Workshop 5: Creating FAIR Outputs Transcript

16 July 2025, 12:01pm

Samantha Rees started transcription

Samantha Rees 0:13

Oh.

Great. OK. I think it's started recording. So I will hand over now to Elf DEET and Jasmine Hunter who are leading this session. Oh, sorry. Can I quickly, sorry, I know I'm having issues trying to figure out how to get over the.

Commissions for the videos because so it seems that most people can access the one that I sent on Thursday about the intro to get. But most people can't access the one that I sent on yesterday on Tuesday morning. That was for the first day.

So yeah, if you if you have been able to access it, could you maybe sort of either e-mail me or just let me know, but I'll presume everyone can if I don't get any emails and thus I need to find a workaround. But yes, sorry I will stop talking now. Thank you. Thank you, Jasmine. Thank you, elf.

Hywel Lloyd 1:08

But.

Jasmine Hunter 1:09

All right. Thanks, Sammy. Yeah, welcome everyone to Workshop 5. Hopefully you can all see the slides on the screen. If you can't let me know, but we'll get started. So some quick introductions. So, yeah, I'm Jasmine. I'm a data steward at ukch, my background's.

Kind of in paleoclimate and proxy data, and then with me is Els diet, who's a bio GA chemist by trade, but also works in the data stewardship team.

So the way this hour will work is we've kind of split it up into five sections. So we've got a general intro on what is fair data. We'll then look at some more practical applications of how we can prepare fair data and then look at publishing and how to do the metadata associated with that.

At the end, we'll then have a tools and resources portion and some time for

questions as well. If you have any questions during that kind of presentation portions, I just chucked them.

Hywel Lloyd 2:07
Nobody.

Jasmine Hunter 2:10
Oh yeah.

Hywel Lloyd 2:12
So that's it.

Jasmine Hunter 2:12
Yeah, just Chuck them into the chat and then one of us will answer as we're going along with that.
And actually you can access the chat as well or you can't see the content.
And everyone else? Is it just one one person?

Els Dhiedt 2:37
Yeah, I can see it.

Hywel Lloyd 2:37
I can see the content.

Jasmine Hunter 2:39
OK.
Well, carry on then. And maybe if you leave and rejoin, whether I couldn't see it, that might solve that issue.
So we'll get started with kind of just what are the fair principles some of you might have heard of these before, some of you might not have, but just for a general quick history, they're kind of this set of principles that were made by this group of scientists for a bunch of different.

Brook Legese 3:03
Yeah.

Jasmine Hunter 3:10

Fields within the scientific community back in 2016, and they've really evolved kind of as we have gotten more and more into, you know, endorsing open and transparent data into all this different stuff now. But it really was trying to fill a gap of people hoarding their own data and working in silos and moving towards working. Together in an interdisciplinary kind of way to make things more accessible to everyone.

So we're going to start with a quick little poll in the chat just for a bit of fun to see everyone's awake, so else is going to put into the chat the first one. So what does F stand for? Is it a findable, B feasible, C functional or D friendly and just want you guys to like?

Using your emojis, which one you think it is?

And there's no we're not marking this, by the way. It's just purely for a bit of info.

So be brave in your your answers.

And to like it, you just hover over the text and then the thumbs up button should come up. Or yeah as you said right in the chat, if you can't use that.

Function.

Yep. So we've had two for a.

Couple more. We'll carry on that. So yes, the first one is findable if we go on to the second one, we've got automated, accessible, accurate and achievable. And I can see Kunis is way ahead of the the game there, going all in one.

Yeah. And I so integrated into operable, intelligent and innovative.

Got one from.

Hi again with BI actually hadn't heard this is the first time I'd heard that word was in relation to the fair principles, which is quite interesting.

And finally, ah, we've got recyclable, reproducible, reliable and reusable.

Yeah, else is working furiously on the copy and paste in into the chat there.

Yeah, this is one of the ones we've got a few different ones. Probably the hardest that we made out of them. In this case it is reusable though, both reliable and reproducible are good to have in your data as well. So yeah, those are the four kind of main fair principles. Findable is of course can that.

Data be found easily both by humans and machines, and is it discoverable? So if you go out looking for data, can you actually find it easily? And then once you've found it, how easy is that data to access and download? Can you see clearly you know what

the terms of use are, the licencing and how to actually use that data.
Then the third one we have into operable. So how do you integrate this data with other data sets, workflows, other applications and really so it can be used across a wide range of systems and finally reusable. How easy is it for the data to be recreated and understood?
And this really looks at the metadata and supporting information around that data set to help provide the context of using that data set in a new way.
Point here just about fair, that's open. Obviously we can't have all data completely open. You know there are confidentiality reasons of maybe personal data involved or commercial, but we can have it as open as it can be.
I love this quote from AIDC, so data should be as open as possible as closed as necessary, and a data set can be still be fair while not being fully open if it shows all those reasonings why it can't be, it has those clear terms of use.
And requirements for reuse in the same way you can have a completely open data set. You know you can launch it onto the web and have it available, but if it doesn't have that metadata and supporting information around it, it's not reusable. So therefore it's not fair. Even though you can.
Find it.
So we'll just go into kind of the why it's important here because it's also one of the main things that we want to look for when we think about why we're going to add these extra things, starting with more the stick than the carrot to look at some obligations, all of you will have different funders that you're kind of supporting your research.
And they'll have their own data policy and requirements for the outputs. From your research, you know this can relate to accessibility. It may have an embargo period. So when you need to publish your data to buy, it may have a set cut off, like within two years or something.
And how you need to publish that data.
As kind of we become more mainstream in this push towards open and transparent science, you see, publishers are also getting on board with this and a lot of them actually have a requirement for you to have your data published in an Open Access repository.
So you know, on top of your other research, you need to have your data available, which just in really improves the reliability of your research in general.
And finally, there may be some legislation, depending on your where you are in the

world you have, you know, different legislation that you'll have to meet to a few examples from the UK and the EU. So personal data, we have the GDPR, so general data protection.
Regulations, I believe, which just relates to if there's, you know, personal data about someone, there are things you need to anonymize and that sort of thing. If you work with maybe a rare species of wildlife or plant or something, you may need to hide that location.
To meet certain legislation requirements, and there's also the Inspire legislation in the EU which just sets about a framework of sharing the spatial location and the themes of data that's been made in the EU.
I bet you have specific ones, so it's always good to keep abreast of where you are with that.
But on to the more fun things. The benefits. We'll start off with you guys. The data creators. One of the key things is of course you get credit for the actual work you've put in to creating these data sets, both as authors and promotion of the outputs that you're putting out. This also gives more trust in your research.
If you've got all the metadata there, people can clearly see how you're how you've made your data, and that you've done all the necessary cheques and balances and QAS in that. And finally, you can build your network by through promotion of your work and hopefully form some new.
Collaboration opportunities.
Looking towards the data user, which some of you will be as well, you don't need to be reproducing this data set that's already happened. You can focus on just using it to get to that next scientific question and we can really, really push the boundary of where we're going with science.
By not having to redo work and spend time and money doing that, it's also easier to access and reuse other data and gives a big opportunity for collaboration as well.
Looking towards the scientific community in general, obviously we want to build on existing science. We really want to be the top of where we can go and being able to find and use data can really help with that and inspire new, you know, new interdisciplinary research projects.
Yeah.
Yeah.
And then just stepping back from the scientific community, if we look at society more broadly, a lot of our work is funded by the public and we want to give back to the

broader society in that. And, you know, impact policy impact through outreach and scientific communication. And that access to data is one way to do that. So being really clear and transparent about our.
Science and the data outputs that come from that.
So we're going to stop for a second and just do some interactive exercises. So if we look at this one, this is how do we actually achieve fair. So we've talked broadly about what to do, but now we want to go and actually come up with some practical ways we can make our data more fair.
So else is going to add a link to a padlet into the chat. I want to get you guys all to go onto. There'll be 4 columns, one for each letter of the principles, and then you can just add ideas using the plus symbol. A black box will come up. You can add your text.
And put it on there. You can also like other people's ideas and expand on them through the comments.
We'll set about 5 minutes to do that and I'll share it on my screen as well.
Now, if you can't access the padlet, just let us know in the chat as well.
I can see a few people are typing, which is good.
And for this you can just add quick ideas or longer paragraphs up to you, but add as many as you like across all four.
Yep. So I can see ones popped up about code on GitHub, extremely important to share. You know the the methods behind the analysis you're doing and GitHub is a great open source way to do that.
Yep.
Publishing it through a databank and reposit, there are some repositories already that you can put your data into and also keeping a track of a kind of a broader database of all the stuff you're working on is really good.
Yeah, clear description of data is hugely important. It's one of the kind of main things I think particularly for actually may be able to find something through keywords. You know, if you think you're searching, you might search groundwater and if it doesn't have that one keyword in, you might not see that data set.
Does anyone have any ideas for interoperable how do we make data that we publish kind of usable across different systems?
Yeah, metadata wise, you definitely want to include the how in that, so publishing your code on your method is really important and having it actually with your data set rather than over in a, you know, random published paper actually having it all in

one spot.
Can really help.
OK. We'll give it one more minute and then we'll carry on. So quick get any last minute typing in.
Yep, standard format. Fantastic and yeah, making that one that can be used across different programming languages and in different software.
Nothing worse than you know. If you find some data you really want to use, and then you need to buy a licence to a specific software to be able to access it.
Yep, and a read me file can be hugely important. All right, I'm going to swap back now, but we'll leave this padlet up if you want to add more to it later. Or take a look at it. It'll be up for the rest of the week.
And yes, so please add any thoughts you have kind of during the session?
And I'll just jump forward to where we were.
So you should see up to the how to achieve there. So these are kind of the specific principles that come through from the original set of guidelines. So you can see having a web accessible record of your data set, hopefully in a online catalogue or something that you can find including.
A unique identifier so ADOY, which we'll talk about in a bit. And also I think someone mentioned this about having keywords and description of that data set. It's hugely important to make it findable.
If we look at accessible again this persistently working links, having something that you know when you click on it, it's always going to be there, at least you know for the long term, including those constraints on your data to be you know as minimal as possible. We want to be as open as we can be and any constraints that you need to have include that in your licencing information.
And you may want to use various APIs and use interfaces if you have a lot of data that people might want to access, you know, like just the chunk of it. You know one month or something like that, including the user interface is hugely important.
Interoperable, we had yet standardised formats, also standardised languages and vocabularies and units, so these are really important so you can compare across different data sets. You know what you're looking at and keeping it in machine readability as well.
And reusable. Again, discipline specific vocabularies that people know what they what their definitions are following all the metadata standards and having appropriate supporting documentation and then clear licencing for terms of use.

So these are some of the kind of key considerations that we want to go through. So obviously you guys are all in, you know, various subfields, but you'll have kind of community standards that are specific to your area. A few examples here. So for techniques.
You can see you want to have clear descriptions of your sampling methods. If you've used a specific method that's common in your field, you know link to that reference. That and say that's the method you're using. If you work in a lab geochemical reference material and lab standards that are cross.
Can be used across different labs. Are really important to include in your metadata so that you can easily compare and combine data sets from different labs and then of course uncertainties and errors we need to know what that data can actually be used for. You know you whether it's A1 kilometre spatial grid or A1 centimetre.
Gridded data. You can't. You need to know those units in order to be able to access that data and use it effectively.
Obviously you need to know what the standards are going to use them, right? So there's two tools here and they're in our link of resources that I will share with you after one is the fair sharing repository of standards. It is really good for like a detailed search. You can apply philtres.
If you want to just search for hydrological standards, you can just add that into the search bar. The other is the DCC list of metadata standards. So that's the digital curation centre. This is more general standards that are common across different fields.
But can still be useful if you're looking for general standards for your meta data.
So controlled vocabularies are kind of a key way to do this, so if you haven't heard of that term, they're basically just a list of standardised terms. So if you think like a dictionary or a thesaurus, these would be examples of controlled vocabularies. They're just used to enhance searching by using the same wording.
Across different research fields and across different studies to make it easier to search. And there's two examples there. One is the Nerk vocabulary server. This is kind of oceanic data initially, and it has a lot of stuff on different units, so you probably can't see in that example, but there's one on milligrammes per gramme, as a defined unit.
And how to wrap that in your in your data set so that it matches search terms? The other is the climate and forecast standard names parameters. This initially was just for atmospheric data, but it's kind of been broadened out for any net net CDF.

Data sets in and just includes kind of different metadata standards that you can use for that and what terms to use.
Just looking at data formats, so there's lots of different formats. You can use. Lots of things to consider. The first one is what are you going to use it for. You know what software do you need to be able to analyse it? You know do does it do you need to do programming or can it be done just in Excel? And what's the easiest way to do that?
Does the file include embedded metadata? So net CDF you can obviously add a lot of the metadata into the file itself, whereas CSV you'll need a separate supporting document to have all of that metadata in there. Software and skills are important to consider. You know. Do you need to learn some programming to be able to do this? Can you just pick it up?
Up and run with it. How intuitive is it? And then how are you gonna version control your data as you're transforming it and processing it? So whether that's through a separate document or just in inside the file itself? It have. If it has an inbuilt version control such as, you know, word does and that sort of thing.
And then finally, when you get to publishing, do you need to change the format so you might have needed this specific software to run your analysis. But then when you get to sharing it with everyone, turning it into a more Open Access format that can be used.
You know, in different systems is a really good way to go and this could be as simple as converting, you know an Excel workbook into a CSV table as that could be used, you know in different languages.
And finally, this is just a plea to take notes as you go. There's nothing worse than getting to the end and having to try and remember everything you did. You know, I can barely remember a month back. There's very specific things. The stuff I did in the lab. But if you do it as you go, you can create your supporting information while you're actually doing the work.
And this includes adding comments to your code to make it reproducible, and so that others can use it.
I see it make detailed notes even if you don't, you know, share it with the world, you might come back in three years and want to change or want to reuse that data in a new way. So you help yourself later on.
So we're going to pause here for 5 minutes just in case anyone has any questions of

what we've discussed so far.
Yeah. And if not, we'll carry on and I'll pass over to Elf.

Els Dhiedt 24:21

Thank you very much. I think I've taken control now.
Umm.
So the next section I will be talking about data publication, which is a very important part of Fair and starting. I will kind of talk about what is data publication and kind of what does the process involve.
Yeah.
So traditionally and you will often see people will do their study, they will produce a a journal paper, a research paper, and somewhere you will see maybe like a statement saying like if you are interested in this data, you can contact.
Such and such a person. But yeah, that's obviously creates an extra barrier. Possibly the data will be not really in a reusable state, so it's not really what data publication is about.
What it is about is that you can have a research paper and it will have a statement referring with a citation that has an associated DIY to a separate outputs, which is this data.
Sets so the separate outputs that comes out of your study is this published data set which has its own licence associated with some aspects of version integrity. How to access it? A whole heap of methods?
Data and supporting documentation and all these kind of things associated, which you can also think about, which are also often associated with other outputs of research like journal papers and also have its DUI.
So and that kind of comes back to. So I'll just kind of expand a bit on this kind of idea of a DIY, which is a persistent identifier. So it might gives a unique identifier to a piece of work.
Or something that is involved in the research process. So you probably have heard it in journal papers, so they are also associated with data sets and you can you can easily link up these two outputs with each other. You might have also heard it for authors as well.
So this would be an orchid ID. So it's a different type of persistent identifier. So if you for instance have two different people with the same name in the same research fields, it's clearly kind of distinguishes these two people and also easily links them to

different outputs.
And you have the same thing. Also for research institutions, but this is then called an ROR and then finally can also have them for physical samples. So this makes it all very easy to kind of link outputs and kind of aspects of the research process to each other.
So now you know what data publication is and what it is not. Yeah, we kind of want to know why we should publish, because well, it sounds like it could involve quite a lot of work. So I want to kind of ask you, what do you think the benefits are?
Please can you put this in the chat.
Yeah, maybe you have published in the past and maybe you can give a reason why you have published. Or maybe you can think of reasons why you should publish and you can maybe think also back on the benefits of Fair in general because a lot of them kind of link back to each other.
You could think of benefits to yourself or more general to.
The research community.
OK, something's coming in. Yes, makes it reproducible. It's important for transparency. Makes your data invisible, so that's very important for. Yeah, kind of your own research career. So you can kind of show.
That you've created outputs and maybe also create linkages and networks with other researchers. Yeah. Also to educate the public we want. Yeah, general kind of.
Uh benefit to the public? Um.
Others can use your data also. Yeah, the confidence. So the transparency is very important.
Yes, avoiding recreation of the same data so people can actually reuse your data and don't need to reproduce it. So we can build on top of other science.
Umm.
Yeah, these are all very good suggestions. Thank you very much. So I'll just kind of summarise based on that. So as some of your sets, it kind of creates this kind of open science. So that leads to more transparency and also helps.
Other science to build on top of what you have produced, but also helps the kind of public benefits as well, because obviously well one that they can also access your data but also just by producing more and better science.
Then it also gives you sell. It gives some recognition for your work, so it's an output on its own in its own right. So you can actually use this to further your career and show on your CVI have actually produced this data sets.

Then it's also especially important for that kind of reason for data creators like field technicians and lab technicians who might not always be involved in creating of.
A journal paper. So here's an example of a data sets on which some research and a research paper was based. But it's not all the authors are the same in each. So.
It actually gives people who were more involved in the data creation or collection also the chance to get an output for their work.
And it's also important that it can be reused. So here is an example of a soil moisture dry down study that was solely based on data that was published elsewhere. So there was no new data created.
During this project, but because all this data was findable, accessible, interoperable and reusable, it was possible to actually combine these in one workflow to then create this new and exciting piece of work.
And then finally, it's also important to meet policy requirements. So, for instance, your funders might have requirements for you to publish your data, and also more and more journals, yeah, are now asking you for you to publish your data alongside your.
Research paper.
So now you know why you should publish. It's also important to know what to publish, because during the process of your study you will of course create loads of data, but not all of it is US of long term value.
So I will kind of walk you through a little kind of scenario example and I will ask you at the end which of these outputs that are created during this kind of little study example study.
That you think are of long term value and also should be published by you specifically. So this example is measuring groundwater and soil moisture using sensors and this data comes in a.
Comes to you over the clouds and sensor specific software formats for which you need a licence to access it. For that reason you transform it into a CSV format so you can easily use it in your workflows.
And then you also want to kind of combine this data that you've collected in your project with some weather data and some river flow data that you've accessed freely because it's published elsewhere, for instance, by the Met Office here in the UK.
And you use these two sets of data in your model. You put on GitHub to then predict groundwater level soil moisture, and to actually predict create these predictions. It takes 5 minutes to actually.

You run your model to produce these predictions. So now I'd like to ask you these are 5 outputs or potential outputs and which ones do you think are of long term value as should be published by you at the end of your projects. So.
Yeah, this is first this.
Yeah, groundwater and soil moisture in the software specific formats. Then also the same data but in a CSV format. The stir party data that you've used.
Your model and then finally the predictions of your model.
I'll give you some time to think about it.
So it can be multiple things as well, but I see that some people indeed take multiple boxes so.
It will start with one, so no one has chosen that one, and that's indeed correct, because you can access it because you have that licence, but others might not be able to. And indeed 2 is.
As most people here have indicated, it's the same data but in a CSP format so other people can actually use it. So that's indeed of long term value and should be published by you. Then three is a tricky 1.
Because yes, it's of long term value, but it's a third party data, so it's already somewhere published, so you don't have to publish it again because you have not transformed it in any meaningful way.
So this one you shouldn't publish then four is yes, that's most people indeed have indicated this one. So yes, it's good practise to a publisher model, especially if it can be useful to other people as well.
And no one has chosen 5, which indeed because it's so easy to reproduce, it's not of long term value because other people can just use your data that third party data and your model to reproduce it themselves.
So I think most people were right on the ball, so I'll just summarise kind of these thoughts and some general practises. So data of long term value is not all data. So for instance if it can be easily reproduced, it won't be of long term value.
So this is often the case with model data. You kind of have to consider. Is this easily reproducible? Like does it only take me so many minutes to actually run the model? Do I not need expert knowledge? Is it deterministic model, etc.
Then you also need to look, yeah, whether you have generated the data and it can incorporate third party data, but you have to check whether the licence allows you to also incorporate this data.
But yeah, obviously it needs to be transformed in a meaningful way and then we

have, yeah, raw versus processed. So generally say as raw as possible.
That's of course. Here. We saw this kind of format that come out comes out of the sensor which is not reusable or interoperable by other people. So therefore that this transformation was needed.
And then finally, as complete as possible. So this means that if you for instance expect to come back to the same sites multiple times, maybe wait until you came the last time or if you have multiple sites, maybe wait until you visited all the sites and then kind of combine these in non data sets.
But of course, if it's long term monitoring, yeah, don't wait until kind of you're finished, because that might be, yeah, years or never. So then annual or monthly snapshots might be more appropriate depending on the frequency of your kind of outputs.
Then it's also important to think where you will actually publish. So generally this is done in a data repository, so as people already kind of sets in the that let's kind of questions.
So this is a centralised place to store, organise and access data. And yeah, the purpose is for sharing your data, preserving it, analysing and management and there's some kind of things you need to consider when actually choosing the rights repository for your.
Science. So first of all, also check your funder data policy, because they may actually say oh, it needs to be published in this specific repository. Or they may say oh, it shouldn't be published in these types of repositories then you.
I also need to think whether the repository meets the fair principles so and this you can kind of check, oh I don't.
OK, sorry. You can check in certifications. For instance. Yeah, some might say whether they're as fair or whether they will hold your data interpretuity as well.
And yeah, finally you also need to check out your intended audience. So who do you actually want to reach with this data and who do you think will be interested in this data? So you might be very interdisciplinary and then maybe a channelist.
Repository may be more appropriate or it's quite kind of more specific to your research domain. So like, yeah. So then you kind of maybe want to look in more repositories that are specific for your research domain.
There's a link here where you can kind of look for repositories as well.
Umm.
Yeah, so now you know kind of more about data publication and Jasmine has also

explained more about how to prepare your data. I will explain a bit more on how to prepare the metadata and supporting documentation so.
So first of all kind of explanation of both, so metadata is more structured and this is the data about the data and it's structured. So there's some metadata standard. So it's actually possible to kind of be interoperable between different platforms. So that can kind of.
To make it clear that you can search and translate kind of metadata over different platforms and then supporting documentation holds a lot more detail about so it makes it possible to fully reuse the data as well.
So the reasons to create this is of course to make sure that you never forget what you did and then also helps researchers to interpret the data and then and also to find, retrieve and reuse the data, especially kind of make it with a.
Keywords and DOI etcetera. Kind of having a clear sensor where your data sits that is findable and accessible. It's also ensures that you get proper attribution for your research with this DOI and also the licence.
Make it clear how it needs to be kind of cited. That also helps to comply with data sharing policies because some may actually. Yeah, there may be policies that require you to publish your data.
In a fairway and then finally avoids misinterpretation because it's all very clear. So I again want to put the question to you. So I kind of think someone comes knocking at your door and says like, oh, I found perfect data set.
For your study, what would you need to actually reuse this data?
So you can't think about.
Yeah, maybe some data that you've used, maybe it's your own data and kind of information that you need or data that you've used from someone else already and kind of information that you've already.
That maybe you lacked and you couldn't reuse it or.
Yes, permissions and licencing. Very important to check in the beginning because you might.
Do a lot of work and then in the end find out that actually you couldn't use this in this way, so you're not be able to actually use it in future projects.
Other other more kind of practical things, you can also think of that are more in kind of when you actually handle the data.
Yes, units very important. Otherwise you might be misinterpreting the where the data is from, so this can be partly where it's actually located study sites, but also who

actually produced it.
Is also important to know.
And how it was derived methods.
Yes, is it reliable. Also very important. Yeah. Are there some information about QA and QC or?
Did they not check this?
Yes, accuracy for uncertainties and resolution you may need to very fine resolution for your purpose and they might have only daily data or one kilometre data.
And generally the metadata. Yeah, just kind of your who what's where. So, yeah, these are all very good ideas.
I'll summarise these.
Yeah, but has sensitive of private data. Usually it shouldn't be published if it has or, it will be clear in the licence how you can use it. If there is sensitive and private information in it.
And then it will make they make clear how you can actually use it.
So some summary of this information. So yeah, you need information about the data itself. So it's already been suggested, so you need to know how accurate and how like. Yeah, like whether it's been Kuwait, the formats and whether it's whole and complete.
Maybe it's about Europe, but maybe it's country is missing and maybe that's the country that you're interested in. Then also you need some metadata. The Who, what, where, how then?
You also need some descriptions of kind of formats and contents and the structure of the data so you know in advance how you will need to end or how you will be able to access and analyse the data without having to spend hours to actually understand.
And what is in there then? Also some administrative data, so the licence of course acknowledgments. How will you actually attribute the data? And?
Then also some more relationship data or metadata. So this is kind of like is it linked with paper? Have other people used it? Is it versions? Are you using the most recent version or the most appropriate version? And maybe it's also a drive data so maybe they've.
Other data and what was what is information about that as well?
So I'll just quickly go through some additional recommended tools and resources and then we'll go into the questions. So Jasmine and I are associated with.
Environmental Information Data Centre, which is hosted at the UK Centre for Ecology

and Hydrology and this is specific remit, is freshwater and terrestrial data, and there's a lot of information about how to publish your dates, the how to prepare your data and also a template.
For supporting documentation, which can be very useful, research the resource if you're interested in actually publishing some of your data. So I would recommend having a look.
And then some external resources are linked here. So there's a Elixir Fair Cookbook which has like very small bite sized information on these fair aspects.
And then there's also the ARDC websites, which Jasmine has already referred to. And there's a 10 a fair 101 course, which is definitely recommendable if you want to go more in depth about fair and it's free.
And then finally, the Turing way, which is more broad than data management and it's also about the research process and open science and also very kind of more yeah, engaging way of kind of learning about these things.
So yeah, that's it's from us. But we'd love to hear if you have any questions or your thoughts.
I'll stop share. Oh, I can't stop sharing properly.

Samantha Rees 49:44
Thank you so much. I think that was really, really useful and really helpful 'cause I know that. Yeah.
It's really important that we have sort of Open Access data. That's been a question that's come up a lot this this week. And so sort of just making sure that we're also knowing the ways to provide it is, yeah, very important. I think so, yeah. Are there any questions at all?
Possibly not.

Jasmine Hunter 50:17
If there's no question, does anyone have kind of a a main take away that they'd like to share, something they've took from the session that they're going to kind of put forward into their research or things to think about?
And you can put in the chat if you don't want to speak out loud as well.

Els Dhiedt 51:03
Yeah.

Samantha Rees 51:03
There we go. Thank you. So, yeah, we've had a we'll refer back to the first slides when I publish my data, so that's good.

Jasmine Hunter 51:13
Yeah. And it's always something you have to kind of think about as you go. And then when you get to that point and you've hopefully got a bit of a step up and not too much work to do at the end of the project and it often gets quite busy towards the end end section.
Yeah, and no problem, Brooke.
Happy to share.

Subhajit Ghosh 51:44
I have a question, how different it is from ISO system because a lot of people also a lot of companies also follow the ISO system. Is it very different how you build metadata or how you store the repository?
Agree.

Jasmine Hunter 52:09
I think each obviously each repository will have its own kind of specific requirements for publishing, but I think broadly I'd imagine it's quite similar in that you know you always need to provide the context and the methods behind your data set.

Subhajit Ghosh 52:28
Yeah, basically. Generally people use ISO 8601. We also use here ISO 8601 so.
That's good. Thank you.

Jasmine Hunter 52:40
Yeah, no problem.
Yeah, and that's great T if you're publishing a paper including those fair guidelines in it incredibly important.

Samantha Rees 53:05
Excellent. Thank you both so much. Really appreciate it. I really appreciate your time.

We have got a panel Q&A next. So I did send some BIOS round for each of the people who will be there. So the sort of setup of the system is that set up.
Of the session. Sorry. Is that the Gemma? Jamie and Fran will join a link that you know the link that I shared for the for another meeting like this. And then we've got. And hopefully people have got some questions that they'll come with what I can do right now is just.
Post their BIOS in the chat in case for whatever reason people didn't manage to see the the bio that I sent around earlier in the week. And then yeah, if you could come with a couple of questions for each of them or whoever is particularly relevant, that would be really helpful. I suddenly have sort of thought.
Yeah, it might have been useful to have you send your questions prior to prior to this so that we had sort of some things to get started, but it's fine. I'm sure I can think of some questions if if people are feeling shy, but so here are their BIOS.
Oops, it's not copied it.
Oh, OK. It's being funny. I don't know why.
Oh.
It felt like I had a long time to do this and it's because it's a PDF format, so it's being really slow copy.
OK, just doesn't like it. Never mind. Um, I will get the Word document up then.
And.
Yeah.
So and I'll also put the link so the next one does start in at 2:00. So yeah, if you could be there for 2:00, that would be fab. And again that gives you sort of 5 minutes now to have a little bit of a screen break go to the loo quickly.
And yeah, see you.
In well, 4 minutes now, but see you in in 4 minutes and and if you if you. Yeah, it'd be great if you could come with a couple of questions for them and then and then we can get the conversation playing but yeah. Thank you.

Jasmine Hunter 55:37
Bye all.

Samantha Rees 55:37
See you later.

Els Dhiedt 55:38

Right.

AL

Ashling Laffey 55:38

Thanks very much.

HL

Hywel Lloyd 55:40

Thanks very much. That was great.

Samantha Rees 55:41

Thank you.

Samantha Rees stopped transcription