

# Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from Germany

## Interview: os\_004 – Translation

1	<b>Interviewer:</b> Alright, please introduce yourself at the beginning. Where you come from, so which institution and which function you have here and what you are researching.
2	<b>Researcher:</b> Okay, so I am [Name] I'm in my fourth year of my PhD, working here at the institute in [Institute] in the neuroscience group. We are working on microglia. Yes and I was also working on microglia during my PhD, which are the immune cells in the brain.
3	<b>I:</b> Sounds very exciting. And how long have you been working altogether in science?
4	<b>R:</b> Well since I finished my master's degree, it's been now around four and a half years.
5	<b>I:</b> And with what research data do you work?
6	<b>R:</b> Where do we get our data from or with human or mouse or what...?
7	<b>I:</b> What kind of research data is it, so exactly what you mentioned, but also maybe what kind research data they are. This now... now I don't know that much about this area of course...
8	<b>R:</b> Okay, so also what kind of methods we use or such...
9	<b>I:</b> That for example also.
10	<b>R:</b> Well, first of all we are a mouse group, which means we do animal experiments and we refer to... most of the data comes simply from animal experiments. In neuroscience this is unfortunately difficult to look at many processes in a different way, that is in the Petri-shell. We also have primary cell cultures, that's where the data also comes from, but a lot of the data really comes from in vivo experiments. The data, what do they look like? Yes we have... we do a lot of microscopy, that means we use confocal microscopy or fluorescence microscopy to analyze our research questions or experiments, but we also have many//other methods that we use, for example PCR or QPCR or where we are basically looking at the expression of certain genes on the RNA level. We look at protein expression at the DNA level and finally we quantify all of this using software, where we end up with really pure data, pure numerical values, which we then evaluate using Excel and various statistical programs.
11	<b>I:</b> And are there any other relevant research data in your field that you may not be working on right now, but which are generally popular in the field or are being used?



12	<b>R:</b> Yes, so what we of course also use is, besides RNA and proteins, there are also just like kits where you also look at protein quantification, but in a different way. Not in the cells, but what the cells secrete practically, what then practically is in the supernatant from the cells, what is being secreted. For example the messenger substances, that we are then looking at, which are important for the communication of immune cells and what is also in the natural and medical sciences big and important, are the like bigger approaches, where big data is generated such as RNA sequencing, single-cell sequencing or proteomics, where we try to look at the entire proteome of the cell types using mass spectrometry. These large data sets cannot be analysed by us ourselves. This is always in collaboration with bioinformaticians or groups specialising in these techniques.
13	<b>I:</b> I find this very exciting in any case. Have you ever published your research data?
14	<b>R:</b> Hm. Yes.
15	<b>I:</b> Do you want to describe how it happened or...?
16	<b>R:</b> Yes. Yes, so... we just pick out a corresponding journal, according to what it... which readership we want to reach, whether it's more technical or really the scientific community or the neuroimmunological community and yes, we were also rejected several times, which is relatively normal, because you always start relatively high up and maybe you overestimate yourself a bit and then go down and become a bit more realistic. And then I actually was relatively lucky. This review process went relatively quickly for me with these two publications that I now have. They were just a few weeks and the revisions were also relatively small. So mostly they demand even more experiments, which can last for months, but that was actually quite ok and the reviewers were relatively nice. Yes. So it was quite good experiences.
17	<b>I:</b> So you have now published a paper and did you have to publish the data with it, so also hand it in?
18	<b>R:</b> Hm. With... yes I also had to, exactly. So I published both papers Open Access. One is a pure Open Access journal. I think, I'm quite sure about the raw data, whether they were for this first publication really public or gotten published. For the second paper, you had certainly had to publish and release the raw data in any case, so that everything is comprehensible.
19	<b>I:</b> And did you have any concerns, that is about the publication of raw data, of the research data?
20	<b>R:</b> No. Nope, I mean, we do science to move forward and not... Nope, I don't see a problem there.
21	<b>I:</b> Concerning your research data, is it not really person-related or sensitive data or do you also work with such data as well?
22	<b>R:</b> No, in my case there were no human data, so no patient data or other data.



23	<b>I:</b> Do you think that in the future you might work with such data, so is that a possibility?
24	<b>R:</b> Yes, so now for my Postdoc I will go to// to a group that now only works with human samples. After decease, however. How exactly// because there is no other way to get to the brain cell type, so only after dying you can get to it at all. How exactly// That will be also in the USA. That means I don't know exactly how the data is handled there and how// I think the patients just get a number. That will all be encrypted too. Erm, yes.
25	<b>I:</b> That would be my next question. Whether you know how to handle such data... to make them for example also available for other scientists, because those are people and sensitive data.
26	<b>R:</b> Well, I didn't get any particular introduction for that or anything, because I never worked with human data, but I guess// I mean we also get from time to time samples from glioblastoma patients, so high degree brain tumors. We have a collaboration with [another city], then we get the samples and then the tumors just get an index, so they're practically encoded directly. We don't even know, when they'll arrive, what the patient's name is or whatever else. There is then just a number of the tumor and then the data about it, age, sex and so on... Surgery date and so on, but names or anything else won't reach us at all.
27	<b>I:</b> And would you be allowed to publish that, that without a name, the one which is anonymised, which you have already got, would you be allowed to publish that?
28	<b>R:</b> Yes, that will then be published.
29	<b>I:</b> And to whom do the collected data or the data you work with belong? Do they belong to you or...?
30	<b>R:</b> Hm. I think this is different from institution to institution. As far as I know the data we collect at the institute belong to// so there is no way they belong to me. If possibly at all, they belong to my group leader or rather the sponsor. And I believe that here you are signing it in the employment contract as far as I know, that it belongs to the institute, the data here.
31	<b>I:</b> And do you know if there is a research data policy here, which then describes, how to handle the research data, how to publish it and where, etc. etc.?
32	<b>R:</b> So in a very concrete way. I do not know this exactly now. I have not seen it yet, but I am sure, that there is one here.
33	<b>I:</b> Do you feel that the process of publishing research data is too complicated or too opaque?
34	<b>R:</b> Yes, certainly it is very complicated and especially the time factor is a problem, I think. So it takes a long time until finally something is really out. The revision is not transparent... but that is being now discussed with the whole Open Access and Open Science Movement and I am in any case in favour, of the reviewers, that they sign by name. That one knows who is looking at the



	<p>research, because one can also propose some. I have also heard of people who practically sign their name under it, even though it is not required. And this fear that the// that your scientific adversaries want to screw you or something, I think that's just a fairy tale. I don't think there's anything behind it. So, yes. This can be certainly still simplified. Yes.</p>
35	<p><b>I:</b> And in terms of the research data: When you submitted for example your paper and the raw data for it, how did the process work? So you only had to hand in the Excel spreadsheet, was there... did you have to fill in metadata fields?</p>
36	<p><b>R:</b> There was only the option. So with these two journals you could basically... one was a neuroscience journal, this is an open access journal, pretty new. There you could decide how much of your data you wanted to disclose and the other one only required an Excel spreadsheet, but I think you can still do more than you have to. The problem is often just that you yourself are a bit too lazy to prepare the data in this way and then make it public when it is not required.</p>
37	<p><b>I:</b> And would you know where you can publish your data without submitting a paper? Would you know where to search for it?</p>
38	<p><b>R:</b> Hm... I do not think that is the case. So I know where I can publish my data or where I can publish raw data without the paper being properly accepted. On bio archives for example in our field. But unpublished data, I would first have to google or something to find that.</p>
39	<p><b>I:</b> What information would you need to make this process easier, just to publish data without articles? Because there is the possibility to make the research data, as you did it, a so-called enhanced publication. But you can also publish research data in a repository. Without... So there is no article involved, it's just the data with a documentation and the metadata. This is supposed to help other scientists to reuse the data. They don't need to know your research question, because they'll probably approach the data with a different one, but they simply need the data. And... what information would you need if you wanted to do something like that now, because from what I've heard now, you've never published data like that in the repository before. What do you think you would need for that? Would you maybe need some on-site training here or...?</p>
40	<p><b>R:</b> Yes, I think you would need some guidelines... like in general. Especially because it is about different... it is being published somewhat interdisciplinary. That somehow everything gets so standardized that everybody can understand it and knows what exactly we have done without explaining it methodically anywhere. I think there's would be... Yes, a training course is always a bit difficult, but at least a tutorial or something online or guidelines from the institute or in general would be important for that.</p>
41	<p><b>I:</b> Okay and do you think that research data from your discipline is more or less published in other countries than in Germany? Or is Germany a forerunner with the number of publications of research data?</p>



42	<b>R:</b> Germany is not the forerunner I would say. That is then in fact probably still the USA. In Europe, Germany is definitely one of the leaders. Yes, but I think the USA.
43	<b>I:</b> And have you ever reused data yourself? Did you search in any kind of repositories, etc. for data that you could use for your research?
44	<b>E:</b> Yes, my first publication was indeed a meta-analysis, because we said we don't make a// a new dataset that confuses the whole field even more, but look at what is there already, what can we compare with each other, what can we extract from it? And of course, in the process we used other published data sets and compared them bioinformatically.
45	<b>I:</b> And where did you search for the other data. Do you remember that?
46	<b>R:</b> Yes, I mean, in the field there are like key publications basically, which should then be included there in any case. I also searched PubMed and then I specifically searched for these big microglial labs where out of them comes just such a constant crazy research. But then there is also//For example, that was a meta-analysis of RNA sequencing, there is such a database called GO, where just everyone can store their data with a specific number. And all publications refer to this number, so to speak.
47	<b>I:</b> Very nice. That was my last question, but I was about to follow up.
48	<b>E:</b> Otherwise you can also write me that.
49	<b>I:</b> So first of all, thank you very much for the interview.
50	<b>E:</b> Sure.
51	<b>I:</b> It was definitely very helpful.
52	<b>I:</b> Did you think with the publication of your research data about licensing? In other words, which rights you grant for your data or which rights of reuse you also give for your data.
53	<b>E:</b> Yes, so actually not really to be honest. I also have not really read in the general terms or conditions or anything similar from the publisher, so not really.
54	<b>I:</b> But do you know different licensing models? Have you heard anything, or...?
55	<b>E:</b> At the Open Science Workshop of this//what are they called?
56	<b>I:</b> Orion?
57	<b>E:</b> No, what were they called here? Ah yes, right, Orion! That's where I heard something already about this Creative Commons, thought it was interesting, but then it kind of slipped away from me and maybe I'll just read about it again in the course of it.



58	I: I would definitely recommend it if you want to publish more data. Thank you very much.
----	---

