Researchers' Perspective on the Publication of Research Data:

Semi-structured Interviews from Germany

Interview: os_014 - Translation

1	**Interviewer:** …very much, that I can interview you. Thank you very, very much. First of all I want to ask you to tell me about in which field you do your research, what you are doing.
2	**Researcher:** So in the field of learning analytics or educational data mining? Generally speaking?
3	**I:** Generally, but I am indeed most interested in learning analytics.
4	**R:** Yes, yes. That… and in this field I had a project, which still interests me, but is a bit on hold right now, to track the interaction between the trainees within a learning platform with the purpose of also giving them an overview how they are acting. So that they can maybe see their own deficiencies more easily. And mostly also for the teachers as well. That they can recognise where the deficiencies of the students, the trainees are and then to plan their courses differently. That's one aspect. And then I also think that something like that can help, if with this you can change, let's say, the pedagogics and go towards (unintelligible) classroom. That's one direction. And the other direction is still learning analytics, but not anymore on the learning platform level, but on the academic data. So that you better understand how the students develop during their studies. What are they, I don't know, in the first semester? How many courses do they take? How many do they pass? With which grades? How do they develop during their studies? At some time what the curriculum and the module manual says and what the students do are indeed different. You discover also if there are ways that better lead to a better degree. Are there different types and so on and that is a new project that I'm starting now and that I'm really interested in.
5	**I:** That does sound very fascinating. Indeed. And with which research data specifically do you work in that case?
6	**R:** In the first project we got a lot from the interactions, we also implemented a lot so that the learning platform can save a great deal. Such as when a learning object gets opened. When answers are given. When the object is closed. That is all that we have saved. It also is a little unproblematic, but you with this data you have to explain well, why…. why you save them and what you are expecting. Also because we can gain a lot how good it could be for students. The idea isn't that you track, especially not for me, but that you… many students have problems during their studies. And the idea is if you support them a bit better during that time. So that was this data. And in the other project that's just starting with the academic data, that our administration got in this information system, there you really have to go through all channels.

Biernacka, K. (2020). Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from Germany. Humboldt-Universität zu Berlin. DOI: https://doi.org/10.18452/21644

	Data security, privacy and so on. And that requires, erm… small and big talks inside the university.
7	**I:** Oh yeah, definitely. And…
8	**R:** But there is no open-mindedness for that. So there are findings that it could be a plus for the whole university.
9	**I:** Very nice, very nice. And erm… have you ever published your research data or have you previously…
10	**R:** Yes, published. In the [project], yes. This I also have on my website quite a lot. And now the new project is just beginning. The first publication will be made in the workshop on Monday.
11	**I:** Ah, very nice. Very interesting. I'll look that up for sure. Erm, in that case do you work with personal and sensitive data or rather not?
12	**R:** I try not to. Also in this project, where we save quite finely, granular. We have in the… with a hash code, a hash function we always saved the identity of the students or the learners like that. That means, the name… there is only a hash code there. But is still is like that, that the data for a person do fit together. Principally you can track it back, but that would be very complex. And I also don't think for everyone…there is the danger as well, even if the data isn't person related, but very specific, that you can still recognize the person. For example, it happens to me regularly with surveys for university teachers. If you ask: Nationality. I am the only one at my university. Even if I then say nothing I get recognized. But I also think this data is so unique, no one would recognize: "Ah. This person opened that much in the learning object.". You couldn't do that. With the other project, with the academic data it's the same. So we're starting this project very carefully and we don't have the identity, so we got a person to extract and they don't belong to the group. They extract the data from the database of our university and anonymizes all the IDs we have got. Running IDs… we don't know. But there was still was the fear that if you definitely know: "So a person got (unintelligible) for every course. Took it in that time. With which grade." and so on. That you then for all courses know they made this and this and all that, you still can identify them. So there are methods for that. And to reduce that even more, first of all we got a declaration: Everyone working on the project obliges how to handle the data and so on. And then we also aggregated, so we do not save 1-0-1-3, that what we receive. We aggregated the data, so that some grades are amalgamated. Our results are a bit less precise, but I think in this first run precise enough for us to know, where we want to go deeper later. And if we need more data, or more detailed, then we will do that. But for me it is much more important in the beginning to analyse the situation and even more central: Trust. So not some… our university should have a good feeling concerning this project. The students as well. So it is much more important for me to handle the data very cautiously, maybe even a bit more inaccurate in the beginning, but that we can detect… which advantages we can gain from this. How we can use them, that we… yes, the trust is so important. If we don't have the students trust, the colleagues trust, then projects like

Biernacka, K. (2020). Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from Germany. Humboldt-Universität zu Berlin. DOI: https://doi.org/10.18452/21644

	this can't happen. So there I try very hard.
13	**I:** H. And could you publish the aggregated data or the data from your first project, which are already anonymised, in a repository? To share it for the use of others?
14	**R:** Yes, I did… In the beginning I did not do that yet. If I take that step, then I would only do it with the partners. Yes that again is a problem. I am still very pro Open Data and Open Science. That's just because… but mistakes are easily made. I now realise myself. Only when we can publish the data and the algorithms, can others review and find mistakes that we made without intent, but yes detecting would be important. Because we use what we conclude from the projects and use that for concrete procedures. The results have to be at least correct. Therefore I am very much in favour of this, but I still see this split between this privacy, in the sense that our students have the right of not having their data open for the world to see and that one could go: "Ah, Müller, you were like that!". I don't want that. I'm not that far yet. I don't see a problem concerning the procedures, but the data in an open repository I would want… I will only do that, if it secured with others, that you can't detect the people.
15	**I:** Are you of the opinion that a complete anonymisation exists? Will you have that certainty?
16	**R:** I am not sure. I want to investigate that. I am not sure. And even then… then you have to basically consider if someone… how probable is it, that someone gets recognised? Who would have interest in that? And what harm could it cause? And you have to weigh all that up. Erm, yes. What's possible as well is not the full data, but maybe the data partially for example. So still enough to maybe review the procedures. But if we got only parts of the data, you can not necessarily draw the same conclusions, that's also correct. Yes.
17	**I:** Just asking out of curiosity: Have you heard about the FAIR principles already?
18	**R:** Yes. It's also important to check there. Yes.
19	**I:** Yes, because there you can publish also the metadata, to then only give the complete data to people who pledge to not share them further, etc. etc... That would also be an option to check for those sensitive data…
20	**R:** Huh, no. What did you… Maybe we're not talking about the same thing. What do you mean with FAIR data?
21	**I:** I'm almost sure we're talking about the same thing. Findable, accessible, interoperable and reusable.
22	**R:** Ah, no. Didn't have that in mind. Aha ((affirmative sound)).
23	**I:** Okay, good. These are principles to make data available in a better way, make it findable, but one of the main principles simply is: You don't have to publish the data itself, but at least the

Biernacka, K. (2020). Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from Germany. Humboldt-Universität zu Berlin. DOI: https://doi.org/10.18452/21644

	metadata, so that one knows that something like that exists, so you can reach out.
24	**R:** Right, that's true.
25	**I:** Especially regarding data, which…
26	**R:** Yes, yes. Of course. And that would be actually very important to me. I mean I would love to know what I'm doing now, can maybe be generalised somewhere else. And only like that you achieve this.
27	**I:** Yes, exactly. And the next question I got: Does the data you collect belong to you? The research data.
28	**R:** Me personally? I have never asked myself that question.
29	**I:** So would you be even allowed to publish them? Not regarding the legal…
30	**R:** Ah. Yes, I have already discussed this with the partners in those projects, erm. The results, this yes. I think I already clarified this with the university regarding the academic data. As long, as you can not quite recognise, that it is about us, meaning this is a study course, a German university or similar, has our university management… or just charts and such, what we publish there is no problem for our university administration. For them as well. They really want us to publish. Well, that is also an… for the university, so research is important. And that is also… the university also did… would like to advance science. Publications are one way of doing that, then the projects and so on. As long as… it is important, that I am that I'm transparent there. Also not that afterwards anymore, but the first publication of that paper I did also sent to the university administration for example… before I even handed it in. For them to have a glimpse. Being predictable in the university; Me personally find it important in the work that I do. ((laugh)).
31	**I:** Okay. Hm. Do you have the feeling that the process of the publication of research data is too complicated or opaque?
32	**R:** It's not that just that simple I think. I have… I should give this more attention, but because I am in Journals, so in this editorial board of the [Journal] and we always ponder about if we also make the data sets available, thus not just papers, but rather the data, the methods. Some also do it, but I am not even sure yet, if we have in our journal the full infrastructure for this. There's an open repository in Switzerland. I forgot what it's called, the name doesn't come to my mind that easily, but that would be also an option. But I basically haven't dealt with this so far and personally have not occupied myself enough concerning this. Therefore I can't quite answer that.
33	**I:** That's totally fine. And do you have the feeling that research data in your field, learning analytics, is published more or less in other countries than in Germany?
34	**R:** No, I don't think so. For example if I… speaking about learning analytics maybe there is this…

Biernacka, K. (2020). Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from Germany. Humboldt-Universität zu Berlin. DOI: https://doi.org/10.18452/21644

	the field is less developed than in other countries. In Canada, Australia, there they are really… (unintelligible) further ahead I feel. But when I see that there are regularly paper getting published from Germany also at the conference for educational data mining or the LAK- Learning Analytics and Knowledge conference. So I do not think so, it is also interesting research I think. No, I don't think so.
35	**I:** And which tools and options can be provided to support the publication of research data?
36	**R:** Research data.
37	**I:** Still the data, yes. For the researchers to publish more, what would help you personally? To publish data.
38	**R:** Ah, first of all publishing data. I would really need to know, if I publish my data, what the… what the benefit is for me. Now it still is like that, what counts for my university is for example, when I finally get reviewed, are the articles I published and also where. But not the data. To really publish the data I would have to… for me to get to new publications, that I can't do without for example. Or that without publishing the data for other projects, I wouldn't be able to do that project. A project with other partners, an EU project or something like that. I would need a really good… a reason and without I won't see any progress or I just like it less. So that would make me publish more, (unintelligible) or similar, or that it would be real fun for me. Ah, that way I gain results or some options that I otherwise would not have.
39	**I:** Have you yourself already re-used data from others?
40	**R:** Ah, one more point! I have indeed provided data for others. That's right. I personally… I don't know if can that easily… but they all were anonymised. One time for a bachelor thesis, hm, anonymized and therefore… and I think you couldn't really tell where they came from. And one time… and just now for a different project. So I was asked for the data and I requested that if there are any results that they should also say, that I have provided the data. Ooh. I realize, for example I was really wondering about that data, okay. Yes, that is simple I thought, so… I thought for… others to think "Oh, interesting". Maybe one can achieve more with that data set than I did. In this case I would also be interested, so that it simply helps to bring the whole field forward, the whole area. So, that's a thing I did. So, what was the question now?
41	**I:** Erm. I just have… exactly. I have just asked if you re-used data, from others I mean.
42	**R:** Re-used? No. Indeed not.
43	**I:** So you also never searched if someone for example published data in a similar field?
44	**R:** Yes, I searched for a bit, but not on a large scale, because the effort was too big and then I got to my own data. In this DataShop of Pittsburgh. But there you can really re-use data. That's a nice

Biernacka, K. (2020). Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from Germany. Humboldt-Universität zu Berlin. DOI: https://doi.org/10.18452/21644

	thing, but now it's actually called learning sphere. But that is really a… I like that a thing like that exists. But then again in the end I didn't use it.
45	**I:** I understand. Okay. And my very last question is: How long have you been active in research, in science? How many years?
46	**R:** Oh. A long time.
47	**I:** Approximately.
48	**R:** I don't know… So I have… not always that intensive, but since I worked on my doctoral thesis and that are quite a few years, so more than 30 years.
49	**I:** Over 30, okay. Thank you very, very much.
50	**R:** You are welcome.

Biernacka, K. (2020). Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from Germany. Humboldt-Universität zu Berlin. DOI: https://doi.org/10.18452/21644