The New NIH Data Management & Sharing Policy

Navigate Webinar

November 1, 2022

The information in this document is current as of December 20, 2022

Questions & Answers

Will ORSP and/or the Medical School check the data sharing plan as part of their usual application review before submission to NIH?

We can't answer this with a complete yes or no (yet). Likely the most that they would do is:

- Check that there is a data management and sharing plan included with the application;
- If they have the time to do so, they may check to see that the specific elements are listed there that are supposed to be in the data management and sharing plan.

Are there some recommended repositories and or data formats?

The NIH does have recommended repositories. The data format is going to be dependent on the type of data that you're working with. But repositories will often have information about recommended formats based on how open they are, and how easy they are to share or reuse.

Are there requirements or guidance for detail in data dictionaries?

The biggest thing is, when you're writing a data dictionary, to think about how someone else is going to understand the data that you've collected. For example:

- Can you provide enough detail to help them fully understand?
- What do each of these variables mean?
- How are they collected?

If a data dictionary is promised, is there any enforcement of this? Would it be assumed to be part of the repository?

Some of that is still up in the air, as far as how much NIH is going to be digging into the details of these plans, and seeing if you've actually done what you said you would do. They are counting on the researchers to make updates to those plans as necessary. They are able to be changed and that's something that the NIH has emphasized - that this is something that can involve a discussion with your program officer when you're checking in on those progress reports. They really want researchers to make sure that those plans are staying up to date, and

accurate with what's going on with the project at the time. So if they need to be updated, if something's changed with the project, definitely update that.

Regarding compliance and enforcement, it looks like it is going to be a question on the progress reports, and then additional compliance, monitoring processes may be implemented in the future.

If you're sharing data with a repository, there is likely going to be a data curator that will review the data being deposited. If a data dictionary isn't already included with the deposit, they will suggest including one to help others to better understand and use the data in the future.

If you use samples collected through NIH-funded research for sponsored research (that would generate new data) - is a data sharing plan needed for such a research proposal as well?

If you mean samples such as tissue samples, or something like that, those are not considered data.

Does using retrospective data also constitute new data, and hence require a data sharing plan?

Any new applications to NIH, beginning on January 25, 2023, need to include a data management and sharing plan. It doesn't matter if the data are going to be newly generated or if you're somehow using existing data - there needs to be a plan included. What that plan contains will be different if you're going to be using data that already exists and you'll need to describe any limitations that the data has. For example, maybe there's only part of the project data that you'll be able to share.

If you're talking about using someone else's data set to do something else, that's where it gets a little more borderline and complicated. You could certainly talk to us, and we can help connect with NIH on the specifics. But retrospective gathering of data would still be a new data set that's being created.

In general, the NIH wants to see researchers really thinking about their data prior to beginning their project and making a plan for how they're going to handle it. It doesn't actually matter what kind of data it is.

This is for NIH proposals, not active research data sharing plans that need to be revised?

Yes. This new policy is explicitly for new NIH proposals. That being said, I have heard that for some of the institutes and centers, some of the program managers are asking researchers to try to think about putting their current research data plans into this new format. In general, it's applying to new applications on or after January 25th.

Do you have a format page that is completed, so we can see which parts were deleted?

Within the Format Page, there is bolded text and text that is in italics. When you're filling out the format page, you can delete those extra descriptions in italics. They're there to help you figure out what information should be included.

How about non-research retrospective data that are collected as part of clinical care, i.e., secondary use of data?

With respect to that, you probably want to start by talking with the Data Office for Clinical & Translational Research (DOCTR), along with people like Sara Samuel. As far as your plans go, the DOCTR office runs the committee for the consideration of the use of clinical data, and, generally speaking, for research purposes, there's quite a bit of flexibility there. But we are going into new territory here. So if it was collected clinically under clinical consents that did not refer to use in research context, then there may be some real constraints. And again we want to honor the NIH's desire to share more but we can't re-engineer the past, and we don't want to create new versions of the Henrietta Lacks problem, or other cases where people feel taken advantage of, or hurt by things that we're doing.

If you believe there is a constraint, look hard and deep to understand the borders of that constraint. Talk to DOCTR or talk to the Office of Regulatory Affairs - we will help dig into it and figure it out. It's okay to be a little on the cautious side, but just explain that's what your plan is.

What is a persistent identifier?

Usually the persistent identifier that people will get is a digital object identifier (DOI). Repositories will assign a unique DOI to each data set that is deposited in the repository. The persistent identifier will not change over time and is a key piece of information that should be included when citing a data set.

So OSF or Zenodo will auto-generate a DOI for you (some of the others will too)?

Yes, most repositories have that feature like when you deposit a data set; when it's published, you will receive a DOI for your data. That includes Deep Blue Data here at University of Michigan, as well.

Can you talk about qualitative data sharing, in particular data that cannot be sufficiently redacted to protect privacy and confidentiality. I'm thinking about several data sets we have, and likely to generate in the future, where we could redact by redacting the identifying elements like name, location.

This is getting back to the idea of maximizing appropriate data sharing and trying to figure out what is the right tension between confidentiality and protection and data sharing. This is something that I think in general researchers and the NIH are going to be figuring out for a while.

One thing to think about with qualitative data: are there aspects of the study, data collection, or study design that could be shared to help someone essentially replicate the study? For example, what are the interview questions that were used, how was the data coded, things like that. That is something to consider when you're writing a plan.

I think it's important that people acknowledge if there's going to be any sort of issue with sharing the data. That's something that can be written about within the data sharing plan, saying we won't be able to share this aspect of the data. However, what we can share is X, Y, and Z, for example, so that's something to consider. That is something that would be good to have a conversation with a program manager about as well.

What if the researcher is collecting sensitive information, or is nervous about sharing?

The key here is really to think carefully about what the data you're gathering is, what makes it sensitive, and if there are appropriate ways to share it.

For example, suppose you're working intensely or exclusively with a specific population, for example, a Native American group. There is a specific extra document that NIH has posted about what really honors the difference that we have to tribal authorities. If the tribal authority had a limitation on data sharing, then you would identify that limitation, and you would constrain your data sharing to meet and honor those restrictions and obligations. In many other circumstances, it is expected that you would be able to find a repository whose policies are appropriately structured and that are suitable for your research. There are some repositories where you're getting full access to the data, and others where you might be using it in a very sequestered enclave, where you can't actually remove parts of it. If it is human subjects research, you want to be sure that you build transparency into your informed consent

documents with what your data sharing plan is so that participants understand what they are agreeing to.

Won't this timeframe also affect the budget? For example, the costs for the data sharing have to be spent before the end of the grant period?

Any costs have to be incurred before the end of the grant period. One thing that NIH has said is that it's possible that no cost extensions may be utilized to help researchers make sure that they're meeting their data sharing requirements prior to the end of the project.

Will MICHR or CTSU have a professional data manager on a charge per hour basis?

MICHR provides data management mentoring and database development services, which may be related to what you're asking about. I will also bring that question forward to both MICHR and the CTSU to see that as we all move forward, leadership is aware of budgetary resource needs on the part of the university community.

Is this DMSP a document that needs to be updated if things change over the course of multiyear studies? For example, you mentioned we need to be very specific regarding software packages, and we recently needed to update our software version numbers used for analysis.

Yes, that is part of the oversight and responsibility piece. The PI is most likely going to be responsible for making sure to keep the DMS Plan updated. And again, that can be a conversation with the program officer whenever it is needed to make sure that the document is staying updated as needed.

Will researchers be able to share their data on Deep Blue?

It depends. The best place to share data is usually an appropriate disciplinary repository, if it exists. Deep Blue Data meets the NIH's list of desired characteristics of a data repository. But researchers will need to ensure the data are able to be freely shared with the public (consider informed consent, PHI, etc.).

On the NIH website there is a list of repositories, and it is almost kind of a finder tool for data repositories. That is a really good first place to start looking, because if there is a domain-specific repository, that's probably going to already have considered many of those sensitivity questions that you have. If none of those work for you, there are some generalist repositories that are part of NIH's Generalist Repository Ecosystem Initiative (GREI).

What about retrospective studies where the informed consent document did not consider the data sharing plan?

If human subjects data was collected in the past without appropriate informed consent indicating that the data may be shared, then the data should not be shared. That is a limitation for sharing data, and the NIH has indicated that there are some justifiable reasons why data cannot be shared. Again, the emphasis of the NIH is that DMS plans should be able to describe the data that will be generated or collected, describe what aspects of the data will be shared, and articulate any reasonable limitations for why the data may not be able to be fully shared.