Voiceover:

As neuroscientists, our research is at the heart of our everyday lives, from designing experiments to analyzing data to writing grants and journal articles. This work brings us closer and closer to understanding the most fascinating and mysterious part of our bodies, the brain and nervous system. But for neuroscience to have the greatest possible impact on our world, it must be rooted in a strong foundation of rigorous principles. You're listening to Pathways to Enhance Rigor: A Collection of Conversations, where neuroscientists come together to discuss how to embed rigor into every part of the scientific process.

Voiceover:

This podcast is a part of the Society for Neuroscience's Foundations of Rigorous Neuroscience Research program, or FRN. Supported by the National Institute for Neurological Disorders and Stroke, FRN is designed to inform and empower neuroscientists at all career levels to enhance the rigor in their research, and the scientific culture at large. In this episode, we hear from Drs. Letisha Wyatt, Jane Roskams, and Maryann Martone. They discuss the challenges neuroscience has faced regarding data storage, management and sharing, including the historical transition from analog to digital and the complexities of neuroscience data, as well as paths to foster more sound data stewardship for the larger scientific community. Without further ado, let's hear about Dealing with Data.

Letisha Wyatt:

Hi, my name is Letisha Wyatt. I am an assistant professor in the neurology department at Oregon Health & Science University in Portland, Oregon. And I also serve as the director for diversity and research at OHSU. When I was working at the bench, I was looking at creating new therapies for brain disorders. And currently I do a lot of training for new graduate students who are pursuing their PhDs.

Maryann Martone:

Hi, my name is Maryann Martone. I'm a professor emerita at the University of California, San Diego, but I still maintain my lab there called the FAIR Data Informatics Lab. My research is in the field of neuroinformatics. So we have built several large database systems such as the Neuroscience Information Framework. And I'm also a big open science advocate and been working on ways of improving rigor and reproducibility. I currently serve as the chair of the governing board of the International Neuroinformatics Coordinating Facility, INCF, which is an international organization that is dedicated to standards and best practices in neuroscience.

Jane Roskams:

My name is Jane Roskams. I am a professor of neuroscience with my primary appointment at the University of British Columbia, but I'm also appointed at the University of Washington in neurosurgery and computer science. Interesting combination. When I had a research group at the bench, most of my research was focused on understanding cells and signals that are used in development that we might want to manipulate to enhance spinal cord recovery and treatments for brain injury including epigenetic. However, over the last 10 years, I've really been focused my attention on developing new ways, collaborations, whatever we need to do to help understand the brain and mental health and where the interface of big data, open science and citizen engagement sits, including game-based ways to capture and analyze data to help us understand the brain.

Letisha Wyatt:

I would say that data management and storage across my career has looked kind of messy in what I've observed. I, as mentioned earlier, am a junior faculty, so I'm kind of at the earlier stages of my career. And what's been really fascinating and something that I noticed really early on as a graduate student was that there is a lot of variability in how different lab groups manage their data and storage of their data. And what's been a little bit challenging for me to accept is the idea that we all kind of make up our own rules and make up our own rules as we go and how we can make sure that the rules that we're using are ones that are functional across different lab groups and different scientific studies has been one of the things that I've started to focus on as part of my career work in training the next generation, is really trying to get them to understand what FAIR principles are and why those types of things are so important.

Jane Roskams:

And I think I was lucky to have a bit of an edge on how I wanted my lab to be able to operate in a greater environment. I, at the end of my post-doc, ended up being right in the middle of an interesting two body problem. That could have been problematic, but I ended up getting the opportunity to work with Cold Spring Harbor Labs on like three major projects that they had going at that time. Two of which, and this is going to sound like a public advertisement, but were two books that they wanted to develop but really wanted the scientists involved. One is called At the Helm: How to run a lab. One was called At the Bench: How to begin to work in a lab.

Jane Roskams:

So I ended up jumping in with both feet, getting involved with all of these people that they had recruited, running major labs all over the place, and started learning at the seat of my pants what worked and what didn't. What were the common things across really good ... I got the metadata for how good labs operate and how bad labs don't. And helped to pull these things together and editing every single tiny little word so that by the time I landed ready to run my own lab, I had multiple copies of At the Bench to give to every single person who started in my lab.

Jane Roskams:

And the principles in At the Helm, and one of the biggest principles that came out of that is the careful recordkeeping that you need to do, the way you need to be able to carefully link every single piece of evidence from a lab notebook through to what's in the freezer. Which freezer? Which shelf? Who's handled it? So I started my lab off with this. And when my students and postdocs started talking to people around, they thought they had some kind of major dictator running their lab.

Jane Roskams:

But once we started collaborating with other labs, my lab started coming back horrified of what wasn't happening in the labs that we were collaborating with. We started sharing best practices very early and becoming kind of a benchmark for our community in how we did this. And anybody who knows me will tell you that is not me. That is not who I am naturally. But because I knew I wasn't who I was naturally, I knew this was the way my lab had to be. So I had to learn, and the metadata taught me.

Maryann Martone:

What has data management looked like across my career? I can honestly say I think it has spanned every technological revolution since the 1980s because that's when I started my career. So I started in light and electron microscopy. And back then data management was, I won't say easy, but it was very

proforma. Most of work that you did was taking pictures. They were on film. You labeled your film. There was a standard for how you labeled your film inside the lab. And they went in boxes that were then also labeled. And all of the boxes generally were in the same place. And so if you needed to go and get something, you knew how to do that. The filing system was of course physical and manual.

Maryann Martone:

Personal computers had already come into existence by then, and I do remember the first time I had stored a lot of data on the computer. It was an old Mac SE, I think. And I was doing my thesis work, so I was doing analysis, and I was using the spreadsheet program, which is often people's entry into computers, word processing and spreadsheets. And I remember one night actually screaming at the computer because the way that the Mac saved things. If you were used to using a PC, you worked inside of a directory, and anything you saved was in the directory, but the Mac used to go put them in various places because it had sort of these pre-configured folders, and I never could find anything. So that was my first experience with digital, and I actually think I called it evil. I'm like, "I cannot believe that you're doing this," because I was trying to get my thesis out.

Maryann Martone:

But I'd been very fortunate where I was in San Diego that very early on in my postdoc in Dr. Mark Ellisman's group, we started to collaborate with computer scientists over at the San Diego Supercomputer Center and so we became very involved in digital acquisition of images. We became very involved in digital acquisition of data. And the challenges that happened there were very profound. This idea that you are physically connected to the data that you produced and that you had to do active labeling and put it in a location meant now things were on computers. And back then of course it was before the cloud. There were some central servers, but which computer was it on? Right? on this computer. I had to take my computer home, and I can't remember where I stored it.

Maryann Martone:

We already had a lot of floppy drives. And going from 350 megabytes to 500 megabytes was a big deal. But we also experienced pretty early that no matter what system we put in place in terms of floppy drives and little Zip drives and Jaz drives, the technology would go out of date very quickly, and so you would end up with a lot of archival media that you really couldn't read anymore, unless you ended up with a lot of readers that still worked. And so many of us really remarked that the old way of managing it, where you put your physical resources in a box, and you labeled it, and you put it up on a shelf, was actually vastly superior, right? You could understand what had been produced by any given individual. All their records were in their lab notebooks. They belonged to the lab. All of that was clear.

Maryann Martone:

Now things were on all kinds of different media. They could leave the lab. When you opened them up, you saw these very cryptic names. There are often limits on what you can name things. When you were starting out, you could only have so many characters, and you couldn't have these characters. And so you realized that people were starting to lose control of what was in their laboratory. In fact, I had a couple senior scientists come to me during the early days of neuroinformatics saying, "We've lost control of our laboratories because now things can leave on little drives, and we don't know where they are. We no longer have the machines to read the media."

Maryann Martone:

I also remember working very early in what was a sort of a predecessor to the cloud, something called the Storage Resource Broker. And this was being developed over at the San Diego Supercomputer Center. And this was a project called the Biomedical Informatics Research Network, BIRN, that we were on where we were supposed to put everything on this central service. And I remember the researcher pitching this to us going, "You're going to be able to access this from any computer that you happen to be on." And I'm like, "Oh. That's magical." Of course the early days didn't work very well, but eventually it became the cloud, and the cloud became ubiquitous.

Maryann Martone:

So that helped to solve some of the problems of out of date archival media, but it really didn't solve the other problems of proper naming conventions, understanding how to put links to what it is that you have in the laboratory in a central location. So, over the past few years, we actually in our laboratory have things called write-a-thons where every quarter for three hours, the lab documents, and everything they document goes into a central location in GitHub, something that's going to stay around and is easily accessible, and links to whatever you do get put in there. And over time as we've learned a little bit more about what are effective names and organizations and effective use of cloud resources, now I can see the light at the end of the tunnel, though it is still a challenge.

Jane Roskams:

Where can people begin today to really plan for the future in using current technology to be able to follow the bouncing ball of where their data are going? Pretty much the state of the art right now is the Jupyter Notebook format, which is gradually being adopted I think more by informatics and data science type labs, but they are now communicating to the people around them that they collaborate with that these can be used in multiple formats, so that you no Longer need to have the core notebook the lab manager has to be able to know what's in each freezer and who uses which boxes. It's going to make ... It's going to be a big effort to make individual labs begin to train people in using that kind of a format. And many institutions themselves or faculties or departments have their own guidelines as to how they prefer that kind of thing to operate.

Jane Roskams:

So, we're pushing a boulder a little uphill trying to adopt, especially I have to say older people like me. People who've had labs for a very long time and a very specific way of doing things aren't super keen on the postdoc who comes in who's already on top of using Jupyter Notebooks, but it's time to just get with the program and trust that the next generation can help your lab move into the new age of reproducible science, and Jupyter Notebooks are one way of doing that.

Maryann Martone:

I think that this navigating the physical world and translating it into digital, a lot of things are designed, again, when you're innately digital, once your data's digital, there are environments, Jupyter Notebook, other sorts of things which really help people to keep track of things and manage things much better than in the past. There are a lot of electronic lab notebooks. It is this transition from what goes on in the actual physical world into a digital format that is often a challenge because it involves ... I mean the thing we like to say about a pen is a pen is cheap, and you get formaldehyde on your pen, you're not that worried about it. You contaminate your iPad or something, there is a concern. I remember spilling acetone on a timer once in the lab and watching it just dissolve and going, "Oops," right? So there is a challenge there.

Maryann Martone:

But I also think that online services, things like Protocols.io where you actually manage your protocols, your experimental protocols, through a central cloud service and start there. And then take that, create something that you can then use in the laboratory, and go back and modify it rather than starting everything in sort of a physical paper form and then having to translate it is a way to start thinking about that all of the things that you do eventually are going to have to be translated into digital format.

Maryann Martone:

And so I think when you set this up in the first place, you should start thinking about where these things are going to go, what is optimized for them. If you are going to use an electronic lab notebook, you want one that allows you to connect to these easily to these types of services. That's what a Jupyter Notebook does so that you can say, "Well, I don't have to do everything here. What I need is actually a link," and I can go that. So I think that there are new strategies that are coming about that will really help people.

Letisha Wyatt:

That question is really interesting, about kind of making the transition from analog stuff and physical stuff to digital stuff. I feel like it's a multi-layered problem. We've already touched on a few things. I've seen kind of the older generation of faculty who are like, "I like it physical because then I know where it is and I can know what's happening to it." I've heard some faculty who have expressed concerns about their digital resources and things that they've invested a lot of data, that they've invested a lot of money and time into being seized and held for ransom by ransomware or being hacked into and stolen.

Letisha Wyatt:

And I think just having maybe very little understanding out security for those types of data and those platforms makes some folks resistant. But also this sort of barrier of time, and the time it takes either to learn how to adopt these things or even just to, if you already know how to use these technologies, the time it takes to transition everything and recapitulate, as somebody said earlier. So I'm not sure how we can overcome that other than building in more incentives for people to take the time they need to learn these new tools and technologies or take the time they need to utilize them, if they already know how to use them, but I feel like that's kind of one of the major hurdles that we face too is incentivizing folks.

Maryann Martone:

I think so. But I think that the number of things that are starting to come out makes it impossible not to pay attention to this now. You can try to go completely physical and, honestly again as I mentioned before, there may be some truth to that, but the fact is most stuff is digital. It's going to be most of your data. And your lab notebooks and everything are going to be in a digital format, and so you really do need to think about this. But I do very much agree with you that I think at some point it has to be a concerted effort across institutions, even across funding agencies, to just start to upgrade and train and give people training on how to manage these digital assets. I think the time has come for that, and I think the technology, unlike what I went through across the decades in the past with obsolete tools, I think there is an understanding that these things need to be built on open standards.

Maryann Martone:

The technology's become stable enough that you can rely on it. I mean, how often now do we backup things on our home computer? We have them in the cloud, and we consider that our backup, right? That we're seeding backups over to others, which brings your security point right to the forefront. But I

do think it's really time for a concerted investment to upgrade the IT capabilities of individual laboratories, otherwise all the new data sharing mandates and things are going to be extremely hard to adhere to because all of that starts with good management of digital assets in your laboratory.

Jane Roskams:

I just want to pickup on one thing that Maryann just said because I think it's at the core of this. Training people in the appropriate ways to do this, in much the same way there was always different ways to set up a PCR reaction, there are successful ways to be able to bridge people into being able to handle many of these resources, but most universities are not set up to be able to provide that kind of training universally across the board. So you do get labs right next to each other that do things in entirely different ways, even use GitHub in different ways, even use their core university IT resources in different ways. So I think there are roles to play, whether it's at the funding agencies or even at societies, quite honestly, to promote and help develop ways where people can find the best way to be trained, to use the tools that will bring their lab into the digital age so that it becomes less intimidating for you to be reliant on the one person in your lab who's the person that you get to tell you why the printer isn't working to then also try and train the rest of your lab. Externalizing that in a way that will be acceptable and used commonly across multiple labs is something that I think is really critical that we do right now.

Maryann Martone:

Yes. I agree. And I think there are some resources that are coming out, like the research libraries often have programs in data management, but I just don't think that it is ... At least my experience and in talking to them, they're still having a hard time reaching the research community because I don't think, again, that this is something that people think, "Oh, I have to manage this actively." It's like, "This is the way we've done it, and this is the way we'll do it, and we'll continue to complain that we can't find data and that data's been mislabeled or labeled cryptically," right? That's why I do agree that this needs to be a concerted effort to recognize that this is important. And I really do like the idea of societies and others getting involved because they generally have a better handle on the types of data that are being generated and the challenge that are being faced in a particular community.

Letisha Wyatt:

Yeah. With regard to societies and funders and maybe even publishers, there is some work, as probably many folks already know, about creating some standards around what is accepted and shared. But I definitely do think that one of our ... another kind of obstacle that we need to get around is creating standards that we all sort of adopt amongst our communities, whether that's by data type or by discipline. I know some of those things already exist and are pretty substantial in different areas, genomics and whatnot, but there's a lot that needs to be done I think in a lot of other fields, when you're talking about like FMRIs and imaging data and all kinds of different disciplines who don't have these sort of preset standards across their data types.

Jane Roskams:

I think neuroscience might be the field where that is the biggest obstacle right now.

Maryann Martone:

Yes-

Jane Roskams:

Largely based on a lot of the issues that are coming up with reproducibility in neuroimaging, which has I think been the big thing, the elephant in the room that we've been trying to deal with for quite a long time. Genomic data has four letters. It's relatively simplistic.

Maryann Martone:

Maybe five, yeah.

Jane Roskams:

A single MRI scan has multiple parameters, different machines, different operators, different capture methods. I mean the number of differences in how you get a single MRI based on it being used to specifically look for the answers to a given question, the variabilities are enormous, and understanding those variabilities is something that our field is grappling with. I think Russ Poldrack at Stanford has been really spearheading efforts to help us in this, in the reproducibility side of things. And what we're getting is paper after paper now being published of, "Oops. We thought that was this, and now it's this." And, "Oh boy. Look, we scanned this same person in two machines in the same building and got very different interpretations back to back."

Jane Roskams:

So I think now we're becoming more aware of the variability and the problems. That's where our better standards are going to come from. The good will of people and community to begin to understand the necessity to get involved in open standards and data sharing is there. We have many of the pieces of that jigsaw, but we don't have enough people trained in bringing the pieces together and building it. And I think we're at a moment where, if that training is available, and the will is there, and the data are available, the next 10 years could be transformative in how we can bring those data streams together and understand who we are.

Maryann Martone:

We are starting finally to see some traction of some standards at least. So, in neuroimaging, the BIDS standard has been extremely successful. And it's interesting because it was designed with the bench scientist in mind instead of the informatician in mind, which is what often happens for these. And again in that sort of physical world where you have to ... You're in a laboratory. You have your colleagues. You have whatever systems are there, and it may not be what's necessary to fully comply with some big computational environment. The reason I think it was successful was because it said, "Well, if you're a bench scientist with no really fancy IT, you can still do this because this is how you organize your folders, and this is how you name your files," and it's the same every single time. And once we can start to align on some of these open standards, then the software can follow.

Maryann Martone:

But what I was very impressed with with BIDS is that it really did try to think about, if I'm sitting in front of my computer in the lab, "What can I do to conform to this standard." And I think when we start to apply that mindset to others, and BIDS is now expanding into multiple different types of standards for other types of data, it's really a way of just managing data in a standardized way that comes off of an instrument. And once one does that, then you open up a world of software that can be developed on top of it because software loves standards, right? And software is built on standards. So you really can see that there's this virtuous cycle that starts to happen. That once the laboratory people start to do just some things, you can then make it a lot easier for them to do it.

Jane Roskams:

So, Letisha, you're on the frontline. You are the spearhead of the next generation who've grown up being able to do a lot of these things. We had to learn it by the seat of our pants. What are the things that ... not Maryann and I because we're with the program, but the establishment, the older investigators, what are the obstacles that are in the way, do you think, that we might be able to help you clear out of the way to help make this kind of future we envisage a reality?

Letisha Wyatt:

Yeah, that's a great question about kind of what our next generation, how they're feeling about this open science environment and the digital nature in which we are kind of working in. And I have been lucky to kind of keep my pulse on a lot of trainees, as I work with post-baccalaureate students who are going into graduate school, and then I work with a lot of early stage graduate students and at some times the later stage graduate students as well. And what I seem to discover is that there's a need for a bridge between what they are getting from their lab and their PI and what they are kind of experiencing in realtime at the bench. And often times they aren't really able to go to their primary mentor for support in this area because the mentors are kind of like, "Well, this is your work," and this is new for them as well, and, "Part of your dissertation is figuring it out."

Letisha Wyatt:

And so when you are kind of starting off in a grad program, everything is new, and you are trying to basically understand how to become a scientist and how to manage data over a long period of time. And what I try to do with them is give them as much information upfront so that their awareness is high so that their eyes are wide open so that they kind of know what to expect because I find that the more seasoned maybe late stage grad students have been burned at some point, and so they take this more seriously, whereas the earlier grad students are just like still trying to get their sea legs and figure everything out, and then they kind of stumble around until they get to a place where they're like, "Okay. This seems to be working." And then three or four years down the line, they're like, "Oh wait, maybe this isn't working as well as I had hoped."

Letisha Wyatt:

So my goal usually is to just try to get to them early. The more senior graduate students, sometimes when I see them, they're like, "Oh, I'm in this place where I've got this system that hasn't served me well actually, and now I need to like figure this all out and straighten this all up so that I can get my dissertation written." And really it's one of those things where I hope to impart some of this stuff that I wish I had known when I was going through my training. It's just a matter of getting the earlier grad students to really understand how important it is and how serious it is. And also I find that when there's no ... So the other trouble is, when you get to them earlier, they don't really have the context to apply the things that you're trying to direct them into doing. It's a hard road to navigate because you have to kind of ... The timing is really important essentially. Like when you get to them, they need to be able to readily apply what they are learning.

Maryann Martone:

Mm-hmm.

Jane Roskams:

In helping them convince the people that they're working for and around to step into this world, Maryann just touched on incentivizing. And incentivizing is a big word that can mean many things, but do you think the younger generation, if they had access to stories from the frontline of data and software sharing, discoveries that were made because that process existed in a more accessible and palpable fashion, that it might help in persuading that this is a world worth stepping into? That bridge has to contain a number of things, funding to be able to support this. And that can come from the agencies, but being convinced that the world of scientific discovery can be changed by this, and your questions might be answered more quickly, I feel there's a gap there. And I'm just wondering if there's room for the Twitter feed or the webpage highlighting, "Guess what just happened because of data sharing," right? "Guess whose community might be helped because this software was shared across all these labs, who then were able to answer the following questions." Is there anything like that that exists?

Letisha Wyatt:

I'm not certain. But I do think, I mean aside from Twitter, I think that that is really a really powerful tool. When I was going through my training as a grad student and postdoc, I didn't really have ... I mean, Twitter existed, but I don't really feel like it was as robust with a lot of conversation within scientific communities. And I wish that I had that as a grad student because when I'm on Twitter ... I'm on Twitter now, and I see so many grad students who are like, "Ah! I need help with figuring out this patch situation, or this western blot, or PCR, or whatever," and you have this just giant international community of people who are there to kind of help whenever people put out a call, which I think is really amazing. So, harnessing that ability for us to be able to like really amplify those stories and to highlight those things I think could be a really powerful way to get people onboard, if they're not already. It's just a matter of getting more people into those spaces.

Jane Roskams:

Maybe we just created a whole brand new project for BrainFacts.org.

Maryann Martone:

I do think also though that, since the topic is data management, we have these communication forums, but again that is a conversation between individuals, and then it just disappears, and that's always been one of the challenges is to get these conversations happening in places where they're searchable and people understand that they're available to you. Because data management requires the stability that things are documented in a lab where you can go back in the future and get them. Good luck trying to go back and find a Twitter conversation you had two years ago, right? It doesn't work.

Maryann Martone:

So I think the mindset here is always this ... And that sort of world of asking colleagues and getting advice is something that we know. We used to do it just physically. We'd go down the hall and ask somebody. But I remember early on, as we were starting to transition to online, some of us realized that these email threads where people were having these conversations, if they were done in online forums, they then became available for everybody to read, as opposed to just doing it inside of one email thread that was only available one on one. And you started to realize the possibilities that happened out there, but still there's this need to make sure that things are identified, that you can find them. The FAIR Principle cite has to be findable, accessible, interoperable and reusable. So, just because it happens digitally, as we know, doesn't mean that you're going to be able to find it and access it later.

Maryann Martone:

But picking up on a point that Letisha made before, and I think also you mentioned again that often times in laboratories every graduate student and postdoc is left to their own to decide how they're going to manage their data, and some of them are going to be very good at it, and some of them are going to be very bad at it. But the thing about that is you are thinking about how to manage it for yourself. You are not thinking about how to manage it for the lab, and you're certainly not thinking about how to manage it for sharing later on. And I think that mindset that data products, software products, all of those are now research products. They're digital assets that are going to be shared and are meant to be shared.

Maryann Martone:

New policies and otherwise journal requirements makes you now have to think, "Does the system work for everybody, or does it only work for me?" And I think that's one of the very first steps. And if you're a PI or others listening to this, whether you're starting out or others, having a consistent strategy in the lab or at least signaling that data management is valued I think is extremely important because you can't ... If that's the strategy you use, then you're going to end up with as many different ways of managing as people coming through the lab, and you may even end up with more because often times people change as they go through their career.

Maryann Martone:

So I do think that a mindset that says, "When I start acquiring this data from the get-go, at some point either it's going to be audited, I'm going to need to create a figure out of it and give that to the journal, I'm going to have to give it to the agency, whatever it is, might cause people to think differently about how they manage IT inside their lab as a whole instead of in this sort of very ad hoc way.

Letisha Wyatt:

An example of making data findable and reproducible on the lab side, I really like Maryann's comment about a good place to start is with something like Protocols.io because I see so many times where folks are creating these protocols and maybe iterating on them on their hard drive or in their lab notebook, and then there could be several different protocols across the lab for the same technique. And so just kind of adopting this mentality, as Maryann mentioned, about opening up your scope and thinking that, "What I'm doing is not just for myself, but it's also for others," which is probably what we should be doing already, would be a really wonderful first place to start.

Letisha Wyatt:

I know a lot of folks think like, "Okay. If I'm working on a new technique, and I'm kind of developing a protocol and iterating on it, should I be keeping notes and keeping records for every single iteration? Are these data valuable and meaningful?" And maybe some folks would argue that they're not, but I can certainly see how something like Protocols.io could make it easy to keep versions of things up to date and make it more accessible to people even just within your lab.

Jane Roskams:

It's absolutely essential, otherwise the buck cannot stop with you, and that's what the funding agencies expect. So I strongly believe in putting the funding agencies on the mark now that so much is digital. So much data that we're collecting including all the photos off the confocal, or the two-photon, or whatever the fancy microscope is or imaging technique that we're using, that every single one of those

things does matter and that funding agencies need to make sure that we're not just filing a data management plan but a data management plan that comes with a cost and that the cost can be covered, whether it's AWS fees for the volume of data that we're collecting or essential software that might not be open that we need to use. And these are not things that are necessarily taken care of at the level of funding decisions right now. I think it's time. It's past time, and it will change the way that all labs can interact in the future with each other and within a given lab.

Maryann Martone:

I think when it comes to data management in the laboratory, I already mentioned I think in the very beginning that technological flux often kind of confounds your efforts because if you invest in a system and that system goes away, then you're in trouble. But I think if you understand conceptually what you're trying to do, then you can swap out the technology as needed or bring it in as needed. So it doesn't matter what your storage is, right? You need to have some storage. And now with Dropbox and Box and Google Drive, there's a lot of good cloud based storage where you can put things.

Maryann Martone:

But if you look at the FAIR data principles, the word metadata appears over and over and over again, right? It's really metadata which makes all digital assets useful. And without that, then they're pretty much useless. There's nothing you can do about them if you don't have the metadata. So getting in the habit of doing things like putting manifests or other things or read-me files where you provide that metadata anytime you have a digital asset is always extremely important.

Maryann Martone:

One of the reasons I like BIDS, I've mentioned that before, but it's just file hierarchies and naming conventions, and even those can get you very far if the files that belong to a subject are named according to that subject ID. Even if you don't have every piece of metadata that you possibly need, you can look at the order. Your machines can process it. It's very, very powerful. I think, thinking again about the longterm and data stewardship, right? Who's in charge of your data? Often times, again, it's a postdoc. It's a graduate student. They are transient, and so they're going to leave, but the institution itself may have infrastructure available to you.

Maryann Martone:

And the benefit of turning stewardship over to someone else is, A, generally you have to put in some metadata, at least some basic metadata about who produced this, what lab did it come from, and what's it about. And I will say that any metadata is better than no metadata no matter how far you go. But that also means that they're going to handle the technological flux, right? If they have to swap out the backend system, that's not going to affect you, right? That's going to affect them, and they take charge of that. So you find these trustworthy stewards of your data. It doesn't mean you have to do everything yourself.

Maryann Martone:

So I think that there are, even though it's daunting to sort of field a fully configured system in the laboratory, the tools have gotten reliable enough and good enough that if you start to implement just some standard practices, you can go far. And I know there are online courses coming out of ... I think it's University College London. I'm not 100% sure ... of just, "Here's how you manage your data." And it starts with things like file names. It starts with things like folder names, right? It starts with things like

basic metadata. And if you do that conceptually, then the technology that you choose to use is a combination of what's available and what you can afford or what you like to use. It almost doesn't matter because you've taken care of the universals about it which can exist no matter what system it's on.

Voiceover:

Thank you for listening to this episode of Pathways to Enhance Rigor: A Collection of Conversations, brought to you by the Society for Neuroscience, the world's largest organization of scientists and physicians devoted to understanding the brain and nervous system. You can hear the rest of this series on your preferred podcatching app.

Voiceover:

Be sure to visit Neuronline.org/FRN to explore the other resources and materials created as a part of the FRN program. That's N-E-U-R-O-N-L-I-N-E.org/F-R-N. We'd like to know what you thought of this episode. Please take a moment and check the episode description for a link to a one minute survey.

Voiceover:

Supported by the National Institute for Neurological Disorders and Stroke grant number 5R25NS112922-02. Drs. Os Steward and Lique Coolen are the principal investigators and senior producers. This episode was written and produced by Maya Sapiurka, Tristan Rivera, Emily O'Connor, and Taylor Johnson. Audio engineering and post-production services were provided by Human Factor.

Voiceover:

Thank you for tuning in, and we'll see you next time.