

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

Interview: os_013 – Translation

1	Interviewer: First of all thank you very much that I can conduct this interview with you. Could you first tell us in which field you are doing your research?
2	Researcher: I work for a company that I also own. We are occupied working on chronic diseases and we help patients learning about their disease. And correlate the data about the over-all health of the patients to find out where... yes, correlations are, which we can then trace back, to help the patients and where we can help them in their ordeal. To quicken their healing process.
3	I: Very interesting. And since when are you active in science?
4	R: For this company about a year. Before I worked for another company where I was the CTO for five years and there in my last year I founded a machine learning team, which worked on laparoscopic recordings and there did image processing. So since my student days now again about two years in science.
5	I: Nice. And now there are in your current field... With which research data specifically are you working there or you yourself are working with?
6	R: So on the one hand we got the sociodemographic data of the patients. Beyond that we got their clinical picture, which is partially structured, partially unstructured. Based on the information the patients.... give us in the form of dialogues. Furthermore (unintelligible) image information. And... we got sensors that come from mobile devices. So for example we know where the patient has been and can correlate this then again with weather data.
7	I: And have you ever published research data?
8	R: Erm, not yet. I was involved in a project but... So with my former employer it was about image information and these we have published anonymised. Yes.
9	I: And where? Do you know that?
10	R: I think it was only on request. We didn't publish it on a platform. But there was the opportunity that other interested... yes, companies and other parties could approach us. And then we made them available.
11	I: Your data also contains personal and sensitive data. You already mentioned anonymisation. Would that be an option to make your data more public, if you anonymise it more thoroughly?



	Would that be an option for you?
12	R: Yes, the question about anonymisation has to be looked at critically. Because... at the beginning of the research you don't really know, what are the important factors. Of course you can generalize. The name will not really be relevant for an illness. But when talking about the age, it gets more interesting. And if I remember the legislation correctly you have to remove all factors for the anonymisation that could give conclusions concerning the patients. When there is a patient whose illness only happens in 0,01% of all cases then you can still find out quite precisely who that is. And this, there it gets problematic. If I anonymise the data now, I also lose a lot of information that I need for my evaluation. Erm, that means that anonymisation can't always be considered as easy, because the amount of information also simply gets reduced a lot.
13	I: What, in your opinion, are the biggest concerns when publishing research data in your field?
14	R: Solely in the research field I see it quite as unproblematic. I'd rather see it in the business sector. Research in that area just simply costs money. I have to pay people and I also have to sometimes make partnerships to get data. All of those are of course financial expenses that I got. When I then publish the data afterwards, maybe even for free. But what (unintelligible) comes afterwards. Publication has to be for free. Erm, then I am disadvantaged compared to my competitors. Erm, I don't know if that system has already been established, that you can just publish and receive a monetary equivalent value back. I have not seen this yet. But there I see the biggest challenge for companies that operate in that field.
15	I: And does the research data that you generate belong to you at all?
16	R: That's an interesting question. Erm, I have discussed that topic with different parties, also with some clinics that I was in contact with for the last two years. And... Erm... I am no lawyer, I can answer (unintelligible) like that, but my impression is that the data, that is the basis for the research, belongs to the patient. Or rather they can give their consent that we are allowed to use it. But the results on the other hand, the now already the trained models, they indeed belong to us. As long as we can at all time ensure that we can remove the data. Well, that we can remove the training data if the patient tells us that they don't want us to keep using their data anymore. That's how I see this model. Meaning that what we extract from the information as structures does indeed belong to us, the company.
17	I: And if the company then owns the data, you can also give rights of use for the data if you somehow decide to publish it for example. Theoretically speaking.
18	R: Yes. So the models are now missing here. So if we are speaking about a neural network for example, then that would be the trained model. You can of course license that in form of a product or also make that available. The data it's based on, as far as I know, we can't, I would say, if a patient allows us to process their data. That doesn't mean that we are allowed give the data



	to third parties, for the processing, but rather this has to be specifically written in a contract that we are allowed to do that. Currently the model here is that we want to process the data directly if at all. What we do with the models, yes, there we're already talking about the possibility of publishing it in form of OEM-contracts for others. But also for publications, because there is the opportunity that we become active in the academic field and when someone for example wants it for their algorithm or for their thesis that would be something we are open to. We would likely give the data, so without the actual training data, we would only make the models derived from the data available.
19	I: And do you have the feeling that the process of publication of research data is somehow complicated or hard to see through?
20	R: I got the impression that it is a big uncertainty factor. Everyone wants to work with data. Many have learned that it isn't always beneficial if big companies like Google and Facebook own data and that they got sovereignty over it. On the other side, especially because media is big on (unintelligible) about data security and data usage, everyone is very, very insecure: What can I do? And then there are two aspects. On the one hand no one wants to be vulnerable. I think that there is just being badly informed. Because I think you can find a good model in which you can publish without stepping into.... yes, legal grey area. The other problem is just: If I want research results and my data be the foundation and publish those to give others the opportunity to maybe reach the same goals with less effort. That is the question, I think, how altruistic you yourself are. I do understand it quite well if you don't want to. On the other hand, I don't see any problems from the legal perspective, but much rather a personal decision.
21	I: And is there anything that would help or what should be provided to ease the process somehow? To remove those insecurities. What would help you for example?
22	R: For example if there was maybe some kind of scheme that says: You have to remember the following things. Basically a checklist, that says: The following data you mustn't publish in any case, other data is okay. Then there surely are some where you have to discuss it, but that you can categorize at least 80% of the data somehow. That's allowed, that's not.
23	I: And do you know if research data in your field is published more or less in other countries than Germany?
24	R: As of yet I only had experience with Switzerland. Ah... and there I had the impression that it's quite similar. Especially when it's about medical data, you always have the ethics commission in between. It's actually very structured. And I think in Germany it is the same. Actually there isn't much of a difference. I think in the European region we aren't all that different.
25	I: That was indeed my last question for you. Thank you very, very much.



26	R: My pleasure. It was fun.
----	------------------------------------

