existing laws, regulations and policies continue to apply. And we encourage researchers to plan prospectively for how they will implement the DMS policy across all stages of the research process, so this includes during informed consent and communicating how data generated from the research will be shared and used and also, with regard to data submission, thinking about and developing the Data Management and Sharing Plan, how data will be accessed or controlled or publicly available at repositories and whether data will be de-identified or not, so these are all important considerations to plan for in the early stages of research. And on the scientific data-sharing website, we have some additional guidances to help support this planning process. I mentioned that there are a number of justifiable technical, ethical, legal factors that might limit the scope of sharing, and just to give a little bit more information about these examples, such factors include informed consent where there's limitations that might not permit or might limit the scope of data sharing or reuse. Additionally, there might be applicable federal, state, local or Tribal laws, regulations or policies that prohibit disclosure, and these should be described in the DMS plan. Additionally, there are reasons that aren't generally considered justifiable for limiting data sharing, and examples of these include concerns that a data set might be too small, if researchers anticipate the data will not be widely used or considerations that data are not thought to have a suitable repository, and generally these are not considered to be justifiable considerations. We do have additional FAQs speaking to this and additional resources to help with selection of repositories. We have over 25 FAQs that give more information about the policy, and these really cover a range of areas including the scope of the policy, managing and sharing scientific data, FAQs related to considerations for data from human research participants and FAQs related to compliance and enforcement. As an additional resource, we conducted two public webinars over the summer and early into the fall that spoke about implementation of the DMS policy, so we invite you to explore these recording and their associated materials on the scientific data-sharing website. And now, I'll turn it over to Cindy to describe in greater detail information about Data Management and Sharing Plans.

Cindy Danielson: Thank you, Julia. So with that, once you have determined whether the data management and sharing policy applies to an application that you are working on, then your next step would be to develop a Data Management and Sharing Plan. And so we wanted to remind you about what information we recommend including in that plan, and at the time the policy was released at the end of 2020, NIH also published several supplemental policy information documents. And this information about elements of a Data Management and Sharing Plan comes from that document. And so the types of information that you want to include in your plan, I will first start with data type, of what data types will you be generating that you are proposing to preserve and share? Some other information to include relates to tools and software that might be needed to access and manipulate any data that you will share. You'll consider what standards to apply to your scientific data and metadata and then data preservation, access and timelines. You should be indicating if you'll be depositing your data in a repository, what repository you will propose to use, if there will be use of persistent unique identifiers and when and for long any shared data will remain available. If there are any limitations to sharing data such as those that Julia talked through in the FAQ, then you would describe any sort of factors that might limit the sharing that you are able to do, and then finally oversight of data management and sharing, how compliance with your Data Management and Sharing Plan will be monitored and managed. We have received a lot of questions over the past couple years since the policy was received about looking for more details about the format of a Data Management and Sharing Plan. So with that information in mind about what content you want to include, we do have some information to provide on what format you might consider. First, we recommend that they generally be no more than two pages, but there might be some cases where you need more length to describe that if you're, for example, generating many different types of data. And it's not a limit. It's only a recommendation, so it may be appropriate to have a longer plan. We have put out an optional format page, which is kind of a basic template. It is not required, but it does give you somewhere to start, especially if you've never developed a plan like this before. And this format page is available in the list of NIH format pages, and it's also accessible through the application instructions, and it just leaves out those recommended elements to make sure that you are thinking through what information you might need to include for your particular project. And while for this initial implementation, we are leaving things open because we realize that what a data management sharing plan will look like is likely going to vary depending on what area of science you're working in, your specific research project and what data types you're working with. And so we want to learn from this experience what format works best for different types of projects. And for the first year of policy implementation, there is a pilot project, which is under the Federal Demonstration Partnership, or FDP, and through this pilot project that will be testing structured templates and tools for DMS plan submission. And so we hope to learn from this what formats might work best for different areas of science that NIH supports. Another area of feedback and questions that we've been getting are, well, what does a DMS plan actually look like? And so thanks to the efforts of some colleagues at our NIH Institutes and Centers, we have made available some sample DMS plans, and by sample, I mean filled-out examples that provide, for educational purposes, an idea of what information might be relevant to include and how this could be laid out in a plan. And there is a variety of sample plans available on our website covering different types of data, and I will point out that there are some examples specific to genomic data, and so I'm going to turn it back to Julia to talk more about some changes to plans for providing genomic data-sharing plans.

Julia Slutsman: Thank you, Cindy. So NIH has worked to harmonize elements of the genomic data-sharing policy with the data management and sharing policy. And we have a detailed notice about this harmonization, but the biggest takeaway is that there will be a single plan that's required for applications that are subject to both DMS and GDS policy, so there will no longer be a separate GDS plan, just one single Data Management and Sharing Plan. I want to add, too, because I know that we've received some questions about it, about other resource-sharing policies. And so for research that's subject to multiple policies, it's possible to have a separate research and sharing plan in addition to the Data Management and Sharing Plan. And if you're interested in going into great detail about other NIH sharing policies, there will be a session focused on GDS and other policies tomorrow at 11. So throughout the presentation, we have mentioned the NIH scientific data-sharing site, sharing.nih.gov, and we're curious how many of you have had an opportunity to look at the site and to check out some of the resources that we have available, and that's the topic of the next poll. So we're going to go ahead and close the poll. And I think we've been able to get a lot of responses and a lot of yeses, so if you haven't yet had a chance to explore the site, please, please do take a look. And for those of you that have been using the site, we are constantly updating the site as new materials, FAQs and resources are developed. And with regard to resources for writing a Data Management and Sharing Plan, the site has a number of guidances and resources and FAQs that will help further elaborate on expectations, providing a Data Management and Sharing Plan and also describe instructions for submitting plans and budgets, and I'll be turning it over to Cindy now to discuss in greater detail the instructions for submission of DMS plans and budgets as part of applications.

Cindy Danielson: Great. Thanks, Julia. So I want to offer an overview of how you'll submit this information in your application, but I won't get into all the details. I do want to point out that there are more detailed instructions in the application guide, and there are also other sessions at this conference that will cover more about application forms and submission. But the main change is that your Data Management and Sharing Plan will be submitted to a new field that has been added to the PHS 398 forms. We have updated our grant forms to accommodate this, and this new field is called Other Plan(s), and this will accept a single PDF attachment containing your Data Management and Sharing Plan specifically. As Julia mentioned, if you're also subject to the genomic data-sharing policy, you will no longer be providing a separate plan but covering that information in your data management and sharing attachment, which will go into this Other Plan(s) field. I noticed I've seen on the Q&As a lot of the questions about the Resource Sharing Plan(s) field, and so we do want to clarify that the existing Resource Sharing Plan(s) field will remain. That's not being removed. And it will remain to support other policies, mainly the Research Tools and Model Organism Sharing Policy. So if your specific research project is subject to one of these NIH-wide sharing policies, then you will need to include this information in a separate Resource Sharing Plan field. In addition to those NIH-wide policies, certain funding opportunity announcements might contain specific instructions on specific types of Resource Sharing Plans, which would be required. But if there are no additional instructions, and you're not subject to any of these other policies, then you won't need to attach a Resource Sharing Plan. The Data management Sharing Plan goes in a separate location, and that's a separate document. NIH does recognize that sharing data and especially doing so in a way that makes those data useful to others may incur costs. And so one of these supplemental policy information documents released at the time of the policy covered allowable data management and sharing costs. And as a brief summary, you may request direct data management and sharing costs for certain types of activities such as curating data, developing supporting documentation, preserving and sharing data through repositories as well as local data management considerations, so things that are needed to get your data ready to share. In all of these cases, the cost must be incurred during the performance period. This notice, this document, also outlines costs that are considered unallowable for direct data management and sharing costs, and that would include infrastructure costs, which are included in your indirect costs, as well as costs associated with the routine conduct of research. So switching gears a little bit, but when you're thinking about your plans and what sort of budget you'll need, one key piece of information is understanding what approach you're proposing to sharing data. And we have put out supplemental information on repository selection, and NIH does encourage the use of established repositories because this generally improves the findability, accessibility, interoperability and reusability or FAIRness of the data. This information also includes some tips to help investigators help identify appropriate data repositories, pointing out desirable characteristics to look for in a repository. And in some cases, a specific funding opportunity announcement may designate a specific data repository that should be used, but if that's not the case, then you will have the flexibility to determine what repository makes the most sense for your area of research and the data you'll be generating. We do have some information on repositories, and in general, NIH prioritizes data type and discipline-specific data repositories when those are available, and we have made available a list of NIH-supported data repositories. This is not a comprehensive source, and you're not limited to these, but it is a helpful starting point to see some information about what's out there. So once you've determined this basic information about how you'll be managing and sharing data and what that will cost, you will then need to submit your DMS budgets, and the direct cost to support the activities proposed in your DMS plan must be indicated as Data Management and Sharing Costs. If you are submitting a detailed budget with the R&R Budget Form, this will be the line item that you're adding, and in a Modular Budget, that will go into your budget justification attachment. In terms of budget and cost, if you do have specific cost questions, we'll note that our grants policy experts at NIH are the best source of guidance here, and any budget and cost questions can be directed to grantspolicy@nih.gov. And if you're looking for general information on budget and cost, there are also some sessions at the conference here that you may want to check out. In addition to indicating the dollar figure of what you'll be requesting for your data management and sharing costs, you'll also need to justify the budget in a budget justification, and so you'll be including a brief summary, as this data management and sharing justification is part of your larger budget justification and attachment. And I also saw some questions in the Q&A about what happens if you're not submitting any budget request for data management and sharing. And in that case, do take a look at the application instructions and the application guide, and what's indicated there is that you should indicate $0 on that line item, and then you'll also make note of that in the budget justification. So once you have submitted this information and in preparation for an award, NIH program staff at the institute, center or office funding your award will be assessing those plans to determine if they are acceptable, and applications will only be funded once the plan has been found to be complete and acceptable. If there are any issues with the DMS Plan that you have provided in your application, then you will be hearing from NIH staff. They'll reach out to you to resolve any issues and to work together to make sure that you can have a plan that is acceptable to NIH as well as you. And that will occur through the Just In Time process. After an award has been made, once your project is underway, if things change and you need to update the approaches that are listed in your approved DMS Plan, then you will be able to submit a revised DMS Plan. For example, if you are going in a new scientific direction, perhaps a different data repository was identified that's a better fit for your data compared to what was listed initially, or the timeline needs to be revised, and in those cases, those revised plans would, again, be assessed and approved by NIH. In terms of compliance, the approved DMS plan will become a term and condition of your award, and you will report progress on implementing that approved plan in the Research Performance Progress Report, or the RPPR. And those instructions will be updated to clarify what information you'll need to provide. NIH will review compliance with the Data Management and Sharing Plan in the same way that compliance is reviewed for other terms and conditions of awards, and again, for more details on compliance in general, there are sessions of this conference where you can learn more about that. And so finally, I guess next steps, of course, some of you have already prepared DMS Plans and submitted this information if you had applications that were submitted for that January 25th due date, so that's excellent. For those of you who have not done so already, now is a great time to take some steps to get prepared for when you might have to submit a plan. First, it can be very helpful to identify what resources you already have within your institution that might be able to assist you. For example, data librarians can be an excellent source of knowledge in this area. You can take a look at examples of any sample plans that might be relevant to your research. Again, depending on the type of research you're conducting, things will certainly vary, and so it might be helpful to see how others have put together ideas of how to cover this plan information for your area of science. If you've not already tried to, then we'd recommend trying to draft a Data Management and Sharing Plan for your work, even just as a practice exercise if it's not for an application that's coming up anytime soon. And this might not be the first time that you've had to comply with other expectations related to data sharing. There might have been other funder requirements or publisher expectations, and so you can take a look at your past data-sharing practices and think about what you might need to update for this new DMS policy. And then, finally, we covered information in general here, but please do read your funding opportunity announcement carefully because there might be additional institute, center, office or program-specific requirements related to data management and sharing, and you'll want to make sure that you are responding to any sort of requirements that are outlined in that announcement. And with that, before we switch over to live questions, I just want to point out that, after today and beyond this conference, when you have access to us at the booth, you can always look at the website, the central NIH Scientific Data Sharing website, for all of this information including a number of FAQs, which might be especially helpful. You can reach out to sharing@nih.gov for any general questions related to the DMS policy, and you can reach out to grantspolicy@nih.gov for any budget or costing questions. And with that, I will turn it back over to Elyse.

Elyse Sullivan: Thanks so much, Cindy and Julia. This was a crash course in DMS. Thank you so much. We have about 15 minutes for questions from the audience, so thank you all for putting things into the Q&A box. We're going to welcome back Cindy, Julia, and we're going to welcome Taunton Paine from the Office of Science Policy to help answer some of those questions, as well. So why don't we bring you guys all back? And I'm going to start asking some questions from the Q&A box. So there were quite a few questions, and you addressed this a little bit, Cindy, about, what should an applicant do if they're actually not requesting any money, any budget specifically for DMS? So you can write that there's zero. There's $0 going to it. A follow-on question was, must you justify that? Must you include a budget justification for that zero? And a few folks noted that putting in zeros, at least using assist, was giving them an error. So, do you have anything to elaborate, Cindy?

Cindy Danielson: If you are experiencing any sort of errors, then please do reach out to, for technical assistance, to the eRA service desk. In particular, we want to make sure that we're able to help you with any specific situations. But in general, please do look at the application instructions, and that does have some instructions for the budget justification portion explaining that you should justify those costs and that you'll need to include this information when you're required to submit a DMS plan.

Elyse Sullivan: And so, to clarify, Cindy, if you're requesting no dollars, do you need to justify no dollars? Or, do you need to add that part?

Cindy Danielson: Yes, you'll still need to justify that.

Elyse Sullivan: Okay, great. Thank you. Let's see. We have another question, let's see, about the elements in a Data Management and Sharing Plan. So there's, we know there's templates. There's samples. There's six elements. Can you speak to us about, is it required to address all of these elements? Is there leeway in addressing those? Cindy, I'll defer this one to you.

Cindy Danielson: Sure, so good question. And in terms of what information will be provided in response to those elements, that will certainly vary depending upon your proposed strategy. If there are, say, limitations that prevent sharing entirely, then you might not be able to provide, say, the name of a repository. So it is certainly context-dependent, but in general, we do recommend that you address all of these elements and so that we can ensure that they've been adequately addressed and that we do have enough information that the NIH staff will be able to determine that what's represented in that plan is acceptable.

Elyse Sullivan: Great, thank you. So more on plans, when plans change after you've submitted your initial plan, what should folks be doing, I guess, at the time of RPPR or otherwise if they need to update their plan?

Cindy Danielson: So the RPPR instructions will be updated so that by the time these first batch of applications are coming in to report progress, it will be very clear what you need to do. But in general, what will happen at that time is, if there - even before then, if you identify that there is a need to revise a plan, then you should go ahead and reach out to NIH. Make sure that that is done appropriately at the right time, and before anything changes, the NIH is aware of what changes you're requesting. And then also just in general, in terms of progress reporting, that's an opportunity to look back and see, have you met the milestones that you've outlined in your plan? Does anything, maybe something does need to be changed moving forward for the next budget period, and that would be an opportunity to consider at that time.

Elyse Sullivan: Great, thank you. Can you talk a little bit about what might happen if there is noncompliance with a plan? What kind of things might happen? I guess, Julia, do you want to speak to this one?

Julia Slutsman: Sure, so I mentioned earlier that noncompliance could result in - can be taken into account in future funding decisions, and so that's the main thing to look out for. So we encourage staff to make changes, as you and Cindy were discussing, if needed, if the research evolves and if the plan does need to change so that there is an accurate, current plan that will be approved and can be complied with.

Elyse Sullivan: Great, thank you. We have some questions about timing. So we understand that the timing milestone about when data should be shared. We've got a couple questions about, what if you weren't planning on publishing until after the end of the award period? And furthermore, what about if the award period is extended? What do you advise in those situations? Julia, we'll address this one to you.

Julia Slutsman: So if the award period is extended, it's possible to update the DMS plan and request approval for sharing to happen at that later time point at the end of the new performance period that's been negotiated, and we do have some FAQs on our website that speak in greater detail to this. It's a great question.

Elyse Sullivan: Great. And so if somebody wrote in their plan, "My plan is to not share," or what if someone was saying, "I'm going to publish after the award period"? Must they still share by the end of the award period?

Julia Slutsman: Yes. Thank you for that clarification, Elyse. They do need to share by the end of the performance period.

Elyse Sullivan: Great. Let's see. There are some questions about research that only accesses data and does some secondary analyses. Can you speak to what are the expectations for sharing? Or, must you re-share data that presumably has already been shared if you didn't generate it?

Taunton Paine: I'm happy to take that one.

Elyse Sullivan: Great, Taunton.

Taunton Paine: So, yeah, so there's actually a FAQ on the sharing.nih.gov website that goes into some detail about this and indicates that, in general, if you are doing secondary research where you are accessing data, for example, that have been made available through an existing data repository, you are not expected to propose maximizing sharing of those data. However, there may be expectations to share any scientific data that could be generated from your secondary research. But I do want to indicate, and this is also what it says in the frequently asked question, that we've provided guidance sort of suggesting that appropriate limitations may include where you have to enter into agreements that are necessary in order to get access to data or types of materials that might impose certain limitations on your ability to share or disseminate the results, so, for example, a data use agreement under HIPAA or things like terms and conditions for accessing the repository in the first place.

Elyse Sullivan: Thank you, Taunton. This one might be great for you, too, if you could help us out there. If PIs are using data from cores, do they need to include that data in their DMS plans, or a core might need to have their own plan? Can you walk us through that?

Taunton Paine: Sure, so in general, what the policy expects is that each award that is subject to the Data Management and Sharing Policy will have its own plan that includes, in some cases, cores, especially if they are generating data that they are themselves using in scientific research that would meet the actual definition in the policy. I think what we've indicated in frequently asked questions is that we really encourage coordination between the different parties that might be involved here, so if a PI is relying on a core to produce data for them, they would still be expected to share those data. The PI, that's subject to the policy that has a Data Management and Sharing Plan, whose scientific data these are.

Elyse Sullivan: Great. Thank you, and I think we have time for just one more question before we close. There are a few questions about budgets, and I know that our OPERA colleagues are really the experts there on the budgets. But one piece was asking about how effort should be devoted and accounted for with regards to the data management and sharing. Is it accounted in the overall effort, or do you need to account for effort specifically for the data management and sharing piece? And I think, Cindy, would you like to take this?

Cindy Danielson: Sure. Yeah, so just in terms of a general answer, all direct costs that you are requesting for data management and sharing activities, including personnel costs, this must be included in that single line item if you are having a detailed budget, and this would include salary and other information corresponding to the time it takes personnel to undertake the data-related activities. But for more detailed questions on that, we do encourage you to reach out to our grants policy colleagues that will probably have the most updated source of guidance on this.

Elyse Sullivan: Great. Thank you so much. And I think with that, we would like to thank all of our presenters and our panelists and all of our attendees.