Data Resource Sharing:   
Firming Foundations for Future Frontiers

Megan Columbus: Thank you for joining us today. I see that folks are starting to come into the meeting. As we're waiting for a few more people to get here, if you feel like it, get into the chat and introduce yourselves. We're about to introduce ourselves. We're here for a data-sharing presentation. Anybody out there? Let us know! Oh, you guys are shy today. All right. Well, we're going to go ahead and ... Oh! Hello. Thank you, there. Keith, I appreciate it, and Dawn and Tajee. All right. So we do have people in the audience. You know, one of the things about doing virtual webinars is that sometimes you can feel very lonely when you can't actually see your audience. All right. So now that we've got a few hundred of you here, I want to introduce Dr. Julia Slutsman. She's the Director of Genomic Data Sharing Implementation here in NIH's Office of Extramural Research. And J.P. Kim, he's NIH's Extramural Data Sharing Policy Officer. They're here to talk to us about all things data sharing. My name is Megan Columbus. I will be your moderator today helping with the Q and As and that kind of thing. One note: I know I just asked you to introduce yourselves in the chat, and that's great. We will be doing the formal questions and answers through the Q and A box, and so that's what I'll be monitoring. We may be answering a few questions during the presentation, but the presentation is not very long, and so we're hoping that most of it will be Q and A at the end, and I'll be pulling out the themes and answering as many questions as we can. All right. With that, let's get going. Julia?

Julia Slutsman: Thanks so much, Megan, and welcome, everyone, to this presentation on, "Data & Resource Sharing: Firming Foundations for Future Frontiers." J.P. And I will be tag-teaming this presentation. Next slide, please. And what we're hoping to cover this afternoon is to give you some background on a number of NIH data-sharing initiatives with a focus on three specific policies. So we'll discuss the NIH Genomic Data Sharing Policy. We'll give an update, hot off the press, on the development of the NIH Data Management and Sharing Policy and then provide a refresher on the NIH Research Tools Policy. And at the end of the presentation, there are numerous resources so that you could read more and learn more about each of the policies that we discuss. Next slide, please. Data sharing is part of NIH's longstanding commitment to improving public health through research and a mandate to make available to the public the results of biomedical research activities that are supported and conducted by NIH. NIH has put forth multiple policies and research initiatives related to data sharing, starting back about 20 years ago with the NIH Research Tools Policy, and then in the early 2000s, with the beginning and evolution of a number of policies related to genomic data specifically, and then now taking us into 2020 with a new data management and sharing policy. There are a number of benefits, of course, of data sharing and sharing of research resources, which is why this is such an important part of NIH's mission, and I'll just highlight some of these benefits. The first is that broad sharing of data allows for an accelerated generation of research inquiries and the products of research. Data sharing facilitates research integrity and the ability to promote reproducibility of research results and the improvement of rigor and research. Sharing data allows us to maximize fully the contributions of research participants while protecting the data that they provide. It allows NIH as a sponsor of research to be a good steward of resources to minimize redundancy and the kinds of research shared. It maximizes the ability of NIH to invest in diversity and a variety of valuable research. Finally, data sharing allows for data that are difficult to generate or come from limited sources to be used most broadly and allows for stronger analyses and greater statistical power. Next slide. There are a number of ethical considerations, of course, to be taken into account when considering how data is shared, and so with the goal of making data broadly available, it's important to do so in a way that protects research participants' privacy, minimizes the identifiability and maintains privacy and respect for autonomy, and there are a number of mechanisms that NIH policies have established to allow for protection of research participants' privacy and autonomy. These include controlled access for individual-level patient data, giving participants control of how data may be used in the future through the informed consent process and other mechanisms, tools such as merging of data and other mechanisms for minimizing reidentifiability of data and then a process of data access reviews by qualified data access committees to ensure for appropriate provision of data and to maintain public trust. While data sharing has many benefits, of course, it is a challenging thing to do for a number of reasons. First, it requires significant time and effort both on the part of researchers submitting data because it requires data cleaning, preparation, preparation of metadata and tools and then also on the part of data repositories. There's a large infrastructure that's required to allow for proper secure archiving of data and for resources necessary to provide a robust data access and review of data-access requests. Data sharing also requires a high level of policy coordination across the multiple stakeholders, including federal agencies, funders of research, academic institutions and publishers, and there's a number of considerations that need to be agreed upon and developed with regard to better incentivizing data sharing. Additionally, data sharing requires human resources to conduct training, ongoing training in data management and sharing as these practices are well-known to some investigators and may be newer to younger investigators, and so training is important. Data sharing is also something that needs to be socialized and part of the culture and really come to be seen as an activity that's integral to the conduct of research with an acknowledgment that there's needs and times to allow for ways to allow advantages to investigators generating research. Finally, compliance and enforcement is an important part of a fair data sharing, with efforts needed in different points across the life cycle, and at times this can be a challenge. Next slide, please. So shifting now from a broad discussion of the benefits and some challenges of data sharing to a review of some specific policies. The first policy that we'll discuss is the NIH genomic data-sharing, GDS, policies. This policy puts forth principles for broader responsible sharing of large-scale genomic data and associated phenotypic data. The policy contains a number of principles that are really communicated as expectations, so there's expectations related to obtaining informed consent, to IRB review, for how data may be shared and appropriate data-use limitations and then de-identification standards per the common rule and HIPAA and also expectations around data access requests and the use of data committees. The policy has scope such that it applies to all NIH-funded research generating large-scale human or non-human genomic data and the secondary research using this data and applies to all funding mechanisms, regardless of cost. The policy went into effect in January 25th, 2015, for the broader research community and for the NIH Intramural Program in August of that same year. Next slide, please. The GDS policy allows for two levels of access for human data. The first of these is unrestricted access, and that's data that can be made publicly available to anyone. The second level of access is controlled access. Just give a minute to see if our tech will cooperate. Thank you. And controlled access is data that goes through a data-access committee process for being accessible to investigators, and this slide shows two processes under the GDS policy. On the left side for data submission, without going into detail, data submission involves prior IRB review at institutions to allow for any data-use limitations consistent with informed consent to be specified and then an institutional certification, acknowledging that the institution is sharing this data, and then data can be submitted to an NIH-designated repository and under the appropriate level of access control, as we discussed previously. On the right side of this figure, there's a data-access process, so for controlled-access data, there's a data-access request that potential users will submit. It's signed by institutional officials. It allows for particular terms of use in a data-use certification form where the PI agrees to a specific code of conduct, and these data-access requests are reviewed at NIH by one of a number of data-access committees, DACs, as they're known, which review the purpose of the research use compared with any data-use limitations, identify credentials of the investigator requesting the data, and consider the ethical implications for possibility of harm associated with the data use and then make a determination on whether data can be released. Next slide, please. The data-use certification that I mentioned is something that's completed by investigators but signed off by institutional officials at the institution, and it asks that institutions and investigators agree to terms related to the use of data. For example, for using the data only for the research use that's approved, for protecting confidentiality, for not attempting to reidentify individuals, and for reporting any policy violations or data-management institutes to the appropriate NIH data-access committee. Next slide. And I'll turn the presentation over now to J.P. He'll be reviewing the next two policies and sharing some breaking news related to the data management and sharing policy.

J.P. Kim: Good afternoon, good morning, wherever you may be. This is J.P. Kim. Hi. How are you? Hopefully you can see me. So I am actually what I like to refer to as a recovering patent attorney and a recovering licensing specialist here at the NIH. I've been here for over 20 years, and I've been working on data and resource sharing for most of that time, so hopefully we'll be able to share things with you and help you to, sorry, help you better understand, actually, the sharing ... what is it ... policies and precepts here at the NIH. So one of the things you should be aware of is, actually, although as Julia mentioned, the sharing of research, resources and data and such has been something that has been a longstanding commitment of the NIH. It's been for many decades. But one of the interesting things that's happened is that in February of 2013, which seems so long ago because it's about 7 years ago, the White House Office of Science and Technology Policy actually released a memo. It's called the Holdren Memo, basically, you know, requiring agencies to actually require more sharing of data and resources. So basically what happened is, in response to that, in February of 2015, NIH released the NIH Plan for this. I'll talk to you a little bit more about the memo and what was specifically in it in a moment, but just first an overview here. The NIH Plan was actually released in February 2015, and it was broken up into publications which were actually under the NIH Public Access Policy, which many of you may be aware of because that actually has to do with journal articles and things that result from NIH-funded research. And, of course, the other part was about digital scientific data through a plan for a public access to these. We needed to consider how to require data sharing, consider how to require and evaluate the data management and the sharing plans, encourage the use of existing repositories and the like, and promote the FAIR principles, which was FAIR standing for, as many of you probably understand is, Findable, Accessible, Interoperable and Reusable. So this was in February of 2015, but just to emphasize, the plan was not the policy, actually. That was just the plan that we were required to actually come forward with under the memo. So NIH had planned to make established priorities for data sharing and the like. So here we go, a little bit more about that particular memo. It was ... Excuse me. It was actually released, as I said, in February 2013 from the Office of Science and Technology Policy and called "Increasing Access to the Results of Federally Funded Scientific Research." Basically what happened was, the memo required the development of agency plans of various agencies that actually funded federal research, and it applies to peer-reviewed publications and to digital scientific data. Some of the objectives of the Holdren Memo? Well, actually, I'm going to actually focus on the right-hand side where it talks about scholarly publications because as you can see, there's actually a number of things that need to be addressed under the Holdren Memo in scholarly publications, but fortunately for the NIH, we already had our NIH Public Access Plan which has, I believe, it's been in place since 2008 or so. And so I think in many ways, I think the memo actually reflected a lot of the sharing precepts and policies of the NIH. So with respect to the publication aspect of the Holdren Memo, we were fully compliant with that. When it came to digital data, we actually were quite far along the curve because we had been promoting data sharing for many years now. And as we said earlier, it is actually one of the earlier foundational longstanding commitments of the NIH when it was first established and it's been used on. So apparently, as you can see here, we want to maximize free access, but, of course, we need to recognize that we need to protect the privacy and confidentiality and national-security issues. We also need to recognize intellectual property rights because we do want, actually, things to be developed as well, ultimately, products and services that actually help the community, and, of course, balancing costs and benefits of long-term preservation. So some of the objectives of the Holdren Memo was that you require data-management plans as part of your applications, allow the inclusion of costs in applications for funding, ensure the appropriate evaluation of these plans when you're funding applications come in. They did require the monitoring of compliance by the investigators, basically compliance and enforcement later on. Also under the memo, to encourage deposit of data in public repositories where possible because public repositories actually takes the burden off the researchers to do all the sharing and, actually, it provides a consistent resource and perhaps a streamlined way that data and other things can be shared. Actually, also said to cooperate with the private sector in achieving these and develop approaches for data citation and attribution, such that basically there's a consistency, that it's easy to compare apples with, well, apples with oranges. And so you don't want to compare apples and oranges. You want them to be consistent so you can compare apples with apples when you're doing your research and trying to do citation and attribution. We also support training education and workforce development for this and assess the long-term needs for preservation and options for repositories, and there we talk about repositories again. So one of the things that we should note to you or will note to you is that the NIH Data Management and Sharing Policy has since then, and actually a little before then, even, been in development. So some of the highlights is that in November of 2016, we actually released an RFI, which is a request for information, about the strategies for NIH data management, sharing and citation, basically. It's part of our policy-development process to engage the research community and such who we work with so closely. In October of 2018, we actually provided another RFI on proposed revisions, or provisions, sorry, for the Draft Data Management and Sharing Policy for NIH Funded or Supported Research. Basically, that was we actually had a draft policy and some guidance, and we wanted to get the feedback from the community, actually, on these aspects. And so having taken that into effect, we actually, then, we, in November of 2019, we did another request for public comments, and we issued a DRAFT NIH Policy on Data Management and Sharing and Supplemental DRAFT Guidance. And that, actually, we had from November until January 10th of 2020. And since that time, we've been working on developing the NIH Data Management and Sharing Policy. And as you can see, the fourth bullet says, "And now," because, breaking news, actually, as Julia was referring to. The Final NIH Policy for Data Management and Sharing, otherwise known as the DMS Policy, was released just yesterday, October 29th. Now, just so that you all are aware, don't worry. Even though it just came out, you have time to understand, and, actually, we have time to actually help interact with the community more because it doesn't have an effective date until January 25th, 2023. That's over 2 years from now. So what happened was, if you're looking for it, you can actually go into the NIH Guide, and you can see the policy itself, the Final NIH Policy for Data Management and Sharing. It's Notice OD-21-013, and the link is here, and we will be providing you with this information, and the slides should be available to you after the call as well. But I'm sure you're actually all wondering, "What does the DMS Policy apply to?" Well, basically it applies to all research funded or conducted in whole or in part by NIH that results in the generation of scientific data. The reason, actually, you should note that it's in the generation of scientific data so that, for example, certain things it wouldn't apply to, like construction grants and basically saying some of the training grants because they weren't actually supposed to be focused on generation of scientific data. However, other mechanisms and the like would be applicable, too. And what else would it apply to? It applies to extramural grants, contracts. It actually also applies to our intramural research programs, so intramural research projects and other funding agreements regardless of NIH funding level of funding mechanism. The reason the, "Regardless of NIH funding level," is underlined is because our previous 2003 policy or, as you may know it as, the 2003 NIH Data Sharing Policy, that actually had a 500,000 threshold level, and so this particular policy does not, and so it applies to actually any funding level that basically would be generation of scientific data. And I should also note to you that this new policy, the DMS Policy, actually replaces the 2003 Data Sharing Policy. So then, of course, you might be wondering, "What's going to be required in funding applications starting January 25th, 2023, under this policy?" Well, basically in 2023, what will be required will be the submission of a data management and sharing plan that outlines how scientific data and any accompanying metadata will be managed and shared, taking into account any potential restrictions or limitations. In other words, there may be valid reasons for not being able to share, and so that would actually be something that would be addressed. So with data, actually, there could be data that may not ... There may be some restrictions about sharing, but we would still be managing it, so basically the data-management plan or data management and sharing plan would address that. And, of course, the other aspect of this would be that you submit it and you comply with your plan and whatever plan is submitted and approved by the NIH institute, center or office, basically. So we should note to you, if you have any questions about the new policy since it just came out, you should actually direct them to the NIH Office of Science Policy. The e-mail address is sciencepolicy@od.nih.gov. That's sciencepolicy@od.nih.gov. And basically what the plan is, actually, also, to develop a frequently asked questions and actually have that available over the next 2 years as well as we develop and find more questions that need answering, and we're trying to provide that type of input back to you all. But as I said, again, the policy just released yesterday, an effective data of January 25th, 2023. And, of course, we actually also have some additional things that were issued yesterday in the NIH Guide. For example, things you're probably thinking about. What should be in a plan? Which costs are allowable? How are repositories selected if we're using a repository? Well, again, to help you along, there was supplemental information to the DMS Policy which was also released yesterday. As you can see here, one of the things is basically the, excuse me, "Elements of an NIH Data Management and Sharing Plan." That's notice OD-21-014, so that would actually help you address what should be in a plan. We also have addressed another notice which actually talks about allowable costs for data management and sharing, so that addresses or could address for you which costs are allowable, and that's notice OD-21-015. And finally, how are repositories selected? As we've told you, repositories are actually a very good resource to be able to share data and resources and the like, and so there is an additional notice called "Selecting a Repository for Data Resulting from NIH-Supported Research" which is notice OD-21-016. So all of these four, the policy and these three supplemental information notices, are there to help you, and as I said, because the policy actually doesn't have an effective date until 2023, there is time for asking questions, and we are going to try to, as I said, develop FAQs and additional guidance and the like basically for how the policy is implemented and how you all would work with that. So now having told you the breaking news, I was going to get back to a little bit more general aspects, is that basically why we think that data and resource sharing is so important is basically data and resource sharing does, as you can see from this pyramid, it forms the foundation for the research community, and then the sharing and dissemination actually leads to accelerated scientific inquiry. And so it advances scientific research which can lead to products and services which will benefit the public health. So one of the policies that we want to talk to you, because we talked to you about data-sharing policies, well, I want to talk to you a little bit about research tools policies because this was something that actually happened in the 1990s, but it's one of those ... "An oldie but goody," as they say. As you can see, the title is, "Sharing Biomedical Research Resources: Principles and Guidelines for Recipients of NIH Research Grants & Contracts." A long title, and we actually affectionately just refer to it as the NIH Research Tools Policy which is actually known probably ... That shorter title is known throughout the academic community, basically, of NIH-funded recipients. Basically one of the things is that the policy itself is made up of principles and guidelines. What happened was, when it was first announced, basically the principles were announced, and we got feedback, and, again, so the policy developed, but it was interesting because the guidelines provided examples and ways to implement the policy. And so the community said that these really need to be or should be a part of it. So what the NIH did was, they issued the NIH Research Tools Policy as principles and guidelines. So as you see here, the basic principles, very foundational for research: ensure academic freedom and publication, ensure appropriate implementation of the Bayh-Dole Act, minimize administrative impediments to academic research and ensure dissemination of research resources developed with NIH funds. And as I said, the guidelines basically provide an appendix of different ways how the research tools can be shared. And I want to give you one example, basically, because this actually happened back in the 1990s when model mice were actually proving to be an interesting issue because people were actually having trouble accessing them. So I'll give you an example, is that ... I'm going to give you an exaggerated example, basically. It's where basically a mouse was developed, let's say, by a researcher, and let's say that the researcher was willing to share but would charge, I'm just making this up, $1 million for the mouse at each of the active institutions. And so that would actually be an impediment, an administrative impediment to academic research. So basically this is ... Reasons like this are the reasons that the Research Tools Policy actually came about. And one of the things that you should be aware of when it comes to data and research resources for yourself is: don't forget, when you're doing research and you're actually sharing data and sharing resources, you're actually ensuring that academic research can continue for not just everyone else, but yourself, because we've even had an instance or two where a researcher actually made certain that there was limitations, shall we say, of sharing of data and resources actually at their institutions so that they could do their research, but then the research moved to another institution and then couldn't have access to those resources any longer. So basically by protecting the integrity of research, by sharing research resources and data, you're actually protecting the public health basically, but you're also potentially protecting your own research in going forward. So basically, here are a few select NIH grants policy excerpts. By the way, as I told you earlier, well, I should say earlier, the NIH Research Tools Policy applies to the grants, contracts, and cooperative agreements. One of the things you should know about in our grants policies excepts is that grantees do own the data they develop with federal funds. That's in Grants Policy Statement 8.2.1. With respect to our public-access policy, final peer-reviewed manuscripts, they should be shared upon acceptance for publication, must be published at PubMed Central at pubmedcentral.nih.gov. And finally, as I talked to you about earlier, the NIH Research Tools Policy requires sharing of unique research materials and biological materials, otherwise known as research tools, and that's also noted in the Grants Policy Statement. So as Julia had noted, there's some key resource links for you to gain access to learning more about these policies, and as you can see, the Genomic Data Sharing Policy. I won't recite the whole length of them, the Genomic Data Sharing Policy, the Research Tools Policy, the Public Access Policy and the Data Sharing Policies. However, I will tell you with respect to the last two, and basically, again, the NIH Public Access Policy, that's an easy reference, publicaccess.nih.gov. And as for that NIH Data Sharing Policies, that particular website actually is somewhat accumulative of all of the policies that have been issued--sharing policies--that have been issued by the NIH up until, let's say, until yesterday because they haven't added the data management sharing policies yet. But the shortcut for the Data Sharing Policies is sharing.nih.gov, just so you know. Okay? And finally, as acknowledgements, we'd like thank Kristofor Langlais, Erin Luetkemeier, Cheryl Smith, Ellen Wann and Dina Paltoo for their contributions to this presentation. So having said that, we will now open it up to questions, and I will now stop sharing.

Megan Columbus: All right. Hey, thanks so much to both of you. That was really informative. We have some questions. I do have a question for our tech folk. If we have somebody who wants to ask a live question, Peggy, are we able to get them on live?

Peggy: Yes, we can.

Megan Columbus: Okay. Could you ...

Peggy: People raising their hands or ...

Megan Columbus: Yeah. Could you put Pradeep Ramana live so she can ask a question on data redistribution after the first steps of data sharing?

Peggy: Yes. You are live.

Megan Columbus: He, I'm so sorry.

Pradeep Raamana: Thank you. Can you guys hear me okay?

Julia Slutsman: Yes.

J.P. Kim: Yes, we can.

Pradeep Raamana: Perfect. Thank you for the opportunity. So this is wonderful to see NIH takes such a great care and think about all these in an equitable fashion. So one of the things I found rather limiting was that although many great data sets are shared, I come from the neuroscience area, they have a clause prohibiting redistribution. For example, I get the data set on Alzheimer's, I can use it for myself, but once I process it and produce something useful, I cannot give it to another person. Right? And some of these outputs are produced with years and years of hard work that cannot be easily produced elsewhere, not just because of lack of computing power, but also lack of expertise in actually doing the same thing. Right? So this prohibition of redistribution is rather very limiting not just to us, but also is actually limiting the potential of the return that the NIH can get out of the same data set. So I personally feel that NIH has an integral responsibility to remove the clause going forward, or at least not letting the PIs of these multi-consortia data set-producing grants to put such clauses. There may be some good reasons in terms of putting the restrictions in how they can use it, but I think so long as ethical and privacy and other policy concerns are addressed that only ... I think researchers should be able to share the data anonymized and in an ethical manner with others to maximize their returns.

Megan Columbus: All right.

Pradeep Raamana: Yeah.

Megan Columbus: Let me cut you off there just because I want to make sure we have time to get to other questions. That's certainly ... I understood that point. A response?

Pradeep Raamana: Right, so ...

Megan Columbus: Julia?

J.P. Kim: May I, or, Julia, would you like to?

Julia Slutsman: I can start, and then we'll tag team, and thank you for that question. Something that is a challenge with data sharing is how to maximize these data while protecting participant privacy and minimizing the potential for reidentifiability, and one challenge is, as different data sets are put together, as there are different uses that may not have been imagined and approved in the initial use request that was approved by a data-access committee or similar mechanism, that potential can sometimes increase for reidentifiability. And it's hard always to know in advance how that plays out, so that rationale of trying to strike a good balance between protecting participant privacy and maximizing usability is one that underlines, often, these clauses and the control data access process that has been set up with a data-access review of every request. But obviously there's always that tension because it does have limitations on, then, the usability of the data, so that tension is well taken. J.P., do you want to add to that?

J.P. Kim: Sure. Yeah, so, hey, Pradeep, that is actually a very excellent question. Yeah, so the NIH does actually understand, I think, probably what we were talking about, sort of like secondary-analysis availability or secondary analysis and actually moving forward and sharing that data or information that you come up with afterwards. But of course, there are, as you know and as I think you referred to as well, potential limitations from the informed consents and other things that actually may have an effect on, even though you have access to it, there are limitations that must still be consistent with the informed consent. So if, however, if you're talking about getting data from other institutions as opposed to from dbGaP and the like, well, I would actually recommend you considering doing ... I should ... Remember, it's institutions that are actually responsible, ultimately, for sharing of data and such. So if you can talk with your sponsored research office ... I don't think you should actually address it directly with the PI necessarily, but I think you should talk to your sponsored research office about the appropriateness of the limitation. And then what'll happen is that that sponsored research office can talk with the other PI or institution as well, and they can actually discuss why these limitations are there and perhaps maybe the limitations may or may not be applicable. But I would recommend that ...

Megan Columbus: All right.

J.P. Kim: ... and if that doesn't work, then basically, I think you should actually write our office, or write me, jpkim@ni ... Oh, gosh. I'm telling you my e-mail address. Oh, it's on there anyway: jpkim@nih.gov.

[ Chatter ]

J.P. Kim: ... follow up with me, so, Pradeep.

Megan Columbus: I'm sorry. So we have a bunch of questions that we need to get to, so if I can be a little bit rude and cut you off. I think there's an essential question that, Julia, you answered some in text, but I'd like to get back to it, and it's about, "What is a large-scale genomic data-sharing set?" I know that there's some variations between institutes, and it can be a little bit hard, right?

Julia Slutsman: Yeah, and that's a great question. There was a series of interrelated questions related to what is meant in the policy by large-scale genomic data, so thank you for those, and I have a kind of frustrating answer to that, which is that the policy was intentionally written broadly recognizing that science would evolve, genomic analyses would evolve and it would be difficult to put a number on that that would be meaningful in the future. So NIH has a supplemental guidance to the GDS Policy with some examples, some use cases, to help investigators think about that, and there's a link in the Q and A that I shared to that. And then, as Megan mentioned, individual ICs also have additional guidance for their awardees about that, both about the size and often questions about the kinds of specific analyses and data types, but it's an excellent series of questions.

Megan Columbus: So another question that has to do with comprehensive, "What about the RNA sequence analysis of particular cell types within an animal? Does that count as comprehensive?" And there's a follow-up to that as well that leaves us asking, "Does performing RNA sequence analysis in one model organism in one cell type qualify? And if not, how many cell types push you over the threshold?"

Julia Slutsman: And again, for those very, very specific questions, I encourage you to look at the supplemental guidance which has some very detailed examples and to look at the specific IC for the award to see what their guidance is because there isn't an easy sort of "one size fits all" answer for that, and we're happy to point you to the right resources if you have a few further questions. Reach out to us.

Megan Columbus: So somebody from Europe is asking about advice to approaching data sharing within European institutions with respect to the EU's GDPR.

Julia Slutsman: So that's a good question, and I can start, and we'll tag team. So obviously NIH is aware of the policy and has been thinking and talking with institutions internationally about that and how best to interpret that, and I think that's an area for evolving guidance, since certainly at the time of some of the earlier policies, the European act didn't exist, so that's a great question, and I think their guidance will come. J.P., do you want to speak to that as well?

J.P. Kim: No. I'd have to agree with you. I'd have to agree with you basically because of the global environment and how everything is being shared, yes. That is actually something that probably hasn't been fully addressed yet, and so that is actually something that is evolving over time, so ...

Megan Columbus: "So what about the sharing of negative data that are not published? Is there somewhere to do that?"

Julia Slutsman: So in terms of the newer policy that applies both to negative data and positive, it doesn't differentiate between the two. So the guidelines for the policy in terms of what mechanism would be proposed in the Data Sharing Plan would apply to both types of data.

J.P. Kim: Mm-hmm.

Megan Columbus: And to follow up on that, then, "What's the role of the data descriptors journals, the scientific "Data in Brief," those kinds of things? Can or should they be part of the Data Sharing Plan, and does the NIH link these data-description publications with PubMed since the articles or the sources of grants and reporter?"

J.P. Kim: I'm not certain what data descriptors are, Julia. I'm sorry.

Megan Columbus: "Data-descriptor journals."

Julia Slutsman: Data-descriptor journals? So I'm not sure what that refers to either. So in one of the notices that J.P. pointed to in our presentation, there's a supplemental guidance that talks more about the Data Sharing Plans and what those need to ...

J.P. Kim: The element, right. The elements of those plans, right.

Julia Slutsman: ... involve, although it's not at this level of detail. But these questions are very helpful to us, especially at ... And they're well-timed as well because, as J.P. mentioned, we have 2 years as NIH will have lots of opportunities to interact with all of you in the research community and to hear the kinds of questions that need to be addressed with regard to the new policy. So we're certainly listening, and we'll be sharing these questions with the Office of Science Policy and other stakeholders across NIH to help inform future guidance, so thank you for that question.

J.P. Kim: Mm-hmm.

Megan Columbus: Well, and throughout the conference, I think we're going to be taking all the questions that we're getting and completely trying to beef up all of our FAQs and other things so that people will have answers at hand. We only have a few minutes left, so I want to make sure to get a few more questions in. "If an RO1 grant required the submission of an extramural invention certification under a PI who is no longer living, will a new extramural institution certification need to be submitted under the new PI who inherited the RO1 grant?" Did you get that or should ...

J.P. Kim: Well, okay, no, no. I'm sorry. So this isn't exactly data sharing, basically. It's extramural invention, so ...

Megan Columbus: I'm sorry. I guess I heard, "Inventions." So you know what? This one, I think this is probably a better question to go to if you go to the Ask a Policy Officer booth, they've got policy officers who are ready to answer this.

J.P. Kim: And basically ... But one thing we have to remember, though, is that an award is given to an institution and generally not to the PI directly. So if it's an institutional thing and basically they're sort of substituting in a PI for that particular thing, I think that it ... Well, I would defer to what Ask a Policy says, but that's just something that it may actually be sufficient.

Megan Columbus: Great. "What about an individual who wants to get access to a genomic data? Is that available to anybody?"

Julia Slutsman: So any individual can submit a request to access genomic data, and it'll be reviewed through the data-access committee process, and determination will be made.

Megan Columbus: "For data attribution, would sharing of data be seen as a plus, like publications are in a grant review?"

Julia Slutsman: So I think certainly that's part of the culture change that NIH is hoping to continue to spark and to facilitate, that data sharing broadly will come to be seen as both expectation and as part of what's expected broadly.

J.P. Kim: Right, and I think that basically it's as institutions, academic institutions, look at this and start to look at it more as the culture changes that the sharing of data and the like and how it contributed to other publications may be considered in tenure or tenure-track considerations as well depending on the institution.

Megan Columbus: Yeah. So we've got a couple questions from small businesses who are concerned about intellectual property and data sharing and what this means for small-business grants. Any thoughts?

J.P. Kim: We love small businesses. They're fantastic. So basically, the main question I think is, "What about SBIR, STTR grants where IP issues might arise?" from an anonymous attendee. That's really fine. "Do they have to share data?" Well, actually, I think what we'd note to you is that, yes, with respect to SBIR and STTR grants' data actually, that could be a reason to ... It could be one of the limitations, actually, for why sharing may not be possible. However, one of the things that we need to emphasize to you is that when you're applying for a grant, if it's the solicitation that the NIH put out under an SBIR, you have to look at what the goals are. What needs to be achieved? So if you can't meet the goals of the SBIR FOA, then you have to actually ... That would be probably something you would need to talk with the program officer ahead of time. However, with respect to the ones that you initiate, R41s, 42s, 43s, 44s, with respect to data, there is actually ... The SBA actually does provide, I believe, a data-exclusivity period for the small businesses, so you would actually be able to cite those. But again, bottom line, you are doing work that's going to benefit the public health, so see if you can find the balance between sharing data and actually still achieving your commercial goals in getting a product or service out to the market.

Megan Columbus: All right. I want to be sensitive to the fact that we're now a little bit over time. Thank you for all the questions, and thank you for all of the answers, and we'll be looking at this as we look forward to continued implementation of these policies. All right? Thanks, all.

J.P. Kim: Thank you, everyone. We appreciate your questions.

Julia Slutsman: Thank you. Goodbye.