

Q: Okay. Yeah, in the first question block, I think you have already read the interview guide, right? #00:00:11-0#

R: Yes. I'm opening the (...). #00:00:22-3#

Q: There we are mainly interested in your (...) as data user, so how often have you reused datasets from your lab and other labs in the past? This would be the question. And I would also like to ask (...). #00:01:00-9#

R: For...for my lab (...) is not very common. So probably around 10%, yeah. Usually I...I took computer simulations, so I generally (...) artificial data and compared different models and see how it works, so we (...), we used this kind of data because we...we have the programs, then we can generate the data again. And there are cases that I may use...re-use existing data. #00:01:44-5#

Q: Okay. #00:01:46-0#

R: They...it's not my project, it is my collaborator's project. So that (...) meta-analysis, so they...then they may, yeah, archive existing data and re-use it to...to do more sophisticated analysis. So me personally, I think I occasionally analyze others' data that some (...) archives, some (...) primary, the other are secondary, for my own (...) re-used existing datas. #00:02:23-8#

Q: Okay. So you would say that you used or re-used primary data both for new purposes, so on the one hand for the development of these models that you use to generate new data, and on the other hand for meta-analysis? Or would you just say, you re-use it only for meta-analysis and not (...) computational models? #00:02:48-7#

R: Yeah, mainly for meta-analysis or for illustrations. For...for...I...I use quite a lot of secondary data from others but not my own. So (...) a sample, I may try to locate papers that publish meta-analysis that include the...the effect sizes and the details, and I can (...) illustrations. So (...) the focus is not to...try to see whether the original authors are correct or not. Usually the idea is to extend new methods by using some existing data just for illustrations. #00:03:35-3#

Q: Yeah. #00:03:37-5#

R: Yeah. #00:03:40-4#

Q: (...) Okay. Good. We can come to the second question. For which purposes have you used secondary data in the past? Particular (...) we (have already) talked about this, so you used it for meta-analysis mainly, right? #00:03:56-8#

R: Yes, mainly for meta-analysis. #00:04:01-0#

Q: Yeah. And, yeah, I would like to know what specific additional information you would need to optimize that work, so to optimize doing meta-analysis in the future. #00:04:17-9#

R: Yeah. Then the most important ones...is the effect sizes. So (...you?) are surprised, many published meta-analysis, they don't include the original effect sizes. They...they may put the study characteristics, like mean age of the participants. But for some reasons, whether it's good or not (?), the authors usually choose not to record the effect sizes as (...) or as online data. So difficult to verify or even try to...try the analysis. Yes. #00:04:54-5#

Q: Yeah. Is this getting better or or (have it) become better in the last ten years perhaps or...? So...so is it mainly a problem of...later years, so to say? Not providing the (...) #00:05:09-2#

R: Based on my observation, I would say it's improved slightly compared to ten years ago, yeah. Because many of the...I think, now there's many journals, some of them are aware of the issues, so they...and also from the authors' perspective, many of them are more open-minded. They use online (?) archives to...to...to store the data, so...so then in the review process, the reviewers can look at the data. So I also review meta...meta-analysis and I find some of them, they do include the...the data in online archives, so...and it...it helps me as a reviewer because I'm...I'm more confident that the findings are real, they are not just some...yeah, so...but I...I would say it's still less than 50%. So the majority (...) don't include anything. So when I review papers using meta-analysis, then usually one of my recommendations to reviewers is to ask them to include the data in electronic form, so that others can...or a list of reviewers can verify some of their arguments. #00:06:33-2#

Q: Mhm, okay. And related to this, (...) information on the main analysis (...) more easily, yeah, re-analyze the data and (...) effect size (...)?) Would that also be some metadata that you think (are useful)? #00:07:04-1#

R: Mm, besides the effect sizes, the...usually the...what kind of meta-data (...) I would say for...for many published many-analysis, they do include study's characteristics, that's why I find (...) surprising why they...they...they included the study's characteristics (...) but not the effect sizes. #00:07:30-8#

Q: Ah, okay. #00:07:32-3#

R: So...so that...if...asking for meta-data, I...I would prefer something more...instead of presenting them in two...in...in the format of PDF, if they can present it in machine-readable data, and it (...) #00:07:55-8#

Q: The more (...technical files?...)? #00:07:59-5#

R: Yeah, so (...) for most authors, they...in...in terms of the study's characteristics, I...I'm...personally, I'm quite satisfied with them because usually they need to include the study's characteristics in appendix because for some reasons, the editors and the reviewers, in some sense they may insist that they need to report the study's characteristics. But for some reasons, I'm...I don't know, then some of them may skip the effect sizes. #00:08:33-9#

Q: Yeah. #00:08:32-1#

R: Yeah. Yes. #00:08:36-8#

Q: Yeah, I just know in...in some journals, it's not...yeah...not generally, you do not generally have to provide...I...I just know from a colleague of mine, he has also published in a...in a journal where they only wanted him to provide the (p-values) he used and so he did not have to provide means or (or the sample sizes...). #00:09:07-3#

R: Mm. #00:09:08-6#

Q: May also be a problem from the journals. #00:09:10-9#

R: Yeah. (...) It's mainly (...) by the journal requirements. To my experience, when...when we asked the authors, all of them will follow. So next round of reviews, then they will submit at least appendix in PDF format about...on all the (...) and all... They may not use (...) the computer code, but they...at least, they will (include) the effect sizes and everything we ask. So it seems to me, if the journal editors don't ask for...for it, then they will skip it. #00:09:51-7#

Q: Yeah. #00:09:53-7#

R: So I think the editorial policy is quite crucial in (proving this standard?...). #00:09:57-8#

Q: Yeah, yeah, good. Okay. Then the next question would be: Are there other methods of secondary data use that you perhaps have not used in...in the past but you know about these methods? And what metadata would be useful for these methods? #00:10:21-0#

R: Mm (thinks). What kind of metadata you're referring to? #00:10:28-0#

Q: Yeah, metadata that would be useful for the respective methods, so...so, for instance, planning a systematic review, what metadata would you need to...to optimize this work? Or for a re-analysis of existing data, what would be metadata which would help you to...yeah, understand more quickly, for instance, the dataset? #00:11:01-3#

R: Then...I think one ideal scenario is the authors, they report their own effect sizes and...and the conditions and other study's characteristics, and in (...) machine-readable metadata. So I...so (...) reviewers of (...) systematic review instead of going through to the method-results section, then we can directly extract some basic study's characteristics, like mean age, gender, racial, number of participants, and the key effect sizes, so that I...so the authors have already calculated these and (they did this) in a standardized way. #00:11:54-2#

Q: Mhm (agrees). So sample characteristics on the one hand and effect sizes on the other. #00:12:01-5#

R: Yeah. So, in...in a standard format, so that when we extract the (...), then...unless we are asking for or looking for more detailed information, we can just directly read the metadata and then retrieve all these information. Yeah, so I...I think it also helps the authors because the effect sizes are calculated by the authors. So it's less likely that the meta-analysis will make mistakes. #00:12:35-1#

Q: Mhm. #00:12:33-4#

R: Because you need to check the mean, the standard deviation and calculate it by the reviewers, then the reviewers make mistakes. #00:12:42-5#

Q: Mhm (agrees). And would you also think that it is useful to provide the power of the (...study...)? So... #00:12:57-2#

R: Power of study? #00:12:58-5#

Q: Yeah. #00:12:59-6#

R: Yes, if it's the...if the focus is for primary research, (...then...) I think, yes. So to me, the metadata or the kind of data the...the authors have to consider is not only for primary researchers. So for primary researchers, we can have means, standard deviation, effect sizes, power analysis. And then for people who are going to do review or meta-analysis, then that would...they should also include all this, so... Yeah, so ideally, they...we should have some things standardized, so we can easily retrieve from the studies and...if (...) haven't gone through the details, then we can get some rough ideas what the (...data...) looked like and how the effect size is. #00:13:53-4#

Q: Yeah. #00:13:56-3#

R: So we can add to the list and think power analysis, sample size or some of the pre...some of the preregistration information, for example preregistration, they estimate a (...) and this kind of information can easily be added in...into some standard way. #00:14:24-2#

Q: Mhm, okay. Perhaps we come later to the question whether you think this is the task for the authors or (laughs) for some repository staff. #00:14:36-0#

R: Mm, okay. #00:14:39-3#

Q: Because first, I would like to know what kind of data you are going to be using for these different purposes, so in your case especially for the meta-analysis, whether you rely more on physiological data, behavioral data or other data. #00:15:00-8#

R: Mm, any...most likely behavioral or observational data...the meta-analysis (...) because usually I mainly use correlations and mean differences, yeah. #00:15:14-1#

Q: Mhm, okay. And what would you say, are there differences in the quality of their documentation or are they equally good documented? #00:15:32-2#

R: Mm, you mean for the data or for what? (...) Usually, for the data there I have...they're not well documented because usually when we do meta-analysis, then we are, as I have said, for most primary researchers, they don't need to write the papers for meta-analysis. So they (...) to write to the general readers, they are primary researchers in their own field. So we have...sometimes we have to second-guess what the data look like for some (...). They...they...they may not provide enough data, (...) they have a sample size, let's say, then this is the total sample size, they may not have the individual sample sizes for the control and the intervention (...) for...yeah, so...I would say, after documentation probably average (?), yeah. #00:16:48-6#

Q: Okay, good. And from...from which platform do you generally get your data? Do you get it from a platform, so from OSF or something like that or do you get your datasets from the original authors? #00:17:04-1#

R: Usually from the published articles. So in most cases, I think more than half, the data are attached as a PDF format. So either as a (...) or in the table form. Then probably less than 20%, then they...there are some archives, so they (...) in publishing the data as appendix, they may point to, they may link to a website that includes the machine-readable files and the syntax, yeah. #00:17:45-7#

Q: Yeah, okay. Good. Then I think we can switch perspectives. So to you as a data provider. And yeah, I would like to know what sorts of metadata do you generally provide about a dataset when you upload it? #00:18:08-7#

R: So for me, usually...now it's computer code. #00:18:13-1#

Q: Yeah. #00:18:14-5#

R: So I may use R in my analysis, so I think in the past few years, I...I started to post the R-code (, let's say,) to run the analysis and also to (...) simulation for my studies and then I post it in GitHub and then the code is available to others who like to check my analysis or try to use my...my methods. Yeah. #00:18:44-3#

Q: Mhm. And do you provide the code also in...in the archive when you upload your data? Or only in GitHub? #00:18:53-8#

R: It depends. For example, for some journals, they...they may expect us to provide an...an appendix of...to illustrate how to run the analysis, then I prepare a separate R-code for that particular examples. Other...otherwise I prefer to put it in GitHub because in GitHub then I can update it and revise because once you have published and then there are...then the code cannot be changed. Yeah, so... #00:19:28-2#

Q: Ah, okay. So...so versioning is easier in GitHub. #00:19:30-5#

R: Yeah, (...). #00:19:35-4#

Q: Yeah. So that would also be nice if...if research repositories would provide such a function for...for the users of the repositories? #00:19:44-9#

R: Yeah. #00:19:46-0#

Q: Yeah. #00:19:47-3#

R: Definitely. And I...I still prefer to just GitHub for another reason: Because I can (...) try everything in code formats, so (...) so I...I, yeah, so (...by articles...) by years, so everything is available, so the readers want to check my (R code), they can go to my GitHub website instead of going to individual articles. #00:20:16-3#

Q: Yeah. True. Okay. Nice. Good. Do you think that these metadata that you are providing are sufficient for re-using your data? (...) for everyone or just for your community? #00:20:41-3#

R: (...) I think it's good enough but also depends on the users because I...I occasionally (...) asking how to run the analysis. So clearly they...they have never checked the GitHub or my...my website. And usually I have to point them to...to the website and say that...that all the R codes are available there. So mainly it's...to me, it's the users who...or many users, especially in Psychology, they are not familiar with an archive like GitHub or other Open Access platforms. So when they have a question, they just drop an email to the authors. So...so...so they, many of them, they haven't done their homework before asking questions. #00:21:39-0#

Q: Yeah, okay. That's a problem. And do you also, yeah, annotate your analysis, so that you're writing lots of different steps when you are doing the analysis? So here I'm doing (...). #00:22:00-9#

R: Yes, I try to but I...I must say, I would document everything because I think it's...it varies based on how familiar the data analysis is to me. I would try to document the key, let's say function, what this function means. But I (expect that you...) have some basic ideas, let's say, they...they should know R, they should have read some of my papers if they are using my method. So...yeah, so...yeah, so I think it's (...a mental degree...). And I won't say I...I provide very detailed (...elaboration?...). I think one exception is for general articles. I think that's the one...one minor difference between submitting (...) R code to...as appendix in journal articles, they...then I tend to write more, to be a little bit more (...). Because I think that (...) they go through the appendix, so they...they may ask some questions. But I...it is just in...in my archive, in GitHub, then I can explain to the readers or to users if they ask me questions. Yeah. #00:23:26-5#

Q: Yeah. Okay, so in...in the latter case, you are, let's say, less motivated to do the documentation stuff? #00:23:36-2#

R: Yeah (laughs). Sort of. #00:23:38-8#

Q: Okay. Yeah, I understand that (laughs). Okay. Have you used certain metadata standards for annotating your data in the past? So something like DDI, for instance, or the BID-Standard?. #00:23:56-5#

R: Not exactly. #00:23:57-8#

Q: Not exactly. #00:24:00-5#

R: (...) I used just standard comments, remarks to...to remind or...to...to let the readers know what this function or this (...) means. #00:24:14-2#

Q: Yeah. But do you know about certain metadata standards? #00:24:21-9#

R: (thinks) I think... (sneezes) Excuse me. I'm not...not exactly because it looks like in my field, we...we don't need to involve lots of metadata yet. #00:24:37-0#

Q: Yeah, true. That's more...more libraries (laughs) at the moment. They know about these standards. Okay. #00:24:47-9#

R: Yeah. #00:24:46-0#

Q: Yeah, but we want to check this. #00:24:52-2#

R: Mm. #00:24:53-9#

Q: So let's come to the last question. If you were to create a metadata standard (for Psychology...), what do you think is the most important information that should be included in such a standard? So perhaps you can think about this question just in terms of the JARS from

the APA. Because this is a standard you are generally using and I think most psychologists (...). #00:25:34-5#

R: I think the most important part is standardization. It means everyone referring to the same data or let's say sample size should use the same field, then other than this one, I think it's up to the...any committee to add more or fewer fields in the standard. My...my...my concerns are, if the...the standard is quite vague, different researchers, they may write different numbers in different fields, then there are loss (of human involvement), you need to check it to see what...what (...) is n, means is it the sample size, the total sample size? So for metadata standard, I hope you'll be very crystal clear, so if this...this is the total sample size, then there's one particular field referred to it, and if there's a sample size for a control group, then there will be another field. So when we are talking about it in the future, in a few years, machines are programmed to read these metadata, we can easily extract everything without human intervention. So now the main problem for...at least for me, when I conduct a meta-analysis, is the loss of human involvement. I need to check each study individually to see whether this field, this data really refers to that particular data point. #00:27:10-6#

Q: Yeah. Okay. #00:27:14-4#

R: Yes, I...I don't (might ...). #00:27:22-5#

Q: Yeah. Yeah, it will be hard work to get to this point that we have got the formal standardization because psychology is a field that is so...so diverse, and, yeah, to standarize (will?) be (...), yeah, will be tough, I think (laughs). #00:27:41-6#

R: But maybe you can just standardize it for some cases...for some experiments. So for mental condition, there are some well-defined fields. (...) number participants in control group, intervention group (...), then the other observational data, the sample size, how many males, how many females. There...they are maybe complicated cases, like the...there may be multiple time points. But there (...) instantly can be easily standardize. So it already helps a lot, especially in checking (?). So when the numbers don't add up, then we know something goes wrong. Yeah. In sub-fields, these numbers can be used for validity checking. #00:28:33-8#

Q: Mhm, yeah (agreeing). Good point. And I just (...) that we come back to the question (...). Do you think that the annotation of metadata should be a job for the researcher or more of the repository staff? #00:28:56-8#

R: Mm, what do you mean by annotation of the metadata? #00:29:02-1#

Q: The annotation of the data with metadata. #00:29:04-3#

R: Of the data! #00:29:03-2#

Q: (laughs) Sorry. #00:29:05-6#

R: Oh! Okay. Then I think probably it should be done by the researchers. #00:29:10-5#

Q: Yeah. #00:29:11-7#

R: Because the researchers know the data, so...and again, to make this as...also part of the requirement of the journal articles. If it is required, then the authors will do it. #00:29:26-6#

Q: Yeah, they have to. (laughs) Okay. And would you say that it's a good idea to rely on the JARS (...for) developing a documentation standard for research data in Psychology? So relying on something that the researchers already use? #00:29:47-0#

R: Yes. It's helpful, let's say, just for one journal or one small society to do it should be something bigger, like APA or APS, for that kind of societies. Because they...they control...for example, APA controls much (... majority of) journals in Psychology, if they implement, then there are already a large enough people who will do it. And if it is done by a few journals, I...I don't see any impacts. #00:30:25-9#

Q: Yeah. Yeah, true. Okay, then I would like to thank you for...for doing the interview with me. And do you have any further questions or concerns or perhaps further ideas? #00:30:44-7#

R: Mm (thinks), I think probably that's all I have because I...honestly, I seldom do metadata (laughs). So probably, yeah, that's why I usually prefer to use some existing datasets like, let's

say, a small dataset published in the appendix in journal articles. Then that's already good enough for me as...to...to use it as an illustration. So I...I have fewer issues with how the...how the others metadata (...store, what other issues, problems...) To me, there are fewer issues.
#00:31:29-2#

Q: Yeah, okay. But when I get...get you right, it would be more easier for you to do your work if there would be some metadata...some more metadata, right? #00:31:41-4#

R: Yes. #00:31:39-8#

Q: Okay. #00:31:43-0#

R: Exactly, yeah. #00:31:43-6#

Q: So then I think that we can contribute to this problem and solve it perhaps in the future with our standard. And I can keep you on track if you want to. #00:31:58-5#

R: Sure. Looking forward to your findings. #00:31:59-6#

Q: (laughs) Yeah. Okay, thank you. #00:32:04-9#

R: Okay, thank you, see you. Bye. #00:32:06-9#

Q: Bye. #00:32:08-9#