

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

Interview: os\_031

1	<b>Interviewer:</b> Okay, so my very first question is: how long have you been working in science already?
2	<b>Researcher:</b> It has been since my... it's like four years. Yes, four years.
3	<b>I:</b> And, okay, and what is your research about?
4	<b>R:</b> I do research about, erm, artificial intelligence in education and learning analytics.
5	<b>I:</b> Mhm ((affirmative sound)). And what is exactly the research data you are working with?
6	<b>R:</b> There's many topics. Two data sets. One is from an online platform from a commercial company. So they collect the data of the students doing, mathematics in junior erm high school. And they collect data for five years and they send the data to us to analyse something. And the second data set here is MOOC data, massive open online courses. And that data is, the participants are teachers in China... they are involved in a national//national training project we call [name]. It started just as a training project and the students would come to the platform learn something there related to the proficiencies and tutor them in teaching.
7	<b>I:</b> Hm. And have you already published your research data?
8	<b>R:</b> I'm sorry?
9	<b>I:</b> Have you already published your research data?
10	<b>R:</b> I published some papers, but I did not publish the data.
11	<b>I:</b> An when you have published your papers, didn't you have to//or do you have to publish your data as well or not at all?
12	<b>R:</b> Not at all, because they are too//too large.
13	<b>I:</b> Mhm ((affirmative sound)), and...
14	<b>R:</b> Do you think reviewers are interested to see my data?
15	<b>I:</b> Well, there are a lot, more and more, journals asking for the data as well when you're publishing your paper. So, it's becoming a very common thing now. That's why I'm asking. Erm, would you publish your data if you would be asked for it? Or do you have any concerns about it?



16	<b>R:</b> Yes, I do have concerns, because we have some protocols with those platforms. So, we should not share the data to anybody. And we're kind of protective about that, but I don't know if, if the journal asks to publish my data. I don't know how...
17	<b>I:</b> And do you work with personal or sensitive data?
18	<b>R:</b> ...circumstances before... Erm, I'm sorry?
19	<b>I:</b> Do you work with personal or sensitive data?
20	<b>R:</b> Erm, yes, we do have some personal data, but not sensitive. Personal includes fields like their names, their gender and their age and nothing else.
21	<b>I:</b> Mhm ((affirmative sound)). And do you think you would to be allowed to publish personal data, erm, personal research data in China or are there legal regulations that would prevent that?
22	<b>R:</b> There is no legal regulations to prevent that, but we will not publish such kind of data or we just omit some fields.
23	<b>I:</b> And why wouldn't you publish this kind of data?
24	<b>R:</b> Erm, because I do think it's kind of the norm in the scientific field that we shall kind of protect their, some kind of sensitive and their personal information, so that stuff.
25	<b>I:</b> And, erm, does the research data collected by you belong to you? Who is the owner of the data?
26	<b>R:</b> The company.
27	<b>I:</b> The company, okay. That was a very clear answer. Thank you ((laugh)). So, if you have never published your data before, do you have an idea how the publication process of data looks like or have you never tried it out or never seen it before at all?
28	<b>R:</b> Erm, what kind of publication process?
29	<b>I:</b> The publication process of data.
30	<b>R:</b> If a reviewer asked me for to share the data to him or the editor... That kind of thing?
31	<b>I:</b> Well there are different possibilities to publish research data, for example in a repository or in a journal together with the paper or you can publish it in a data journal. So, there are different processes as well, so for example in a data journals you have a peer review as well of the data, while in a repository you can just upload your data with your documentation. But you have never tried it out before?
32	<b>R:</b> Yeah, I have never tried that, because... I just have a little question here.
33	<b>I:</b> Yes please.



34	<b>R:</b> Have you ever met anyone that has asked you to publish data or if that kind of data is from a small like (unintelligible) or like a small experiment, with a small amounts of data? I think that they would ask me to publish that kind of data, but if your data size is too large, like have trillions of records in our database. It's hard to publish that kind of data to the journal or to the repository even if you have it.
35	<b>I:</b> Yes, that's definitely a barrier or a challenge, but there are possibilities to publish this kind of data as well. Like, erm, in repositories you usually don't have huge... missing a word now... like limitations of the space where you can upload your data. So, there are different possibilities where you can look for it. But of course, it is a huge challenge if you have big data. That's true.
36	<b>R:</b> Yeah, some might take the motivation for them to// to see the published data. They have... they want to confirm if your process of this kind of data processing is scientific or integrated. So, it affects only...
37	<b>I:</b> Well, the motivation is of course research integrity, so you can, erm... in the transparency in the research. So, everybody can see your data and, erm, re-analyse the data to see if your results are good enough.
38	<b>R:</b> So, it's to tell for them (unintelligible).
39	<b>I:</b> Exactly. And they can combine it with their own data to answer new research questions. And for you as a researcher of course the motivation is that you can, if you publish your data, you can be cited because the data gets a DOI and then you have more credits, right?
40	<b>R:</b> Erm, yes. So, I think that's a dilemma. Because the data do not belong to me, they belong to the company and we are not allowed to share it to anybody. So, if I want to check if my conclusion is correct or scientific it is really hard.
41	<b>I:</b> That's true, in this case it's...
42	<b>R:</b> What I can just report is these processes and the... some kind of plots and graphs and not the raw data.
43	<b>I:</b> Yeah. It is really difficult in this case if you have agreements with the company, that you are not allowed to publish at all. But maybe, what you could do is to talk to your company and try to make the research a little bit more transparent and to go a bit more in the direction of Open Data. So, in this case, you as a researcher, as you said, you don't really have the possibility to do anything, it's yeah... It's a little bit sad as well.
44	<b>R:</b> Erm yes.
45	<b>I:</b> And have you ever used data that was published somewhere? Like re-used data from other scientists?



46	<b>R:</b> Erm yes, we do have some. In China we have a... a large database called the CFPS, but I don't know the exact name, I just know it says CFPS. And... All those data are conducted by the famous universities and they publish them online and anyone that wants to use it should apply for the accessibility. You should like do a submit a form to them and after they agree, so you think you accept. But you have... I have seen some other... other data like in, in GitHub. And they are free.
47	<b>I:</b> Sorry, I didn't understand you. The last part.
48	<b>R:</b> Erm, sorry.
49	<b>I:</b> What was the last thing you said, because I couldn't understand that?
50	<b>R:</b> Erm, which part? I'm sorry.
51	<b>I:</b> That's the problem, I couldn't understand. So, I cannot really tell you what it was.
52	<b>R:</b> Okay, so let's make it slowly. Hm, in China we can have some kind of access to some public data and those public data are mainly in two types. The first type is: Those are statistical data from the government. It's a official, something like statisticonal features or something. And the other part is that the data are collected by a certain kind of... a group of scientists and in a specific purpose. For example, we have a family investigation... a family survey//family survey and those include all the information of that family and some more sets of better quality. Those are the two types of data that I can access online for free.
53	<b>I:</b> Okay. And have you ever looked for data from other countries?
54	<b>R:</b> Yes, I do. I search data on GitHub. I see some repositories with small stars and they are some free data sets that have never used that. And also try to search some data in India for open... for open online educational sources. I found that thing there and I also searched data from CIA.
55	<b>I:</b> Oh, nice. And do you know if research data in your discipline is more or less published in other countries in general?
56	<b>R:</b> I think it is less published. I seldom see in the field of education there is somebody... I mean in China, in the journals in China somebody would publish his data in the... like in the appendix and I've never seen that. I have just see the results.
57	<b>I:</b> Hm. And what would help you to publish research data. Are there tools or information that you would need, that should be provided to researcher in China to publish data?
58	<b>R:</b> Erm, I think... I should know who sees the data and if they could have a full access to the data. Like they can reprocess it or not or if they could share it or not. Or here is another kind of important thing that's: If in China I try to publish data and I say: "And now you cannot share it with anybody.". But I think this kind of statement makes no sense.



59	<b>I:</b> Mhm ((affirmative sound)). It's true.
60	<b>R:</b> Now that's contrary like in the United States, so it's kind of a statement. Like: protect my data. But not in China I think.
61	<b>I:</b> And have you heard about the Creative Commons, where you can put the license on your data for example?
62	<b>R:</b> The CC right?
63	<b>I:</b> Yes, exactly.
64	<b>R:</b> Yes, I know.
65	<b>I:</b> Yeah, so this is one possibility how you can give the access rights to anybody, but you, of course, have to be the owner of the data to give the licenses, so again we are here at the problem when the company doesn't allow you to publish the data.
66	<b>R:</b> Erm. In a sense that's mainly a problem of the company.
67	<b>I:</b> Yeah, that's the problem.
68	<b>R:</b> (unintelligible) the data file or I would like to make it published and to the public. Like with open source, that kind of thing.
69	<b>I:</b> Mhm ((affirmative sound)), that would be great.
70	<b>R:</b> I would like to do that. Yes, I would like to share my data of my own, but now I only have data from the company. I cannot do that.
71	<b>I:</b> Mhm ((affirmative sound)), I see. Okay. Thank you. Thank you very much. That was very informative for me. It helped me a lot.
72	<b>R:</b> Oh, really yeah? I agree.
73	<b>I:</b> It is exactly what I need to know. How the researchers work with the data and what possibilities they have to publish it or not, so it's exactly what I wanted to hear. So, thank you very much.

