Q: How often you reused datasets from your lab and other labs in the past? Could you quantify your specification by providing the relative frequency for reusing data compared to producing primary data?

R: Regularly during the last 2 years: 10 times re-using code to reproduce results during the review process; 3 times for educational purposes (e.g., recreating pictures); for simulations and papers on methodological topics (4 papers in the last two years)

Q: For which purposes have you used secondary data in the past? What specific additional information on that data (i.e., metadata) would you need to optimize that work?

R: Link to doi of the paper should be provided in the paper and vice versa (i.e., link to the doi of the data in the paper); Questionnaire type; Manipulation should be provided in a standardized form (for instance by making use of something like an ontology); Things you generally have, like demographics should be standardized.

[Hier startet die Audioaufnahme]

R: We can search the literature in terms of meta-analysis, for example, for certain key concepts. So if datasets are available and I want to know if they use for example the PANAS which...I don't know if your background is Psychology but... #00:00:18-1#

Q: Yes, I know it. #00:00:22-8#

R: So let's say we're interested...now, just this morning, *[person 1]* (...) talked about this, we're working on a project where we need the ... have a large amount of this type of data. #00:00:29-2#

Q: Mhm (agreeing). #00:00:29-2#

R: So if you could search based on questionnaire type, sort of questionnaires get some sort of index that's known. And there I know there are some databases that try to do this, standardized questionnaires, and then have specific, yeah, labels to reference, validated questionnaires like the PANAS. But that would be very useful. I think...and then the same

could be said for manipulations. Those are typically the things that you want in a meta-analysis. So if the manipulation also is standardized and it's also indexed and you can find it. That would be kind of useful. Other basic concepts like demographics, things that people commonly use which are in datasets, of course, naming might not always be consistent, cross-datasets. But that you know which kind of commonly used things are available. I mean, there will always be huge category things that nobody has done before or, you know, that are not used widely enough to make it into these standardized datasets. But I think that would be kind of useful. #00:01:41-1#

Q: Okay. Standardizing the manipulation, how would you imagine this? A catalogue? #00:01:49-4#

R: Yeah, exactly, you would need some sort of database where this exists. So for example you have...I don't know, there's this public speaking there…it has a name of a researcher, I don't know, but there's an example where we induce stress by telling people that they will have to give a public speech, right, that's common in Psychology. Or for example social exclusion. There's a ball-tossing game invented by somebody, Kip Williams invented this. I know many people use those manipulations for social exclusion or anxiety, very common. So meta-analysis would likely also focus on these kinds of manipulations. Or maybe more interesting, if I want to do an a-priori power analysis, then, and I use these manipulations, I want to know what kind of effect size they have maybe in specific populations, right, so then a search could be for this manipulation, and I am performing this in the U.S. or in Europe, so I have demographics, I have a certain age category, and I could just go in and find these data points, and then I would know what effect size I can expect, which is useful for a power analysis for example. That's one of the main issues now where people need data but it's difficult to find. #00:03:05-7#

Q: Mhm (agreeing), true. So would it also be interesting for you to have the effect size by itself as a metadata? #00:03:14-8#

R: Sure, yeah. I mean, that's...but I don't see that as an issue of the primary data. I like findability, so...again, in a perfect world, what you would have is, I could go in and I have the dependent variable, let's say, self-esteem. Rosenberg's self-esteem scale has an indicator, social exclusion has an indicator, and I could go in the dataset and just compute this because I

know which column indicates which condition people were in, and I have the main dependent variable, which is also standardized. Now that's probably far, far removed. For now, I'm not sure if this should be findable or...I think I believe a bit more for effect size calculations, right, they're error-prone, people don't always know what they're doing, so...I would like some curation of this, so I like platforms like "curate science" where somebody goes in and manually takes the effort to curate this information, like effect sizes. And it would be nice... Sorry? #00:04:23-9#

Q: Someone who knows more about statistics or who... #00:04:30-3#

R: Just not the original, just yeah, just not the original author most of the time. Most researchers don't have statistical skills that are good enough to...I would say, maybe indenpendently report statistics, so yeah, so there are too many mistakes. #00:04:44-7#

Q: Okay. #00:04:45-7#

R: And there are also some variations in how you can calculate effect sizes, so I think some curation is important. And then if you have a platform that takes information, ideally it would be reproducible maybe, right, so this platform would be a layer on top of the lower level of the raw data. And then this platform says, well, this is basically called re-run these original datasets, and if you have a problem with the code you run or you want a different variation, you can change the effect size calculation. If you don't want this, you can have a different variation. But this level of code running on the original raw data, that is a bit curated, so people looked at whether this actually makes sense to do for individual studies. Yeah. I think it's important. And also in terms of reproducibility, you don't want this number, like, where would this number come from, right, it has to come from the raw data. So, so I would prefer something where the raw data is like the baseline of everything, and then we have different code extractings, and that can put it in a database, but I don't want to independently have a number somewhere that says, oh, this number, which is the basic level. Exactly, exactly. #00:06:01-4#

Q: Yeah, of course. Okay, so it's more about code activity (?). So that you can reproduce the aggregated data file, for instance from the raw data. #00:06:13-4#

R: Yeah, I think that makes more sense. #00:06:16-9#

Q: Okay. #00:06:19-0#

R: And it also...because it also allows people to adapt the code, right. If I'm interested in exactly these effect sizes but only for men, then it's just a tweak in code. And otherwise, if we have raw numbers, you don't do much with it, you have the raw number. #00:06:34-2#

Q: True, yeah. Okay. Good. And I'm not sure whether there are still some methods which we were not talking about yet. But are there some other methods of secondary data use that you know but have not used on your own and would require other metadata than the ones we were already talking about? #00:07:05-4#

R: Other methods of re-using data? #00:07:08-5#

Q: Yeah. #00:07:14-7#

R: Mmm (thinks). Other methods of re-using data... I mean, yeah, we didn't talk for example about things like citations or, you know, things that are related, I don't know if you mean those kind of information, like that data as well. Very often, you are going to link things together, so some projects, we do this, we get DOIs and we link it to additional information. We have a project that's ongoing where we download all open reviews that were associated with articles. Right now, they are not easily linked for most cases. PRJ does it pretty well in the sense that it's just...after the DOI, there's like review or something...so it's really (...) not after the DOI, I think. But everything has an article number. Yeah, so it's not perfect. Both systems, I think, are not perfect. Both the Royal Society has a lot of open reviews and PRJ. Neither of these seem to be directly findable through the DOI. But if you know the article number...like, PRJ just starts with 1 and counts up, anyway...but if there are open reviews available and you can extract them in text formats, that would also be useful. Yeah. So I think ideally that should just be linked under DOI if they are openly accessible, so... That in essence, if I have a paper, and I can just, let's say, in Zotero, I can write a plug-in, and just, say, show me the review of this article that I have, it would just pop up. #00:08:56-7#

Q: Okay. So you would imagine something like chunks in a perfect world. Like chunks of reviews plus article plus data plus code. #00:09:07-7#

R: Yes, I think ideally what you have is...because now data is on the OSF, e.g., right, I mean, I know you have your database, there are many others, but let's say, let's say there's the OSF. Then there's no direct link. We have the article, the published article, and then you have to look in the article to find this location of the data, which is not ideal. So in an ideal world, what you have is, you have a DOI/data or /materials or however we name things eventually, but let's say there is /data. And this will bring me to a folder that has a predefined structure. #00:09:45-8#

Q: Mhm (agreeing). #00:09:46-3#

R: There's a raw data folder, there's a processed data folder, there's a code folder, whatever, so that from wherever I am, I can access the raw data, so again, if I would be in Zotero and I wanted to see the data, I know that it has to be under /data/raw data. And if I just want to grab the, you know, whatever it is, take a look at it, that should be a click away because eventually what you want is to be able to write software that takes the DOI, it knows, it goes to /codebook. It figures out that the Rosenberg self-esteem scale is in there and that's what you want to know. It checked this and then it grabs the raw data from /data and then I can extract the relevant information without having to go in and doing any manual work. I mean, ideally, right, we're far removed from this but since you are shaping the future of our science in this respect. Yeah, it's a good take (sort in mind...) see where we want to end up. #00:10:56-5#

Q: Yeah. But this idea sounds to me like an extension or a more general form of the BIDS standard which is already existent. #00:11:03-8#

R: Yeah. Excactly. Yeah, yeah, yeah. I think that is a very good standard. #00:11:09-2#

Q: Mhm (agreeing). #00:11:06-3#

R: It's of course narrow in the sense that they have scans of people's brains, so, you know, everything is the same. And in Psychology, if you go beyond this, it becomes much more variable. But that is a perfect example, yeah. #00:11:25-6#

Q: Mhm (agreeing), okay. Good. Then I think there is only one question left for the data user perspective. Regarding the kind of data you are generally using for the different purposes, perhaps you can shortly elaborate on this, whether you use more physiological data, behavioral data or text-video material... #00:11:51-3#

R: Mm (agreeing). In my case, most of the data ends up as a CSV file, like simple questionnaires, reaction time data. Most of the meta-scientific research we do, yeah, we would analyze citations or we code things from raw text, so sometimes the underlying data is just analysis of papers e.g., right, that we look at. But no, yeah, no huge data sets or...everything is typically kind of small. I mean, unless we...yeah, I mean, we can have, like, couple of gigabytes of papers, like PDF files of papers that we extract and text-mine. That could happen. Yeah. And sometimes the data we analyze is simulation data, right. We simulate data, e.g. So this can be quite substantial in terms of size because if you simulate across many conditions and scenarios, you can end up with a couple of gigabytes based on data, text files, but these can typically also be recreated, right, so you can run the simulation for a couple of weeks, and you can recreate all the gigabytes based on text data if you want to. But very often we try to share, yeah, so people don't have to do this simulation step and wait for... #00:13:15-8#

Q: Good. And do you perceive any differences in the documentation quality of these different data types? So are for instance behavioral data better documented than (...) data? #00:13:31-0#

R: Well, so again, we just did these projects where we computation-reproduced registered reports. I think the preprint is already online. And it was kind of difficult to do for most...for a lot of studies. So I would say, yeah, at least for my field, I mean, if you look at empirical papers, there the quality of data documentation is very low. I would say it's very difficult to figure out what's what. #00:13:58-4#

Q: Mhm, okay. #00:13:59-9#

R: Codebooks are very often...I mean, I could look up the numbers or you can look up the numbers again, the preprint is somewhere on the (...) archive. But, yeah, codebooks are very often missing, so figuring out what's what in a dataset takes a lot of time, if you really want to

reproduce results. Sometimes it's very smooth and people, it depends on what kind of code people use and how else document it. Sometimes from the code you can figure out what's what. You look at the original paper, you look at the code, you're like, well, this code reproduces this, so then these variables must be the variables you're analyzing, you can get it from the paper. But without the paper, it's very often (...), no, no, it's quite difficult. So, yeah. #00:14:45-5#

Q: And where do you think are the reasons for this, yeah, bad documentation? #00:14:51-0#

R: We don't force people to do it and then they don't do it because it takes time. Yeah, I think that's the main reason. It's just the individual time for collective benefit problem. If we would force people to do it because if they want to get published, there has to be a good codebook and their data needs to be reproducible. I think it's a minimum standard but we don't enforce it, so... #00:15:14-6#

Q: Yeah. So we should cooperate, of course, with the journals. #00:15:16-3#

R: If you want people to do it...no, no, exactly, publication is a very good point where you force people to deliver things in a certain...that has a certain standard. Yeah, and I think, really, this is something you have to force and maybe, you know, this is not a carrot kind of approach. It has to be just a rule. A rule with some fines, maybe, if people don't do it. No, honestly, I mean, it's just a super boring job, let's be honest, I mean, you know. It's not the stuff that we entered science for. It is important to do well, I think we can, you know, probably develop tools that make it a lot easier for people to do these things. But beyond that, it requires some work and it's a little bit boring. #00:16:10-0#

Q: Mhm (agreeing). True. Okay. Then we come to your perspective on secondary data use as a data provider. What sorts of metadata do you generally provide about the data when you upload? #00:16:25-7#

R: Yeah, I mean, this is a continuously sort of improving cycle as well. So it's difficult to say what my current standard is. Like, for this paper where we had the computational reproducibility. After we've done this, which is very recent, we submitted it in the summer, then e.g. we used Ruben Arslan's codebook package to create a codebook that has all the

required metadata of our variables. But that package didn't exist earlier. And before that, yeah, maybe we had an Excel file that said these variable names, that's what we typically have, like, a spreadsheet e.g. that basically is the codebook, I mean, the raw data would eventually be a CSV file somewhere but probably we worked in an Excel spreadsheet and then save it as a CSV file, and then the second tab of the CSV file would be, well, these are all the variables on the first tab, and here's a typed explanation of what everything means. Very often just an explanation of what it means, like, not even a range of possible values or any other thing. So we would, we would have a very pretty minimum standard, minimum level of saying what every variable is. If you're lucky. #00:17:41-3#

Q: Okay (laughs). (...) labeling (...) of all the variables. #00:17:46-3#

R: Yeah, yeah, I think so. And then there would be, like, a READ-ME document, a READ-ME file in the folder somewhere that has some written explanation how to reproduce the results in the paper. #00:18:04-6#

Q: Okay, so where you explain how to use the code or...? #00:18:07-6#

R: Yeah, exactly, so run, you know, this file, this is the raw data, then we run this file to create processed data. And then you can run this file to get the results that are reported in the manuscript, yeah. #00:18:25-2#

Q: Okay, so that they can reproduce the whole process, so to say. #00:18:28-7#

R: Exactly. Yeah, exactly. And sometimes now, it depends on how easy these steps are. But sometimes we just put all of this in our markdown file, so we write a reproducible manuscript. And then there's not so much information of how everything is done because, well, yeah, you can see it in our markdown file which is annotated but otherwise, yeah... Our markdown file just reproduces everything, so. But then in terms of metadata on the underlying data sources, there isn't a lot. There isn't a lot of what we really use. Yeah. #00:19:00-8#

Q: Okay. And do you think that these provided metadata are sufficient for re-using your data? #00:19:08-3#

R: Not in the perfect future version, I mean, we don't code anything...but there are no standard names of things anyway and we've had meetings on this. There's this PsychDS project, I don't know if you know this, it's very strongly based on the BIDS work trying to think of variable names. I have a PhD student who volunteers some of his time to work on this. But it doesn't exist, so, I mean, there are no standards, so what are we supposed to do? I mean, yeah, we describe things a little bit and if you would go in and you would spend some time and effort, then I think you should be able to figure out a lot. But probably, you know, probably from the last year, it's been easier. If you go back two years, I didn't know a lot of this, so it will be messier. If you go back five years, I mean, there might not even be data shared. I mean, you know. So there's a huge change. #00:20:05-4#

Q: Yeah. I think it's also difficult to develop such an ontology, let's say, for Psychology because there are so diverse concepts. #00:20:20-0#

R: I agree. But even then...so in the worst case, everything has its own number, right. So in the absolute worst case, every variable in your dataset has like a 12 string of letters and numbers and a unique code, it's just like, the question you asked in September 2015, and that is then the label. But at least it has a variable description that I can approach and I can retrieve rather automatically, like a codebook that I can just retrieve and then the codebook has unique items. So I can't search but, you know, then I could retrieve all items and in the description I could say, well, let's search for self-esteem in 12 different ways. #00:21:04-7#

Q: Yeah, okay. #00:21:02-4#

R: I mean, at least it's better than nothing. #00:21:07-8#

Q: That's true, yeah. #00:21:05-3#

R: So, yeah. So I agree, the ontology is difficult. We will always end up with extremely peculiar items where you say how much do you like the temperature in *[city 1]*, whatever, when the window is open? I don't know, like, a question where you're like, nobody is ever going to ask the same question. We will always have those items. So they will always need unique identifiers. #00:21:35-3#

Q: Mhm (agrees). Okay. #00:21:37-6#

R: So that will never be...like, and that is not the case for the BIDS project, right? It always focuses on the brain. It's always, you use this machine or this machine, it's so extremely narrow in terms of the data that you can describe it. But psychologists do so much stuff. So you will always have some rest category which is huge, probably like, I don't know, maybe 70, 80% of the questions we ask are unique, unique questions that will never be asked again. I mean, maybe it's a bit high but something. Anyway, so the ontology is different but also the structure, I mean, I like the codebook package because it does a lot of stuff that I didn't have to think about. And I'm not really sure what kind of, you know, maybe I sort of understand now what a JSON file is but I didn't two years ago. I want to know how to create it, so it's lovely that I have my codebook in a format the codebook package will understand. And it provides all this lovely information. I mean, so that's a super useful tool. And those kind of tools I think are the way to solve this problem. #00:22:39-0#

Q: Mhm (agrees). Yeah. So we have to invent a tool. (laughs) Well, we already have one but... #00:22:48-1#

R: Yeah, yeah, no, but I mean tools that are usable and that save people time. Like, a codebook thing that just, you know, I run it, it just does a lot of stuff for me. I mean, of course, I also need to still have my Excel, but anyway, it creates something that actually is much more reusable than it otherwise would be. #00:23:05-9#

Q: Yeah. Okay. So then: Have you used certain metadata standards for annotating your data in the past? So, something like DDI or Dublin Core? #00:23:21-1#

R: No. No, not yet. And I know, I'm on a grant, I have a grant where I have a data management plan. and I should do this but I'm already failing, to be honest. I remember that originally in the data management plan, I looked up what these things are, so they're not a complete surprise that they exist. I think if I would run down the corridor and ask colleagues about it, most of them wouldn't know what these, you know, what this means. At least I read about it, I know that it exists, but to be honest, no, no, I'm not using it. I wouldn't know, I wouldn't know how...where to get started, basically. No training in this, and, yeah. #00:23:57-7#

Q: Yeah, I think they are not relevant for us at the moment, right, because these standards which exist are mainly from the librarians, yeah, or from the social sciences. DDI is from the social sciences. Which can be applied to psychological research but of course it is not very suited for every dataset. #00:24:22-9#

R: Yeah. Yeah, and also the question is: Where would I...what would I do with it, right? Where would I upload this metadata file then? So that other people actually can use it? I mean, if I, if I make it and then I put it, yeah, I don't know, just also on the OSF next to the data, it has no use. So it feels like something that's good but we need to do the groundwork to make all these efforts actually useful. Right, you can do a lot of stuff, you can spend a lot of time on a lot of things, so everything is a cost-benefit analysis, and here, yeah, the cost, the time, the benefit, I don't even, I don't even see any benefit in doing it now, to be completely honest. Except checking a box, like, okay, data management plan, my file's now in some standard that nobody uses anyway. I mean, that's of course...but, but we need some infrastructure where we then can upload it. I would love to, like, I don't really particularly like the OSF, to be honest, in terms of a solution. But it is the best... #00:25:20-2#

Q: It's bad documentation. (laughs) #00:25:18-2#

R: Yeah, and it's horrible, I mean, you know, you know we will have to restart completely from scratch in the future somewhere. That's why I like things like the PsychDS project and those kind of, you know, innovations. We know we're uploading all this stuff and it's all useless. #00:25:39-8#

Q: Yeah. #00:25:41-5#

R: Well, I mean, pretty much useless. I mean, it is there, it's better than nothing. If you put in a lot of effort, you can go in and try to figure it out. But it's really like the, you know...and most likely I will have to go back in ten years, I'll have to go back, download everything from the OSF and reformat everything into a way that actually it makes sense. #00:25:59-2#

Q: This is hard (work) #00:26:04-5#

R: But...but now I don't know what to reformat it in and how to do it and where to upload it. We don't have the systems. So there's also no other way. So it's better my data is not gone, it will stay, you know, be found, and... #00:26:19-6#

Q: Yeah, true. Okay, so in the end I would like to talk about the gold standard, so to say. Which additional information on your data would be most important for you to optimally re-use those data? So perhaps you can just think about this question in terms of the sections included in the JARS because I think this is a standard we already use and which we know and perhaps... #00:26:54-0#

R: What are the sections in JARS? I don't know them by... #00:26:57-6#

Q: Yeah, Hypothesis and Method, Results, so the typical sections in a paper. #00:27:03-5#

R: What do you mean, JARS? I don't, I'm not even sure if I know... #00:27:08-6#

Q: (...) Journal Article Reporting Standards from the APA. #00:27:10-7#

R: Ah, okay. Mm (thinks), we have a..., so we have a project that is very related to...I'm just looking it up now. Mmm... So we had this kind of crazy ambitious idea to create, like, a grammar of science. *[Person 2]*, (...), *[person 3]* and I worked on this, we called it *[project 1]*, *[deleted for data protection reasons]*. And I'm just taking a look at what we, yeah...alright, so we thought about the gold standard, right, I mean the gold standard...but what we would put in there. So there's some information about the study, like a name of the study, then there's some information about the project, so this could include things like the authors who were involved or even under the authors you could subdivide it by the author roles that people had. #00:28:20-5#

Q: Mhm. (...) #00:28:26-2#

R: The rights? #00:28:25-3#

Q: Yeah. So which rights the different persons in the project have? #00:28:31-4#

R: Yeah. Yeah, so, I mean, so, what we were thinking of...there are a lot of these projects where people are working out, like credit e.g. or what is it called, like, author credit system or...do you know? So there are a lot of these projects but then there is no way to combine all these projects, right, so the goal of this *[project 1]* thing is, you have a top level structure, and then credit would go on the authors/credit, basically. So all this information people generate would just be one of the parts that go in there. Anyway, I don't know what all the things are that would go in there. But some description of this study that we deem relevant. Funding, conflict of interest information. That kind of stuff. Stuff that we report somewhere, right, in reporting guidelines. *[deleted for data protection reasons]* So all these aspects are basically made quantifiable and machine-readable. So that's the part about...huh? #00:30:49-7#

Q: It's a nice idea. #00:30:51-6#

R: I think so, yeah. We're writing up a small paper about this. And again, you can just see there's one of the very tiny parts, like, many papers wouldn't even have a hypothesis section, like, it would be empty. It's not relevant for what they do. But I mean, and you could think of the data section. Where did your data come from? Who's responsible for, yeah, so where can you find the data? If it's a public dataset, ehm (thinks). Yeah, alright, anyway (laughs). *[Deleted for data protection reasons]* which is like a sub-section of the questions you're asking, right? So it's a tiny part of a tiny part, but nevertheless, that in itself is a project, right, and something we can think a lot about. And so there would be many of these tiny parts. Then there's an analysis file, which probably would have a pre-processing and a main analysis script. #00:31:51-2#

Q: Mhm. Also for the additional analysis? So sometimes, you have exploratory analysis. #00:31:59-3#

R: Yeah. Well, basically, we would list all analysis. Some are related to a hypothesis, so they are clearly identified: This is an analysis that answers hypothesis 1, which is labeled in the hypothesis section. So it refers to the hypothesis section. And some are not linked to a hypothesis, and then automatically that would make them exploratory analysis. #00:32:22-1#

Q: Mhm (agrees). #00:32:22-4#

R: But they would be a way that you can say: What are all the tests down in this paper? Which were exploratory, which were related to a hypothesis, right? So it's just a label within the system. We would like to have an archive file of the raw data that you can link to. And then, yeah, I think that's...yeah, and the results would basically be generated by combining the data in the analysis scripts. So that's...that's what we have worked on for this *[project 1]* thing. Yeah, what else would be in the metadata? I mean, there would be things like the reference list, you know, that you can retrieve all the references easily. I don't know. Like, all the sub-sections basically that we have in a paper. #00:33:12-4#

Q: Yeah, it's a question whether it make sense to include all sections we have in a paper, so we were asking ourselves whether this makes sense, for instance, to include the discussion sections. I think this makes no sense because most (...) rewrite things in the discussion that the reviewers want us to write, so. #00:33:32-7#

R: Well, in...yeah, but in...yeah, but in our view...the entire manuscript, like, all the text in the manuscript would also be...so there would just be, like, paper. And it would just have the machine-readable version of the entire paper. Like, you know, like...many publications have this. And then I need to go to a website, you know? And I can click the whatever, XML link or whatever, all these things. But the...so all of this, like, if you have one data structure, and I have, let's say again, the DOI, then it would be DOI/papers/XML. And everything is just organized in a way that you can (...) yeah, so I would say the entire paper is part of the final archive version as well. And then if I want to search, like, we had a project recently where we searched for...test the hypothesis, right? It would have been lovely if we could do this on the entire literature. Instead, we have to download every PDF file and then search through every PDF file. I mean, you know. Just kind of weird way to do things. Makes sense, I mean. It's already good we don't have to go to the library but...yeah. #00:34:41-5#

Q: Yeah but it's very time-consuming to do these things. #00:34:45-4#

R: Yeah. So if there's better structure, so there (...). And then I think you don't have to discuss: Should we have a sub-section for the discussion? Maybe. Maybe. I mean, there's always a level of granularity where you're like, argh, yeah, you know. If people want it, eventually they'll have to go in and code stuff. But, but on the other hand, like, it really depends. If the software that the publisher uses...they link something as discussion and it automatically tags it

as "this is the section starting the discussion", for whatever reason, if the software does this, great! Right? I mean, if I can go in and I can say, material/survey/Qualtrics, and Qualtrics provides a lot of metadata, then I could just say, well, I know that Qualtrics has, whatever kind of stuff that it stores with it. So any software package that creates metadata in itself, yeah, put it all in. You never know, like, if it's free, I would say, the more, the better. You never know if somebody needs it or not. Yeah. Okay. So I don't know, there's probably, like, a lot more. But I do think that this JARS kind of sub-division, that does make sense, for many empirical papers in Psychology, it makes sense. #00:36:08-6#

Q: Yeah, and I think it is a standard which people already know, so it is not that time-consuming to learn the new standard then. #00:36:16-6#

R: Yeah. And things like, you know, having hypothesis results, methods results, I mean, there's things in which most papers break down. Not all, but many of them break down in these sub-sections anyway. And in the methods part, you will have materials. And in the results part, you will have raw data. So it is a logical subdivision. #00:36:38-0#

Q: Yeah. Okay, then thank you for your time and for these many information, it was really interesting, thank you. #00:36:46-2#

R: Alright. #00:36:47-7#

Q: And, yeah, perhaps we see (...) #00:36:51-0#

R: Yeah, exactly. Exactly. Yeah. Good luck! I mean, this is important stuff to work on. I think it's...it will take a while to develop a standard but I would say if enough people give it a try, then we will get somewhere. And I think this would also be a nice topic, I mean, for your institute, to think more about...like, now you are working on things and you do interviews, that's great. But as you start to develop prototypes or things...I mean, a lot of people are working on these kind of issues. We have the Society for the Improvement of Psychological Science, and there is like, yeah, many subgroups in the Society that have like, this PsychDS project is one. There was stuff, like, a meeting of people who create software, like codebook and other types of software like this. There was an ontologies-related subgroup. You know, these hackathons. So there are a lot of people interested in these kind of things in Psychology.

And basically, what we agreed upon, well, it's good that we all know that we are all interested in these things, and people are trying out different things. I think one of the nice things of this *[project 1]* idea, like, a grammar of science, is that whatever people are doing, the individual projects should go into some other place. That makes all these individual things much more valuable. Right, so if I know, well, I didn't do everything but I did make a very nice list of the authors and what contributions every author had, okay, well, then you have this thing and you can put it somewhere. And if you have a metadata standard, it would go under study/study metadata, right? So what we discussed about earlier with the DDI or whatever these kind of standards are. If I make it, where does it go? So thinking of this overarching structure, and then say well, whatever you have, you can put it here. That is useful. #00:38:55-5#

Q: Yeah, this is useful and you also need a structure where you can translate these metadata files so that you can read it, really. Because many researchers do not understand an XML file or something like that. #00:39:07-9#

R: No, no, of course, no, no. No, of course. I don't understand. Yeah, yeah. No, of course, one of the nice things of this *[project 1]* we were working on, and honestly, this is just, you know, a couple of, I don't know, a couple of weeks at most, so it's like...a prototype. *[deleted for data protection reasons]* So, these use cases, anyway, so there's a lot of people thinking about all these different aspects and these kind of nice use cases and...I think it's an interesting community that doesn't have a home. But you would be a logical...yeah, place to...yeah, yeah. So in the future, I mean I know you have many conferences on sort of, yeah, all sorts of related topics but...this, yeah, maybe you discuss this, I'm not completely sure, I don't know everything you do but...but I don't think this community has been brought together in a way that is like a couple of days. Like, at the *[meeting 1]*, we had one meeting where everybody was together for, like, two hours or something. Yeah, we should definitely keep in touch and think about what we're doing. People know each other on Twitter but...yeah, there's room for community-building as well. #00:43:00-2#

Q: Mhm (agrees). Yeah, I will talk about this with *[person 1]*, for instance. #00:43:03-8#

R: Yeah. Yeah. Alright, good. #00:43:08-3#

Q: We can do something. #00:43:10-3#

R: Yeah. I mean, it will be a long process but...we can work on it slowly but surely. Then something will happen. Alright. Nice to talk... Sure. See you later, bye. #00:43:19-2#

Q: Yeah, bye. ##00:43:23-##