

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

Interview: os\_009 – Translation

1	<b>Interviewer:</b> Well, then again thank you very much, that, that I can do the interview with you. Erm... First of all I'd like you to ask you to introduce yourself again. So which institution you come from. And, uh, what you do research on.
2	<b>Researcher:</b> Right, my name is [NAME]. I am an agricultural economist. I received my doctorate in agricultural economics at the [University]. And I run at the [Institute] the [department]. [One of the departments]. And I am mainly engaged in computer simulations, future projections on the topics of climate change and agriculture in both directions. The impacts of climate change on agriculture and the contribution of agriculture to climate change. And this on a global scale. So totally on a large scale.
3	<b>I:</b> Very interesting. So how long have you been now in general working in science? Like in years?
4	<b>R:</b> Erm... Exactly... [23 years].
5	<b>I:</b> That's a long time, one has to say. Ok, and with what kind of research data do you work with?
6	<b>R:</b> We//erm, mainly use, we//in my working group erm... we do not collect our own data, but use... erm, statistics from UN organizations or from national offices. Or also, erm, remote sensing satellite data. We use data from other sources. From other work groups. And then integrate them into our own computer models.
7	<b>I:</b> Ok. Erm... So how do you select this data? Are they all freely available? Or... what do have to you look out for?
8	<b>R:</b> Exactly, we//we are striving nowadays towards a strong... open... access... culture, with our model. We actually intend to... half a year ago... our land use model, what we developed in the group, erm... as... we did provide it as an open source version. With, erm... Global Common License and so on. On Github. And... we are also in the process of completely shifting the data foundation to open available data. But it is still till now the case, that some, some economic data that we need are not unconditionally, openly available. So these are data//There exists a consortium, that provides for global economic modeling such a data set and is also constantly updating it. And... they have the concept, that in each case the newest version of their data set... erm, erm... they are selling it. So selling in the research field. And then in each case they make the second last version of the data set, but slightly older data also, openly available. So we are trying



	there now, to find a way that we can work with the open data. But we haven't quite gotten there yet.
9	<b>I:</b> Aha ((affirmative sound)). Have you yourself already published research data?
10	<b>R:</b> Erm... I would say, no. No. We, erm... Exactly. Because we also don't publish our data... the results from our models directly, but rather in the form of... papers, of articles... for the scenario analyses. Erm, I'm just thinking right now...
11	<b>I:</b> Maybe I ask briefly in between: When you publish the papers, are you not obligated or...? Are you allowed to publish data with it, was there this possibility?
12	<b>R:</b> Yes, I am thinking right now. Actually, so... erm... exactly. But for that you have to say, that these are then the//the results from the models. Which are not directly related to the initial data. So then we publish them, in order to make it comprehensible. We then publish, uh... not the complete results from the model, but the essential data, which are necessary for the understanding of the articles. This is also, by... demanded by a number of journals.
13	<b>I:</b> Yes, Aha ((affirmative sound)).
14	<b>R:</b> Erm... But we do not publish the complete input data set and the complete output data set. You also have to say, these are very complex models with very large//large amounts of data. Erm... well then I would say, partially. So... I would say... parts of the results from the models, that stand in relation to the articles.
15	<b>I:</b> And these parts are then... open to others, or...? Do you also give licenses there?
16	<b>R:</b> Erm... Puh... Well I would say... open, in the sense, how the//for example the journals make it then available. So generally then yes under some kind of Creative Commons or... what is the other variant?
17	<b>I:</b> There is still... an Open Commons or...
18	<b>R:</b> Yes, or, or... from the GNU. Context. Yeah, well... I, I don't know about that part that much unfortunately, because we//because we found out, that there are also different variants. Erm... and erm... and that this also is partly related to the journal. So, exactly. But under//under Open Access lice//licenses. Which then also, the//then in turn regulates the use of the data. In the event of doubt.
19	<b>I:</b> And are your data person related or sensitive data?
20	<b>R:</b> No. They are not personal data. Erm... If, it is only about//it's more about the fact that//that we//that we partially have to buy input data. Which we then use in our model and of course can not publish the input data just like that. Could. Erm... partly we are also still negotiating, with



	the//with the//with the data owners. Whether we can then, for our purposes, publish aggregated data sets, processed data sets. Erm... because they are then important for understanding of the model. But of course not the original data.
21	<b>I:</b> Aha ((affirmative sound)). Ok. Erm... Are then the models, so that's also research data, right? Because that's also what you are working on. If you publish this... or want to publish. Can you even do that? Is that then, your data? Or who owns the research data at your institute? Is it somehow regulated?
22	<b>R:</b> The//the research data belongs to the institute.
23	<b>I:</b> The institute.
24	<b>R:</b> Right.
25	<b>I:</b> Do you then have a research data policy?
26	<b>R:</b> (unintelligible) The model... the model code, as well as the data. So everything that's being developed at the institute. Erm... that is, it's being organized as a registered society. And, erm... the property rights, they belong to the institute.
27	<b>I:</b> Is there a policy, on how to treat this data or is it up to each researcher individually?
28	<b>R:</b> Erm, we actually do have a//a modelling strategy and also a ... research data strategy. In the sense of... recommendation. Good practice, recommendation. So... Erm... I would say it's not regulations, but... erm, yes, there are like internal guidelines. Erm... how we should handle research data and also model, model code.
29	<b>I:</b> Aha ((affirmative sound)). Do you have the feeling, that the process of publishing research data is somehow, complicated or non-transparent?
30	<b>R:</b> Erm... I don't actually think so, if you occupy yourself a little bit with it... Even because there is now, erm... Because there are also new special journals. That purely publish research data. And that is of course a nice development. Until a while ago, as far as I know, it wasn't that common and eh... eh... we haven't used it in my research group... yet. But other colleagues at the institute with us, have... used it. Erm... where they... Yes. Their own data sets... which they again... No, mostly generated scenarios, data from the model results. Erm... for the use by others then, simply. I'm just thinking now. Eh... From Nature there is a journal only with, erm... scientific data or something. And then there is something from Geoscientific Model Development. There is also a data journal. Well this, especially, that we can publish data sets with DOI... number, so digital identifier. That matters for us, just like any other publication also does. This is a very nice development.
31	<b>I:</b> Is it really already considered that way in the community? As a publication, as real, as a



	reference when you publish data?
32	<b>R:</b> Puh, yes, how it is viewed, I don't really know at all. With us it's more like, that we//they are usually also, erm... reviewed in some form. They go through a quality process, of course. And, for us, we interpret it as... erm... a relevant publication. Just like a... erm... a peer-reviewed publication in a magazine. Yeah. So we look at it this way. I guess it hasn't really been that widely accepted yet. But just the fact, that you... that you have these journals now. That these data sets are citable. Is for me the... So it is already a big step towards, that it is seen in the same way. Because often yes, in the data, in the position of the data, there is at least as much work as in the creation of analyses or articles on it. Right?
33	<b>I:</b> Yes. Definitely. I find that very interesting. And erm... do you have the feeling that in your discipline, research data in other countries is published more or less? Can you describe it really by intuition...?
34	<b>R:</b> Well, I can not really distinguish it between countries. I also have to say that... that we, in our institute, work very interdisciplinary. That means, erm... So I don't really have such a close disciplinary link. When I say I'm an agricultural economist. But I have physicists, mathematicians, geo-ecologists in my group. So... We're quite a mix in terms of disciplines. Erm... What... my experience rather is, that certain working groups and this is for me now... We also work very, globally interconnected with partners. That it's more the culture of the specific workgroup, like... I say, proprietary or how open they are with their, model code and also with data. Because the question is always, with the//with open data and open science, erm... How long do I keep my thumb on my own data, to generate publications from it. In your own interest. And at what point do I make the data public and open it up naturally then for others to work with. And, erm... the same of course applies to model code that you publish. You have developed something, then you want to. We call that harvesting a bit, yes. You also want to harvest the publications from it, especially in the form of dissertations or similar. And that needs a certain amount of time. And... I think there is this discussion in all the working groups: If I make it right now, this things available to the public, then maybe there are others, who will publish faster on it. Erm, well... I, and//And the discussion is, I think, with all the working groups who say: Ok. How long//how long do we do it for anyway? How long do we wait? Let's say, after half a year or a year. Let's make the stuff available. Erm, so It's a kind of bounded openness. And I think this discussion is held by many groups. Internally. And then they also have to come to an agreement. Because, if it's a larger group, then... Erm, they have to have a common line. Find one.
35	<b>I:</b> Aha ((affirmative sound)). Do you know the FAIR principles? Have you heard of them?
36	<b>R:</b> Erm. Yes. Erm.
37	<b>I:</b> I...



38	<b>R:</b> Free accessible.
39	<b>I:</b> Findable. Accessible.
40	<b>R:</b> Inter, Inter//Interoperable.
41	<b>I:</b> And reusable.
42	<b>R:</b> Aha ((affirmative sound)).
43	<b>I:</b> Why I'm bringing this up, I did not want to question you right now. It was really just out of interest. Because, concerning the FAIR principles, it is also the point, that one does not necessarily have to publish the data per se. But... It is also enough or at least to publish the metadata.
44	<b>R:</b> Aha ((affirmative sound)).
45	<b>I:</b> Would that be especially for these discussions, where one is afraid that the data... Let me say, being stolen. For others... so that someone else works on the data, eh, faster. That you would only publish the metadata. Would that be a solution? Would that be something, that could maybe do more...
46	<b>R:</b> Puh, yes...Yes, but that wouldn't do so much in this case, I think. Then you would kind of know, what kind of data, is being generated. Erm... Nope, it is also, it is also not about data theft or something. It's just about that, to say, that you have spent some time into the model development and into the generation of scenarios. And you just have to accept, especially with the younger scientists... that it often takes a bit longer, for them to produce a publication. And it would also somehow unfair to say, you just put them easily on the net. And some experienced, I'd say fast-writing colleagues would then there... write articles based on it. So it's also a bit of protection, for individual... colleagues. And especially for the younger scientists. That they have enough time. To make something out of their work. And then later others can still, base their further analysis on that.
47	<b>I:</b> And does the code in your case... You've mentioned it a few times now... Is the code also considered a research datum?
48	<b>R:</b> I wouldn't say so. So... I think, we would probably differentiate between model code... erm... but for which//for which there are now also... special journals. So like this Geoscientific Model Development, it is such a GMD. Is such//such an... open journal, here in our... community. Where model code can also be... published. So for others, maybe that's//maybe they are... other data. For us it's more like... one thing is the data with which the model is operated. And the other is the model code. So, for us they're two different things, but... maybe it is seen from the outside... or, actually, similar.



49	<b>I:</b> Yes, okay. And, are there any tools or... erm... Would there be any information that would help. So that you publish even more data. In principle the movement is obviously there. But anything what would help further, that you could disclose your data even more openly.
50	<b>R:</b> Well I think, we are essentially profiting right now from the fact, that we have a disproportionate number of many physicists at our institute. And I believe that there has been in physics... for a long time a very strong movement. So there//there//that also shapes a certain culture here at the institute. Erm... So the journals, they are important. This whole question... erm... source code management. So GitHub and similar things. So our model is now up and also running there via GitHub. Version management of models. Erm... That//that we have in the meantime already quite professionalized. I'd say so. Erm... I am not unfortunately, I'm technically not that strong at all there, in it. So that's why I can't judge, where there still the gaps are. I have a feeling that with these constructions you can get pretty, far. Erm... what//what a problem was. And probably still is//is the question of//are the legal issues. Erm, which licenses are for what. And we have researched for a long time, which... open source license is now good for our model. Compatibility issues between different licenses. Erm. We also partly do model coupling. That we couple our model with//with let's say, for example an energy system model. And if you then have such common//so if//if you have then two working groups. Although they both pursue an open source strategy. But//but somehow, for some reason, choose different licenses. So, then it can lead to, compatibility problems. And... that seems... And//and then also the question: What are really the consequences of such an open source license? Well, you do that for, also in order to prevent abuse or as an prevention. But we have yet no experience at all, what that is like. If now really once, someone//if someone doesn't follow the rules. Erm... I don't really know, would we pursue it? How would we pursue it? And suchlike. So, ah... All that legal background, that's taken us quite a bit of work. And, it'd probably be good to have, if you had also... lawyers... who... specifically know about these things. Especially some from, in the research... concerning the research context. Because the boundary conditions in research are of course a bit different, than if//let's say you are in some kind of start-up company, or in the big companies, or... elsewhere. So, therefore. Erm... yes. So, legal questions I think, could... there is certainly some...
51	<b>I:</b> Do you have, do you have a lawyer at the institute?
52	<b>R:</b> No, we do not have a lawyer. We//we are not, big enough, for that. And we//we buy legal aid so to speak, per hour. Erm, but I know from my colleague, that//that it was a lot of work to get... erm... an well educated reference.
53	<b>I:</b> With//also with these compatibility problems between the licenses, did you also then seek... legal assistance or how did you solve these problems then?
54	<b>R:</b> We have a, yes... Well, for one thing, we have a few people at the institute who are pushing this whole strategy forward. That is why we then have developed such guidelines... for the//for all



	the groups. So as an guideline. And for us a colleague specifically took care of that intensively. And compared advantages and disadvantages. So, that was simply the commitment of one colleague.
55	<b>I:</b> Very exciting.
56	<b>R:</b> Who you have already also interviewed.
57	<b>I:</b> ((laugh)).
58	<b>R:</b> The corresponding one.
59	<b>I:</b> Very dedicated.
60	<b>R:</b> Right. But there are two, three others as well. Yes, well.
61	<b>I:</b> That's good. Well, that was my last question. Then I would like to thank you very, very much, again.

