Transcription of the webinar: "What is the Forest Service Research Data Archive and how to prepare data and metadata for submission"

Description of acronyms used:
APHIS = Animal and Plant Health Inspection Service
BDP = Biological Data Profile
CSDGM = Content Standard for Digital Geospatial Metadata
DOI = digital object identifier
EFR or EF&R = Experimental Forest and Ranges
FGDC = Federal Geographic Data Committee
FY = fiscal year
GTR = General Technical Report (Forest Service publication)
ISO = International Organization for Standardization
JFSP = Joint Fire Science Program
R&D = Forest Service Research and Development
RGEG = research grade evaluation guide
RNA = research natural area
SRS = Southern Research Station
USDA = U.S. Department of Agriculture

Speakers:
Speaker 1 is Stephanie Laseter
Speaker 2 is Laurie Porth
Speaker 3 is Dave Rugg

NOTE: Presentation officially starts at 00:30

Speaker 2 (00:02)

Making my screen full to see more of a slide.

Speaker 1 (00:05)

You may have to make your screen full to do that.

Speaker 2 (00:08)

Okay. Sorry. I apologize – you just hit record.

Speaker 2 (00:14)

If you're monitoring in the background, I might do that just to make it a little bit bigger.

Speaker 1 (00:18)

Yeah, that is no problem. Well, go ahead and do… we have about a minute, folks, coming into the room. So good morning or good afternoon, depending on where you are joining us from today.

Speaker 1 (00:32)

And thank you for being here.

Speaker 1 (00:34)

We are recording. I just want to let you know that and we will post that off online and I'll share the link where that will be posted in the chat pod. I appreciate everybody putting who they are and where they're coming from in the chat so we can see who's joining us today. So feel free to do that. I'm going to introduce myself, I'm Stephanie Lasseter. I'm with the Southern Research Station, and I am a science liaison with Region Eight out of Atlanta, Georgia. I'm going to introduce Laurie, who's doing our presentation today. Laurie is an information specialist with Research Data Services, and she's at the Rocky Mountain Research Station. And I used to be at Coweeta Hydrologic Lab, affiliated with that EF&R and submitted a lot of data through Laurie up to the archive and have worked with her for years. And she is a joy and makes this process painless and easy. And so I really look forward to having you learn more about the archive today. If you joined us last year, Lindsay Rustad was the chair last year of the EF&R Working Group. I'm the chair this year, which is why I'm sitting on screen here with you.

Speaker 1 (01:45)

And we introduced this idea at the town hall. The first few slides you may see that those are familiar if you were on that town hall, because that's going to be an introduction to the archive and then get into kind of the nuts and bolts. Also, reach out and introduce Dave Rugg. He's probably on the call here today. Can help field some questions in the chat with me also. And Dave is the Research Data Archivist. If you need to make the screen bigger, at the top of your screen here is little four arrows, and you can make the Zoom that out to full screen so that you can see the slides. I did a quick audio check and got some good feedback, so hopefully you are able to hear Laurie and I no problem.

Speaker 1 (02:30)

Please put your questions in the chat.

Speaker 1 (02:33)

You all are in listen only mode right now, and we'll adjust that at the end for questions, but go ahead and drop in anything during the chat, and Dave and I can help field those and get them to Laurie at the end so she doesn't have to try to watch that.

Speaker 1 (02:50)

Okay.

Speaker 1 (02:51)

I think that's all for the housekeeping. Laurie, anything I missed?

Speaker 2 (02:56)

I don't think so. I think we're ready.

Speaker 1 (02:59)

I'm going to turn off my camera and turn it over to you. And so thanks, everybody, for being here and take it away Laurie.

Speaker 2 (03:07)

Thanks so much, Stephanie. Thank you guys all for joining. We're excited to talk to you guys about the Forest Service Research Data Archive and what we do and how we can help you get your data published. As Stephanie mentioned, I'm actually a member of the Research Data Services staff so I work directly with Dave Rugg, and one of my main duties is production manager for the Forest Service Research Data Archive. So if you are someone who's going to be publishing data with us, you will be interacting directly with me. The purpose of our archive is to publish digital research data that's been funded by Forest Service R&D. We are also the recommended repository for the Joint Fire Science Program, so anyone who receives JFSP funding, it's recommended that they put their data or their metadata in our archive. We also have helped some other entities publish their data, such as APHIS, the National Wildlife Research Center. What do we publish? We publish all different sorts of data. We publish long term research data, such as experimental forest data. We publish data from specific studies, short term studies, data that are related directly to a manuscript for a journal article that you might be submitting.

Speaker 2 (04:20)

We also can publish supplements to Forest Service publications. We do have some scientists who are publishing GTRs, and there's maybe some additional information that they can't, for whatever reason, put into the GTR or research notes. And we can actually publish that for you and directly connect those two items. And recently we've added the ability to published image libraries. I'll talk a little bit more about

those in a few minutes. So to find our archive, the URL is listed above. The other way to do that, and probably the simplest thing to do is to just go into Google, and if you just put in “Forest Service” and “research data” or a couple of those words, you'll find us pretty easily. There's also a link to our site on many of the different Forest Service station sites and others as well. What are we going to talk about today? I'm going to show you how to find and obtain data from Forest Service Research Data Archive, and then I'm going to step you through how to actually prepare all of your files for submission to us. And then finally we'll step you through the final how do you submit and get to the point where your data are published?

Speaker 2 (05:25)

So the first thing I want to show you is what users see when they come into our archive. We have the menu items at the top, and we tried to keep our website very simple, so that is very easy for people to find data. You can see that the very first item menu option at the top left is Data Catalog. So folks interested in data can simply click on the data catalog button and that's going to bring up a list of publications in our archive. Right now we have just under 600 data publications that are available to folks. The easiest way to narrow down that list of 600 is to either limit your search with some of the menu options that you see on the left. You can look for things published by year, by funding entity, by location, by authors. We also have the ability for you to look for data that are GIS related. So geospatial versus standard digital tabular data. You can also just put in a key word or phrase. So in the center of the screen you can see we put in the keyword “Penobscot” just to show you how it works.

Speaker 2 (06:27)

And as an example, we have nine data publications in there right now that are related to Penobscot. I also want to point out on the top right-hand side of the screen is a search catalog options. And that option shows up on every page in our archive. So no matter where folks are, they can easily just write or type in their keyboard or phrase to find data. Once you do that search, it's going to bring up a list of all the titles that meet that criteria. And then below the title you will also see the list of authors. The user can then just click on the data of public interest and that's going to bring up what we call the data publication details page. So on this page we're going to give the user all the information that they need in order to access that data. They have the title and authors at the top. As you can see, we also have right near the top how to cite the data. This is very important. We want to make sure that anyone using this data very easily can see how they should properly cite it.

Speaker 2 (07:28)

I have an arrow pointing at a URL that is a digital object identifier. If you are submitting your data and we are housing your data in our archive, we will assign a DOI. This is a persistent URL. It's the URL that will not change. So for instance, when the Forest Service web pages changed from fs.fed.us to fs.usda.gov, we were able to redirect all of our DOIs so they will always point to where the data actually are housed. So these DOIs are great for journal articles and many of you probably know that a lot of journals are

actually requiring the data be published and the DOI provided before they will publish the manuscript. That is something that we can help you do. Further down on publications page, you can see that we're going to provide folks with an abstract telling them what data are included. And then towards the bottom of the data publication page, you are going to see a data access section. So in this section we are telling folks how to find the data. For every data pub, we have the view metadata option. Metadata is documentation for your data, so that allows folks to get a quick glimpse of the documentation.

Speaker 2 (08:43)

If we are including your data as a downloadable zip file, we're also going to provide you with a file index, which is the second thing that you see listed in the data access section. This file index basically is going to list all the files in your publication and a short description of what they are so that folks can get just a real quick glimpse at what's included in the package. Now, the download options are going to vary a little bit depending on your publication, but we have four main download options: the zip file, query database, the image library, and external repository. I'm going to show you just a little bit about each of those briefly. Downloading a zip file is the most common way that we provide access to the data publications right now. If the user downloads that zip file and unzips it, you're going to see over on the left-hand side, a look at what they're going to get. They're going to get at the root of that folder a file index, which if you look at the right-hand side, you can see an example of that. We literally list the files, what folder they're in, and then a description of what they are, which also includes the format.

Speaker 2 (09:49)

We're also going to provide them with the metadata in two different formats. We're going to give you the XML, and then we'll also give folks the HTML, which is a little bit more user-friendly version of the documentation. You're also going to see a data folder. Every date of publication that comes through our shop is going to have the data put into a folder called Data, and anything additional to that would be put into a folder called Supplements. One of the main things that we're trying to do when we're providing these data to folks is to make it very easy to understand what they're getting in their package, hence the name of the folders. We do have the ability to generate subfolders, so depending on the complexity of your product, we can certainly have subfolders below those two folders as well. The second way to provide access to folks is via our query database. This is something that we've done for a couple of our experimental forests. When you come into the query database, you'll get some summary information. So up here on the left-hand side, you'll see the Penobscot research data. We just kind of provide an introduction to that area.

Speaker 2 (10:54)

And then as the user scrolls down that screen, they're going to see a list of data that are available for that site, which is what you're seeing on the bottom and left-hand side. Under the Penobscot Silvicultural study, we have five different data sets. Folks can again use the metadata, and then they can choose the query download option. Once they're on that screen, they can limit their search by choosing different

variables, maybe different years, or in this case, different management units that they may want to download. And they can also choose to access summary data as well. One thing I want to note is that even if folks enter the query database environment, they still have the ability to download the full zip package. The third option is the image library. This is fairly new. You can see it's a slightly different interface to the archive, but basically we provide access to different images, and most of them right now are historical images that we provide, like this really cool picture that you see here. But one thing I want to note is that we don't just publish the images. We actually have to have documentation or metadata for those images.

Speaker 2 (11:58)

And we realize that could vary. You may have a lot of information about some, not much about others, but we are publishing more than just the images themselves. And the last download option is what we call the external repository option. We know that some folks have to publish their data in other archives. Tree ring data or genetics data may needs to go into GenBank, or maybe your data has already been published in another repository, but we still want those data to be discoverable within our Forest Service Research Data Archive. So what we do is publish your metadata in our archive, and then if you see the publication details page as you look down under data access, we're going to allow them to view the metadata, which we do host on our site. But then when the user wants to access the data, once they click that button, it's literally going to take them to the other repository where they can then access the data and additional documentation there. So now we've looked at how to actually get data out of the archive. Let's now talk about how we can help you prepare your files for submission to the archive.

Speaker 2 (13:05)

Step one is going to be to determine what files you want to include. Typically, we are publishing the raw research data files, so these could be data directly associated with a particular publication. They could be observational historical data that are related to a short-term study or a long-term study. We also have the ability to publish secondary data if it's something that's been substantially modified from its original version, as long as we have concurrence from the originating source that it is okay to do that. And then as you look at files for the supplements folder, we obviously are providing access to additional files as well. And we can really include almost anything that you think is important to understand the data that is beneficial for the user or things that you want to keep with that data package. So that could be maps that could be photos, it could be unpublished reports, study plans, data analysis, documentation. We've also had people include R code or MATLAB code, anything like that that you want to include, we most certainly can. So the second thing, once you've decided what you want to include in the package, the next thing you need to do is prepare that data and the supplemental files.

Speaker 2 (14:15)

When it comes down to the data, the most important thing is to make sure that the data have already gone through rigorous quality checks. We know this is something that people are typically doing on their

own before they even think about submitting. But we found a lot of just minor issues and people send the data to us. So we definitely recommend that you do some quick checks just to make sure that everything looks ready to publish. Now, we do have cases where people need to publish data that perhaps have not been through a full QA/QC process. Sorry, quality assurance, quality control process. And there may be a myriad of reasons for that. And while that's not the ideal situation, we can still publish those data, but we need to make sure that the documentation very clearly note what level of QC or not, that data have been through. And a couple of things to have you double check. A lot of common errors that we see are outliers. We see a lot of typos with longitude and latitude as an example or other little things. So definitely take a look and make sure that your data doesn't have any outliers.

Speaker 2 (15:19)

We are not going to do any big checks on your data. However, we are going to very closely compare your documentation to the data itself. So as an example, if you have an outlier, we're probably going to see that. And if you know those values are valid, and we want to make sure that we are putting that information into the metadata for the user so that they then in fact don't have to ask that question as well. Another big thing that we see is categorical variables. When you tell us that you have a variable that has values of ABC and D, definitely double check to make sure that you really only have values ABC and D. Make sure there isn't an E and an F, because very frequently we find that. And you also want to make sure that all of those values are fully described. Another example - species. If you've got ten different species that you're providing us data for and you're using species acronyms, double check that those acronyms are correct and that they're all fully defined. Another thing to think about is missing data. If you do in fact have missing data, we need to make sure that we're noting why you don’t have that data if it's known.

Speaker 2 (16:26)

For instance, if you didn't collect data at one particular station for one year for some reason, we definitely want to provide that information to the user. And then the last thing to think about when you're looking at your data is making sure that you are clearly defining and consistently using zeros, how you're using blank cells versus - 999 to denote missing data, or if you've got like “<” signs. So very frequently we will see people using multiple ways to denote missing data. And we just want to make sure that one is consistent and that we're also explaining that to the user in the documentation. As far as the files themselves, we like to use transparent file names. You can submit any type of file names that you want to us, but we are going to consider we definitely don't want to have 300 character long file names. But again, we're thinking of the user and trying to make things very clear, easy to use and simple. So if you have a data file that has streamflow, and if you have another one that has precip, it would be beneficial, obviously, to have that information in the file name.

Speaker 2 (17:35)

As far as file format, you can submit your data again in virtually any format, and we can help you convert it. But when we are publishing data, what we're thinking of is we want people to be able to access the data today. We want them to be able to access it in ten years and 30 plus years from now. We also want to make sure that anyone can access the data. We don't really want to require them to have certain software or certain operating system in order to access it. So those are things that we're thinking about when we're publishing your data. Some very common formats that we use are commonly limited ASCII text, which is CSV, Excel, Text, JPEG, GIF, all those standards. And so we will work with you because we realize there's some cases where we need to have your data in a different format, but just know that those are the things that we're thinking of. We want those formats to be stable and usable to the best of our ability. So you've got your files prepped and ready. Now we need to write the documentation. This is probably the biggest step in the process for most folks.

Speaker 2 (18:37)

The data documentation is required for all of our submissions. This document is going to provide the user with everything that they need to know to completely understand the data. We're telling the user why it was collected, how it was collected, the quality of the data, who they can contact if they have questions. Metadata can be written in many different standards. The standard that we use is the CSDGM and the BDP. In 1995, the Federal Geographic Data Committee created the standard. And for the Forest Service, and I shouldn't say they created it in 1995, but the Forest Service made it mandatory in 1995 to utilize the FGDC standard for all geospatial data. Then not too long after that, the Biological Data Profile was created, and this is a superset of the FGDC. So what they did is they took the CSDGM standard and they added a couple of elements so that the metadata standard would actually work for biological data. What we found is that that standard actually works for any type of data. It works for more than just biological data and actually works for almost anything that you want to publish that is data or supplemental file related.

Speaker 2 (19:54)

In the inception of the archive, we made the decision that we wanted to utilize these two standards. Since then, multiple other metadata standards have transpired. We have the ISO 19115 standard. At this point, we are still using these standards. We could potentially move to another standard in the future. But if we did, anything that's published in our archive, we would go ahead and convert that metadata for you. So let's just talk a little bit about the standard and the different sections. There are seven main sections in the BDP and the CSDGM standard. The first is identification. This is the section where you are going to tell us what data were collected.

Speaker 2 (20:38)

Or I should say, tell the user why that data were collected, where specifically it was collected. We're actually going to ask for boundary coordinates and a description of the location, and then if there's any particular tools that are needed to work with that data. The second section is the data quality section. This

is a section where you are going to tell the user specifically how the data were collected. We have a methods section in here. You're going to talk about the reliability of the data. So how accurate do you think those measurements are? Did you do any quality assurance and quality control checks? If so, let's provide that information in the documentation, and then if you had any omissions, or if you had certain selection criteria when you were collecting that data, you're going to provide that information in that section as well. Sections three and four are spatial, and those you don't necessarily have to enter by hand. If you are dealing with geospatial data, you're probably using ArcGIS, and so some of that will be filled out automatically for you. So really, you just need to know that this section is going to tell the user if it's raster data, if it's vector data, is it point or line data, is there a projection.

Speaker 2 (21:46)

That type of information is going to go into those two sections. The fifth section is very important section, that is the entity and attributes section. So in this section, we are going to fully describe the files. We are literally going to list all of the files that are included in the package, and we are going to define all of the variables that are within that file. Now, one of the things I keep saying that we want to make it simple and easy for our user to understand and utilize the data. Every submission is very different and we realize some folks are going to be submitting metadata with one file and five variables so it's very simple in one metadata document to describe a file with five variables. But we also have submissions where we have folks literally submitting hundreds of files, and there could be 30 or 40 variables. And so putting all of that information into one metadata document maybe isn't the most feasible and easy way to make it available. So we do have some other ways that we do that. We can sometimes provide a list of variables in an Excel file, or we might have what we call a user guide where we do that.

Speaker 2 (22:53)

So I just want to let you know that depending on your data, we do have the flexibility to do some other things, but we really our key is to make it easy for the user. And of course, we want to make it easy for you as well. Section six is the distribution section. This is where we're going to tell the user about the formats that the data and the files are in. And the last section is the metadata reference. And there we're just going to tell the user the last time the metadata were updated and who they can contact if they have questions. So now you're probably wondering, how am I supposed to create the metadata in the standard that you just talked about? If you go on our website to the Metadata and Tools tab, we're going to give you a little bit of information about metadata and what it is. And then towards the bottom of that page, we're going to provide you some details about the tools that are available to you. There are many different ways to write metadata, and we're just going to talk to you about the three main ones that we use.

Speaker 2 (23:51)

ArcGIS, as I mentioned, if you have geospatial data, you probably are already familiar. You can write your documentation directly within this software. The default standard is ISO. If you are submitting that data to us, you can simply submit the data with the ISO 19115 metadata. That's no problem. We have the ability

to export that and convert it to the CSDGM standard. We don't need you to do any additional work there. We also have Metavist software that was written by our very own Dave Rugg, and it's free. It is user friendly. We have a link to download that on our site there. And it requires a little bit of knowledge of the FGDC standard. But literally me just kind of stepping through the sections right now, as I did, is probably enough information for you to understand how to jump in and use Metavist. It's really easy to use. And I'll show you on the next slide what it actually looks like. Metavist will work for geospatial data and non-geospatial data, so it utilizes both of the standards that I mentioned, and it's going to generate the metadata in a raw XML file.

Speaker 2 (24:57)

And then we have a style sheet that's associated with it in Metavist so that you can export it to the HTML format of the metadata that we provide, which again, just makes it a little bit more user friendly for people who aren't familiar with XML. And then the third option is the Microsoft Word form. This Word form is very easy to use. It doesn't require any knowledge of the FGDC standard. We try to provide some little tips and tricks in there on what type of information should go into the cell. And it works for all data. It works for non-spatial data. It also works for geospatial data. However, we don't have the ability for you to fill in the sections three and four, which are those two spatial sections, but that's information that we can always add later. You can certainly use the Word form for that data as well. So now let's take a look at what Metavist looks like. So over on the left-hand side, you're going to see seven different menu items, and those are the seven sections of the metadata that I just showed you. Basically, what you do is you click on the menu option, and then it will bring up on the center screen a series of tabs that allow you to enter additional information.

Speaker 2 (26:13)

So right now I'm on the identification section and basic information tab. And so here you can see that we have the ability to put in the citation for the data. You can put in the abstract and the purpose, time period, and then if you were to click on spatial domain, then it will allow you to put in the boundary coordinates and information about your study area, etc. It's very simple just to page through all these tabs and you can just type directly in it, or you can copy and paste from Word as well. One thing that I want to point out, just because we happen to be looking at the metadata right here in front of us, one of the common mistakes that we see from folks when they're submitting documentation for us, very frequently, the data are directly associated with a journal article that someone is working on. We don't want the abstract and the purpose to be that directly from your manuscript. We don't want the manuscript abstract, which is talking about your results. Basically, we want you to focus on the data itself, and then we will point folks to your manuscript in the metadata.

Speaker 2 (27:20)

So it's definitely going to be accessible to folks, but we want the metadata documentation to focus on the data itself and what's included. Okay, so let's take a look at the Word form. I hope you're able to see this.

It's a little bit small, but I wanted to try to get the majority of the form on here. So basically we've done it with the same sections that you've already seen. You can see on the top left is the identification information section. And so literally you're just going to tab through this form and again, type directly in it or copy and paste from something else. You can see under originator it tells you that you need author names - please include the middle initial. Under the description of the data publication, it tells you the abstract should be a narrative summary of the data. You can see in parentheses, we try to provide a little bit of information to help you better understand what goes in there if we think that that might be a question or a stomping point for anyone. Over on the right-hand side, you can see about part way down in section two.

Speaker 2 (28:21)

So that's the data quality information. So again, here's where you're going to be providing your methods, anything that you did to the data, information about its accuracy. And then you can see it goes at the bottom - you go straight to section three, which is entity and attributes, because again, as I mentioned, we don't have the two spatial sections available in this file. Okay, so we've prepared your files, we've written the documentation, and we just need a few more things from you. We have to have signed submission forms in order to publish data in the Forest Service Research Data archive. There are two forms. One is a Forest Service author form, and the other one is for non-Forest Service authors. The main difference is that the Forest Service author form requires a signature from the program manager. It's basically just to ensure that you have the ability to note if the data are influential or highly influential, just in case it requires some additional checks. We need the first author to sign a submission form, depending on if they're Forest Service or non. However, if the first author is not forest Service, let's say it's a University collaborator, but you are as a Forest Service employee are the third author, we need that first Forest Service author to also sign the form, simply because this is the Forest Service Research Data Archive and everything is going through us.

Speaker 2 (29:48)

Also want to note that we now accept electronic signatures. If for some reason people do not have the ability to use a lincpass and officially sign the form, a lot of times people will just kind of copy and paste in the JPEG of their signature. And that's fine. We accept those as well. But we like to have the email showing that form being “signed” and forwarded just so that we have additional confirmation that that person actually signed it. Okay, so now you're ready to compile your submission package. You've got your data, your metadata, your supplemental files. A lot of times, folks have many different files that they're submitting, so feel free to organize your data in whatever folders you want when you submit them to us. But do note, as I mentioned already, that we're going to put everything into a Data folder in a Supplements folder. But again, we don't require that when you submit. We just want you to know that that's how it'll end up looking to the user when they download the files. And again, you can also generate subfolders. We just ask that you try to keep those names simple and very clear and also useful.

Speaker 2 (31:00)

You don't want to have more folders than files, so just use those to help make it easier for the user to see what's included in the package. And then the last little thing that we like you to include when you're submitting the data to us is the file index. You saw an example of that early on. Basically, we just need to know or have a short description of the files that you are including in the package. You can submit this to us in any way that you want. You could do it in Excel, Word, you could use a text file or email. It doesn't really matter. We just really need that information from you. But again, going back to keep it simple, we do have folks that again, are submitting maybe hundreds of files. We don't necessarily need to list every single file in the file index. For instance, if you're submitting 50 images, and perhaps the file name is the description of where they are, or the date or anything like that, we can generically describe those files in the file index. So again, we're trying to keep it simple. We just want to make sure that we're getting the information out to the user and making it easy for them to use.

Speaker 2 (32:04)

So if anything seems like it's taking too much time or it's not making sense, please contact us before you dive into spending too much time on something that we might be able to simplify for you. And that's it. You are now ready to submit your package to the archive team. So now you've got everything prepped, now how do you actually submit? On our website, we have the Submitting Data tab, which is going to basically step you through all of what we just talked about and what's included. The first thing that we ask you to do is to contact a member of the archive team. If your files are small enough that you can just email them to us, that is totally fine. That's no problem at all. But more and more frequently files are too large for people to submit via email. So that's why we ask you to contact us, and we can work with you to perhaps utilize Pinyon or we've had cases where people have actually had to mail us an external drive because the files were too large to actually easily upload into Pinyon. Right now, when you click on the Archive Team on our website that's going to bring you up to information for Dave and myself.

Speaker 2 (33:11)

So either one of us can accept submissions. Once you submit that package to us, we are going to respond to your email with a track ID so that you have a number assigned to your submission. And then we are going to put you into our review queue. We thoroughly review every submission that comes through our door. Unless it's an external data publication, where perhaps the data are available on another site. But anything that we are internally publishing, we're going to thoroughly review. That's one thing I want to point out that is a little bit different than a lot of other repositories. Most repositories, you just upload your data and your documentation and they published it as is. We're actually providing a service. We want to make sure that these data are used and that they're used properly and that the user completely understands what they're getting in the package. We are going to literally step through your metadata and we are going to step through the data. We are going to compare the two. We're going to make sure that the documentation fully describes the data. We're going to make sure that the documentation is understandable and complete and consistent.

Speaker 2 (34:19)

And we also try to make the documentation understandable to non-subject matter specialists. However, we realize that depending on the complexity of your data, that may not necessarily always be the case, but we do the best that we can so that anyone coming in to use those data could read and understand the documentation that they're getting. We will completely review the submission package, and then we will respond back to you with review comments. And then we will also send you a draft version of the metadata and the file index, both of which we will put in Word with track change just turned on, so you can very easily go in and can make whatever modifications you want to the metadata and that file index. And then you send those two things back along with your review comments. And then once you send them back to us, you don't have to go back into our queue. You basically are at the front of the line at that point. And then we make final edits. And then as long as you and I agree on what the final package looks like, then we assign a DOI and voila your data are published.

Speaker 2 (35:29)

And just a couple of additional things. We have a lot of folks, as I mentioned earlier, journals are meeting DOIs, and they're needing them before the data or before the manuscript can be published. So we have a lot of folks who are in the journal publication process and all of a sudden realize they need a DOI. We can assign a DOI early. But what's required is a submission form. And then we prefer the final version of the data and the metadata. But depending on your timing, we realize that you may not be quite ready to do that. But if we have at least a draft version of your data and metadata and the signed submission form, I can assign a DOI immediately. We don't even have to begin the review process. We can assign that immediately so that you have a DOI. It will point to a temporary out of stock page, so it will tell someone if there were to access that DOI, it's going to say, hey, these data aren't ready, check back it should be ready soon. So you have a working DOI that you can put into that manuscript, and then in the meantime you need to get those final files to us so that we can begin the review process.

Speaker 2 (36:35)

We have a pretty large review queue, and we have a lot of folks that are needing early DOIs? We're kind of trying to juggle all of that. We just ask that you work with us to help ensure that your data are published, because we want to make sure that that DOI is pointing to actual data before your journal article goes live. And obviously, ideally, we would prefer that you touch base with us a couple of months before you're submitting your manuscript. We just realized that right now sometimes that's not happening. We're at least providing this option so that we can get folks that DOI for their article. And last, just wanted to tell you a couple of fun things about publishing with us. For those of you that do not know, this is fairly recent. The Forest Service RGEG direction now says that Forest Service Research Data Archive publications are now considered “formally refereed”. So this is great news for those of you that are RGEG. And also just wanted to let you know we have a lot of folks that are asking, hey, are people coming? How do I know people actually going to find my data?

Speaker 2 (37:38)

Do you have a lot of folks accessing your website? And we in fact do. In FY 20, we had about 35,000 global customers, which I think is huge. We had almost 8000 downloads just in fiscal year 20, and that was about a 25% growth from the previous fiscal year. Now we can't keep track of specifically who is downloading and accessing our data. However, we do have the ability to see where they're coming from, and we can tell that we have a lot of government agencies, academic institutions, we have a lot of K-12, and also a variety of business folks that are coming in and accessing our data. And we also have other repositories that are providing access to our data. One of them is Data.gov. That is a catalog for all U.S. government data, so folks in there could find our data. AG Data Commons is a catalog for all USDA scientific data. And then we also have other sites like Science.gov and Data Citation Index that are also providing access to some of our data publications. And that's it for me. Questions. Looks like I saw a lot of chatter going on during the call.

Speaker 2 (38:46)

So, Stephanie, I don't know how you want to start questions.

Speaker 1 (38:53)

So if you guys want to put questions in the chat pod, that would be great. Or raise your hand at the top, that would be great. Or in the chat pod, I'm putting the call on number in, and you can call in. And if you do that, make sure you mute your computer speaker, because you'll be on kind of a two-fer. So you have lots of options here. A lot of good questions and conversation in the chat. Laurie and Dave answered several of those, and I see that multiple people are typing. And yes, there is a recording.

Speaker 1 (39:33)

So the recording. I'm going to put the link here. It's on the archive website in the chat pod. I just posted that. A question from Steve. Is it possible limit access until the manuscript is actually published? Great question. I know you get this a lot.

Speaker 2 (39:50)

Yes, we most definitely can. We have a couple of different ways to do that. So one, we can actually fully prepare for your data for publication and not even announce it or make it available in the archive. But we have it 100% ready to go. And then as soon as you're ready, we can push the green light and make it available. The other thing we can do, and we've done this for a lot of folks. We call it the data embargo. We put the metadata, the documentation, into the archive, but we don't provide access to download the data. So the user can read the full set of documentation and see what's coming. And then we just kind of give them an approximate date. We give them a month and a year when we realize sometimes that can change. And so then we can do that as well. Now, every case that we've done so far, we've asked the author if they would be willing to provide the data to folks that wanted it. If that's the case, we can do that.

So as an example, we have a couple of data pubs in the archive. The documentation is available only, and then we have a note on there that said, hey, these data won't be available until about April 2021.

Speaker 2 (40:56)

Sorry. But if you're interested in these data, you can contact them and then we can let them know who they can contact. So we have a lot of different ways that we can do that for you, depending on your needs.

Speaker 1 (41:09)

Great. Thank you. Laurie Eradellas. Yes, we will add the slides with the recording. I will make sure that that happens. And at the end, I'm going to also put up the link to the Experimental Forest Box folder and talk about that, too. So we can put them in a couple of places. Cherie asks, is it okay to fuzzy the coordinates for ecologically sensitive field sites?

Speaker 2 (41:34)

Absolutely. And we can certainly publish, for instance, if you have information about your site, such as these specific longitude and latitude, we can publish your data without that, but we can internally on our end, keep that information for later if it's needed. But yes, you don't need to give exact coordinates. Really, we just want to get the folks an idea of where those data are. And providing the bounding coordinates is a really easy way to do that in addition to the description. It also allows us to have the ability in the future, we would love to be able to have folks come into our archive and kind of click on a map and say, I want data for this area. And so having those bounding coordinates in there allow us to do that in the future.

Speaker 1 (42:14)

Research natural areas. Do you have information or data about RNAs on the archive already, Laurie or Dave? This is certainly not limited to EF&Rs.

Speaker 2 (42:28)

We're hosting this training.

Speaker 1 (42:30)

But it's open to any data sets – any Forest Service data sets. But can you speak to the RNAs, Laurie, do you already have any information on those?

Speaker 2 (42:39)

We do not. I don't think we have any data for RNAs on our site, and I'm not sure that our site would be the location to provide those documents. We certainly have the capability to do that, but that would maybe be a question for Dave, if that's something that he would want to include. I know some of that information is already available, like in Pinyon. We have a lot of establishment reports available for EFR.

Speaker 1 (43:08)

I will say that I am the RNA coordinator for SRS and we have a coordination call on Tuesday the 23rd at 200. If you're associated with an RNA and you don't know about that call, if you just send me a Ping me and Stan, that might be a good time to have that conversation. Just send me a chat and I'll make sure you have that link because the Team's call. Okay, Jonathan, is there a maximum size data file size? So in other words, you ask me about high res aerial photos or drone data. Kind of big file sizes at this point.

Speaker 2 (43:43)

No, at this point we don't have a maximum size. Now, the maximum file size that we've provided for download at this point has been a little over 30GB. We can provide access to a larger file than that. However, we realize it's hard for people to download. So what we've typically done is if it's too much bigger than 30GB, we usually break that data down into some sort of a meaningful package and allow them to download multiple files. And note, this is when we zip the file. So this is a 30 GB zipped file. I realized some imagery when you zip it, it's not going to get much smaller anyway. But at this point, we can certainly try whatever file size you have. We just haven't played around too much with files that are bigger than that once zipped.

Speaker 1 (43:38)

Dave, go ahead.

Speaker 3 (44:39)

All right. It worked. I called in because I want to do the RNA thing. And so there are two things I wanted to mention about RNA. One, we do have a little bit of RNA data in there already from the Moqua RNA up in Wisconsin. I know we have that because I co authored the data set. We don't have a lot of RNA data. We do know that every station director has a responsibility in the manual to have a repository structure for RNA data, but station directors are tired of maintaining that, or they never got around to building it. All you have to do to get RNA data is just submit it just like EFR or any other data, and we are more than happy to publish it.

Speaker 2 (45:27)

So Dave, they had also specifically asked about establishment - RNA establishment documents, and so that's why I comment. I'm not sure we would really be the place to host those, but do you have any thoughts on that?

Speaker 3 (45:42)

Correct. Yes, it's been on my to do list for a long time. And actually, since Ralph Crawford was doing the establishment reports for EFRS, we have talked on the EFR working group from time to time about having a counterpart site for RNAs, and none of us have ever gotten around doing it.

Speaker 1 (46:00)

So again, I'll put the link in there. For the first time I've been part of a kind of a national RNA call, which is the 23rd on Tuesday. That would be good to maybe talk through some of these questions there. Okay.

Speaker 1 (46:12)

Talking about accuracy, Laurie and Dave, accuracy varies with the variable. And can you talk about how that's handled in the archive?

Speaker 2 (46:24)

So that's one of the sections that's typically lacking for almost everyone who submits it. So really, we can provide whatever level of information you have. We would, of course, prefer accuracy down to the level of every single variable. But honestly, we don't see that very often. So, yes, we absolutely can do that. We can put it directly into the metadata. If the information is too substantial to maybe put into that main metadata document, we can provide it in another file and then include it as a supplemental file. Absolutely. Anything you want to tell the user about your data we are willing to include and also point out, too, if you have substantial methods or other documentation, we don't necessarily have to have that all in the main metadata document. We prefer that. But if it's really extensive, we can actually, again, provide it in a separate document that's included in the package.

Speaker 3 (47:17)

And we have actually done that in a number of cases. We had a major carbon database that was published as a collaboration across a number of EFRs. And so we had a standard metadata document, and that, in fact, then referred people to the enclosed user guide, which was much more extensive. It all works pretty well.

Speaker 1 (47:38)

All right. Can you point folks to sources of keywords? Is there a certain source you want folks to refer to?

Speaker 2 (47:50)

We provide three different types of keywords. We let the authors submit whatever keywords they want. We haven't actually developed our own taxonomy for that. That's something that we maybe should do in the future. We have our own internal list, and we'll kind of try to compare and slightly modify like if you put the word tree and we want to be trees, we'll make some of those modifications for you. But we'll do all that on our back end so you can put whatever word you want there. But we also include ISO 19115 keywords, and we also include R&D taxonomic keywords. So when you submit your package to us, we actually on our site have a list of those keywords that you can choose from. But typically we will just send you that when we send you review comments and just say, hey, can you also add some of these keywords as well?

Speaker 1 (48:39)

How about folks that want to send you huge numbers of pictures? Is it possible to do an archive of a big entry – in this case, hundreds of photos that relate to a project?

Speaker 2 (48:50)

Yes, we absolutely can do that. So we can do that as either its own submission. So we have a lot of data pubs that are really just images, so that would be an image library submission, but we can also include them with the data publication as well. And so what we do even though I mentioned that we allow people to download a zip file very frequently, it's more than one zip file. So we could provide your data and supplemental files as a zip file, and it has multiple downloads that allow them to choose what images they might want to download. If you have images for a certain decade, we could zip those up and provide them a separate download. So there's a lot of different ways that we could do that. So, yes, we can definitely include multiple images. If you are submitting for just the image library, we definitely would want documentation for each of those images. What they are, keywords, year it represents, where, photographer if possible, and all that type of information. If you are just including it in your data publication as information related to the data you're publishing, then we just need a short description of what is in that image, or if we can generically describe it for a set of files. Again, I don't necessarily need you to define all 100 images if we can somehow do a simple, more generic description.

Speaker 1 (50:11)

Okay, Laurie, talking about data set that may have PII that relate to people, is there a way to publish something like a data summary rather than a raw data stream?

Speaker 2 (50:24)

Absolutely. We can do whatever makes the most sense for you. We just prefer to publish the raw form of the data typically because that's what it's simplest to do and then allow folks to make their own modifications. But in a case like this where you do have PII or other information you don't want to provide,

we can certainly do that. And we've also done things too where we've published data, maybe down to the person level, but the particular variable that would be information that would be considered PII or enough information for someone to be identified could be removed. So we can do a lot of different things but we are willing to work with you to make it work.

Speaker 3 (51:04)

And then the data about people is also a good example of why we have the alternate repository capability because there are places like University of Michigan that specialized in this kind of data and protecting it. And so no problem from our perspective for you to use those kinds of repositories and just drop a metadata record with us so the people who are looking for services or related data can find it. Our whole point here is to get, as Laurie said, get your data out, get your data out where it can be used and particularly whether it's sensitive people data, sensitive animal data, sensitive archeological data, any of that kind of stuff. The other point is to make sure it's protected just so that you would do if it was just you doing this all the time and responding to the people asking to have a copy of your data. So yes, and if there's anything that comes up after the presentation that you have questions about or would like to do something that's out of the ordinary - always ask us, almost always we're able to do something smart.

Speaker 1 (52:14)

If you guys cannot tell, this RDA team is just awesome to work with. So this is a good question. One from Steve and Jonathan have the same thing. I send you all my rain gage data Laurie, and next year I need to update it with the new file. How do we go about updating data or manuscript updates, etc.

Speaker 2 (52:35)

That's a very good question. Something I probably should have mentioned. Yes, we definitely can update your data and we have a couple of different ways to do that. One way to do a simple update. If you just have changes to your metadata, something was incorrect, you just want to add a little bit of information or you found one or two data points in your file that you already published that were incorrect, then we can just make a modification to that metadata and then what we do is we update the citation and will literally just include “updated on 02 February 2021” or something like that. Now if you have a substantial change so you've added a new variable or you have massive amounts of new data to add, then we would potentially do a new edition. When we do that then we will republish the data and the documentation. We will put like a “-2” in your DOI we'll have the same title and say second edition. If someone were to go in to try to find that first edition there's going to be a note right at the top that says hey there's a newer edition available so we make it very clear that people - for people - to find that new addition.

Speaker 2 (53:40)

Now if you specifically are referencing adding another year of data we have a lot of long term EFR data etc. are on our site. What we do when we have new - if it's just a new year of data, we can also just add

that to the current data publication. We will just put a note in the metadata that it has been added and then we'll also add that updated on date in the citation so that folks can see that. We try to be careful and not put the actual years in the title perhaps because we don't want to change the title for a publication that we're updating, so we just want to make sure that we're careful so that title doesn't have to change and then we can just add that on and that's very simple to do and those types of changes also go to the front of our queue. We usually get those done pretty quickly for you.

Speaker 3 (54:28)

And we can assure you that when we do new editions with a little header on the original edition page people do find their way over to the latest stuff. It doesn't mean that they stop downloading the original edition because sometimes they want to see your original work because they're doing something building on those original articles you published, but people are very good about all of a sudden shifting their download frequency over to the new stuff and we have some pubs that are in their 4th edition and we've got actually one coming up we hope this year that will come out its 5th edition, one of our most popular pubs.

STOPS AT 55:05

Speaker 1 (55:06)

Great, these are great questions, everybody. Thanks so much for putting them in here. So Ken says, great, this is helpful. Now I need to get to work. How can he link photographs with – you know if you've got a project and maybe times series data or maybe vegetation plot data and pictures. Can we link those?

Speaker 2 (55:26)

So, we don't put any hyperlinks or anything like that in our documentation. But again, we're going to list in the file index, we're going to list that those photographs are included in the documentation. We're going to reference those as well. We could if you really wanted to link them, we could create an Excel file that goes in your package that lists those particular images, that has a hyperlink to them within that package. So we could do some fancy things like that if you wanted that. Typically, what we do is we use that file name transparency to help make it clear what's in there, so that if someone was looking through the supplemental files and in your photographs, it would be very easy for them to see which images they would want to look at.

Speaker 3 (56:09)

And certainly, the metadata, the formal metadata standard has a place for describing images that support the data and what they're about. As long as you don't have - it'd be hard to do this in the metadata thing.

I've got 50 images and to document that in there. But if you only have five to ten, that works pretty well. Otherwise, the approach Laurie was talking about, where you put that in as a separate file works really well.

Speaker 1 (56:36)

I'm going to do the last two questions here. Well, Jeanette has one, too, so I only have five minutes left, and I want to make sure I get to a couple of other closing points. Cherri has two. The answer to this is I think it will be easy to say yes, but data be archived that's not connected with a publication (pub)? And if there's support for historic data sets from a retired scientist. So I know Don Bragg is working on a Centennial project, for instance, and that's SRS. So - not connected with a pub and some historic data.

Speaker 2 (57:15)

So absolutely, it doesn't have to be connected with the pub. Most of our data sets are not. So, yes, we can certainly publish any sort of data for any time period that you want. As far as support, I don't necessarily know that we have financial support, Dave can maybe address that more. But basically, if you can send us a draft version of your metadata or a manuscript that contains that information we can help, we are very overloaded with work, so we can't necessarily write from scratch metadata for folks. But we really try to do everything we can to make it easy for you and help you as much as we can. We can definitely interact with you one on one to try to assist with that.

Speaker 1 (57:55)

Okay, now to talk about accuracy, can you give an example of a good description of variable accuracy? And Jeanette, I would say, too, that you can go online and download a couple of examples from the page. And I think Laura Kenneth mentioned this, too in the chat. It's really helpful to look at what maybe somebody else has done as an example to get that take. But maybe I don't know, Laurie. It depends on the variable. But do you want to give an example of accuracy?

Speaker 2 (58:26)

Well, this may be a simple example. If you're providing a tree diameter measurement, how accurate is the measurement that you're providing? If you provide information about the measurement instrument, typically, you know, oh, are you guys still there? Sorry, my whole screen just went black.

Speak Speaker 1 (58:42)

Yep, I can still see you and hear you.

Speaker 2 (58:44)

Okay, good. Basically, you would just want to let folks know how accurate that measurement device is. A lot of times we'll have people giving us variables where there's, like four or five decimal places. And so we want to let the user know, is the data really accurate to four or five decimals, or do you need to round it? So information like that is very helpful. And then the other thing is providing information on how you probably checked it. Did you graph it? Did you compare it against data from another station? Things like that would help us tell the user how accurate that measurement is.

Speaker 3 (59:16)

Now, for this audience, I will also toss in another notion of accuracy and metadata completeness. Many of you, possibly all of you, are running long-term experiments, so you're writing metadata, particularly for the metadata for the data sets that are coming off these long-term studies. Definitely do think about … what would I want to know if I was taking over this study with no prior knowledge? What do you need to communicate to your successors so they can be successful with the data that you have been stewarding all these many years?

Speaker 2 (59:52)

That's a great way to explain it.

Speaker 1 (59:53)

Just in closing, Laurie, thank you so much. It looks like several people had not heard of the archive. That is great. We have a couple of people on here who are longtime users of the archive, so we're hitting a really nice variety of folks. I dropped a link in to the box folder. If you were here from an experimental forest. The working group who organizes this training is not limited to EF&Rs but if you're affiliated with an experimental forest we do have this box folder available to you. Our team and work this year is going to focus on climate across all 82 plus experimental forests. I'm looking for folks who would be interested in doing some climate synthesis work. And that's a vague and broad term, but as chair, I'm trying to pull together some of those ideas, and we as a working group would like to talk with you and bring you into the fold. So, plug for that. We can do more training like this if that's something you would like and just encourage you to go to that box folder, and really glad that you all filmed it this afternoon and made time on your Friday. It looks like really good input and a lot of kudos to you Laurie, so again the recording will be on the archive page. The email that I use to send this out. I will let you know when that's up through that same email and I don't know about you guys, but it was kind of nice to come back to Adobe today. I was a little bit Teams done, but I don't know. I found that kind of a refreshing different looks for a webinar. Happy weekend to all of you and have a great rest of your day and thanks again for joining.

Speaker 2 (1:01:47)

Thanks Stephanie and thanks everyone. Bye.