Researchers' Perspective on the Publication of Research Data:

Semi-structured Interviews from Germany

Interview: os_005 - Translation

1	**Interviewer:** Okay, now it seems to work. So, once again, thank you very much for being willing to be interviewed by me. I would like you in the beginning to introduce yourself.
2	**Researcher:** Yes, gladly. [Name]. I am a junior research group leader [at an Institution], where I work on questions of biodiversity, especially the impact of climate and land use change on species diversity and species management.
3	**I:** Oh, that really sounds very exciting. And how long have you been working in science, like altogether?
4	**R:** Is that now, you mean from the beginning of the doctorate, or...?
5	**I:** Yes, this I would also already count as science.
6	**R:** Yes, yes, erm, exactly. Then I would say, that would be... twelve years. Approximately. Twelve and a half.
7	**I:** That's nice. Good. And with what kind of research data are you working with?
8	**R:** With which what?
9	**Both:** Research data.
10	**R:** With, erm... different ones. Among other things, distribution data of animal and plant species. Furthermore also climate data, land use data, erm... Trait data, so to speak, thus data on characteristics of animals. Yes, those are like the most important ones, I think.
11	**I:** Aha ((affirmative sound)). In your research area overall, is there any other research data that would be very relevant?
12	**R:** Beyond that which I already use anyway?
13	**I:** Exactly, exactly.
14	**R:** Not directly right now, because of course I try to use everything that is available. Or rather what is useful.
15	**I:** Very good. So you also reuse research data, or... do you generate it yourself?

16	**R:** I also use... published data and in fact on a very large scale and in addition we have recently started to generate data ourselves on a small scale.
17	**I:** Very nice. And have you also published research data yourself?
18	**R:** Erm, I am just contemplating now... Indirectly yes... Insofar that we of course, when we calculate models or make analyses, we naturally also publish the results of these analyses. Partly even also the raw data. So, the raw data, which of course are not raw data in that sense, rather they are also processed or result data, so to speak.
19	**I:** Aha ((affirmative sound)). Ok. And in what form do you publish this? Is it then as an addition to an article or completely separate?
20	**R:** Exactly, normally these are appendices to the scientific publications. But recently we have also made a large amount of results data available online via a research data platform, which is operated by the [Institute]... which is in the context of the [project]. It is about like coordinated climate impact research in different systems. So with this they bring different areas together and also offer there an infrastructure where the data can be uploaded accordingly.
21	**I:** And did you have with this publication of the research data, that you just mentioned, any reservations about publishing the data? Were there any problems with this process?
22	**R:** No, not problems in that sense, except for the fact that they are just many, or that it is a huge data set, which//just the uploading and then the quality check and so on... takes many weeks and sometimes even months. Fine, you always have some concerns of course when you have produced such results, that you are no longer the master of it, the sole master. But in this project we had there//from the outset we approached it in such a way that we said that everything there has to be published, so that others can also use the data accordingly.
23	**I:** I find that very exemplary I would say. Do you know under which licenses you have published the data?
24	**R:** Yes, this is a Creative Commons license. The most open one. There are different categories, I don't have them in my head right now. But I think it's the most open, where we so to speak//have to be cited in the usage of the data, but other than that there are no restrictions.
25	**I:** Very nice. Do you work with person related or sensitive data or not at all?
26	**R:** Not really. No.
27	**I:** I see... Do you know whether the research data that you collect, belongs to you?
28	**R:** Could you repeat again, somehow my speaker...
29	**I:** Erm, does the research data you are collecting belong to you?

30	R: Hm, the data we collect? Well, so I have for example for the project, I financed it with a cooperation partner [from another country] or through the DAAD. Erm, then there is the question then is who owns it. Well, first of all I would say that within the project they are ours, the data we collect there, yes.
31	I: Ok, and that means that you now also have the right to grant the rights of use?
32	R: I assume so, yes. For that I would have to look into the exact regulations again, but I assume that this is the case, yes.
33	I: Does your institution then have a research data policy?
34	R: In mine, in our institution?
35	I: Right, in the institutions.
36	R: Yes, there are. But I have to say, I don't have them ready right now.
37	I: That's perfectly fine. Erm, is there//or what information would have to be provided to make the process of publishing research data easier? So have you found the process as a whole somewhat complicated or was the publication itself then really simple. Except that it's just a lot of data.
38	R: You mean now with this project, what I just mentioned, with the uploading and so on.
39	I: Exactly. So really like the software that you used or the service you used. Was it easy, did you have to enter a lot of metadata or were there points you couldn't do anything with or did you give it to someone took this over? So how the process...
40	R: I actually did delegate this to my data manager, who of course is also responsible for it. Erm, well so I have an employee who is mainly responsible for data management and data preparation, administration etc. So I really can't say too much specifically about that.
41	I: Aha ((affirmative sound)). That is really is very convenient that you have such a person in the project, it is really rare.
42	R: That's true and I am very happy about it.
43	I: ((laugh)), I believe that immediately. And do you know whether research data, in your discipline of course, is published more or less frequently in other countries? Is Germany the forerunner or not really?
44	R: I think there is Germany by no means the forerunner, from what I can see. However, in our field so especially in ecological biodiversity research, which is a difficult story anyway, with the publication of raw data. Because of course, logically and understandably, many colleagues say: First of all, of course, we want to work with the data and how we have collected them before

	someone else publishes anything with them. Erm. But as I said, I think Germany is lagging behind in international comparison.
45	**I:** Aha ((affirmative sound)). Which countries would be like the//Or which countries do publish more? Do you know that?
46	**R:** Well that's just a gut feeling, I think....
47	**I:** Yes of course. That is exactly what I am asking about. ((laugh))
48	**R:** Erm. So the in USA and in Great Britain I have the impression that it takes place more. Australia too. Yes. Furthermore, I am now thinking... Hm, yes... Let's leave it at that.
49	**I:** No problem, gladly. Erm... And... Is there... Or what tools do you think you can get other scientists to do that, or what information would make scientists to publish more? Do you have an idea?
50	**R:** So to make more data available?
51	**I:** Exactly.
52	**R:** Yes, that is a difficult story. Something which we also often discuss. Hm. I think it's not so much about tools. Of course it's also one thing that the tools should be kind of, erm operable and not a science in itself. If you want to upload something or similar. But beyond that it's all about the reward system, so to speak, for the scientists. So in the end you don't get the credit for uploading data, but for the publishing in particular. Or for the application for the project. And as long as there is basically no equivalent measure or equivalent system for honoring the submission of research data, it will be very difficult to get people to invest their time and effort in this, I think.
53	**I:** And when you reuse data. Do you quote them also in your articles? So do you use the DOIs that have been assigned and to refer and point to them?
54	**R:** With which data now?
55	**I:** Well, you mentioned at the beginning that you also reuse a lot of data for your work. And erm...
56	**R:** Re... And what do you mean by re-use?
57	**I:** That you there...
58	**R:** So other data, other... so data from other colleagues.
59	**I:** Right!
60	**R:** Yes. Yes, of course they must be quoted accordingly. Erm, either these are then publications that are, ah... equipped with. Or rather connected with. Or there are corresponding, yes... DOIS,

	websites etc.
61	**I:** Aha ((affirmative sound)). Why I asked the question, because that would be exactly where you can get this recognition. If the data have been published, that you can point it out in scientific articles.
62	**R:** Yes.
63	**I:** And if they are not published, you can always point to the other scientific article, but they don't contain the data per se there, because there is mostly then already the result or the research question etc. there.
64	**R:** Yes.
65	**I:** And that... I also find very interesting.
66	**R:** That is correct.
67	**I:** That this is still always missing a bit. That this is also a kind of recognition.
68	**R:** Yeah, that's right. Only if you then look in detail... So of course you can then, acquire there quotations, so to speak. Erm, but if you naturally look in detail into a publication list, then that means in case of doubt... if it is only a data paper, so to speak. Yes, it's just a data paper, it's not somehow intellectually as demanding as an, I don't know, empirical paper or analytical... or a review or whatsoever.
69	**I:** Yes.
70	**R:** That is I think, the in-balance that is a little bit in there.
71	**I:** That's right. I can understand that. Erm... I also think about the open content, that they are afraid... somebody... evaluates their data faster. Or... In that sense... Is this really a big concern for many of your colleagues?
72	**R:** I think so.
73	**I:** Yes or. Sort of.
74	**R:** Well I do think so. That's what you hear over and over again, that you say. Yes, so we can also for example**:** If a doctoral student somehow collects data for a year or two, which he then//which he then has to work with and he makes three or maybe even goes into extra time and works on it for four years. Erm overall, that... then it is of course clear that it would be fatal for him if someone else suddenly downloa//downloads them. And then uses it for their own study, which practically then steals the doctoral student's doctorate away.

	I: Aha ((affirmative sound)). And what if the doctoral student only publishes the data completely after the doctorate? Then there would be no risk of that.
75	**R:** Then you would have... exactly, then we wouldn't have the problem. There are often like blocking periods or like... delayed publication dates. I think this is one way how you could solve this problem. Of course, you can't solve with that the//the problem, that some people still sit on their data, until the very last day. Because they say yes, but I collected it myself and I want to protect them now. Yeah.
76	**I:** Ok, so. By the way, I hear from this a little bit as well, that you actually on one hand would have to communicate more, that it would really have an advantage. To publish data and also... you can then start new cooperations and make the work more visible or also handle new research questions with it. Very interesting. I really find it very, very exciting. In principle, that's about it.
77	**R:** Ah yes.
78	**I:** Exactly, these were all of the questions that interested me very much and I find yours answers really very useful, very helpful. And I would like to thank you very much once again. I would have like one... question, which is no longer part of the interview.