Transcripts researchers: I5

***Interview date: 02-06-2020***

[...]

2. On a scale from 1 to 10, where 1 means unexperienced and 10 means very experienced, what is your experience with RDM? I would assume you’re quite high up the scale as you’re a data champion.

I don’t think I’m very high, because really, I don’t have that much experience with it. You know, when you start researching, you start it slowly, you are not starting with many projects or anything. So I would say maybe… 4 at most. I try not being too modest about it, but really, I think there’s so much still that I don’t know. I’ve had a few projects where I am trying to think about it of course, but yeah, let’s keep it at 4 I’d say.

3. Alright. And a bit of a more open question: what would your opinion be on RDM? Do you mostly do it because you think it’s important, or for compliance with policies of whatever kind?

Well, whatever you do, you do have some sort of management, and you can do it better or worse, right? There is always management, even if you don’t publish anything, even if you don’t do anything, you store it on your computer and that’s part of management. I think it’s important to have your data, well, managed as good as you can, as possible. To me, all of this open data as well, and if possible publishing data and making it available, it is for me like a quality sign. Something that you have done your work very properly, and you are also open about it that you say, you know, “you are able to check my work, I let you do this, it is okay for me, and even better if you can learn more from the data than what I have”. So that is my motivation for doing this. The data management is, well, that’s the way how you get there, and it’s for few people only the primary task that they would want to do with their research. It’s always something extra that you want to do, so it requires some discipline and effort. It’s not always easy. I try to follow the good practices, I want my work to have this quality sign.

4. It’s kind of a guarantee that the work you’ve done is of good quality, because you can say “well, you can check it for free”?

Yeah, you can see the data for yourself. It gives you some certain credibility as a researcher. Also, it helps people to understand your work. You always try to be self-contained in the paper, but it’s not always possible. There are certain aspects perhaps, maybe you forgot to describe exactly the methodology, then the person can see the files and can see the codes ideally, to see exactly how you did that.

5. How often have you published your data openly?

That’s what I’m saying, I have as a first author published once. My professor, he published data where I was a co-author, so he did it the second time. Right now I am in a larger project where I will publish data in the coming months I think, when I submit the paper. That will be my personal second time.

6. What infrastructure will you use for uploading your data?

4TU.

7. And what are your reasons to choose for the 4TU repository?

It’s a default, I didn’t look much beyond that. I think why not, it is the TU Delft participating there, why should I look for other places? I happily associate myself with TU Delft, so as long as I don’t have a very strong reason to go elsewhere…

8. When you uploaded your data the first time, was that also to the 4TU infrastructure?

Yeah.

9. And what kind of support did it offer you? Like, did it give you guidelines on how to format your data?

Yes, well, there were a few explanations for the fields you need to fill in. Although some things were still confusing, I admit that I actually did not do it in an optimal way because I couldn’t understand everything, but it’s done anyways.Altogether, it achieved the main goal, which is that the data is published. Some of the descriptives are not maybe a hundred percent correct, but then it's again very difficult to change because you need to contact the person and say, you know, like, update the whole thing. So I just thought "OK, this is my first try, second time I will know better!" But the biggest support, and actually I think you will probably ask me later, why actually I ended up publishing the data and why I ended up becoming a data champion was because my promotor, my professor asked me to contact the data steward of my faculty at that time. And he just said "talk to him and let's see what we can do with this data". And he was basically the one who was, we had a longer meeting, he talked about publishing data as well as all the management things about data storage. So he got me into it, and then I started to think "yeah, OK actually, I think I want to do it as good as I can". I mean, as much as I can afford my time doing this.

10. And if we go a bit back, you said you cannot edit the earlier submissions for 4TU, is that correct?

Yeah. As far as I understand you can't.

11. Would you say that you would have improved the description of the data you uploaded if that were possible?

Yeah, well, the problem is I didn't understand that the authors-fields should indicate all my co-authors, so I only put myself. But that's of course not correct. I mean, they were co-authors and they should have been in, but I understood that field would be just like a 'creator' kind of field. Hence they end up not giving credit to my co-authors with this dataset, which is a bit silly! But I mean, I think I will not change it anymore because I have already referred, I think, to this data source in some other work. And then it is the reference as only my data sort of, which is not fully good, but I mean, if I had known that I could edit it at that moment, I would have done it. And I think that's in general...

I do not have many desires for improving the infrastructure, but I have one, which is that it would be really nice if you could have some sort of versions in the data. So you have the submitted version... I understand the reason why you don't want to allow people to change easily, because then somebody else accessing the data potentially will not know whether he's seeing the same version as somebody else who accessed it earlier, right? But then you could solve that problem by having a certain version, version zero or something is the submitted version, and then you have a version one and that shouldn't be too difficult, you know, you can easily do that in all of these programming structures like GitHub and anything. So similarly you could, you know, allow that to be then in 4TU I think. Well, this kind of small mistake, that's one thing. But the primary reason I think it would be nice is that in that case, I would be confident to already upload my data when I submit the paper for the journal. Right now I am a bit hesitant because I think "what if the reviewers ask me to do another analysis?" and then I would want to add another analysis file in the data, right? So I mean, ideally, you put the data itself as well as the program you used to produce results. So when they say "OK, I want you to estimate more models", then I will want to modify or add more files, but that is very difficult. So it's always some sort of timing problem: when is now the good time to say "I can publish the data"? It's now that the reviewers can see it, it's now that the people can see it and even more so, I have right now this data for the bigger project which I mentioned, I try to have two publications for it. And one will come first, of course, and the second will come maybe a few months later, and the second will also have some sort of model files. But since I want to publish the data already with the first, and the model files will not be ready yet, then it will not be complete. So...

12. So it's not quite sure when the definitive version of your data would be ready?

It'd be nice to, I would like to publish it already because then I can refer to it in my first paper. But then it's not completed, so there will be another addition, maybe, maybe not, maybe it's not necessary. Maybe the model is so easy that you don't need to add it in the data. Some sort of solution for that would help. And also maybe a guideline for the researchers about what the best timing to publish the data is. That would be, I think, important.

13. So what is the rest of the usability of the platform like? Because you said you can't edit your datasets, but can you navigate it easily? Is it responsive or slow? Can you find what you're looking for easily?

I can find back my data quite well.

14. Right. So that part of the platform works?

Yeah, yeah I think so. I'm not sure, this DOI number, if I were to search it, would I find my data? That I'm not entirely sure, but if I have the link then I can follow the link.

15. Do you get a separate DOI for the dataset or just one for the publication?

I think I have it for the dataset. But I think, I'm not sure about that one. I think I tried to search the DOI once and I didn't quite get it. But the link works for sure.

16. Have you heard of Zenodo? It's like another data infrastructure.

No. I like to use TU Delft as much as I can. So there is a similar story about repositories, about publishing your papers. There is only one time when I thought "OK, TU Delft repository is not the best choice for me". So I also tried to put my papers, the author's version in the repository. But one paper I put in [another institutional repository] because it has certain reputation. It's a repository which is very well... I know that people of that field to which this paper belongs to, they would often publish in that repository, even if they don't have an association with [the institution], it's odd. It's some sort of historical thing. And then I thought "OK, yeah, this paper belongs to that field and I think I should put it there". I don't know this kind of subject-specific data repositories. I've never heard, I mean, it's much less developed than the paper repositories, right?.

17. So would you say you would switch away from the TU Delft repository if you found a repository that is more fitting to the domain?

Yeah, if it's more fitting to the domain, yeah. That's important, because then the audience would really be the people I want to address. The people I want to address are my peers, most of the time, right? Fellow researchers who might want to use the data. Yeah, I would like them to know there is a paper or something. So then that would be a strong motivation. But I doubt that there is, at least for my field, I don't think there is a field-specific repository.

18. Still on this infrastructure: have you ever used it to download research data? So to use data of other researchers?

No.

19. And is there a reason for that or would you just say you...?

The only time I did that was to check their formatting, to see how to format my own data and the helpme-files and that kind of thing. So that is for learning only. But for research? No, I haven't, because it's really not common at all to publish data in my field, so to find something for my questions that I want to know, it's very unlikely. I think that's a far future, that would be a nice future, isn't it?

20. We've talked a bit about it before, since you said the 4TU repository didn't have a lot of support for it in terms of the description with the fields, I believe you said?

Well, at least some things, the time I did it, it was already more than a year ago. I don't remember the details about it. I only remember the problem that I eventually caused for myself, which is that I made a mistake, but otherwise it wasn't a big deal. I mean, it's more or less standard, you know, you enter descriptions of your files. So it was not a huge thing. But then the authorship, yeah, that was a bit confusing.

21. Have you talked to any support that the TU Delft offers for research data management, like the data steward?

Yeah, I have talked to the data steward a couple of times. It's nice, we know each other well now! Yeah, when I have a question I ask him, and it's easy when he's often available. Also, I recently became data champion and then there's not much that I have done. But then, some colleagues have asked me some question and I say "I don't know, let me ask the data steward" and I go to the data steward, but it's good! It's also learning for me, you know, like "OK, this kind of situation. What do I do?"

22. What are the most common subjects you discuss with the data steward?

Well, the other day I asked him about publishing my code. Yeah. It was a similar question as what we discussed earlier, data that might change later. So I was submitting a paper and I wanted to publish the program code and I said "OK, it might change later. What should I do about it" and he said "Maybe publish in GitHub or Gitlab, one of those". And so I did. And then we had a brief discussion like, is it double-blind, single-blind. It was single-blind in that case, so it didn't make a difference. I could have my identity there and then the reviewers wouldn't know me anyway. But now I actually was rejected, so I resubmitted to a double-blind journal so hence I had to delete the reference. But it's a bit silly because, of course this is also, I mean, it never works optimally, this double-blind single-blind thing. Most of the time you can figure out some things about the authors, even if it's double-blind. I have a working paper of this paper in a repository where you have the link to my... It's basically the same thing, except that the link is to the code in Gitlab. And then I have submitted the journal where I said "oh, this link is anonymous. I cannot put it yet because it's a double-blind review".

23. That sounds cumbersome.

Yeah. I guess you can't help that. So I basically asked him about what to do about this program code which might change later and so forth, he suggested this option to me. And I think I have asked a couple of other things. I can't recall. Yeah, whenever some question comes up.

24. It's nice that you have a low bar to step to him and ask questions, that sounds good.

Yeah, certainly.

25. And have you used other types of institutional support? Like people at the university library?

No, I haven't asked for their help. I have met to talk to them, because they wanted to interview me as well. And now I think I received too much attention for the little that I do, but they wanted to interview me and ask about my experience as well sometime ago. So you know each other as well. But for all my questions I have always bothered the data steward.

26. All right. And are you aware of the funds they offer for data refinement?

Yeah, I heard about those. Yeah. For formatting data. It briefly crossed my mind, but I did not use it after all. The current project that I'm working on, it requires a lot of data cleaning. But that is quite a specialized thing for exactly the way I need them to be cleaned, and I could have perhaps applied for it. So I asked a BSc student to help me do that, and that is sponsored by my project. But perhaps that could have been reimbursed by this data management, right?

27. I'm not entirely sure what they precisely offer in the fund, but I've heard not a lot of people have used it.

Yeah, it might have been an idea to... Because the reason is that it is not exactly for publishing the data. The cleaning I need also for my analysis, I need to clean it in order to have the correct data. It's not only just formatting and putting it in nice tables and something like that. So that is not the biggest share of the pie.

28. So you would say that the fund doesn't really apply to what you want done with the data?

I am not sure, but I think it may not, I think it's maybe some sort of grey zone but I think, yeah.

29. Yeah, it's complicated and has a lot of sides to it. To me, the fund is more like when you're done with your research and you have a dataset and you just weren't working with the idea of "well, I have to publish this afterwards" in your mind that you would then have to still clean it and make it shareable, but I think like you describe, if you want to clean the data then maybe it's less fitted, but I'm not sure. Well, it's good to know at least!

30. I have another question on how satisfied you were with the support offered by the TU.

Yeah, certainly. Very good, very good! I can always find the data steward, he answers emails and yeah, perfect. It's like colleagues, right. Nice colleagues and then you talk to each other.

31. Is there anything you would like to see changed in the support they offer? I'd assume not because you already have a lot of contact with the data steward?

Yeah. I yet need to make more advantage of the data champions community. Somehow it was always very unfortunate, it's been like half a year or one year since I'm data champion, and there have been meetings, one or two, and I couldn't just make it. So I basically, I have been to one meeting in the very very beginning. But, yeah, there's not much you can do, it's just the way it is. So yeah, it is a good thing.

32. How does being a data champion influence your daily or weekly activities? Do you have a lot of people coming to you for support?

No. I just announced it in our section meeting: "I'm now data champion, you can ask me. I'm happy to discuss with you", but not many people, just, yeah, a couple of people have asked me something. Maybe I need to announce again or something, but yeah.

33. Would you say that maybe that is because right now you're, well, I assume you're working from home a lot now?

It was also before, I don't know why. I don't know. Maybe those people are doing their own thing and they don't want to change too much, or maybe there are not so many problems, that is also great! Or maybe they solve the problems directly with the data steward or with their supervisor, so they already know what to do. So for me, it's like "yeah, OK, I'm available for you. But if not, that's fine."

34. We've already talked a bit about your motivation to become a data champion. Would you say the advantages they list on their website also played a role? They say there are financial incentives, you could learn new skills and there is network building and increased impact. Was that also a reason for you to become a data champion?

I think learning is on the top for me. So the reason why I want to be in the community, yeah, not for impact or anything specifically. I want to know how to do this best and it's still an emerging field, this data management. [...] And all this, the data management plan was not mandatory or it was not mentioned. So yeah. I just learned how to do this thing good. So that I can have that, yeah, that quality sign for my own work. And if along the way I can help others that's good enough for me.

35. I see. I had another question on what your opinion on the advantages were and if they should be good enough for people to stimulate them to become a data champion. But I think I agree with what you said, that just learning should be the biggest motivator for people.

Yeah. So I think there is value in... Yeah, you want to publish data and then there are certain... The ways of doing that are still developing. There are some inefficiencies. So you may say "OK, I don't need to be data champion for that, I can just publish my data anyways, you know". But then if you want to, you know, just chip in a little bit, maybe I talk to you in this interview, I talk to some other people to say "I think these are the problems I experience", and then eventually down the road, it will help to improve the procedure. So then you can do that as well.

36. Now moving on, we've already talked a lot about research data management and I would suppose you're also familiar with the policies in place?

Um, I have looked at it briefly, actually I just looked when I saw your questions, I was like "hey, what are the policies?", but there's like a short document about it for each faculty. So I looked at it and there is not really much content. One thing is that now new PhDs need to make the data management plan, which I didn't have to do. Well, I could do it but I'm also busy with other things. But I think I will perhaps attend that course, the data steward told me there was this course about making data management plans for PhDs. So it might be useful for me later because I will stay in academia. So I will need to know more details about it. I know a little bit, I have seen one by colleagues, but yeah, a bit more would be good for me to know. And what are the policies besides the faculty-specific one… The broad idea which I get from there is to be open, be transparent when you can. Right? You know, that's the policy, that's the goal. And not much more than that, I think. So yeah, specifically this FAIR, findable, accessible, interoperable, reusable, these specific ways how you can be open and that's a bit you can keep in mind. But you know, on the big line it is, "just do it if you can". And I agree, I agree.

37. So it doesn't change a lot for you because you were already focusing on keeping your data shareable, if I understand correctly?

Yeah, yeah. There is also this Ethics Committee thing, which is also... but that is not in the policy. It is a separate aspect, is it?

38. Yes, it's separate. I actually had to go to the HREC myself, because I'm doing interviews. So I had to fill in a DMP myself and then talk a bit about, well, just the general set-up of my interviews. They have questions like "do you use your own prototypes?" or "are there financial incentives in place?".

Yeah, I think I also did that one for a survey, now that you remind me, but yeah. So I have done one data management plan!

39. So that is your only experience with DMPs then? I'm not quite sure since how long it has been obligatory to fill one in, but you've only filled in one?

I have filled in one. I have had two data collections so far and the first one was before this whole thing started to become mandatory, I think. Or what was it... wow, it's been some time! For the second one, I surely went to the Ethics Committee, and the first one I think I didn't. The first one was focus groups, which maybe would have been relevant because the people actually came to TU Delft and we had interviews and it was recorded, and I didn't do it, I didn't do any of the good things that you should do. But then for the second one I did the whole thing, I went through the Ethics Committee and then I must have had this data management plan as part of that. And there were also some other obstacles which is quite interesting. I said from the start "I want to publish the data, the anonymized data". But then we had a survey company, which had an existing contract with our faculty or at university level, and they were according to the old guidelines of no publishing, you know, you also must destroy the data when you don't use it, nobody should have access to the data. But that is, of course, idea-wise completely opposite to publishing the data, even if it's anonymized. So they have a very conservative kind of contract which had to be re-signed for every specific project. So then I had to go to our faculty contract manager and she had to renegotiate this contract with the survey company. A survey company is the one who is just having a panel of respondents in the Netherlands. They just have access to many people and that's what I wanted for my survey. They said that there was a bit of a hassle and it is an interesting experience. This is moving on from the old to some new kind of way of doing things. Yeah. Um. so what was the question again? I explained the whole thing, but...

40. It was what your experience was with DMPs, have you filled in a lot, or... But I think I've got a good answer to that now.

Yeah. So the DMP, it was part of the ethics procedure, now that you remind me. But there are a few things along the way which are not in the standard questionnaire which you need to do in order to manage. Like that the contract doesn't foresee this kind of use of the data, so you need to do extra there.

41. And what would your opinion on the DMP be, if you still can recall? Did you mostly experience it as cumbersome and as something administrative, or did you see the use of it?

I don't remember the specifics of it. It must have been OK. Well, you know, of course it's a standard form for all the faculties, and the research varies so much. And you always, everybody will find certain sections irrelevant, that's unavoidable. I don't remember exactly.

42. So back to the research data management policy of your faculty. Well, you already said that it's quite generic, not too much specific information in it, which is correct I'd say. Is there anything you would like to see changed in the policy?

Something I'd like to see changed in the policy... Hmm…

43. Like maybe it should have more restrictions or be more specific so that more people get aware of the topic?

Yeah, I'm not sure how many people just read policy documents like that! You know, everybody's quite busy and the further you are in your career, the more busy you are and the less time you will spend on reading the documents, except maybe if you are a leader of a section or something. So what happened for us was that we invited the data stewards as well as the library people to present in our section meeting. They gave a talk and much, much more details about how to do the research data management, broadly speaking. And there was a space for questions and answers. That was a much more effective way of reaching the audience, I think.

44. I have one last question, it goes back to the data infrastructures again. We've talked a bit about the 4TU Repository, and that was the primary one for you to use because it's the TU-specific one, and that you did not know of any domain-specific repositories. Let's see how I will formulate this... Because, well, I'm just going to explain what I think is the problem. There are a lot of repositories out there, but you cannot really search through them. Is that something you can relate to?

So, indeed, like you can search through papers via Google Scholar, right? You cannot search through data in the same way. Maybe there is a repository somewhere, who knows where. And maybe they have some data which is interesting for me, but I would never know. And yes, sure, I think it would be a good idea to have some sort of algorithm, some sort of unifying thing that helps you to search through repositories like you search through papers and all kinds of things. I think it's a step in the future. Currently there are few datasets only, I think, in my field. Maybe other fields are different, it's not so customary to publish data. I think it's mostly like a pleasant surprise when you see “oh wow, this paper has a dataset”. I remember one time I gave my paper about these focus groups to a fellow researcher and she's also doing qualitative research. She's a more senior person, so she's like, “ah, okay, another paper, whatever”. And I told her “there is data!” And she was like “there is data…?!” So she was like, OK, a little surprised, and she didn't come back to me about it, so I was thinking “yeah, maybe you want to use my data again”. It was interview data, you can do many things with it and there is always of course large costs involved to collect data. So yeah, TU Delft, the Netherlands is well-off, we are very fortunate. But there might be people everywhere around the world who have great ideas but not so much money to collect big datasets, so it would be very beneficial if they also can access data and use them, yeah. And really why not, I mean, if you have already published it?

Transcripts researchers: I6

***Interview date: 19-06-2020***

[...]

3. What is your role now? How much do you do lecturing, and how much do you do actual research?

I’ll have to define research here. Lecturing is like 60%. Research like 30% and 10% is leadership, management, all kinds of stuff. Research, in fact, I’m doing the more 'consultant' kind of project as research, and I’m also publishing a lot based on cooperation with other researchers and based on my own literature findings; I see that as research. Well, publication is in a strict sense not research, is it.

4. Could you elaborate a bit more on the 10% managing and leadership?

Well, [...] I’m member of the management team of one of the departments of the faculty.

5. Let’s dive into the research data management, the actual topic of the interview. Are you familiar with the topic of research data management (RDM)?

Well, I know what it is, and I know that [data steward of faculty] is now running here at our faculty and has to make policies on research data management and education data management.

6. Since when are you familiar with RDM as a separate concept? And how did you get familiar?

Like, 2 years ago? In fact, the very term of data management I knew before, but to me it was a very abstract thing. However, I was involved in a European project; there was also data management involved. The very idea was that if you want to safeguard privacy, it is not only about your data management, but you also have to trust other actor’s data management, and they have to trust other’s data management, so that’s a chain. There it became more specific to me, like “okay, wait a moment, it’s an issue”.

7. Are you familiar with data management plans (DMPs)?

Yeah, well, that sounds really like bureaucracy to me! So I know that they exist, but I never read anything.

8. Alright. This is a bit of a more reflective question; on a scale from 1-10, where 1 means unexperienced and 10 means extremely experienced, what would you rate yourself on the topic of RDM? So are you unexperienced, or extremely experienced with RDM?

That’s a very hard question, because I can say it’s a 10, because in an implicit way I was always involved with data management; I save stuff and I take care of the management of the data. However, as an explicit activity and using the tools that are available, then I would give myself a 3.

9. Why would you say that you’d give yourself a 3 on the tools available? You know that they are available I suppose, but is there a reason you do not choose to use them?

Hmm, why don’t I choose them… Well, in the first instance to be honest, it’s never been really urgent or necessary. There is no privacy issue involved. I am an open data scientist so if anyone wants to know anything from my data, they can get it from me, there is no problem there at all. I just got away with not using any tool, so if you actually never have time for anything, just to do things right, then this is not the first priority. Now the psychology of the thing: if you do not take a deep dive into this issue, then you’re going to expect or suspect that there are all kinds of tools around that you should have read or get involved into, but don’t really know it. So I have this idea that there are a lot of tools, but maybe there aren’t, you know. Based on that psychology, I would give myself a 3.

10. Alright, I understand. And you just said that you’re an open researcher, so that if people want to know more about your data, they can get it. Have you ever published your research data openly?

Yes.

11. And how often?

In fact, the open data movement exists for quite some time. However, it’s also a matter of budget, because if you want to publish open access, that’s what it is about, you have to pay extra so you have to have budget for that. So you also have to make a self-assessment: is this article worth it? So the first real open data publication that I know of was also 2 years ago; before that, I was just publishing in journals who had a different policy.

12. Did you publish the article open access or was it the data behind the article that you published?

The article. However, then there is a reference under it, like “If you want to have the data, you can find it on or contact bla bla bla”. Nobody does that, huh?

13. What you just said, was that on paper publishing, or on data sets in specific?

Article publishing. What you should know, I guess, is that I’m a social scientist. My data may be different than data of more quantitative or more engineering scientists. My data are, in fact, not figures, but interviews, just like what you are producing with this interview. This is also what I produce with my interviews and that is my data. And of course the coding I use et cetera, that is the data part.

14. Would you say that, because your data is not that quantitative, that is the foremost reason you’re not publishing your data?

It’s very contextual data. I can open it up, but it’s very hard to penetrate for people who don’t know the context of my data. However, and now we’re talking like 20 years ago, people sometimes asked me “Hey, you’ve done that interview, do you have a summary of it?” and I send it; that’s open. But those are usually colleagues who just knew what my research was about.

15. So you’re not frequently using open data infrastructures like the 4TU repository or Zenodo?

No. In fact, the open data repository… No, I don’t use that no.

16. Well then, onto the RDM policies. Are you familiar with the RDM policies that are in place at your faculty?

I’ve talked to *[data steward of faculty]* about it, but I don’t know the formal policy.

17. Well, it’s quite new still. I was going to ask how satisfied you are with the policy, but I suppose since it does not really influence your work you are quite satisfied with it, as it means less bureaucracy?

Uh-huh, good conclusion, yes!

18. Then back to the data infrastructures, like the 4TU data repository we talked about. Have you ever used them to download data? So not to upload your own data, but to download someone else’s data.

No, never.

19. And how about similar infrastructures?

No.

20. Is there a reason? Did you just not need it, or did you not know of the possibility?

It is in fact the same thing. A lot of qualitative data are very contextual. If I find research that is the same qualitative kind of data about the same kind of topic, I can contact that person; usually I know the person. I can get that data. So I haven’t used repositories. I’m just now thinking loudly what would be the added value of that. I think it would cost me a lot of time to find out whether the data would be of any value. There is a lot of categorisation going on of course: what is the link with the article, what is the summary of the data, bla bla bla bla, and I cannot tell before that if it is worthwhile looking deeper. Only when I read it, then I can see “okay, this is nice to me. Or not”. So that is a lot of overhead for me.

There is also another thing. What I do, and also my colleagues, we are really keen to look at niches, something new. So if I would research something that’s been researched before, however I want to sharpen it, test it, then I would really rely on other people’s data of course. But if I find a niche, then I don’t really rely on that, I don’t have to.

21. Because then you suppose that there is no one else that already has data on the topic, I suppose?

Yeah. For example, the topic of [...]; there is none still. So I’m just one to know how [...]. I cannot rely on data from other people for that, so that’s my niche.

22. I understand. Would you then say that maybe… Your data is very contextual I understand, because, well you’re looking at the Dutch [...], so that publishing that for researchers in Germany or France would then not make sense because they have their own [...] and…

Well, if I would do an international comparison, then it would really be relevant, yeah. There’s no principal reason for not using the repository, but if I would do such a comparison or if I would get a grant for a European project on the matter, yes, then it’s very interesting to look at that and I would look at OECD for example.

23. Also another reflective question: would you be more inclined to use research data infrastructures (for either up- or downloading data) if they were incorporated into other digital services you use, like Google Scholar, Scopus, or SURFDrive?

Oh yeah, yeah, yeah, yeah. The reflection is: “Are you a lazy guy?”. If everything is concentrated into one interface, for example Google Scholar, I would definitely take a look because it would reduce the overhead. But the same issue remains: a lot of my data is contextual. If I use Google Scholar for relevant articles, there is never such a hit like “Wow, this is exactly what I want to know”. It’s always a bit like it, you know? So then I wouldn’t delve so deep into their data.

24. I see. Because, again, it’s not worth the effort to get into their data if you think up front that you’re not going to use it, right?

Yes.

25. Now more onto the institutional arrangements: have you ever made use of institutional support or resources for data sharing? This support may be for example data stewards, data champions, or the head of your department helping you.

No, the only thing is, not really the data but more publication itself: PURE. I’ve attended a workshop on PURE, that’s what I did. In fact, on the management level, data is not really discussed. Data management is sometimes mentioned but not really discussed. I talked to the data steward in his first weeks, I just talked him a bit around in the faculty, who should he talk to et cetera et cetera, but that was not really detailed on the policies because they were yet to be developed.

26. And you haven’t spoken with him since, since he was introduced?

Well, I’ve seen a couple of presentations from him.

27. Are you actually familiar with the possibility to go to him for help, or to the data champions?

Yes, if I would need help, I would go to him. I know that because of his presentations and because I met him.

28. Are you aware that the university library offers two funds for data management?

I don’t know it, but if I would look for funds for my data management, I would look at the library. I think I would go there also via the data steward, then he would tell me.

29. Yes, that’s exactly the point of him! Also, another question which is rather straight-forward: have you ever used the DMP/RDM support pages on the TU Delft website?

No.

30. Alright. Well, it sounds like a basic question, but really it is insightful: if they are up there but not being used, why are they up there?

Well, again, maybe it adds more value to other researchers, so don’t worry about it!

31. Then back to the DMPs. You said you’ve never actually filled one in, right?

Nope.

32. Alright. Did you ever do research with another researcher where he or she had to fill one in? Or just none for your research at all?

Um, I’m just thinking loudly… [...] The final European research I had a coordinating role in was three years ago, and there was no… Well, I wouldn’t say it’s a plan. It might be that I had to check a box or something that I would take care of my data or whatever, but it was nothing substantial.

33. Alright. Well, just to inform you, mostly if you go to the HREC committee then you also have to fill in a DMP, so that’s what the form is.

Well, I know that for MSc students; I also direct them to them. If they have interviews with sensitive content, then I direct them to them. So I know about that, but I never made use of it for my own research, it never was necessary.

34. You said that you were not really using research data infrastructures to upload your data. If you would upload it, where would you go? Do you know of any domain-specific ones, or would you go to the 4TU repository?

I actually don’t know. I would just turn to the library for advice, that’s what I would do.

35. And then go to the library for advice, or go to the data steward? Not really testing you, but just going over the alternatives!

I would go to the library because the data steward I see as data management, GDPR, you know? And library more data infrastructures. But it might be that the data steward is also doing the data infrastructure.

36. Well, both should work, so…

Usually, If I would do something like that, I would just go to my support, the department manager. That’s on the department level, and she would say “Oh, turn to the data steward, or turn to the library”.

37. So, she is just for general research-related inquiries, or is it mostly on publishing?

No, she is management support on a department level. She knows a lot of administrative issues and policies, so she can direct me to the person I need for that. Because otherwise, the line manager would do that, and that person is busy enough. She also sends all kinds of protocols and policies to my e-mail, in which I have a map called “Forms” and I put everything in there in case I need it at some time.

38. Is she also the person that informed you of the RDM policy that is in place, or would that be the data steward?

I think both. If there is a formal policy on data management, so the DMP is finished and then it goes into the line, I get that via the department manager because she then acts for the management of the faculty, so it comes in at hers and she sends it around. However, because I’m also involved in all these meetings, and then the data steward shows up and he’s doing presentations, and I save the presentation and I remember, so that’s why I know to find our data steward as well.

39. Alright. Have you ever had a funder requirement on data management?

No.

40. Can I ask which funders you usually have? How many different funders do you usually have; is it always the same one funding you, like the NWO?

No, in fact, it’s a lot of funders, usually governments for consultant-like research, or sometimes the EU as a funder. That’s mostly it. I don’t think via NWO, because I just started to ask for funding like 3 years ago for the first time via NWO. I’m in the process of trying to get funding as well, so I think I might encounter these issues.

41. Yeah, it’s an ongoing development, you keep on hearing more and more funders saying “Oh well, maybe we should get in on this open science move", and they then require researchers to publish their data.

Well, there is a problem because it costs money, and there is still a limited amount of outlets that have good open access policies. So sometimes you like to choose a journal and not choose for open access; there is not total freedom to publish open access. It’s actually quite complicated: there is the green version and the golden version, and all journals have different policies there. Also, you have to take into account that you should have impact, so that means that people refer to your research. It is not that all open access journals have a large impact, so there’s multiple criteria for choice for a journal or outlet, of which open access is one. If they force you to do it, then that goes at the cost of other considerations, that would be a problem.

42. Now you’ve mentioned metrics and impact being a factor; would you then consider publishing your data separately from the paper, if it meant more citations for not only your article corresponding to it, but also for the dataset itself?

Nice question. Yeah, I don’t mind publishing my data. That’s a lot of extra work but I think it’s meaningful work because I can inspire people with my data, so no problem with that. However, there is one consideration there, but I think it’s out of scope. If I see how people refer to my articles, then I almost need a drink after that! Because they are not taking the essence of what I’ve written and sometimes they refer to something I haven’t written at all. I think that would be the same if my data would be published; then they would also go on the road with my data. So I’m not that idealistic on open data. I would open my data for others to verify it, but for really building on someone else’s data, that to me seems very idealistic.

43. Well, it does work both ways maybe, that if they read your paper and then think they’ve gotten the right idea but are not sure, then they can check with the data behind it.

That’s true.

44. But that does mean extra effort for the both of you.

And well, it depends on the motive of the other publisher or writer: why does this person want to access my data? Is it for legitimization of what he or she already wanted to write, or is it because he or she wants to learn from my data? There is a thin line between these two.

45. I have one more open reflective question: you said that you cannot really find data for the topics you write about because it is contemporary and that kind of stuff. If there were like a Google Scholar but for datasets specific, would you then be more inclined to look for open data?

Yeah, I think so yeah. But because I doubt whether I would find a lot… It does reduce the transaction costs, the overhead for looking for it. However, I still have doubts whether this data, contextual as they are, would really fuel my ideas because I am looking for niches. Then, okay, because the transaction costs are lower I would take a look at it, but I… I don’t know. It might be that if we would talk to each other in about 5 years then I’m very enthusiastic about it, but I don’t think so.

Transcripts researchers: I7

***Interview date: 17-06-2020***

[...]

3. How much of your job right now is giving lectures and how much is doing research on the side?

It’s supposed to be 50/50, more or less, but I guess it’s more like 60% education and 40% research.

4. Is it always that way or just because of covid-19?

I think it’s always a bit imbalanced with me, because I teach quite a lot big courses, so it’s quite a lot. But I like teaching, I don’t mind. There are many people who would try to keep the teaching load as low as possible, but I like it.

5. Then onto the ‘real’ content of the interview, so to say. Are you familiar with the topic of research data management (RDM)?

Somewhat, I would say. I know there have been many developments to organize this more explicitly, but I’m not familiar exactly with the term research data management, although I can have an intuitive notion of the concept.

6. And you said earlier [in previous correspondence] you were familiar with RDM since 2020. How did you get familiar with it? Was that because of the policy, or did you hear about it from someone else?

I’ve written a DMP one time, so I assume that research data management is more or less the organisation behind those kind of plans. That’s kind of why I responded in such a way.

7. Research data management is more or less the practice of managing your research data, so that means storing it at a location where you can retrieve it easily, storing it in a good format so that others may also use it if they get the dataset, and also keeping track of the data and documenting it so you can publish it after your research is done. So that is what is meant by it. And data management plans are also tools to get people informed about that by asking questions like “how long will you keep the data” it should make you aware of personal data involved and that aspect of data management.

Alright.

8. On a scale from 1 to 10, where 1 means unexperienced and 10 means extremely experienced, what is your experience with RDM?

A 3. Like I said, I only once wrote a plan. I am also not aware of the bigger visions about open data and sharing. I mean, I know this is an ongoing process, but what this whole movement is trying to achieve is still a bit vague to me. I wrote such a data management plan once and I think it’s good that you keep a track record of the data, that you share the data with others. But there’s also for me a bit of a trade-off, because it is also quite a hassle to get everything in order. But I think it’s a good development.

9. Yeah, it’s always extra time and effort you have to invest into data.

Yeah.

10. And you said [in previous correspondence]you published your research data openly three times. Could you elaborate a bit on that, like when it was and what kind of research?

Normally, for my own research I actually mainly use data from existing sources. Collecting data by creating surveys is not something I typically do anymore. I mainly use existing data. So for me, uploading datasets is also not a common practice, so to say. But the times I was involved in uploading the data, was actually because that was required from the ERC grant. So the professor of our department received an ERC grant two years ago and for that grant it’s mandatory that all data are uploaded to a repository. And so all the PhD-students who work on that project and who I supervise are also required to upload any data they collect or make it valuable in a repository. So that’s why I was involved three times with making those data public.

11. What would you then say was your role in making the data public? Was it mostly because you were supervising those students or did you have an active role in making it available as well?

The former, I helped the students with… I think one dataset we uploaded to the 4TU website and the other was part of the LISS panel, that’s already an online repository. So for that we didn’t have to do much additional work.

12. Have you ever consulted DMP/RDM support pages, like guides published on the TU Delft website?

No.

13. Would you say the reason is that they were hidden, or simply that you didn’t need them, just because you didn’t feel the need for it?

Didn’t feel the need for it. Also, nowadays, when you write proposals, there is always a paragraph on how the data will be stored and how it will be disseminated. So in those cases, whenever I need information, I just contact our data steward directly. That’s for me the easiest way to get answers.

14. Well, that is really the goal of having data stewards in place as far as I have understood, so that is good to hear!And the next question, which is more about the RDM policies. You said that you know there is a website on RDM policies and that there is a data steward in place. Have you read the policy yourself or heard about it from the department supervisor or…?

Yes, I've seen emails on this, but I must confess, I haven't read them entirely. I know that, again, like I said before, I know that this development is ongoing and also support it, but I haven't been very active in it.

15. All right, and what does the policy imply for you? You said that it has led you to publish data openly; was that because now PhD-students are required to publish their data, so that that led you to publishing data?

So, yeah, this whole transition to open data has indeed led us within that ERC project to make all data publicly available. So that's the implication of the requirement, right?

16. So that was mostly the requirement by the funder then, and not really the requirement of the TU policy?

No, it was from the ERC council. They require the data to be... They also require that you write the data management plan and they require the data to be made publicly available.

17. On a scale of 1 to 10, where 10 is very satisfied and 1 is very dissatisfied, how satisfied are you with the contents of the data management policy? For example, are the demands reasonably executable, or is it too cumbersome? And why?

6, I think it is good but also much bureaucracy.

18. Is that mostly because it's just another form you have to fill and just another requirement you have to fit?

Yeah, I guess, yeah. It's not that much work in the end, but it's just... Often these forms are not very clear. So yeah, you end up just trying something and then you get it back, "OK, this is not correct. This has to be done". It's just a hassle. And yeah, any extra form is just a hassle.

19. It makes sense, it's what I'm hearing a lot. This is also what the policy makers say, they know that researchers think it's actual work, which it is, and they don't have a lot of time for it, so they are trying to make it more approachable.

20. And then onto the research data infrastructures. I think you've read my definition [in the questionnaire sent beforehand], it's mostly just repositories for uploading and downloading data, like the 4TU platform you mentioned. So you already said you were familiar with DANS Easy, LISS panel and MPN Panel. Could you elaborate a bit on those panels? Are they domain-specific?

So the LISS panel, it's a large panel, it's organized by Tilburg University and they administer regular surveys in those panels and those data are all freely accessible. And what you can also do is write a proposal to conduct a survey in that panel. And the nice thing about it is that then all the data which you gather regarding specific subjects can be linked to any existing data which is already in the panel. That was also the idea behind that panel. And they've set up their own infrastructure, it's LISSdata.nl from where all the data is accessible, all the variables and surveys are explained. So I think it's a nice initiative. Yeah. So it's not... It's-domain specific in the sense that it covers basically the social sciences. It's not engineering. I think the 4TU Repository is more focused also on image data for example; these are only mainly survey data.

And then the MPN, I should mention the whole name, it's the Mobility Panel Netherlands. It's a specific panel which focuses on travel behaviour and they also have regular surveys each year and they also make their data public. That's also data I use for my research, because I am active in that domain.

22. All right, and the MPN panel, it's only focused on the Netherlands I suppose?

Yeah, it's a panel of Dutch households, 4000 households. So it's quite a large panel.

23. You also said you've used DANS Easy before, could you elaborate a bit on what you used it for?

[...] But for those datasets for example. I did an analysis on [...], I used a dataset from DANS. It's not that I use it often. A couple of times.

24. I'm not sure if this is a question you can answer easily, but how many do you usually use in one research topic or paper? On how many datasets is that usually based, just one, or do you sometimes combine some?

Usually one, but for example, I had a master student who looked at [...] and then he integrated all the [...] datasets from all years. It depends on the purpose of the study. But also for the MPN panel and the LISS panel, the nice thing is that you can combine datasets from different surveys and look at relationships which you couldn't have examined if it were separate data collection efforts.

25. And about what percentage of the datasets you use, are you reusing?

So, you mean, any datasets I didn't collect myself and how often I do that?

26. Yeah, just a generic estimate.

Yeah, I think mostly I use data that already exists. I think nowadays 90 percent is based on the reviews of existing data and 10 percent in which I collect data myself.

27. All right. Well, good to hear, because it means a lot of less effort you have to make.

28. You said that DANS works fine, I suppose the other platforms also work well. You already mentioned with LISS you can combine different surveys, so that also scores high on useability I suppose. Is there anything else you'd like to add to that?

They all work quite well. I think the search engines within those panels also work well. You don't have to use Google to search something, you can just really use the search engine within the panel.

29. And then onto the same questions, but for uploading your data, so depositing it. Which infrastructure for uploading data do you know, besides the 4TU and the DANS Easy, LISS panel, if you can upload there. Are there any domain specific ones?

You cannot upload your own data on LISS panel. It's only, they do the data collection among their panel and that's the data which gets published there.

30. So for LISS you can only propose a research proposal?

Yeah, and then they administer the survey and the data becomes automatically available to other researchers. I'm not familiar with any other repositories other than the 4TU and DANS. I think for the ERC Programme we ended up going for the 4TU, but it's been a while since I was involved in that because the data collection already happened for those PhD students.

31. And do you remember why you chose 4TU?

Because it was part of the TU Delft, I guess. Actually the data we collected was also survey data. So probably it would fit better with the DANS infrastructure, but we chose 4TU because we would get support from the library as well.

32. You just said the 4TU offers you support, the library offers you support for uploading your data there. What kind of support did you receive, if you still remember? Like, was it in shaping your data, or did they tell you "well, you have to add this metadata or change this" or...?

Yeah, I'd have to look it up, but I think indeed they sent us a format for the metadata. Other than that, support... Yeah, also they helped us create a DOI link so that you can also receive citations to your dataset. Yeah, what else... Not sure. Because the 4TU repository does not have a social science orientation, I remember that the metadata we had to provide was not that in-depth. So, for example, for the LISS panel, you have really well or explicit metadata guidelines on how the variables are defined and what value labels were used, et cetera. I remember for the 4TU there was not such specific guidelines because the types of data is much more open.

33. Then onto the next question. We already talked a bit about it, but what led you to not uploading your data? Is that mostly the belief that because you reuse data, it's not really that useful or 'new' to upload your own analysis?

So for me, the main reason is that for my own research, I use mostly the LISS panel and the MPN panel datasets, which are already publicly available. So then it does not make sense to publish that data. For now, for example, we collected some data about [...] in the Netherlands, the reason I would not upload that data is probably because of the hassle. Because it takes time and additional effort. I guess that would be mainly it. Although it would be good, I acknowledge that, to upload the datasets once we have written the publication about it, we can just upload it and then other researchers can reuse it. Yeah.

34. Now there are some incentives in place to encourage data sharing. One of them is a fund that the library offers, so that after you have your publication finished, you can get some additional money to make your data shareable.

Oh!

35. Would you think that that fund would then motivate you to share your data?

It would help, yeah.

36. Yeah, that's funny because somebody at the library also said to me "well, we have this fund in place, but not a lot of people know about it". So that's also just proven again!

I wasn't aware and I guess I'm not sure what it is, but yeah, I can also use some small money for hiring student assistants.

37. And another incentive. If you could get publications for the dataset, like you said you got a DOI for the previous one you uploaded to 4TU, would that stimulate or encourage your data sharing?

Yeah, maybe eventually. Because citations are important, for any academic career it's really important that your work gets cited. And so if it adds to the number of citations, that would clearly be an incentive. But I guess it's also a bit of a chicken-and-egg problem because it's not common nowadays or still not to cite a dataset. Those datasets typically do not receive many citations, so they don't add a lot to your total citations. But if it becomes common practice to indeed publish the data and that researchers indeed start using each other's datasets and citing those, then of course, yes, it could lift off and it could become more of an incentive. But in my experience, researchers, they tend to use their own data, at least if they collected the data themselves, they tend to analyse it themselves and they're not really keen on sharing.

38. Alright. And when you uploaded data to 4TU, how content were you with the usability of the platform?

I remember that was okay, I would rate it an eight.

39. All right. And onto then next question, which is on the institutional arrangements, so mostly the support for researchers that the TU Delft offers. We already talked a bit about... you mentioned that you've spoken to the data steward. How often have you spoken to him? Was that only for the DMP or for other matters as well?

I think it's a handful of questions. I think three times or something, I approached him. So for the data management plan we had some questions, and then for the... I wrote a proposal for research funding and then I had some questions for him. So it's very irregular.

40. And have you ever heard about the other support options I listed? There are data champions in place, and also the library offers not only funds but also training.

No, I wasn't aware. Maybe after this interview I'll become more involved. But, yeah... Take now this example of this [...] dataset. I could publish it, but yeah, honestly, I don't think any other researcher would want to, or will reuse that data.

41. Also maybe because of the temporary nature of this [dataset subject], so to say?

Yeah, and it's always very difficult, if you really want to reuse a dataset, everything should be very clear and explicit. So you should on a metalevel have a very big description of the dataset, how questions were formulated, what order, were they randomised or not. It takes a lot of additional effort to create that metadata. And if you collected the data yourself, you're familiar with the dataset and you're much more likely to use it. I actually was surprised because this MPN panel we talked about earlier was really developed to help researchers, to take away the data collection efforts and try to actually stimulate research in mobility. And I really like these datasets, but it takes some time to get familiar with the data and how things are organized. That's also an effort, and I'm actually quite surprised that not many people use these data because it's a very rich source and it's really a true panel, so people participated in multiple years. It's quite unique that we have that here. But still, people do not use it for some reason. They want to focus on their own data.

42. Yeah, maybe like you said, that you have to not only be aware of what others have done in how they gathered the data, but you also have to trust them in that they did it all correctly, I suppose. Would that also be a barrier for you to use other data sources, let's say there was a third research panel besides the LISS and MPN, would you then be more reluctant to use it if you were not familiar with how they gathered the data?

Yeah. Trust, of course, also plays a role. But it's also, what can you get out of the data that the original authors did not get out of it? It seems like the added value of the research is not that significant, because probably the authors who collected the data already analysed it thoroughly. You really have to have a specific question, which was not addressed by the original collectors of the survey data. And survey data typically are collected for very specific purposes, of course. I can imagine if you have other data sources, like images or text data or whatever, then you can just mine the data and look for patterns, which is different than social science data, I would say.

43. So mostly because social science data has a specific goal for which it was gathered and then reusing it would be very much different from the original goal?

Yeah.

44. Alright. I'd like to take a step back and go back to the policy that is in place at TU Delft. I'm not quite sure if you said you heard about it or actually read it?

The policy which is on the website?

45. Yeah, let me get a link for you. [Link sent]

It's a nice website.

46. You said you saw it come across in an email, right?

Yeah.

47. Did you hear about it from like... I'm not quite sure how your research is organized. Do you have a department head or manager or something like that?

Yeah, I think the secretary... But this figure [in the policy on his screen]... come on! I'll share my screen. I mean... \**sighs\**. This makes me like not wanting to get involved in this stuff, right?

Diagram

Description automatically generated



48. It looks complicated, I have to agree.

A form here, approval there. Policy officer... Also, what does DPIA mean? Okay, DMP I... That's the ethics... I'm not sure, it becomes too complicated!

49. Yeah, I understand!

Yeah, well, I think it's good, but I mean... If I would have to do everything right, all bachelor students who I'm supervising should also go through this process. So bachelor students, they typically develop their own survey, administer it and collect some data. It would be undoable to go through this whole process. I'm supervising sometimes 10, 15 students per year. It's undoable. I think it's also good that there's attention for these issues. But then at least for bachelor students, make something very streamlined. Just, you go to a website, you tick some boxes that it's no risk and then it's OK. But now it's with forms, getting formal approval, et cetera et cetera. But yeah, I'm just also a bit nagging, it's not that big of a problem.

50. And what do you think the data sharing practices and data management practices in your department are like? Do you think that your colleagues have similar attitudes because of the extra work and if they also work with surveys and panels, do you think they also do not have a lot of added value for sharing their data?

Yeah, from what I hear around me, that's still the case. All people who use these same data I use, they tick the formal boxes whenever it's necessary, but really trying to make use of it as an opportunity in itself to get additional citations or additional funds, that's really still not done, I would say.

51. All right. And is the department head or department manager encouraging it?

On paper, he would. He is not informally very active in stimulating this.

52. I'm just asking to gain some insight into the practical workings of all the intentions that the people at the faculty and at the library have. So it's not to tell you how you should do your job whatsoever, that's not my intention at all, but it is interesting to hear the practical implications of what they intend with this policy!

Yeah. And also, if you want to stimulate this kind of behaviour, you need to incentivize it, of course. But it works best if the system does it automatically. So indeed if you could get citations to a dataset that would be helpful. But in the end, yeah, like I said, for the types of data we collect, they're gathered for very specific purposes and it's not really done in the field also to reuse other's datasets because they have already been analysed. So I imagine that the whole idea of open data, sharing, learning from each other, collaborating, it's a really nice vision, but it doesn't work like that in practice. Maybe for other data sources, it works better like that. So I know at Civil Engineering, they have for example imagery data from [...]. Of course, those data you can just publish online, it is very data-heavy and somebody can develop a machine learning algorithm to use those data, and then there's collaboration, right? But yeah, survey data, they're used for specific purposes. You collect the dataset, you write a paper about it. And then what's the use of that dataset?

53. Interesting to hear. So it's mostly because of the domain you would say then that you're not sharing data?

Yeah.

*[Non-related informal chat omitted]*

*The following is the last part of the informal chat, which originally took place in Dutch and was translated into English for this transcript.*

54. Also the library acknowledges “yeah, we do not expect everyone to know and everything is so new and it might not apply to everybody”, so I don’t think it surprises them.

Look, also, in a bureaucratic organization, some e-mails come by in which these kind of policy things are explained. I have to say, not many people really take the time to dive deep into that. You get so many e-mails a day, also from all these students!

55. Yeah, someone else told me “yeah, no researcher reads all the policies, because this does not only apply to open science policies, but also to who knows what. Maybe it’d be better to let department heads tell those stories”, so that is why I asked you about it. Just looking from a more high-level perspective: who can we contact to stimulate this?

But look, it works best if eventually some kind… Look, you can of course have a stimulator, some sort of external incentive. But eventually the system itself should reward it, that you indeed share that data, so the added value should be clear. So you can use this external incentive to give it a twist, but then it must be adopted and eventually self-sustaining. And yes, again, I think it would be good if that would happen. If you could also receive many citations with online datasets. But that is not the case yet. On another note, we did not talk about that, but there is right now this whole transition of course to much more data-intensive methods, like machine learning and artificial intelligence et cetera. So I think that that can also really be more of an incentive for this kind of open data initiatives than, for example, the possibility to hire a student assistant.