

**Luann Good Gingrich:** Thank you. I'm the academic director of Research Commons. We're so pleased to have partnered with the Open Scholarship Department in libraries to bring you today's workshop. Which we've entitled Creating a Data Management Plan or DMP for your research project. This is a regular workshop that libraries hosts and we've collaborated, collaborated with them to gear this specific occurrence of the workshop to faculty members. The workshop will be led by the Open Scholarship Department's acting director, Stephanie Quail, and scholarly communications librarian, Lily Ren.

We're also so pleased to have Professor Randy Lewis from the Faculty of Science share his experience with creating a DMP for a grant proposal. So before we get, begin, just a couple of housekeeping items. Please note as you just heard that the workshop is being recorded and will be sent to everyone who has registered along with the presenters files or slides. I'd also like to. Invite you all to please complete the evaluation form after the workshop. Your feedback helps us tremendously for planning upcoming workshops, our operations manager, Hajer Mirwali, will share the link in the chat now.

And we can share it again toward the end of the workshop. So on behalf of Research Commons, I want to thank you for being here today and I'll now pass things over to Stephanie and Lily.

**Stephanie Quail:** Thank you, Luann. As Luann mentioned, we are recording this session today and so for those of us, those of you who are just joining a little bit later, you may not have heard that, that the session will be recorded and you will receive a copy of the recording as well as the slides. I'm just going to go to the next slide here.

And before we start, I did just want to provide a land acknowledgement. So I am of settler colonial ancestry and my heritage extends to England, France and Germany. And I, like many other settler people, have benefited greatly from living on Turtle Island as a visitor. I recognize my responsibility to acknowledge the grounds on which we are privileged to gather today to have our conversations. I'm on campus today and I will be sharing the land acknowledgment for York University and I encourage others to recognize and reflect on the traditional territories on which they reside. York University recognizes that many Indigenous Nations have longstanding relationships with the territories upon which York University campuses are located that precede the establishment of York University. York University acknowledges its presence on the traditional territory of many Indigenous Nations. The area known as Tkaronto has been care taken by the Anishinabek Nation, the Haudenosaunee Confederacy, and the Huron-Wendat. It is now home to many First Nation, Inuit and Métis communities. We acknowledge the current treaty holders, the Mississaugas of the Credit First Nation. This territory is subject of the Dish with One Spoon Wampum Belt Covenant, an agreement to peaceably share and care for the Great Lakes region.

Alright. So, oh, sorry, I realize there's a slight issue with the slides there. So hopefully, We won't have too many little weird issues there with the slides. But just to briefly introduce myself, so I am, Stephanie. I am the director of the Library's Open Scholarship Department and this department in the libraries is responsible for providing supports in the areas of open access

publishing, research, data management, research visibility, and we also offer, some advanced data services. I guess, support such as if you're looking for finance data or have research questions around that or GIS or a data visualization support. So I'm going to turn things over to my co presenter, Lily Ren.

**Lily Ren:** Welcome to today's workshop. I am co-presenting today with Stephanie and I am the scholarly communications and metrics literacy librarian on the same department as a Stephanie. I also want to acknowledge that this particular slide deck was adapted from my colleague, Minglu, who is the research data librarian within the Open Scholarship Department who is currently on sabbatical. And so we're really looking forward to speaking about this topic. With you along with some of the topics and resources that Minglu have previously developed. In addition to research data management, I also support research visibility and research impact. Thanks.

**Stephanie Quail:** Alright, just to get started with things. We actually wanted to do a brief poll with everybody first just to get a sense of where you're coming from in the university as well as your familiarity with the topic. So I'm just going to actually switch things over here to a tool called Mentimeter. And in case you haven't used Mentimeter before, the way that you can actually access the poll is you can open either a browser window on your smartphone or new tab on your browser and your computer and you can go to [www.menti.com](http://www.menti.com). And then you can enter the code, 38 12 70 53 to join. And we can just see how many folks are joining down here. So we have 3 right now, so we'll give it a little bit more time and see if we can get a few more people to participate. Alright, so let's just, I see we have about 20 people, so that's pretty good. I know not everybody is always going to participate in the mentimeter polls, but I'll just go to the first question now. So this is just a open ended question. What faculty or department are you from? Ecology, as good biology. Great. So looks like we have a nice mix of faculty. From a variety of places across campus as well as people who are dealing with particular grants like the Connected Minds grant too.

So great. So thank you so much for participating in that poll and I'm just going to go to the next question now. And so our next question is, how would you describe your knowledge of research data management? Alright, so it looks like the middle option of 'somewhat knowledgeable' is, kind of winning, I suppose, for this poll here. So a number of you do have, some experience with research data management, but you wouldn't necessarily categorize yourself as very knowledgeable. And then we also have some participants in today's workshop who are new to this topic. So it's really good to kind of see the mix. Of individuals for this particular workshop and hopefully by the end of today's workshop you will know quite a bit more about research data management and in particular the data management plan.

So thank you so much for participating in this poll and we will be using Mentimeter a little bit later on in the presentation too. So don't close your, Mentimeter tab just yet.

Alright, so, our objectives for today's workshop is that we're really just hoping by the end of today's workshop, you'll be able to understand the importance of research data management,

learn key data management plan considerations, and best practices, and also know where and how to use the data management plan assistant tool and also gain some familiarity with the York specific template that's been created. And so our agenda for today is: we will be touching on the tri-agency RDM policy and funder requirements and we have a special guest speaker Randy Lewis from the Faculty of Science who's going to talk about his experiences with that. And then we'll also be looking at library supports and services for RDM, DMP and grants, What is a DMP and why would you need one? DMP tasks and the research data lifecycle and then crafting a DMP using the DMP Assistant Tool and York template.

Alright, so I'm just going to turn things over actually to Randy. To discuss, research data management from his perspective.

**Randy Lewis:** Great. Thank you. If it's possible, maybe I could share a few slides. Is that something I can do?

**Stephanie Quail:** Of course, I will just give you access to sharing Randy. Just one second.

**Randy Lewis:** Great. Just have a few slides that may be helpful, I hope.

**Stephanie Quail:** I think you should be able to share now. Oh yes, okay.

**Randy Lewis:** So if people can see this size. My job in just the next few minutes. Yes, to tell you the approach that I took in creating a data management plan. During the fall term and I had to do this because I was applying for an answer Discovery Grant and I was in the subatomic physics section. And we were used as the guinea pigs to have the first. Requirements to have a data management plan. In our answer funding request. So my basic calls when I made my plan were. First of all, the funding agencies think it's important to have a data management plan.

And I also think it's important to have a data management plan and I wanted what I wrote on the page to convey that. That acknowledgement. And to do that, I try to spend some quality time and learn what this amount and to write something substantive on page. The second call I have in my own mind was to show that I understood at least the basics of what data management meant.

Specifically in my research field. I'm sure this is different for the Wi-Fi of people that we have here. And so I tried not to write generalities, but to be very specific on the page for what was needed in my case. And the third thing on my mind was I know what it's like to be on these committees, you have thousands of pages to read, you have many, many applications.

I tried to be as helpful and clear as possible. I looked at the instructions that were provided by NSERC and I literally used their headings so that they could easily find what they were hunting for. Because I followed their proposal.

Did my approach work? I have no idea because the results haven't come back yet. But feel free to shoot me an email a month from now to discover, Hey, Randy, how did it go with your application? And in particular, did you get any feedback on your data management? I will be curious about that as well.

So I thought I'd just give you one slide on my research. This is not a talk about research, but at least you'll have a context for how data emerges in my world.

I do elementary particle physics. So are you should glance around the room where you sitting now find something you remember that it's made of atoms, ultimately and the atoms have electrons going around the nucleus. Let's zoom in on the nucleus. The new piece is made of some number of protons and neutrons. I want you to zoom in on just one of those and I want you to look inside of it and inside the proton or represented on this slide by the blue circle. You'll see that even at that tiny, tiny distance scale. It's a very complicated place.

There are, there are quarks, anti-quarks, there are blue ones, there are all kinds of forced contributions inside this blue circle. And it's my job to do calculations in that tiny little world. It's more or less hopeless to do it with a pencil and paper. And so what people do is use super computers. You simply break the volume under study into a tiny grid. Or lattice and you do your little quantum physics calculations, cube by cube to you . Every place there's a little cube, I have to do some quantum field theory calculation. So that's the data. That needs to be stored and used in my field.

When I set out write a data management plan. For this topic. I went to the NSERC website and this is a direct cut and paste from what they provided. You can see the link at the very bottom of the slide. On the one hand, this is a lot of words to put on a slide. We won't read them all now. But on the other hand, when you sit down to write a data management plan, suddenly it feels like very few words. They were leaving a lot to my imagination. And so they approach I took. Was to take each of these 7 bullet points that they've provided. And each, bolded phrase, this was a title in my data management plan. And then I wrote them a little paragraph or sentence or whatever I thought I needed for each of these 7 topics. So the remainder of my slides will be just to give you an idea of how I approach each one of these.

It'll just take us a couple of minutes. The first one is data collection. So in my world, I was able to rely on what my community already does. There has been something called the International Lattice Data Grid. It's existed for 20 years. And this is the standard where people in my community share information, share data with each other. Much more recently, I've heard people talk about the FAIR principles. Data should be findable, accessible, interoperable, reusable. These are things that my community is learning along the way. Nobody was using these words 20 years ago. At least in my field.

And so I tried to do my homework. Learn what was already known in my community and I use these phrases when I was talking about data collection. And then specifically in my Management plan, I said. How our research will use the international format. We don't want to

reinvent the wheel. Anyone who has already prepared data that I could reuse, I will reuse that data. And in cases where the data doesn't already exist and I need to build it myself. They will be created in the international format. And we will make them publicly available so no one else has to reinvent my wheel. And so people can check. Hey, Randy, make a mistake in his calculation. Hey, they will have my actual data and they can double-check it.

The second bullet point that NSERC provided was about documentation and metadata. So in this case, I pointed out at the international working group had created a markup language. It's based on XML for those who speak that language. And this is what's used in the file headers. For every file. That means the header of the file contains all the information about how that data was created. And the body of the file has all of the numerical values in a very specific order in for the lattice that spans the physics of your calculation. So I was able to just mention that I'm aware all of this exists. And that I will follow that International format exactly in all the details.

The third of the 7 bullet points in the NSERC list was storage and backup. This is where I said, well, we're running on the Niagara computer at the digital research alliance of Canada. Normally we store our files on the general purpose file system. But we also have a special allocation on the high performance storage system and that's where we have our critical data backed up. And available for others who may need it.

Item number 4 is data preservation. So I assume that the HPSS is going to last a very long time. But if our data is even valuable after the lifetime of what, the Canadian alliance is providing, then we will work with the international community and find a way to host these files elsewhere. Item number 5 is about sharing and reuse. So I wanted to emphasize once again, our data files are publicly available upon request from anyone in the community. And in addition to that, the computer codes themselves that we write to prepare the data. We tend to provide direct links to those in our published papers and people can click and immediately see the code. On GitHub or Zenodo, you know, for example.

Item number 5 responsibilities and resources: how is it that we manage the data management plan on a regular basis? And here I was able to point to a collaborator. Anthony Francis, a few years ago was a postdoc here at York. He's now a tenured faculty member in Taiwan and he's still collaborating with our group. But he's also in this open lattice initiative. This means he's one of the leaders in my lattice community. In terms of data management. And so we're able to rely on him and I could write this in my proposal that Anthony Francis keeps us, aware and drives us in the best practices. Even though in reality, our day to day needs are very minimal. In my research, I don't know about you. I don't think about my data management plan daily. I think about yearly or seasonally and this is good enough for what we need. So finally, at the last topic is ethics and legal compliance and for me this is very simple. We're doing open science. There's nothing special we have to worry about. So that's, my story. Good luck to all of you when you apply. I hope it goes well. Thanks very much.

**Stephanie Quail:** Thank you, Randy, for that excellent presentation. Does anybody have any questions for Randy before we continue with the rest of the presentation? Alright, I don't see any questions right now, so I'm just going to go back to the slides. And I'm going to turn things over to Lily now.

**Lily Ren:** Thank you, Stephanie, and thank you, Professor Lewis, for walking us through your experience with developing a DMP. Like Professor Lewis mentioned earlier, his DMP was a requirement as part of the specific grant, the subatomic physics Discovery grants through NSERC. But like that particular grant, there are actually many grants that are increasingly, requiring DMPs through the Tri-Agency.

So on this slide, you'll get a flavor of some of the grants that are coming soon with the DMP requirement or have already launched. For those funding programs with the DMP requirements, you'll see that many of these are still in the early stages of the new requirement. Like Professor Lewis, a couple of our researchers from SSHRC as well, CIHR, are also waiting for their final results as the first cohort. We really appreciate Professor Lewis taking us through that experience and hopefully best results to you in a couple months when they release the results. Hopefully that gave you a flavor and scope of some of the different types of considerations and data that a DMP will account for.

DMPs are also useful not just to meet DMP requirements through funding agencies , but they also can help you think about the longevity of your research and how you want to manage it, how you want to share it, and perhaps also increase the visibility and attribution of your data to all your hard work that you have gone through to gather and collect that data.

In 2011, the Tri-Agency funding program launched a RDM policy or research data management policy which includes 3 pillars. One of which is the requirement of DMPs for funding opportunities either through SSHRC, NSERC, or CIHR, and the other 2 pillars include institutional strategies for research data management, for instance, what are our strategies at York for, supporting research data management? and a final stage that will be later phase in mandatory deposit of research data for researchers who are funded by the Tri-Agency.

This particular element hasn't been required yet but they're setting up the infrastructure to do that so that is also a consideration later on when you're thinking about your DMP in terms of meeting some of these requirements that are set by the government.

Next slide.

In the libraries, we provide a range of support for research data management, including workshops like the one you're attending today. Typically we run a research data management series twice a year featuring 3 workshops that cover best practices for managing research data, creating a DMP, which you're attending today, and in 2 weeks we'll be also launching a workshop on publishing your research data. So if you're thinking about data repositories as well

as the infrastructure that is available through you through your university. Please feel free to register and take a look at that workshop. If you need a quick reference point, there is also a lovely guide that has been developed by my colleague, Minglu, that details some things that would be mentioned today that are documented in a DMP, but additional considerations for research data management, including the institutional repository, the York University, Dataverse, which is a generalist repository for York researchers that is shared by many other Canadian universities on a national network called Borealis, as well as the DMP assistant tool that can help you craft a DMP. We have a York-specific York template on there to help you guide through the different elements that Professor Lewis has alluded to earlier.

Additionally, we also offer support and customized consultations to address all of your research data needs. We can help you deposit and curate your data in the York University Dataverse, as well as support the creation and review of a DMP. If you are playing for a grant that has a DMP requirement or you're looking to just create a DMP, we can sort of take a look, walk through the steps, as well as review your draft prior to submission. And if you're looking to apply for a grant, the Libraries has recently expanded our services to include a grant support services that Stephanie will touch on.

**Stephanie Quail:** Thanks, Lily. So earlier, I guess in September or October, 2023, we did launch a new web page on the libraries site to just explain how researchers could get a support letter for any grant funded research projects that they're engaging in. And so our, I've just detailed the steps for accessing or getting a support letter provided to you. So the first step is to email [libgrants@yorku.ca](mailto:libgrants@yorku.ca), ideally at least 2 to 3 weeks before your faculty's internal review deadline to arrange a consultation.

And then after the consultation. And this also is dependent on library capacity as well. Because depending on the number of people who ask for a library support letter, we may have more or less hours, to, or other types of in-kind support to offer to researchers. But basically after that consultation, a formal letter will be prepared and then signed by the Dean of Libraries and then on the, actual web page itself, we've listed the different types of in-kind services that are available to researchers. So things such as data visualization consultation and workshops, at digital preservation consultation. You also see that you can add in your library support letter, research data management consultations and workshops and that kind of thing as well. And so I'll just show what the library support for Grant funded research projects web page looks like so you can go and explore, what the different, I guess items are in the menu here. But definitely the first step is to reach out to [libgrants@yorku.ca](mailto:libgrants@yorku.ca) and then we can set up a consultation and then have a meeting with you and discuss the different types of supports that are available for grant funded research projects. Alright, I'll just turn things back over to Lily now.

**Lily Ren:** One way to help contextualize the importance of a DMP and some of the considerations for a DMP is a typical scenario that researchers are increasingly finding themselves facing. For instance, you publish your manuscript either in a preprint server or in a journal and then a few years later someone stumbles across your paper and they're really interested and they want to look at your data and code to either replicate your work as well as

build on your idea. However, it's been so long that you might not have access to the data. Maybe perhaps many people were working on this project or the data was stored on one of your colleagues computers and you don't currently have access to that, or some other reason why you might not actually have that data readily available to share with that particular requester.

Instead, a DMP can help you sort of think about all these different ways which you can share your data, as well as manage your data internally within your team as well as when others would like to see it. In this scenario, which involves the DMP, you publish your research either again in a preprint server or a journal article, but you also publish and share the data in either a generalist or disciplinary data repository with a ReadMe file and other information to help others understand that data. And then years later, when someone stumbles across your paper, they can actually directly download your data and code from the repository and use the ReadMe file to understand that data and use that data and of course also acknowledge and attribute that particular scholarship to you so that you can get a citation. You can get recognition for that work that you have done.

Within that sort of contextualization of how a DMP might be useful is that a DMP can help you think through about how you want to manage your research during your research project. And after it has been completed. It is essentially just a document that guides the collection, documentation, storage, sharing, and preservation of your research data. And essentially supports more efficient, proactive and systematic project management of your research.

Whether a DMP is required by funding program or not, it is integral to conducting responsible research and can help you and your team save resources by ensuring your data is complete, understandable, and secure.

Even in the broader research data lifecycle and open science or open scholarship, DMP is guided by the FAIR principles that Professor Lewis previously mentioned and essentially it aims to make data findable, accessible, interoperable, reusable (when possible) because we recognize that not all data can be shared to maximize the research potential of that data that you have worked hard on. Recognizing that you all may engage with different forms of data, these 4 principles will manifest a little bit differently depending on your research approach and discipline.

Generally, a DMP will save you lots of time and prevent data loss by helping you keep track of your data throughout the active phase as well as the final completed project and helps identify research data management opportunities, resources, challenges, and issues.

During the earlier stages of the research, you can think about these things as well as plan for it. It also ensures correct attribution, rigor, reproducibility, and accountability of your research, which makes it easier to access, share, reuse, and repurpose the data for the broader research community.



Again, increasing the visibility and impact of your work. And of course with many funders, institutions, as well as journals, increasingly requiring DMPs or increasingly requiring the sharing of data, it helps you meet those guidelines and requirements if they are part of your research project. Next slide.

These are a couple of examples of DMP documents. The first one is a template designed by the National Institute of Health (NIH) in the US in its design for survey and interview data. So you can see that there are many different types of templates available and there are different ways that you can develop these DMPs but later I'll touch on the components that are shared across these different templates across these different templates. Stephanie is just providing a sort of quick overview. If we can move on to the second example. This particular example is developed by the Digital Research Alliance of Canada and this particular Alliance is essentially a nonprofit organization funded by the government of Canada, that integrates and champions and funds the infrastructure and activities required for advancing research computing, research data management, and research software, and they have many, many different templates, so the one you're seeing here is mixed methods template.

And the last example is one that is being produced by Portage, which was initially launched by the Canadian Association for Research Libraries but has now since been a part of the Digital Alliance as well and they have a series of templates and examples that you can use for different types of research and in particular this example that relates to digital humanities and the use of secondary data.

So hopefully that gives you an idea of what a DMP looks like. We will walk through the sort of components that each of these templates or DMP templates share in common so that you get an idea of the type of information that you can include. But depending on your discipline as well as research data type, you may choose one of these templates or some other template that you have come across and discuss considerations that are particular to your research needs.

Next slide.

So just to get an idea of the type of research data you're working with, we have a Menti poll that asks about the type of research data that you took to create or use in your field.

I know earlier we got a flavor of the type of data that Professor Lewis used in the sciences, but perhaps there are other types of data that you engage with more often.

So we have measurements, demographic data, experiments, human participant data. We will definitely be touching on that, as well as Indigenous data, images, observations, records, transcriptions of manuscripts, interviews, survey and secondary data, qualitative, focus group. Text analysis, visualization. We will give it a few more seconds for anyone that would like to add to this list as well. As you can see, there's a wide range of data that we use across these different research fields that are represented.

And some of these data are pretty common across all fields, like for instance, survey data, secondary data, qualitative data, but there are also data types that are unique to specific fields and specific research methods. In thinking about your data management plan, those are considerations that will need to be fleshed out as well.

As you saw, research can take many forms and research data create and use can manifest in many different ways, and so this is just a snapshot of some of the different types of data you're working with as well as the ones that have been submitted to the Menti poll.

Next slide.

Within the wider research lifecycle, there is actually a smaller component called the Research Data Lifecycle, which you're seeing on the screen right now. The research data lifecycle is a model that just emphasizes the continuous nature of data creation, collection, and use in the process of scholarly discovery and research. Here the data is considered to have a long life span that extends beyond just the creation and completion of a project, where data can be preserved and shared and open when possible and closed as necessary, recognizing the sensitivity of some of the research data you might be working with and the limitations that you have around sharing that particular data.

In the management of research data, there are several data principles that guide our practice, including the FAIR principles, which we touched on earlier, as well as there's also care principles for Indigenous data and considerations for sensitive or confidential data. So for the next couple slides, we're going to walk through each of these different components in the research data lifecycle as well as the various components of the DMP that match to these different elements.

Some of the things that you might be thinking about is if you're sharing your data, perhaps you'll need to actually get consent from your participants if you are working with participant data.

If you are working with a data team, or a research team, it's always important to have file naming structures around the file naming conventions, as well as how you want to store and manage your data, including an inventory of all the different data files.

And as you move through the cycle, there are other considerations for dissemination and reuse and preservation as well such as licensing, embargo, applying different terms of use, as well as the different types of day formats and where you actually want to share your data.

Next slide.

As you just saw there are many different formats and templates of DMPs, but generally they include these 8 elements which we will walk through including responsibilities and resourcing,

project description, data collection, documentation metadata, storage and backup access sharing and reuse, preservation, and ethics and legal compliance

With the DMP, the first consideration to plan for is the responsibilities and resourcing of your project. Prior and during the data management planning it's important to factor in resources both in the short term of your research project as well as long-term infrastructure and data needs as well as roles and responsibilities of yourself as well as the team you're working with.

In considering human resourcing, you'll want to identify specific RDM roles for each team member and account for changes that might happen throughout the data lifecycle. You can consider who will create the data, including individuals or teams who are involved, we're actually, you know, collecting, gathering, analyzing the data, as well as the physical location of where that data is going to be stored and who will actually access and edit the data, either immediately or at different stages of the research process.

And in preparing for data management, you might also want to assign roles for data management responsibilities so that there is a little bit of accountability and you know who to go to at different stages of your research project in relation to the data you're working with. Depending on the scale of your research project, you might actually even have a required dedicated RDM position or small committee that oversees this process. You may also want to appoint a main contact for the data should there be an inquiry or request for the data after it has been published and designate a backup contact if that individual leaves the project or the institution that you're currently with.

Other costs that need to be accounted for during the research data management process and in a DMP include training and skill development for yourself as well as your team, generating documentation, proper de-identification of data, which may involve external tools, and high quality preservation for long-term access internally or externally to your team. Also, there are certain costs that are associated with research data management that go beyond the active research project stage that need to be accounted for as well, including software that you might use as well as data repositories because some repositories are free but some repositories do have a fee associated with them.

Next.

In any DMP, you'll definitely want to include a description of the project and data to provide context and understanding of the type of data that is being presented. If you are gathering your own data, you'll want to consider the different types of data that will be used, collected, or created, how much data will be created for storage purposes, if you'll be collecting any identifiable information or confidential data, because if you are collecting those, there are different considerations as well. If you are using secondary data as part of your research, data that has been produced by someone else.

You'll want to consider a data use agreement if that has already been applied to the secondary data and if there are any restrictions in place for you to use and share that data. And if there are, you'll also want to document that in your DMP. It's also important to identify all the stakeholders of your research data and account for their specific requirements to enable your data sharing in later stages of the data lifecycle.

So for instance, if your research involves human participants and you plan to share and deposit research data after the project has finished. You'll need to actually include in your consent form when you're soliciting for participants about this particular sharing of data and of course when you're applying for ethics approval. If you're looking to publish a specific journal, you can also check for data sharing requirements or data availability statements that they may have around publication in those journals.

Next.

Data collection provides an overview and description of your data types, formats, and data size which will later support the storage and preservation of your data. If you plan to communicate clearly to all the team members about the type of data and the data collection that will be happening, this can really help your future data management tasks when you need to track and trace down your data files. And allow your team to work more efficiently. This also includes the consideration for privacy, as well as sensitive and confidential data that you might be working with or gathering. It's always important to have an inventory of all the different types of data files that you'll be creating and using.

With that, I also like to ask, how many here on this call typically work with sensitive or confidential data or perhaps secondary data as part of their research.

So on the screen there are a couple of options: the different types of data that you and your research team might engage in. We're just curious to see how often these come up in your own projects. So I'm seeing that a lot of you typically use participant data and sensitive data in human participant research. There's some of you that use secondary data without restrictions. And some of you also use secondary data with restrictions. So meaning there's restrictions around the use and sharing of that particular data. Now, I also see that there are a couple of you that are using sensitive data with non-human participant data as well as Indigenous data. So we have a quite a variety of different research data types across our participants here today.

So for those of you working with sensitive or confidential data, such as either insider knowledge or business data or direct or indirect identifiers in participant data, Indigenous data, data collected that have additional layers of sensitivity and security, you'll want to consider some of the ethical and privacy related regulations that might impact your data management decisions. You want to consider the mechanisms to manage these types of data, including secure storage, encryption, or anonymization or de-identification if the data is to be kept or accessed by others at any point.

At York University, there are a couple of policies and guidelines for researchers working with sensitive data and confidential data through the Office of Research Ethics and Data Security guidelines that we have developed here. I recommend looking at those resources to help you decipher the type of data you're working with and how you want to manage it and of course if you do decide to share if that data is actually shareable.

Next slide.

As you move along toward the data processing and data analysis, is it also important to document the changes that happen.

This is crucial for research responsibility and important for data use and reproducibility for internally for your team, but as well as if you do decide to share your research. At this stage you want to engage with your research team and decide on a common strategy for a data organization and this could include setting up, filing naming conventions, folder structures that will accommodate data growth over time, and meaningful file variable naming conventions for the team to follow consistently and have version control strategies. As you document major changes between these different versions and determine if a date of inventory is needed to account for all those files.

Depending on the scale of your research data, you could design or even automate a metadata system as standardized or structured set list of descriptive fields or tags to provide relevant contextual information to describe the data. Your instruments and tools might be generating this kind of metadata file already for you on the back end, but it's important that you follow the standard languages and metadata scheme that is used in your discipline to ensure that your research data is machine-readable by those fields, interoperable with other researchers in your field and is easily discoverable.

But the goal here is just to be consistent. As a team, you want to use the same template, the documentation structures so that information is always captured properly and in a consistent manner so that others can understand that.

Next slide, please.

For data to be readable and interpreted correctly, both within your team and if you decide to share your data, it is best practice to always include a ReadMe file that will be saved together with all your relevant data files.

This gives others and your team context as well as guidance on how to understand that data and reproduce and use that data if needed. This includes an accurate description of the steps that it takes to process and code the data. Labels that you use in your files and, of course, you don't want to use 2 and biggest language or specific acronyms that are hard to understand or misinterpreted. So you could, if you do decide to use acronyms to include a legend of some

sort. You could also provide data or variable level documentation such as a code book for numeric tabular data as well.

And there are different, I know there was a question that came in about examples of good file naming conventions. There are different ones out there that are both general as well as discipline specific and I'm happy to share those resources as well to give example of what that looks like. On the screen you'll also see an example of a specific file naming convention that is being recommended for general disciplines as well.

Next slide.

Storage access and backup of research data can vary across different projects and storage choices depend on the status.

Of the data collection, the data type, and nature of the data to probe actively safeguard your research. So based on your data needs, you might want to account for a backup and growth when calculating total storage requirements that is needed for your project. Especially if you're using an external storage that has limitations or that has fees associated with certain storage sizes. And you'll want to store the data for either long term or short-term storage during the active as well as the completed stage with your research.

If the project is active, you might want to consider short-term storage solutions that support active research so that you can go in, edit your research data, you can do version controlling, you could do collaboration.

For the completed data or a completed project, they're also longer-term storage options. That you might want to explore for your research data and these provide long-term access as well as preservation. So depending on the actual platform and tool you're using. And you always want to consider storage that has built-in backup and synchronization options so that if something ever happens, there is backup on the tool. On the end of the tool that you're using as well.

Additionally, for securing sensitive or confidential data. You also want to ensure the storage options support the safe handling of the data and the type of data you're working with. And if you are working with this type of data, some common tips that we recommend is to host the sensitive data on infrastructure physically, located at the university or on a secure Canadian network or server, avoid working directly with the sensitive or confidential data. Once you have received that confidential sensitive data, you want to code it or anonymize, or de- identify or encrypt it at that particular stage before it gets shared with other members of your team for further analysis and processing. You can also utilize secure. Secure storage options that are chosen for the safe handling of that data.

Certain tools are made for handling sensitive or confidential data and there are many other tools that are not, so you want to sort of look into those if that is the type of data you're working with.

Next slide.

So we're also curious how you typically store data during the active stage of your research. Perhaps you're using Google Drive or Onedrive. So we're just trying to get an idea of the type of data storage that you use. Perhaps this is also a place for us to gather ideas for storing data during the active stage of the research as well. I see there's Google Drive and One Drive. Digital Alliance of Canada. I would say Digital Alliance of Canada, depending on the specific tool you're using through the alliance, typically they tend to be for the completed stage of your research. But for instance, OneDrive, Google, external, external hard drives can also be good choices for, active storage, including Dropbox.

Because York has an institutional license with Microsoft and OneDrive, typically OneDrive is considered safe for most of the data that you're dealing with during the active stage of your research. But depending on the nature of your data, there could be other physical options as well. Like for instance, paper files and a filing cabinet, as well as hard drive. I'm glad to see that there are many different storage options that we're all using today.

Moving on, after all the hard work of data analysis and writing your manuscript and perhaps even finally publishing your research. You can share your data and assign a DOI so that others can build on your research and cite your data, increasing the visibility and impact of your research. It's also recognized that some of the data simply just cannot be shared for either contractual, legal or ethical reasons. For best practices and sharing research data, you'll want to consider sharing their data in a recognized data repository as they typically tend to be more secure, sustainable and discoverable than say sharing your data on your own personal website or on an external website. There are different types University, York researchers have access to our institutional data repository called YorkU Dataverse. And this is a free, generalist data repository for York researchers to deposit relatively small sized data sets and projects. Now if you're working with large data sets that require 500GB or 1 TB of storage, the Canadian, Canadian researchers all have access to something called the Federated Research Data Repository and essentially it's a national infrastructure that is actually hosted by University of Toronto.

That will allow that sort of preservation and storage of extremely large data sets. If you're interested in these repositories as well as perhaps options for you to publish your research, the workshop that will be hosted in 2 weeks by the library on publishing your research data will touch more on these options and how you can actually go about and publish and share your data.

But when you're sharing your data, you also you always want to determine data restrictions, licensing, embargo, and use agreements to term and how others can use your data and this is something that you can assign or that is mandated by the funding. The funding program or the journal itself and you can communicate how you would like these restrictions to take place. So

for instance if you are licensing your data you can select the Creative Commons licensing of how others can use your data as well as how they can attribute or cite your research.

Thanks. Next slide.

In preparing your data for long-term retention and archiving, sometimes this process can be a little bit time consuming and resource intensive because you have to prepare the data but also ensure that it is archived.

It'll be easier to plan for this ahead of time. And so, if you are sort of looking for long term retention or storage options or looking to archive your data over a long period of time you're welcome to consult the library about data deposit archiving services that are available.

Some repositories also have long-term preservation and retention built into that particular repository. So you'll want to just visit their preservation policies when you're looking and considering these different types of repositories. If they do, don't have this service you might want to find a separate archive to use for preserving your research data. When planning for long-term archiving of data, consider if all the data are to be kept or if some of it will be destroyed once the project is completed.

Not all data needs to be preserved forever. So you can determine which data you will like to preserve. You can also consider standards and formats that we use for documentation, data, and code for future reuse and discovery, as well as where the data will be stored long term and who would be responsible for it on your team.

Next slide.

Once you decide on the part of the data to be preserved, you'll need to prepare your data. And typically it's recommended that you preserve your data and save your data in open formats. Where possible so that the data could be read by variety of processing programs as well as across different disciplines. This ensures that the data is accompanied by documentation of the data. So that is understandable and usable.

And this might be easier to do along in terms of documentation rather than doing everything towards the end. And on this screen there are a couple of different open data file formats for different types of data that are considered to meet the FAIR data principles. And a lot of these data formats are also the format that is being accepted by data repositories.

Not all data repositories will be able to accept all these types of file formats, especially if you're using proprietary data formats. So that is also something to look into prior to sharing your data and preserving your data on an external platform. Thanks.

And last but not least, I know we walked through many, many different components of the DMP, the last section is to address ethical and legal compliance. If there are any legal or ethical or



intellectual property issues that must be addressed, and how you would apply to the specific requirements, whether it's coming from the institution, the funder, the journal or perhaps if you are using secondary data, the secondary data owner. Research ethics is a common issue. So, if you are working with participants and you're looking to share participant data later on you always want to make sure that you get permission early on in the project such as through the consent form, and follow all the security measures to protect data according to the level of the exposure risks of that data and the York University Research Ethics Board provides a couple of guidelines around the use of human participant data as well as templates on how you can incorporate these types of language into your consent form and into the public facing aspect of your research.

I also want to emphasize that although you want to address each of the questions or components of the DMP that we covered earlier, you do not need to have a perfect DMP. This is typically a living document that will change continuously. It will be developed and adjusted over time. This helps to serve as a template and for various considerations that you might have at different stages of your research project.

As you saw, there are many different templates and there are lots of flexible customizations that can be placed on top of these templates to meet the needs of your own research. Thanks.

This slide just provides additional templates as well as exemplars for different types of research that you might engage in. There is also a York University template that is available on a tool called the DMP Assistant Tool that is available to Canadian researchers and this specific tool allows you to craft DMPs directly, so that you, you don't need to remember each of the components I talked about today because the tool will guide you through each of those processes through that DMP assistant tool. And Stephanie will speak a little bit more on that shortly.

But in addition to that particular template that is available to you, there are also other templates as well that are developed by Canadian institutions, including the Digital Research Alliance of Canada, and they have several different, different types, including one that we saw earlier on the mixed methods.

But there are also DMPs that are shared by Canadian researchers. These are actual DMPs created by researchers who have engaged in research and those DMPs are typically called public DMPs. If you are looking at it through a DMP assistant or if you're looking at it through another repository for DMPs such as Zenodo, just keep an eye out for those public DMPs because they also serve as good examples, but just keep in mind there's a lot of variety there and not all of them are vetted. But they do serve as great examples, depending on the field that you're working in and the type of data. Next slide.

**Stephanie Quail:** Okay, so thank you, Lily. And, as Lilly was mentioning earlier, we do have the DMP assistant tool that helps you create a data management plan. And so the DMP assistant is

a free bilingual open source tool that helps you prepare the data management plan. It includes a variety of different templates, but there's also the York specific template too. And so there's a link here on the slide just to go directly to the DMP assistant, but I'm also going to show you how you can access that from the libraries' site.

So on the slides here, we've just actually created few slides that I'm not going to go over right now, but it's just for you to refer back to for when you decide that you may want to use the, the DMP assistant and access the York specific DMP template. But I'm also going to cover how to do that right now.

So if we were starting on the Libraries homepage here, if you want to get to the RDM specific section of the library's website, you would just start on the library's website and then go to research and learn and then under publishing support, RDM and ORCID, you can click there.

And then we have, a variety of things available on this page. Such as open access publishing, York Digital Journal, YorkSpace, etc. But then you'll want to go to this section here on research data management.

And so we've been referring to number of different resources here, but we also have the link there on that page for the, workshop that's coming up on March 18th that Lily was talking about for publishing your research data in a data repository. But if you do want to get to the DMP assistant, we have a link to the DMP assistant right here. So if you've never used the DMP assistant before, what you'll first want to do when you get to this web page is you'll want to create an account and just fill out all of this information here and then put in your university as York for your actual home institution there. But if you do have an account setup you can sign in. And then if you want to create a data management plan, you can go to create plan here.

And then you can actually check this off if you're just doing a mock project for testing, practice or educational purposes. So if you are just testing things out, I definitely recommend checking that off. And so, I already have a test one, but I'll say my second test DMP. And then here you can indicate your primary organization, just start typing in York University. And then here is where you can actually access the DMP template that you'd like to use.

So there's a few different templates within there that Lily was talking about, but if you want the York University one, it's just listed alphabetically. So we're right, it's right there at the bottom. So now we have that's inputted into that dropdown and then we can click here to create the plan. And so the first section is just your project details and you can fill out all these bits. Including funding, the funder, funding status, that kind of thing or if there's a grant number. You can also add contributors to your plan as well. So you can put their name, email, ORCID, their affiliation. If it's, if it's different from York University, you can obviously change that. And then you have the plan overview here. And so, this template version was from March 13, 2023.

So you can just kind of get an overview of there and then you can go into the sections where you can actually start writing the plan. So if you want to write any of these sections here like in

data collection. What you'll see is as you work through each section, you will have your specific guidance on each side here. So if you scroll down, there's guidance and it will actually link out to different things that could be useful and help you better understand that piece of the data management plan that it's trying to help you with.

And so you can see for each section there's some nice your queue specific guidance here for every single section you can go through. And same with documentation and metadata. So metadata standards and tools. Just some different assistance there and then also it will point you back to the [yul\\_rdm@yorku.ca](mailto:yul_rdm@yorku.ca) if you do have any questions or need help. So you'll see that as well. And so it has some examples here of ReadMe files and examples for documenting content there too.

So once you've filled out everything, you can then add the different research outputs that you anticipate, you can also, set your plan visibility and manage collaborators here as well. And then you can also download all the documents here. So it, it, automatically defaults to PDF, but you can change that to CSV, text, Word doc, JSON, etc, Whatever would be the most useful for you for when you're exporting your data management plan. So we do recommend taking a look at that tool and using the tool because it can be very helpful and the fact that you can add collaborators to the plan is very useful too. And so definitely if you are deciding to use it, you can always go back to the slides here where we have a couple things just to walk you through it. But overall the, the site itself is pretty simple and clean and easy to use, especially if you're using the YorkU template where it has all of the different York guidance as you're filling out the sections there. Alright, so, this is our last slide for the presentation, so we'd be happy to take any questions now. We've also included the information on the slides for if you have questions around research data management services and we have the links to the RDM website here, as well as the [yul\\_rdm@yorku.ca](mailto:yul_rdm@yorku.ca) email and then if you do also have questions about library support for grant funded research projects, we have the contact information for that as well, as well as the link to the website there too. So we'd be happy to take any questions or comments that people have for the the last 10 min or so of today's session. And please feel free to put your questions into the chat or you can also unmute and ask your questions too.

### **Q&A section removed**

**Stephanie Quail:** Alright, well, it looks like there's no other questions, so maybe we'll give everyone 5 minutes back to their day. Thank you so much, everyone, for attending today's presentation. We really appreciated your participation in the Mentimeter polls. We also appreciated Professor Lewis' presentation as well on using the DMP, or preparing a DMP. So thank you so much and please do feel free to reach out to [yul\\_rdm@yorku.ca](mailto:yul_rdm@yorku.ca) if you do have any other questions about research data management. So thank you so much and have a great day and also the link to the evaluation form has been put into the chat. So please take a few moments as well to complete the evaluation for today's session. We really appreciate your feedback as it will help us improve future iterations of this workshop.