

# Open Science Talk No. 52 (2023): Responsible Research Assessment : a computer-generated transcript <sup>1</sup>

00:00:06 Per Pippin Aspaas

Open Science Talk, the podcast about open science. I am joined today by Felix Schönbrodt. Welcome to the podcast!

00:00:17 Felix Schönbrodt

Hello!

00:00:19 PPA

We are now in Munich in Germany, at Ludwig Maximilians University or LMU. This university – LMU in Munich – is alongside the Technical University of Munich the highest ranked university in Germany, according to these Time Higher Education rankings, so you are in a place where many people would like to come – a competitive place – and you are by now a professor of psychology since a couple of years. So this means you must have done something right?

00:00:56 FS

Maybe I did something right, but I also had a lot of luck, I guess.

00:01:01 PPA

Well, that's a modest way to put it, but to succeed in running for a position, that is at the heart of what we're talking about today, isn't it? We'll talk about an initiative that you have been part of, where you try to find responsible research assessment in psychology specifically. Could you tell us something about it?

00:01:30 FS

Yeah, sure. I mean – for me, that's the most exciting topic in in the world of open science at the moment. So we we approached this topic of research assessment from two different angles. The first angle is – I often give workshops on open science for early career researchers. And we often do an anonymous survey before the workshop, where the participants can hand in questions or comments – anonymously, that's important. And then we collect them, and then we discuss them in the workshop. And one of the main themes in these comments is the felt tension between doing good research and having a career in science. So the typical comment that we get there is, for example, how much of open science can I afford to do – or do I need to get tenure first, before I can do open science? Because when I do open science too early in my career, well, maybe I don't have enough publications, the publications are not spicy, sexy, clean enough, and then I don't get tenure. So at least there's a felt contradiction between doing good science – open science, reproducible science – and having a career in science. So that's one of the angles. And from a second angle – which also leads to this responsible research assessment – is from the publishing industry, or the publishing structure that we have. Recently, we published a paper together with Björn Brembs as the first author, on replacing academic journals. And we suggest to replace the traditional way of publishing with commercial publishers with an open, decentral infrastructure. And actually, everything is more

---

<sup>1</sup> This is a computer-generated transcript of the podcast episode Open Science Talk No. 52 (2023): <https://doi.org/10.7557/19.7344>. The automated transcript has been proofread by Per Pippin Aspaas and is included here for the sake of Universal Design and improved discoverability by full-text search engines.

or less in place. We do have the repositories, the peer review already is done by the community, so that's already, like – it's not within the commercial system, so the academic community provides them, and so on. And there are actually publishing systems that already achieve that like the *Peer Community In*.

00:03:47 PPA

How about yourself? If you look back, I mean, you took your PhD back in 2010. Ten years later, you were in a full job – a permanent position – as professor. In those ten years, did you embrace open science or did you do things the old fashioned way?

00:04:08 FS

Well, I changed my research style completely. We had this big replicability or reproducibility crisis in psychology, and that started around 2011. So I did my PhD in 2010. 2011, we had a couple of failed replication projects with heated discussions between the replicators and the original authors who were very prominent, older researchers. And it became so evident for me that we have a problem that I was personally at a point where I thought: what sense does it make, what we do here? It never was my goal as a researcher to produce a lot of hot air, and I had the impression that the field – including my previous research – didn't contribute a lot to knowledge. It was more about optimising publishability and having great, sexy results and so on. And I was at a point that I said: either I leave academia, because that's not this research style I want to proceed doing, or I try to change something first in my own research style and then also, like, systemically in the community. And in that year, with a couple of colleagues from the LMU, we produced *Our Commitment to Research Transparency* where we wanted to, like, point it down what we wanted to change. And we said: OK, we don't want to proceed the old way and we want to have a vision of our new research style and we want to commit to it and then we developed these 12 Commitments and put it on our website. So it was really just a personal commitment, and then more and more people wanted to sign the Commitment in some way to say: well, that is great, can I sign anywhere? And then first, we started to maintain, like, a local Word file and we had the signatories, but soon that was not very sustainable. And then we created this website – where people could, like, submit their signature via a formula on the website. And yes, since then I committed – and I think I really did it – to always publish the raw data when I have a publication, unless there are good reasons not to do so: sometimes we have privacy issues and then we cannot publish it, but as a default, I always publish it. I will always provide the computer code to reproduce the analysis that I did and a couple of more commitments.

00:06:49 PPA

Yeah, just to stop you there – because you have done this initiative as you say, and it's been anchored hasn't it, within the Deutsche Gesellschaft für Psychologie – the German Society of Psychology – where they sat down a committee a couple years ago, a special Commission for Open science, with about 10 members, one of whom was you. And then you decided two things: first thing, you signed the DORA Declaration on Research Assessment, which is about journal impact factors are not such a good thing, first and foremost, and of course other aspects are there in the Declaration as well. And then up came something called CoARA, which is on the European level. Could you just tell two words about CoARA and how this relates to the Deutsche Gesellschaft für Psychologie of which you are part?

00:07:48 FS

So CoARA is, well, it's an acronym for the Coalition for the Advancement of Research Assessment. It has been signed by more than 600 organisations by now, and signatories commit themselves to really do steps towards more responsible research assessment – so it's not just, like, lip service. So I think

they really are committed to do something. And my hope is that CoARA is the solution to the collective action problem that we face at the moment. So if only one institution is the first mover and changes the assessment procedures, they probably have disadvantages – either as an institution or the individuals within the institution – and so there needs to be a collective action, that multiple important big players probably move at the same time towards another way of assessment.

And the idea of CoARA is to get away from what we perceive as invalid metrics like the journal impact factor – and that is the link to the DORA initiative – but also the h-index, and move more towards qualitative assessment, in combination with a responsible use of metrics and indicators.

**00:09:06 PPA**

Yeah, so this has been, then, anchored in the committee that you are part of in the German Society of Psychology – how you can implement, if I understand it correctly, CoARA, as psychologists. Is it just within Germany, or do you have broader ambitions than that?

**00:09:24 FS**

Yeah. So it starts within Germany and also within Psychology. So it's, like, the scope is limited to a field and to a country, but at least concerning the field, I think that our ideas apply very well – to other empirical sciences, at least. And I think, also, that the German context is quite generalizable to other countries.

**00:09:49 PPA**

Sure. So these two papers that were published in PsychArchives – it's the psychology equivalent to BioRxiv or things like that. So in PsychArchives, you published exactly one year ago, in late 2022, two papers in English on responsible research assessment. And one of the things that struck me there is that you're quite specific about two phases in assessing a candidate for a job, for instance. What are these two phases?

**00:10:28 FS**

So the goal that we wanted to achieve is to have a feasible system of assessment. So there is one party says: OK, we need qualitative assessment all the way. But at least in psychology, we sometimes have hiring committees where there are one hundred, one hundred twenty applicants, and the committee will never read, like, three to five papers from every single applicant. This will simply never happen. So we needed a system that is both efficient and valid, and we think that a combination of quantitative and qualitative assessment promises to achieve that goal. So the first phase is where we reduce the long list of applicants to a short list, and the second phase is where we have an in-depth evaluation of the shortlist candidates and we want to employ more indicator-based assessment in the first phase when we have many, many applicants and then, once they pass the hurdle into the short list, we focus more or less exclusively on the qualitative aspects of assessment. And indicators should play no more role – or a very little role only – in the second phase. So the question is, how can we operationalize quality? And I think that that's very difficult. So it's if you ask researchers: what is a high quality publication, the most typically answer that you get is: well, sorry, I cannot define what quality is, but I know it when I see it. And, well, could be! The problem is if two different people look at the same paper, probably they come to different conclusions, so they see different things in it. And from a measurement point of view, well, this of course is a problem. So we said, well, quality is a multidimensional construct, so there are more lower-level properties of quality and more higher-level. So what do we mean with that? We say the lower-level quality standards are about doing good research practises according to the standards of the field, so our specific samples. So when we do computations in empirical research, it happens that we make errors in computation. So if a paper has

an obvious error – a numerical error, which changes the conclusions of the paper – well, that is a problem, and this is not high quality. And if there are papers with many, many errors, we would say, well, this is not high quality, independent of the actual content. Nowadays, from an open science perspective, we would say also: well, papers that refuse, or authors that refuse, to share their raw data, or who do not give access to their analytical scripts, like the computational scripts where people can verify and check what they did – nowadays you would say this is also a quality criterion. And this is very basic. So we don't know if the content is high quality, actually. We could have very trivial hypotheses for example, and trivial models, which are presented in a very transparent and reproducible way. So reproducibility does not guarantee high quality, but it's a necessary condition for high quality.

**00:13:53 PPA**

Just to check that I get you right: you're saying that when you apply for a job, you should list, of course, your publications and the journals in which they were published, as before. But you should also, then, add your code and your research datasets that you have archived openly, and so on. Is that a way forward to get past that first hurdle, the first initial screening, and then ...?

**00:14:19 FS**

Exactly. That's the idea. And you shouldn't only do it in the application process. Ideally, you should provide the code and the data with the publication itself. So when you publish, it should be provided with that.

**00:14:32 PPA**

What then about the reactions? I mean, when reading up for this interview, I looked a little bit into the official journal of the German Society of Psychology, the *Psychologische Rundschau*, and this summer – the summer of 2023 – you presented, in German, the gist of your work and invited others to comment upon it. And you've got several comments which were very interesting: short articles providing criticism or embracement of your work. Do we risk something when we create this new world – that we miss some good measures of quality?

**00:15:16 FS**

Maybe just as a little context, so we received a lot of comments – actually, we wrote three target papers overall, and received 40 published comments from different people. On a general level, there's always a risk in doing assessment quantitative assessment, because we all know Goodhart's Law: once a measure becomes a target, it ceases to be a good measure. So of course one risk is that people of course will adapt to the new evaluation criteria and well, then, if they know they get bonus points for open data, of course they will start publishing open data, which on the other hand is fine. I mean, we motivated them to be more open, more reproducible, which is good, but they could also cheat the system. For example, if you get bonus points for, let's say, pre-registration, they could start pre-registering nonsense stuff and then we have exactly the situation that you get a bonus point for a superficial behaviour, which under the hood is not really improving the quality of the science. So I think humans are very creative in gaming any system and I have no illusions that they will start to game the new system, if that gets installed some day. So that is one risk. Of course, I'm also thinking about the risk that we are creating our new Metrics Hell, with the new system. So at the moment we want to get away from the h-index, we replace it by other metrics, and then, after some years, you realise: oh no, we replaced one bad system with another one. But I'm – still, I think it makes sense to go that way and to have a constant evaluation of the new system. So this is not like a one shot thing and then we change the system. I think that needs constant evaluation and updating according to our experiences with it. Do we lose something else by focusing on these, what I call low level

quality criteria? Generally, I don't think so, because in phase two of the assessment system we go deeply into the content of the papers. So it's not just replacing peer review with these indicators, but it's just having the first phase indicator-based and then we will do really an in-depth peer review, so I hope that that happens in phase two and will be still a selection criterion on the shortlist.

**00:17:54 PPA**

Have you had any reactions yet from beyond Germany on your initiative?

**00:18:00 FS**

So we have these 40 published comments on our paper and I think – I don't have the numbers, but at least half of them were from international researchers, not affiliated with Germany. So yes, it's visible in other countries. And as far as I know – it's hard to, like, have an overview of the complete scene, there's so much going on, but as far as I know – our suggestion of this assessment system is one of the most concrete that we have at the moment. So there's a huge consensus in the community what not to do: for example, not using the journal impact factor – there's nearly no consensus about what to do instead. I think that is one of the main problems of the DORA initiative: it has so many signatories, but actually not a lot of implemented action, because we don't – we didn't know what else to do? And then even if we have some consensus about alternative ways, the consensus shrinks as soon as you have a concrete implementation, because as soon as you have, like, concrete indicators or concrete procedures, well, people start seeing how they themselves would perform on these new metrics, and have objections. And they say: oh well, but my research doesn't fit into that. And then the discussion is, like, about very nitpicky little details. And then it's even harder to reach consensus.

**00:19:33 PPA**

But one question from the humanities would be, then: do you actually need metrics at all? Maybe we could have people saying that, OK, this position is for this particular kind of psychology in this particular lab where the focus is this or that. Maybe you could have more narrative CV's and more narrative, also, applications where they try to explain how they fit to that job. Would that at all work, and would it be transparent?

**00:20:06 FS**

Yeah. I have absolutely no objections against that way of assessment. So if the pool of applicants is small enough, go for it! So our system works well in empirical fields. It does not work at all in purely qualitative fields. So I would never suggest to apply that to humanities, for example. And we also say if it doesn't fit, don't use it. So it's just a suggestion, it's not supposed to be normative, to be put, like a one-size-fits-all solution.

**00:20:37 PPA**

Could we now turn it a little bit to your other role? I mean, besides being a professor of psychology here at LMU, you're also the Managing Director of something called the LMU Open Science Centre. What is the LMU Open Science Centre and what are do you try to achieve there?

**00:20:52 FS**

OK, so our centre is an interdisciplinary centre and we want to promote reproducible science, primarily within the LMU, but also beyond the LMU, if possible. And we see ourselves as a broad centre, so most of our members are from empirical science at the moment, but not exclusively. And we want to expand into the humanities. But that is challenging because most of the open science movement happen within empirical sciences, and it's not clear at all whether and how these

concepts of, for example, reproducibility or replication are transferable to the humanities. So this year, we had an event with short talks and a panel discussion where we tried to explore this applicability of open science and all the concepts to the humanities. And we have seen, well, a mixed picture. So some concepts can be translated in some way, some others are probably hard to transfer and I think at the moment it's important to get the discussion started, in what way this is possible or achievable or not.

**00:22:11 PPA**

One thing about this recruitment of people to positions at a modern university is that you perhaps need other competencies than before – that is at least something that you state very clearly in your position papers. Could you elaborate a little bit on that?

**00:22:29 FS**

Yeah. So our target paper has four principles of responsible research assessment and the first one is that academic contributions are multidimensional. And at the moment, at least in psychology, we mostly only look at the research output. But there are much more contributions, many more – for example, teaching is very important, but there's also service to the field, to the scientific field, there's something like supervision and leadership. So when you are a professor, you have to manage a team, you have to supervise your PhD students, your team – you have to be, like, creative and so on. So that's another competency, leadership competency – and finally, societal impact. And I think it's impossible to excel in all of these five dimensions, and we might have different profiles within universities. The people who excel, maybe, in societal impact – they create podcasts and speak to the public and communicate science – and others specialise, maybe, on teaching, and others do research. And we should cover all of these dimensions in research assessment and ideally, the institutions that look for people should state upfront what kind of profile they'd like to see. Do they want to have hard-core researcher, who probably is not good at teaching and has no societal impact at all, or do they want to have people who excel on the other dimensions.

**00:24:02 PPA**

Yeah, the old Humboldt ideal of a professor, who should provide teaching based on research – but perhaps this is not one person anymore, it's several persons working in the same institution. Is that right?

**00:24:19 FS**

Yeah, maybe, maybe. And we shouldn't have the illusion that you can cover all of the dimensions on the highest possible level. So we need probably a specialisation and we need recognition for the other dimensions of academic contributions.

**00:24:38 PPA**

Thank you very much, Felix Schönbrodt for pointing to these several dimensions of research assessment that are coming to the future.

**00:24:47 FS**

Thank you very much. That was a nice conversation!

**00:24:50 PPA**

Open Science Talk is produced by the University Library of UiT the Arctic University of Norway. Thanks for listening.