Transcripts policy makers / support staff: I1

***Interview date: 26-5-2020***

1. I did some research on your background; [...]

At [...], my role was a data steward-like kind of job, meaning I was more in contact with researchers, providing support to them, we implemented the policy of the university regarding RDM, helped them with DMPs, and so on, and established also the data archive. Then, in [...], I started at TU Delft as [...]. Now, my role is more on a higher level. I am not in daily contact with researchers. I do meet them through the services of the data archive. And then for the library I do coordination of certain services and also work on a policy that will come up. More a coordination role.

3. What was your reason to start as a data steward, more or less, in [...]?

[...] Also the background as a researcher helps a lot when you need to talk to researchers about their data, because you are familiar with research workflows and the problems. Probably you know more how not to manage data as a researcher, and now you have to tell them how to do it correctly!

4. What motivated you to take on the more high-level policy role in Delft?

[...] One important thing for me is that TU Delft is really one of the frontrunners in the field, so it was kind of going to a major league club on the field. That was also very motivating.

5. That’s funny, because I did a little bit of research into RDM practices at other universities in the Netherlands. I was surprised because there are only 3 other universities I believe that are somewhat at the same level as Delft is right now.

I think many universities are looking at TU Delft as a model how to implement the services. Now there are also other universities that want to have data stewards and a defined role of data stewards. In many universities, people from the library, librarians are taking the RDM support by themselves, not necessarily have a research background, but now more universities are looking for those people with a research background that understand research workflows, provide advice and so on.

6. I found that RDM support at TU Delft at least includes data stewards, trainings offered by the library, that kind of stuff. Could you elaborate a bit more on the types of support at TU Delft that are offered?

As you said, 3 years ago, the data stewardship program started, where the aim was to have one data steward per faculty. The data steward should have a researcher background so that he could provide advice to researchers and understand the research workflows. This is since 3 years. Before, there were also some generic… This has a long time in TU Delft, but I think since the data stewards are there it kind of formally started to develop more and more.

We also have the 4TU Research Data Archive, which is also administrated by the library, where researchers from everywhere actually, not only TU Delft but all the universities in the Netherlands and also other countries can deposit research data. There we have a big team in the back-office that will receive the dataset and curate them a little bit, so improve them or add suggestions to them such as how to make them more reusable, which licenses to use, to tell other people how to use the dataset. So there is lots of support there in the back-office of the 4TU. We have a front-office of the 4TU Archive. There are 3 universities who are members of the 4TU archive, which is TU Delft, TU Eindhoven and Twente University. There is one representative for the front-office in each university.

Then, the university also provides a tool called DMPOnline, which is a tool that allows researchers to easily draft their DMPs. That is also hosted by the library. The advice on the DMP is done by the data stewards, but the library provides the infrastructure so that there is a kind of registry of all the DMPs and data stewards can provide feedback on them and so on.

Also as I said, last year a vision for research data management training was created (*see* [*https://openworking.wordpress.com/2019/10/25/vision-for-research-data-management-training-at-tu-delft/*](https://openworking.wordpress.com/2019/10/25/vision-for-research-data-management-training-at-tu-delft/)) where the university has big ambitions; one is to have a centrally organised RDM course, an introduction course for PhD students.

We are also running the Carpentry workshops at TU Delft. Software carpentries deliver regularly, this year a bit less because of the online setup, but also data carpentries. The carpentries are very hands-on workshops, where you learn a little bit of programming, a little bit of data analysis, and best practices on data management and code management. We started that because the 4TU archive is a member of the Carpentry, so we can provide the workshops but we can also encourage researchers who have some knowledge in programming, for example, to become certified instructors of the Carpentry. They could go and teach wherever they want for the carpentries.

7. If we go back to the 4TU Data Centre and support you talked about, then I assume that if you want to get to the support people at the back-office, you have to go through the front-office first.

Yeah. So, what the front-office is doing, for example: researchers don’t know how to start it, so the front-office can guide them a little bit, also represent the archive at conferences for example, where people can get to know the archive. The archive also offers what is called the data funds: you have the data refinement funds and data paper funds. The front-office takes care of those, researchers can apply if they need help with improving the quality of the dataset that they want to make openly available through the archive, so they can hire a student assistant for example to help them to document the data properly or to anonymize the data properly before they make it available and so on. So the front-office is there for any questions, but when it comes to uploading datasets and publishing the datasets, the front-office will always send them to the back-office for their questions.

8. Alright. And what is the cooperation like between the university library and the support there – the front- and back-office – and the data stewards at the faculties?

The data stewards are coordinated by the library. The data stewards are sitting in the faculties, they are hired by the faculties, and they respond to the faculty secretaries, so their official boss is at the faculty. With the faculty secretary they coordinate which activities to run at the faculty. But the whole group of data stewards is coordinated by the library, by a data steward coordinator. Basically, what she does is to make sure that data stewards are always talking to each other to find maybe common challenges at different faculties. Some challenges that faculties have, if they are repeated in many faculties, there is probably something that could be done at the central level. That is the role of the coordinator: to bring in all the feedback of the data stewards, and then say “Okay, we need to talk to IT to solve this problem for all the faculties” or “Okay, we need to talk with the legal services to solve this for all the faculties”. So she does a lot of that, the contact between the data stewards and other administration offices that are relevant for establishing the services, the policies, the infrastructure.

9. So basically if researchers want to deposit their data somewhere, and they’re not sure on how they should do it or in general want help with that, they should be at the library for help?

Yes. For data publication, well, they should contact the 4TU e-mail address and then it will depend on the kind of question if the back-office takes it or the front-office takes it. But researchers could also contact their data steward. I mean, the first contact point can always be the data steward: the data steward will know who to contact or who the researcher should be talking to. For the researchers at the faculty, the easiest is always to have the data stewards as the first contact point: then they can contact people in the library, in IT, somewhere else. But of course, if it’s only about data publication, they could also go directly to the archive e-mail address. I think it’s good to have different entry points for the different questions.

10. Yeah, maybe contact with data stewards is a bit more low level for them, so that might lower the barrier.

Yeah, exactly. They have somebody sitting there at the same faculty and with some knowledge on the research done there. They know the administration people at the faculty, they know the people at the central administration, they know the people at the library. So the data stewards are there really to guide researchers to get the best information about data management.

11. How well do you think the availability of the support, so on one side by data stewards and on the other side by the library with training and stuff, is known with researchers? Do they know that the facilities are in place?

I think they know more and more about the data stewards. I don’t know if you have seen, but the data stewards have run already twice a survey at the TU Delft about data management. I think, according to those surveys, researchers are getting more familiar with that the data stewards are there. For some researchers, I think the library still does not mean immediately ‘data’. For them, maybe the library is still books, journals, articles, and stuff like that. So they relate it less to data, I think. I think there are also many researchers that don’t know about the 4TU, that there is an archive for them. I think the data stewards are doing a great job on making visible all the other services that are there. Whenever they have a presentation with students or researchers, they mention the archive, they mention the DMPOnline tool, and so on. They have been very crucial for researchers to know better about the services there, which is fine: that was the idea. Data stewards are kind of the ambassadors. But I think there are still a lot of researchers that don’t know about all the services offered around data management. I think there is still some communication work to be done.

12. How do you think we could do that? Is that mostly just waiting a bit, because there is also a lot of word-of-mouth spreading of the possibilities, or do you think we should take a more active role, in sending out e.g. newsletters?

There is also something that the data stewards started, which is called the data champions community. The thing with newsletters is that you receive many and may not read them all. I think it’s always good if researchers can see a face associated to something. I think the data stewards definitely, they will continue improving this. The data champions community I think is also doing a great job on that, because they know about the services and the support that is there. I think maybe researchers also trust it more if a peer tells them “hey, you know there is an archive and I deposit my data there” or “yeah, I used DMPOnline and it was a great tool”. I think, for me, what would help a lot is to continue building different communities. We have the data stewards, we have the data champions, we might want to, maybe around data publication we could create communities of practice, I don’t know. Maybe if we have a contact… We are trying to do that, we have contacts for example with researchers in wind energy. We tell them “Well, you know, we can create a collection of your type of data, let’s try to engage with other researchers in that community related to that discipline.” I don’t know. For me, I think the work still needs to continue on this word-of-mouth contact or peer-to-peer contact about the services.

13. With the peer-to-peer aspect in mind: if we say that the support of data steward and the library is more “formal”, so to say, and then being helped by your peers or colleagues or department head or something would be called “informal” support; which type do you think researchers would prefer?

I think, as I said, researchers appreciate the informal. The thing is that since the data stewards also have a researcher background, they are not per se peers but somehow they can talk to researchers in a way that is understandable. I think many researchers see the usefulness of the data stewards and they still feel that they are close enough to researchers to talk to them, kind of a peer relation. But I think definitely researchers prefer to talk to peers than to administration. I think there is still a barrier there. It could be whenever you tell researchers “oh, you have to talk to the lawyers”, it is always a little bit of a pain for them because they know it’s difficult to communicate with them so that they understand the problems that they have. In that sense, I think the data stewards are a great buffer. So yeah, it’s difficult to define the data stewards because you could also say it’s kind of peer-to-peer but more formal support. The informal support is for these recommendations, like “hey, you know, I tried this tool provided by the library, it’s great”; then others would use it also. To solve more major problems, more official problems, I think data stewards are still considered kind of peers close enough to researchers. I think researchers also see a lot of value.

14. You just said that trainings at the library were not as popular, if I remember correctly?

No, it’s that we don’t have at the moment a central RDM course. We were not really providing courses for researchers about data management. Now we are planning one, and it will start in September. We started with Software Carpentries at the end of 2018. They are very popular! We have a huge waiting list for those kind of workshops. The thing is that there was no clear vision of what we should be providing and what not. Now it’s clear that because of the data policies we have in place and the requirements that are there for PhD-students, we have to have a central RDM course. Besides, whatever training the data stewards can provide. Then, at a next level, on more practical skills: how to really analyse and document data analysis properly; how to publish your data, what do you need for that, in a more practical way. That is a great addition to the carpentries. The vision I talked about earlier has different scales. At the moment, what the library offers is more about information literacy, which has to do more with how you manage your literature references, how you look for publications relevant to your research, for example if you are doing your PhD-thesis and stuff like that, and not much on the data part. They are more generic courses. There is also the MOOC on open science, which is very generic, it also talks about data management but in a broader context. What we try to do is to define the different levels of knowledge that researchers need. Then you have the very generic ones for somebody who just wants to know a little bit about it, then you have a more basic course that all PhD-students should know or all researchers should have. And then you come to the more practical part that it’s more tailored to what kind of research you do, because different tools will be useful for different disciplines. Not all tools are useful for everybody. So we don’t know how popular this central course will be. We hope it will be very popular but we don’t know yet!

15. You talked a bit about the data funds that the library has available. What is the demand for that; do researchers know about the possibility to apply for the grant?

I think they are getting a little bit more popular, but I think they are unknown for many researchers. As I said, the data stewards always try to talk about 4TU and the different services; I do the same whenever I give a talk and so on. Last year I think we had maybe 4 applications for the data refinement fund, which is more popular than the data paper fund. With the data refinement fund, you can ask for up to 5.000 euros to hire somebody to help you make your dataset better before you publish it. The data paper fund is to pay the fee of a data journal: there are journals that specialize to publish datasets, where you don’t need the same structure as a normal scientific publication: you don’t need to have conclusions, a hypothesis and so on. You just need to have a complete description of the dataset. Those are less popular, but I think it also has to do with the knowledge that researchers have about it. This year, we had more applications for the data refinement fund. I think we have 7 at the moment and we have a maximum of 10 per year, the budget is for 50.000 euros. But, I have to say, they are also, or have been more popular, in the faculty of Architecture than in other faculties! I think it has to do with what we were talking about: this peer-to-peer communication. One of them applied and they say “hey, I applied to this fund and I could do this and this and this with the data” and then the next one applies and the next one applies. So I think it’s also about this peer-to-peer communication. Because of the type of research they do, they do use these refinement funds a lot to actually generate the data they need to use. Because they reuse data that can be in forms like pictures, books, so they need to gather information from those and extract those to do the research they do. That is what they have also received the refinement funds for; to ask for a student to go through these books and extract and document this and this and this. And then they can do their study and publish their books and reports.

16. The people that have made use of these resources, have you gotten any feedback from them on the resources? Like, does the grant have enough money in it, or is the duration long enough?

Yeah, every time that a dataset is published, they are interviewed. You can find the showcases on the 4TU website (*see* [*https://data.4tu.nl/info/en/about/researchers-about-us/*](https://data.4tu.nl/info/en/about/researchers-about-us/)). They are asked why they applied, how they got to know, why they are important, and so on. In general, the thing is that RDM is still not very incorporated in the workflows of the researchers. They gather a lot of data, not thinking about how to properly structure them, document them, so that they can find them at the end and make a nice publication. The data refinement funds specifically have helped with that: if PhD-students finish and realise “okay, we have all this data, we would like to publish it but we are missing, for example, the documentation of the code, we are missing a structure, the right file-naming so the different versions make sense”, and all these kinds of things. It can be a lot of work if you have lots of data. If your PhD-thesis is already finished, then you don’t want to stay there to do that, but move on to the next job. Your supervisor would not do it because they don’t have the time. So then to have this money to hire somebody to do it has been very useful. But the idea would be that in the future this is not really required, but that the researchers, through training and so on, do really incorporate this habit when they are doing their research. So then at the end you have your publication, your research article, and you already have all the data well-documented, so that you can make it findable for reusability.

17. So that it becomes less of a formal restriction? Like filling in a DMP for when you want to go to the HREC; that is more an obligation. But you want to change the mindset so it becomes more voluntarily?

Yeah, I mean, RDM has a lot to do with a culture change on how we do research. When I was a researcher [...] years ago, nobody was thinking about the data. You always focused on the article. You want to have a publication, because you are evaluated by citations and the number of publications and so on. Where the data was, I don’t know, my paper is out, that is what it’s about! But the thing is that also journals are changing their policy, because there is spoken a lot of fraud cases. In some disciplines they talk about reproducibility crisis and stuff like that, like in psychology and so on. Funders, journals and the universities want to see the data available so others can reproduce what you have done or build upon what you have done. There are demands to have that. Now, the researchers are not yet used to think about data. Then they see “oh, the journal is asking for the data. Where is the data? Who has the data? What is the last version of the data that we used in the publication?” and all this kind of questions start to rise. This is still not in their mind. Training, lots of support, is needed. Also incentives I would say, are needed for changing this culture, I feel. In the meantime, the data refinement funds could help them at least to ease the pain a little bit when they realise they need to organise all the data to make it publicly available.

18. Then onto the RDM policies and the open science policies, which is also one of the incentives I suppose, to stimulate researchers to do their RDM properly: how well would you say that the data management policies of the university, and more specific the faculties - if they have one in place of course - are understood by researchers?

The framework policy, the central policy, is very generic. It only tells who is responsible for what, and what are the minimum requirements for PhD-students, and which other administration offices have a role in all this. It’s really a framework policy. Probably what the researchers can extract from that is that PhD-students need to prepare a DMP and PhD-students need to deposit data. That is the more practical requirement that the policy has. It was a decision on purpose, because workflows depend or differ a lot by the type of research they do; at different faculties there will be different workflows. Hopefully the policies that are now being drafted - not all the faculties have them ready - are a little bit more specific on what is reasonable to ask depending on the workflows. For example, in Applied Sciences, they want to have as a minimum requirement that all the data behind publications should be openly available in a repository. But then, other data that can be also used for others, or used in other kinds of publication, reports or something, might be optional because maybe it requires much more work to make that available. So hopefully that kind of indication will help the researchers a little bit to say “okay, this is my minimum, how much do I want to aim for?” Also, hopefully they have a little bit more information on specific knowledge, for example for some faculties personal data is very relevant, because you run interviews, collect personal data in many of the departments of the faculty. Hopefully, then in those faculty policies there will be more information about that: how to deal with that practically and who you can ask for help for this kind of thing. But there again, the data stewards have a very important role on how to communicate these policies with the researchers. That remains to see, hopefully they are all in place at the end of this year and then there should be monitoring on how they are implemented at the faculty.

19. Do you have an image of to what extent researchers really comply with the policy guidelines?

I would not know, I think that the data steward could answer that question better, if they are complying or not with the policies. The requirements for data publication are from this year on, so until we have a better overview of… If we are, at least if 4TU is receiving data from PhD-students, it would take a while, but they are free to publish wherever they want, they do not need necessarily use the 4TU. It’s a little bit difficult to track. I do hear data stewards saying much more that researchers are requesting feedback on DMPs. They are getting more requests for feedback, so I think, at least at the PhD level, these requirements are working. Now, that’s different to the culture change we were talking about. Filling a DMP does not really mean you will change your habits when you are doing research. That is much more difficult to measure. You say that policies might be incentives, to be honest I think policies are more like, if you think of sticks and carrots, I think it’s more the stick: they are telling you what to do, what you need to do. But there are very important incentives that for example we should be working on in the future. I know TU Delft is very much interested to work on this in the Open Science Programme about rewards and incentives. Researchers that invest the time to make their data FAIR or to publish their datasets should be rewarded for that, and for example that citation of datasets should also count together with citation of publications. So not only impact factor should be a way of measuring the productivity of a researcher – if you can measure productivity that is, but that is another question. The consideration of the work, the effort they make on improving their data management routines or habits, should be somehow rewarded by the university. It should be rewarded by for example funders, it should be rewarded by journals also, allowing citations of datasets and things like that.

20. Is there anything you have in mind that you would like to see changed in the data management policies right now?

Hm. I think, well, I don’t know. I think the framework policy of TU Delft fulfils the purpose: it has to be generic, and the requirements should be the minimum requirements for researchers to ensure that the university produces good datasets and trying to implement good practices using a DMP and so on. What I would like to see more than policies, is maybe more… We could do a better job on practical guidances. In policies we tend to talk about open science, FAIR data. With FAIR data specifically, there is a lot of room for interpretation of what does it mean, FAIR data, depending on the research you do. Sometimes that is not really understandable for researchers but you need to try to find ways to guide them, to give more information how they can make their data more findable and which data should be more findable. Not all data needs to be stored and published and kept forever, in my opinion. I think there should be guidance on what to keep, and for how long to keep it and so on. I don’t know if those guidelines should be in a policy or should be really as separate guidelines, but that is something I would like to see on the policy level. More practical guidelines for the researchers according to the type of data they work with for their research. But then you need the research communities to work with. The deeper you want to go into best practice, it gets more complicated, because not everything fits for everybody. You have to be careful there. But I think our framework policy kind of serves the purpose, actually. We have the minimum requirements and the roles and responsibilities. It remains to see what all the faculty policies will look like. I bet they are very different.

21. I suppose you are familiar with data infrastructures for uploading, browsing, finding and downloading data. Of course you have different types, like the 4TU Repository you mentioned, but also Zenodo, Dryad, Figshare… How familiar do you think researchers are with the proper infrastructure for their research? Like, in terms of only uploading data, how well do you think they can find the proper infrastructure to upload their data to?

I think that really depends also on the discipline the researchers work on. Some disciplines, they are doing this already for more than twenty years. If you think about people working with genomic data: genomic data is available there for a very long time. They also have their repositories, there are two main repositories and they are interlinked between each other. That is in the mind of anybody who works with genomic data: they know they need to deposit their sequences on any of these repositories. But if you go to the more technical, engineering disciplines, they are not used to this. They might have more difficulties to find the right infrastructure. It’s difficult to say in general. I would say it really depends on the research area, how researchers know how to do this. But I would say that in many areas still, this is something new. It is difficult for the researchers to know what they can publish, where they can publish it. Licenses is also something that a lot of researchers struggle with and we could do a better job on informing what type of license they should use. But I think it’s still confusing for the researchers. It depends on the research discipline.

22. Would you say that the main reason is because they are either not used to do it themselves, or because not a lot of people in their domain do it?

Yeah, because in some domains, this has never been done, they really keep their data for themselves. They don’t see the need to share it with others. I remember from the time that I was working in [...], I was working with engineering researchers. And they say “well, I don’t have data” and I say “of course you have data, of course you produce data”; “no, but I work with prototypes”; “yes, but your prototypes, you test them to produce data, and depending on the conditions, the modifications you do to the prototype, you produce different types of data”. The question is, that data that you collect with the prototype, is it useful for somebody else or not? Is it useful for yourself in the future to not test the same thing? Then you have to go to a much deeper conversation with them so that they think it makes sense to organise the data and publish it. And maybe in some parts, maybe digital data as we normally think of, maybe doesn’t make sense to keep it, but maybe it makes sense to keep the blueprints of the prototypes they are producing, which is also another type of data. I think the difficulties with infrastructure has to do a lot with the culture in the discipline on data.

23. Would you say that the infrastructures are easy enough to use? For example, if somebody wants to upload their data, do you think they can find their way independently?

Well, that depends on each repository, but I think 4TU, Zenodo and Figshare are quite easy to use, are quite intuitive to use for researchers. At least for 4TU, that is the feedback given in the interviews. And also because they really appreciate the back-office services that 4TU offers. When somebody in the back-office sees, I don’t know, maybe the lack of certain information, certain documentation, or maybe they think “okay, maybe this license might not be the appropriate one”, they can contact the researchers and then they can have a conversation about it. That is also very much appreciated. Dryad, I don’t know, I didn’t try it myself. But I think all infrastructures aim at making it very easy to upload data. In the Netherlands also, many people use Dataverse, the one provided by DANS, which I think is also kind of straight forward.

24. Now, if we look at these same questions on infrastructures but for downloading uses, how familiar do you think that researchers are with the possibility that they can download data for their research?

That also depends on the discipline! Because yeah, I don’t know, many researchers are used to use databases, for example from banks and other sources of information. Some are really not; they don’t reuse data, because they were not there for many years they are not used to that. I would say that maybe also researchers cannot imagine that there would be a generic repository where you can find all sorts of data. Because, I don’t know. If I think of more humanities or social sciences, they have their archives, repositories, databases, where they normally find their data. I don’t know if they would go to a generic repository like 4TU or Figshare or Zenodo to look for the data. But now more and more people are also putting their data there. Now the problem of Zenodo and Figshare is because there is no back-office, everybody can publish whatever they want. It could be very crappy data, that you need to explore to know if you really can use it or not and so on. So yeah, I think it’s the same about reuse of data; depends on the discipline and the culture.

25. What would you think the general opinion of researchers is about DMPs?

They hate them, haha! Well, I think the first reaction is “Oh no, I have to fill in another form”, or “I have to fill in another document”. They see it more as an administrative thing to do, more than a tool to reflect on things. I guess it is perceived as such because it’s part of many policy requirements: from funders, from universities, and so on. They see it as a requirement, but maybe they don’t see it as a useful tool to use. Basically, a DMP is a tool, not something you fill in for the sake of documentation. The idea should be that the questions in the DMP should help you reflect how you are working with your data. Also, in the time that I was working in [...], after I talked to researchers and went through the questions of the DMP, they were thinking “oh, yeah, I could do this differently”. If they read a question… Also depends on the template they use. I find the template of the European Commission very confusing, or not clear enough. So this is where the support is very important, to explain “okay, this question means this”, and if you can give them examples. For example: “Think about this workflow. How are you dealing with this, and is that good enough to make your data findable later on when you want to write your publications and so on?” And then they start to see the sense of doing this exercise, but because DMP is part of requirements, I think it’s very much seen as an administrative thing to do, or another paper to fill in. But if you do the exercise to reflect better with them, at the end they find it useful.

26. What is your own opinion on DMPs, specifically the set of questions the TU Delft asks?

I think I like the template of TU Delft because it’s very straightforward; it’s very simple and doesn’t have too many questions.

What I miss a little bit more in general about DMPs is the possibility to link it to different services. For example, in a DMP there are questions about the size of the data you will produce. That is very difficult to know at the beginning, but you more or less know depending on the data you normally work with, how much storage you will need. This is a useful question for the IT department to know. In the future, and we are working towards that, an ideal situation would be when a researcher says “well, for my PhD project I will need 5 terabytes of storage”, that automatically when he clicks 5 terabytes, IT receives a message and they can contact the researcher and say “okay, why do you need these 5 terabytes?” and try to make sure that there is 5 terabytes for the researcher to use in the project. Or, for example, if they say “I need to archive in 4TU 10 terabytes of data”, that then the archive has that information and somehow is prepared to contact the researcher to host that data, because we can’t provide 10 terabytes of storage at once! What I’m missing in DMPs in general is a little bit of automatization of the processing of the information. Not that you need to download the pdf and then inform IT, “look there is this and that”, or maybe IT will not know; until now, they will not know, because there is no connection between these services. This is something that we call ‘machine-actionable DMPs’. It’s something that we are working towards. Then it will really be more useful, the usefulness we see will be much more than now, that is, now a kind of exercise for researchers to reflect on how they are handling their data.

I think we have done also a good job to incorporate the workflow on personal data in DMPs, so to have a dedicated section about that and to guide researchers to contact HREC whenever it’s necessary for example, and to keep that information in the DMPs is important for GDPR compliance. I think in that sense we also have done a good job.

27. What do you think the most frequently occurring subjects are that researchers ask assistance for? Is it mostly the DMP, because that is obligatory?

I think at the moment DMPs are a big request, specially for data stewards. I think assistance with personal data, so all the permissions that you need from HREC and so on, which come through the DMP also. Also storage of data. I think those are things that researchers always want to know: “Where can I store my data? How do I comply with data management plan? How do I comply with personal data?” I would say those.

Transcripts policy makers / support staff: I2

***Interview date: 08-06-2020***

*Upon review of the transcript, I2 added two clarifications. These have been displayed in blue text.*

[...]

3. What motivated you to take on your position?

Well, I think it was a very good time to get interested in research data, because it was becoming more and more important for scholarly communications. As you know, previous research outputs beforehand were very much focused on journals. With the increasing interest in open science and reproducibility and research integrity, there’s been much more focus on data, just really beginning to take off around then (in mid-2010s). I really saw an opportunity to make a big difference by becoming [...].

4. Did you have any prior experience in RDM policy making?

Yeah, my background beforehand was in [...]. It was never called RDM, what I did before, but it was all about the relationship between data, research and public exposure of data and sharing data; what data can do to improve how institutions work and how researchers do their work. So there was a broader interest beforehand.

5. What was your role in the designing of the data management policies?

[...]

But what was interesting, at TU Delft we approached it differently from other universities. Other universities tend to create a policy first and then build an infrastructure or build the support that people around can go to. And actually at Delft, we already had the data archive, and we had the data stewards start working before, and then we had the policy. The policy for us is just a few pieces of paper that summarize what we’re doing and gives some guidelines to what we’re doing. That was important.

The other thing that was important was that we weren’t too autocratic, because this is developed by the library mainly. We didn’t want to tell researchers everything that they had to do, researchers don’t like that. There is a lot of space for faculty policies, because we knew that each faculty, maybe even each department, wanted to develop their own policies because they do stuff differently. So we allowed that freedom. You had a kind of top-down part, where this is signed by the CvB and organised by the library, and we also have a bottom-up part which is more researcher-focused. That worked well at Delft. It took time to get that embedded but that worked well.

6. What was your role in designing that?

[...]

A lot was with the faculty secretaries. Each faculty has its kind of senior administrator, the Faculteitssecretaris, and they’re important players in the policy and administration at the management of the university. I worked with them to get feedback and develop it. [...]. The Data Stewards also played a vital role in discussing the content of policies with researchers in each faculty.

[...]

8. Of course there are different types of support. I think I have kind of a good image on it, but could you also elaborate on what you see as all the support offered?

*Information as provided in the list with interview questions sent beforehand: “RDM support comes in multiple shapes and forms: examples are assistance by data stewards, data champions or the university library, or workshops and tutorials on RDM. These can be considered more ‘formal’ types of support. In contrast, assistance by a research supervisor, colleague or head of research department can be seen as more ‘informal’ support.”*

Well, you’ve got a quite good list there, right, in your paragraph. In your question you talk about data stewards, data champions, university library, workshops, and what you also have is national or community-based assistance that works within the Netherlands or even international as well. That’s interesting you call it informal and formal, I don’t really make that distinction. Because we also try and set the data stewards up to be quite informal support. So what we think is key, is rather than have the expertise based in the library and then you put a separation between the library and the research in the faculty, that the data stewards are each based in the faculties and have understanding, experience of the research topics in that faculty. So we set the data stewards up to be almost a way of informal support, right? They're the people next door. So you know, the data steward has an office there in the faculty and the idea is that people can just go knock on his door and ask her or him, set up in quite an informal way. So I don't kind of see that difference, that difference between informal and formal support. What I do think will happen more in the future is that research supervisors will begin to understand more about this topic and we'll be able to get more expertise. But in 5 or 10 years, they will have a better understanding of the answers or they will be able to point you to other people in the university. So data stewards or data champions or data managers.

That also leads on to your next question, right? Which do they prefer. It is my strong belief that they prefer... The two best ways to give information to researchers are: one, somebody nearby in their physical environment. So somebody that they trust, who ideally is a door next to them and they could knock on the door and say "Can you help me with this?", so it happens easily and fluidly. I think putting lots of information on the library website doesn't work so well, because then people have to navigate the pages, they have to understand the language, it takes time. So I'm not a big fan of university websites as a way of getting information across. The other thing that people I think do is just type stuff into Google. So if you have a problem, like "what is data storage at TU Delft", you type it into Google and see what comes up, which isn't always reliable because a lot of the information that's needed is hidden on the intranet at Delft or is in a difficult place. But that's what people do, right. So to answer your question for what do they prefer, like somebody nearby that they trust, or Google.

9. Yeah, that's fun. I've heard people say that they would most probably look for support with their peers, but the Google aspect is new for me!

Yeah, well, peers, I mean, yeah. That's kind of, you know, their group or who they really work with because that's who they trust and that's who they think can answer their questions.

10. And would you say that the researchers are aware of the support options there are? So not only data stewards, but also the library trainings offered?

Well, I won't answer that, but we've done two large surveys in 2017 and 2019, which showed that awareness is rising. The blog https://openworking.wordpress.com has got all the statistics on awareness for research data services. So I would just refer to those. So whatever the data steward said in his interview, just go by that.

11. Yeah. All right. Well, kind of a similar question, but do you think the demand for these services is as high as expected or as low as expected?

It's getting a lot bigger. It's still not as big as I want it to be, but the data stewards are busy, the data stewards have enough work and that was the key thing. But I think this is also the type of support that researchers want, and it's really hands-on support sometimes as well. So if you're running a big project, with lots of confusing data and data sources and there's issues of rights and there's issues of sharing and administration, you actually do not want somebody who gives you advice, but you want somebody who sits in your team and helps you develop the data part of your project. So this is why we're introducing data managers, as a separate concept from data stewards. Data managers are people that a researcher can hire into his or her project for three months or six months to do the real nitty-gritty stuff of data management. We think in certain projects there is a real demand for that. So we'll be concentrating on that more in the future.

12. And is there a plan already for when that will be introduced?

Yeah, so we've hired two data managers in the last month and they'll start in September. And actually we're going to call it something like the FAIR Support Team. And it's also going to include coders as well, and developers, because the other thing that research teams need... Typically Delft is people who can do coding and development and can help with their software development. So we want to have a team of data managers and coders, there's a bit overlap there, and we're going to call it the FAIR Support Team and that will get going in September. And that, I think, will provide an extra layer of support.

13. And how many data managers do you aim to hire?

We're going to start off, there's only two at the moment. And I think you could have, well, actually there's only two that are centrally organized by us at the library, but actually in other departments, they already have data managers. So in Civil Engineering, in a group there, they have a data manager already. And I think it'll grow across the university and other faculties and groups also have their own coders. QuTech has their own base of coders that they want to develop. It's kind of, some of it will happen bottom-up and some of that will happen top-down or organized by the Library as a service. And we want to really develop that in the future. I could see us having 10, 20, 50 data managers in the future.

14. And how are they financed? Because you just said that Civil Engineering already has theirs and the Library is also hiring some.

That's a good question, really good question. So centrally, we're running this Open Science Program at the moment and the Library has special funding for two years to fund the data managers. And then what we want to do is, we come up with a business model where if a research group wants to hire a data manager, they put it into their project bid. Imagine that you're a researcher and you're writing a bid to go to the EU for Horizon 2020 and you're asking for three million euros for the project as a whole, and we think you should ask for fifty thousand euros fora data manager. And that's how the data manager gets paid for, via project funding.

15. So they'd have to know that they need one beforehand, when they're applying for funding?

Yeah, which can sometimes be a problem. If you have a researcher who doesn't really understand data, but uses a lot of data, that's a problem.

16. Yeah. All right. Well, interesting development. And then I also wanted to ask something else on support options. I heard there was an RDM course starting, a central one in September, could you elaborate a bit on that?

Yeah. So in the policy it now says that if you're a PhD-student and you start this year, you have to deposit your data when you finish your PhD. If as a university you make that request from PhDs, you also need to provide the training. So this is why we're developing central training for any researcher, really, PhDs who want to learn more about research data. And it's a kind of generic course. And then we also think that faculties will do their own courses, but they'll have more specific modules suited for the disciplines and methodologies within that faculty.

17. And how long is the course, like how many ECTS is it?

I can't remember. [...] is the expert.

18. All right. Well, I'm not sure if you can answer this question then, but otherwise I'll ask [...]. Do you know the difference between that course that will start in September, and the open science MOOC that's already running right now?

Yeah, the open science MOOC is much more focused on open science as a whole. So it goes into open access and publishing, journal articles, it goes into use of social media, it goes into how you build impact and communication for your research, whereas the research data one is really focused on data and the tools and techniques to manage your data. So the open science MOOC is an umbrella one with lots, and a bit more superficial, whereas the research data one is one topic and goes into a lot more depth.

19. Yeah. All right, I see. And then onto the next question that is actually on my questions list. What do you think is the general opinion of research about the support they can receive?

I still think the, um, well... I think it's changing, right? A couple of years ago, a few years ago, most researchers thought research data management was a bit of a pain and something that was bureaucratic and something that they just had to do for the funders. Now it's moving in a different direction where more researchers think that "okay, data is essential for my research, I need to manage it well" and also the funders are asking for it. So it's getting towards the stage where researchers think "now, yeah, this is useful". But, however, that's still going to take time. There's still lots of researchers who hate bureaucracy, they think "no, no, I can do it all by myself, I don't want to be involved with anything that the services are telling me to do". So there's still resistance. But it's my role, and my team's role to try and change that over. You know, keep on changing and keep on working on it. But I think it's becoming more, you know, I think in 5 or 10 years, for a lot of disciplines, subjects that will be embedded is a thing to do [*recording of this last sentence unclear*].

20. Yeah. And do you think that the policy can be a way to make people change their mind?

The policy is the least important thing. The best way to change people's minds is with people like the data stewards and the data managers. So the informal support, right, just there next door; "I can help you with this problem". Policies, policies are just there, if somebody is arguing or disputing saying "no, I don't want to do this", and then we can say "OK, this is the university policy, you need to follow it". But I hate doing that. The policy is just something in the background which helps in difficult circumstances. But the focus should be on the expertise that can help.

(I also did not mention rewards and recognitions in the interview - if you want researchers to change behaviour it is vital that they are recognised for their work in creating, documenting and publishing data)

21. Yeah, that makes sense. And do you then think that DMPs could be a good way to get people introduced to the topic?

Yeah, some people will like DMPs because it asks all the right questions and makes you think in the right way; other people will hate them because it's a form you have to fill in. So I mean, it's obligatory if you have NWO or EU funding. But the difference again can be if you can turn it from a form into an actual service. There's this idea going around of machine-actionable DMPs. And that is a plan where you can actually, it's interactive and you can click buttons and things will happen. So a simple thing can be, if there are some dates in there, you remind the researcher to upload their data or to update the plan; just a lot more can happen with them. If it becomes less of a form and more of service, then I think in the future it will be very interesting.

22. And do you think that the questions that the TU Delft offers in their DMP are good? What is your opinion on those?

I mean, we ask the questions we have to ask. It's a bit long, it takes time. But that's the ones we have to ask if you want to do good data management.

23. Yeah, I see. And then onto the next question, which goes back to the policy. How well do you think that researchers understand the policies?

I think they understand the headlines. So you need to deposit your data, your students have to, if they're doing a PhD, you have to hand in forms related to this. Not everyone will know that. But most people will understand the headlines.

24. Yeah. Would you say that is most important as well, that they just know the general content of it?

Yeah.

25. And then something else, I also talked to a data steward about it. To what extent do you think researchers comply with the policies? Because the data steward said “well, we don't really follow up on it, we just trust that they keep up with the DMP”.

Well, I mean, there's one thing in the Delft policy which says that all researchers should publish their data. And in actual fact, the compliance of that is about 15 percent. But, you know, as I said, the policy is a background document, right? We want to convince people in other ways, not just because they follow a policy. So I want to convince them it's good for their discipline, it's good for themselves, it's good for the group. And that's how you get compliance. So for certain aspects, you don't have great compliance. It's not a problem. It's just pushing and pushing it in that direction. And the policy isn't the stick. The policy is nice hanging in the background just to remind people where we want to go.

26. That's mostly the process of changing people their mindset and then...

Yeah, you know, cultural change and behavioral change, that's what we are trying to do. I mean, when people don't comply, it's often because they don't think it's important yet or they don't have the time. Occasionally they have a philosophical objection, but that's less so. But there's also, I think, a difference in different disciplines, right? If you're in certain disciplines, then data management is essential. Because the data can be reused or has a head role. Other disciplines, the data will never be reused because you've done a unique experiment and so it's less important. So there's also disciplinary bias in compliance and how that works. That's important, that's why you have faculty policies as well.

27. And then the last question on policies, which is a bit more open and reflective: is there anything you would like to see changed in the policies now?

There is about to be an updated version of the data policy published. But that's just a few amendments that deal with the GDPR and research data, and rules for how you deal with personal research data. It needs a lot more detail there. And then also for open science, we're also going to have a policy for software. So if you're a researcher that has developed software, how do you know how to license it or commercialize it or publish it or archive it, what do you do? So we're going to have a policy, guidelines, later in the year for that.

28. All right, and the data management amendment, was that the university-wide framework that is getting changes?

Yes, the framework as a whole.

29. All right. Well, then we can move on to the next subject of research data infrastructures. How familiar do you think that researchers are with those infrastructure for uploading their data, so depositing it?

I think in many cases, researchers just find the first one that seems to work well for them. And they use that and there might be better ones for them. But you know, they don't care, some people just want to get the data published and get it online. Zenodo is very good for very quick publication. 4TU is better in a certain way because we do metadata reviews, we approve the metadata and we check the data. So it's a different method from Zenodo, we are a bit slower. Some people don't like that, some people just want to publish their data straight away. Researchers don't always use the best infrastructure, they choose the most convenient one, but *c'est la vie*, right? We all do that kind of thing.

30. Yeah. What would you say the position of the 4TU archive is to domain-specific archives or Zenodo?

Well, 4TU is, we're not an institutional data repository as you ask in your question, we are a domain-specific one because we take data from across the world in science, engineering, design. But I think we have a job to do as 4TU to distinguish ourselves a bit more as a domain-specific repository. What that means is, it's quite, the services we provide at 4TU are quite general. We're updating our archive in July, so we're going to be working with Figshare and we're going to have a different platform (*see* [*https://openworking.wordpress.com/2020/08/18/why-figshare-choosing-a-new-technical-infrastructure-for-4tu-researchdata/*](https://openworking.wordpress.com/2020/08/18/why-figshare-choosing-a-new-technical-infrastructure-for-4tu-researchdata/)), so that will allow us to focus a bit more on our domain.

31. And what is the cooperation with Figshare going to look like?

Figshare provides the backend, they provide the platform. And it's our, we have a service on top with our logo, with our workflows and our processes. So it will look like Figshare, but it will have, or it will feel like Figshare, but the interface will have our logo and our processes behind it. So we're moving because we used to do everything open source and in-house, but that was no longer feasible economically. So now we've got this contract with Figshare to use their underlying platform.

32. But both platforms will continue to exist separately?

Well, if you look at... No, it's going to be... It's kind of, you have a platform which provides the underlying technical infrastructure and then you have that kind of interface on top. I'll send you a link to Monash University in Australia, they have the same idea (*see* [*https://bridges.monash.edu/*](https://bridges.monash.edu/)). It says Monash University and it's got their branding, but it's using the Figshare platform, right. And so 4TU will look a little bit like that.

33. Now onto the next question, which is the same but then for downloading data. Do you think that people can find the proper repository for downloading data?

If they need it, yes, right? If data is essential for your research, so for example, imagine you use satellite data for looking at land use, something like that, you know exactly where you have to go through the data. And I think the more tricky, difficult thing is data that could help you look at your research in a new way, I think that's difficult to find, typically with interdisciplinary research, right. If you're a biologist and you need to find data from the social sciences to answer an interdisciplinary question, that can be more difficult. So I'm generalizing usually here, but I think essential data, everyone knows where to go because their community knows about it. That other types of data, that can make them look at your question in a new way, that's more tricky.

I have one thing to say about the general attitude of researchers regarding downloading research data, because I think what's important here is trust, right? If you're a researcher and you come across a dataset that seems to be within your field, can you trust it, right? Because you will then embed that data into your research. And if actually there is mistakes in that data, that means there will be mistakes in your research. So the trust and the documentation is really important. And I think that's a bit of an issue at the moment, because to trust that somebody else's dataset... I'm not sure people do really trust other researchers in a way like that, unless they have a really good connection.

34. Yeah. Is that something that the 4TU repository aims to overcome with the metadata check?

Yeah, but I think that's not a technical thing, I think that's about the actual data and how you structure the data. Some really, really high quality data is well-structured, well-documented and easily reusable. And even if you have a data check by people in the middle, that doesn't necessarily create high-quality data, right? This is where it becomes a community thing. You know, if you wanted to, for example, if you wanted to look at somebody who'd ask similar questions to your interview questions and reuse those data, if you came across that dataset, you would really want to look at it, you'd go "Did this person really ask the same questions? Who was giving the answers? Can I really reuse this?" And you could probably do that, you'd have to play around to work out if you trust that data. And I think that's the same for any researcher.

35. Yeah, that sounds logical. It's also kind of hard to tackle, I guess.

Yeah, yeah.

36. What would be reasons for researchers to not use certain infrastructures, so why would they choose not to use 4TU, but Zenodo instead? Except the example you've already given.

Well, let me answer that in a different way, I think. Why are they using infrastructures? If it gets community approval, right. So if you're studying material science and you find that everybody else in material science is using this database, then you will use it as well. So that community buy-in is really important. Other reasons are convenience, trust, speed. And recommendation from peers. And then people might also not use particular interfaces for political reasons, commercial interests, right? So some of the Elsevier infrastructures, because Elsevier has a bad reputation for open science it might be less used. So, I mean, this is an advantage that Zenodo's got, because Zenodo is open source and it's funded by CERN. So it's a properly open source business model, whereas other ones aren't so open and certain people, not all, but certain people might not like them.

37. Yeah, I see. And then I have one specific question for the 4TU infrastructure. How do you deal with when people upload their data and then later on want to change the dataset?

We can do version control, so they can upload a second version of the data and we make a connection between the first version and the second version. If they ask to delete the first dataset, we only do that in very, very extreme cases. So when some law is broken or when somebody's personal data comes into it and it's been published for the wrong reasons. So we are of the belief, and we state this on our website, that once you publish data it's part of the scholarly record and it shouldn't be deleted unless there's really good reason for it. So that's why we have versioning rather than deletion.

38. And do you think that researchers are aware of that option?

No, but we do get researchers writing to us and saying "oh, I want to delete this" or "this is all wrong" and we tell them what I just told you.

39. Mm hmm. Because, well, the reason I ask you is that I spoke to a researcher who said "Well, I have this big dataset and I'd like to get two publications out of it". For the first publication, she wanted to deposit the data, but she wasn't sure if she should because for the second she would add some new models to it. And then she thought, well, I'd have to publish it again or I cannot properly link to it or have an outdated version of the data uploaded.

You can do versions. So if you look at Figshare, for example, it has really nice version control.

40. Yeah, so I think mostly that she did not have the right image of possibility.

Yeah, it's complex. But she can easily speak to my colleagues at 4TU, we can help her.

41. And then one more similar question. She also said that for the peer review, she would have to upload her data, but she could not do that anonymously in the 4TU repository.

Not yet, but she will when we move to Figshare. When we use the Figshare platform in July, we'll have a way of exposing a dataset for peer review so they can just be sent to peer reviewers.

42. And then onto the next question, which is on the institutional arrangements such as the funds that the library offers. Do you think that the researchers are aware of the options that they can use those funds?

You mean the ones from 4TU? The data funds?

43. Yeah, the refinement fund and the ...

Yeah, not nearly enough. I'm really surprised it doesn't get more interest. Because I reckon researchers sit with loads of data on their laptop, on CDs, on USB sticks, on storage, which is useful data and should be published, but they don't do it. And so we give them money. I think the problem is time, and maybe it's not seen as that important, right. It's something to do that is often at the bottom of the to do-list. If you're a researcher, you're always looking for the next grant, the next money, how to do your next paper. So publishing your data, they know "I need to do it" but it is always at the bottom. So we also think that data managers can help with this. If we have more data managers based in projects, they can focus on the data refinement and getting the money and we can do more of it.

44. And do you think that the people that actually use the funds are satisfied with them? So do you think the money is high enough?

Yeah, I think so. We've done all these case studies (*see* [*https://data.4tu.nl/info/en/about/researchers-about-us/*](https://data.4tu.nl/info/en/about/researchers-about-us/)), I think genuinely people are. I don't know the details.

45. What would your opinion be on those resources, the data funds?

Yeah, it's, I mean, it's a great idea. I want to see it used more and embed it more in the university.

46. And how long are you guys planning to keep the funds in place?

Well, it's part of the budget for 4TU. That's a few hundred thousand every year. And so we take fifty thousand of that from the budget for data funds.

47. All right, and then for how many years... Because I spoke to another policy maker who said "yeah, those funds are mostly transitional, so we want to keep them in place for a while people are changing their mindset and getting used to the idea of proper data management". Do you have any image of how long that should take?

It's a decision we'll make every year. We just review it every year.

Transcripts policy makers / support staff: I3

***Interview date: 26-5-2020***

[...]

2. Alright, I see. And you started as a Data Steward in [...], right? What motivated you to take on this role?

Yes. Data management is something that I kind of struggled with [...]. It took me time to organise everything and try to share it reasonably. Even though I had done it, I was not completely satisfied with the outcome, in terms of documentation, this kind of instability of the stuff that I created; in general that was not ideal. Yet, [...] in the best possible circumstances, I struggled, so I thought, well, okay, this is actually worth spending some time on – and everyone must be struggling in their own ways.

3. How big a part of your daily job/work at the TU Delft is your data stewardship? Do you do it full-time or do you also do research on the side?

I do full-time data stewarding. Data stewarding does imply, I think like 5% of research on how we do data management. So with the data steward team we publish a few papers per year, on how we get organised, how the community changes, and very general information. But it is mostly a support role.

4. Could you elaborate on the types of support that is offered to researchers?

From just the data steward team itself; our idea is to provide hands-on support, well actually *not* hands-on support; we provide *guidance*, *direct guidance*; absolutely not hands-on support. We offer guidance to researchers, and want to do that face-to-face. We try to be available for the sudden burst of questions that people have once in a while. The objective was to cover as many questions as possible. So I think that’s how it worked out. The idea was to start from pure data management problems*,* so the FAIR data principle and the archiving, and use the DMP as kind of a way to ‘trigger’ questions. It’s really a kind of how researchers are supposed to organise the data; that is what we focus on. We are in collaboration with the library for more publication-related processes, because they are in charge of that. So we have to interact with all the different support teams, and everyone offers something that is slightly different but still related to data management. Like the data storage is still data management, but it is ICT’s responsibility. We are on everyone’s little chair!

5. So if I understand it correctly, the university library organises workshops and tutorials on RDM. Is it correct that you as data stewards offer more ‘specific’ guidance, for when researchers have ‘specific’ questions on their own research, as opposed to more ‘general’ support the library would often?

I think that’s reasonable. I think the main difference is that if you go to the data steward, we are training the very generic stuff. We will take the time to unwrap your project. Whereas the library, you go to the workshop and can ask a couple of specific questions, but you do not necessarily get the long-term support or the follow-up to the process, which is something that we actually do when people write a data management plan. It takes a little while, there is a bunch of additional services you may have to go through (so ICT, privacy and all those guys), so we try to do a bit of follow-up on that, or at least orient people to the things that really matter to their experiment. Whereas from the library, they say “oh if you’re in this situation, do this or that”, we actually take the time to look at your document and say “yes, you are in this situation that we are talking about; this is what you have to do”.

6. Alright, so almost like a bit of competition between the two departments then, I guess?

Well, historically we have kind of a different mission. The library was really about publishing and storing the papers, rather than storing the data. So we are trying to define exactly where everyone stands, but it is mostly good collaboration to be honest.

7. Alright, well that is good to hear. Do you think that the researchers at the TU Delft know about your assistance and the kind of support that data stewards can offer? Is it well-known that they can come to you for help?

It’s starting, we ran a survey in 2019 to figure out what people were doing and how they saw data management and the support associated with it. I think about 80% of the people are aware there is a data steward around – they might not know my name, and the same goes for other faculties. I will send you a link with the result of the survey if you want, so you can look at the questions that we asked (*see* [*https://public.tableau.com/profile/nicolas.dintzner#!/vizhome/RDMSurvey2019/Dashboard1?publish=yes*](https://public.tableau.com/profile/nicolas.dintzner#!/vizhome/RDMSurvey2019/Dashboard1?publish=yes)). But so we’re getting a bit of traction, I started talking recently to students that I’ve never seen come to me, so for the past 6 months or so, people are spreading the word among themselves: “oh, you should do this, you should do that, you’re going to have challenges, we know someone in the faculty”, so it’s happening!

*[On the questionnaire:]* The FAIR principle is not that popular yet, but data steward, we’re getting there. At some faculties the response was not that high for whatever reason, so it is not really representative of all TU Delft's researchers’ field in general, but it is a good indication of where the people that were interested in the survey are.

8. Is the demand for your services as low/high as expected? Do people make use of the opportunity of the support you offer?

Not so much the opportunity as the obligation: because there are constraints on data management plans for the ethics committee, so they actually don’t have a choice!

9. So is the reason they come to you then more formal, because of the obligations, rather than because they need help with saving their data or are concerned with how they can archive their data? It’s mostly just because of requirements?

Yes, that would be accurate for more than half of the people that come to me, it’s because someone asked for a document.

10. What do you think the general opinion of researchers is about the support you provide? Let’s say for mostly the DMPs.

Well, the fun thing is that despite being formal things that they have to do, I get rather positive feedback after the meetings that we have once we’re done working with the DMP. Usually, people say “oh, well that was actually helpful”. Mostly with PhD- or master students, we really take the time to sit down with them, taking the experiment apart, and they really benefit from that, for data management related tasks. For project proposals, they don’t care so much. They just want to have the document so that their proposal is complete, and that it cannot make sense, and then in any case there would be more data management related tasks throughout the project, so they are less worried about will the data be properly stored, they care less; they just want the proposal to be done.

11. Since you say they just want the process to be done, to what extent do you think researchers actually comply with the standards they set themselves? Because of course in the DMP you promise a lot; do you think they stick to their goals and ‘promises’, or that they let loose once they have gotten their approval?

I think you would have to check with other data stewards, but I think the official position is that we do not follow up on that. So we help them build the plan and try to give them as much information as we can, so it is as easy as possible for them to follow, but we do not really check with them at the end of the project whether things were archived or not, so I don’t really know. I think as long as they don’t face any issue, they are kind of happy to comply with what’s written in the plan, because we try to make things feasible in the plan; we avoid reaching for the stars, so it usually works. But I’m pretty sure that if anything happens that is slightly outside the expected process, then I think people just do what they think is best, and not necessarily in line with the DMP.

12. How well are the data management policies of the university and the faculties understood by researchers? I suppose most people may know they need a DMP for the HREC approval, that is rather straight forward, but what about the RDM practices laid out in the policies? Do researchers have knowledge of them, so do they know they exist, and do they understand them?

Knowing that they exist, for my faculty itself: I did a round of communication about the policy itself, so I know for a fact that there is at least, at the very least, all department heads, and I think all group heads, a good 50% of researchers or 50% of the group, were present at the presentation where I said “we have a policy document, this is wat it contains”. The only thing that was missing was where to find it, because the official website was not ready yet. So they should know that the policy exists, so I would say a good 60%, like two-third, know that the document exist, and a third are somewhat aware of what’s inside.

13. Is there anything you would still like to see changed in the data management policies? Like, with hindsight, do you think some things should be formulated differently, or make it a little less complex (as far as you can call it complex)? So in general, any feedback on the policies?

So, the way that our faculty-specific policy is written right now is actually super, super generic. Because of the different activities that there are at our faculty, it’s hard to have something that is specific to a domain. We have researchers that have no data whatsoever. So when you want to write a rule that says everyone should archive their data, then they come around and say “I’m not going to do that, because I don’t have any data”. So that explains why it is very generic. My job would be to update this document in about a year now, to see if anything needs to be changed. The overall goal of the policy is “don’t lose your data and archive stuff at the end”, so it could be a two-line thing and then everything else should come in the form of a manual on “how do you do that for your specific type of experiment”. So I don’t think the policy is such a key document, I would say that it states very well what the objective is.

14. We could make a distinction between two types of support. The type of support you offer, I would call ‘formal’ support, because people have to come to you for DMP approval, and you’re the ‘formally’ appointed data-steward. Support from colleagues, or a department head, or just peer feedback in general, I would say is informal support. Which support type do you think researchers prefer? Do you think there is a barrier for researchers to come to you, for instance if they do not know you or if it is their first time of asking for help?

I think there will be a strong difference depending on the position of the researcher within the faculty. We should’ve asked that in the survey… But I think the more senior people especially are used to have support services they make use of, they tend to go for more formal support, because they are actually already highly autonomous in their work so they just need the formal, the kind of the ‘university stuff’ to go well. But for PhD-students and master students, I think the informal support is actually much more interesting, because otherwise they come to me and they say “oh I have no idea what I am supposed to do”, so I can give back and help them out, but if colleagues say “oh, we already have this platform and this is the tool we are familiar with”, I tend to favour that. I think that researchers should do first what their group is doing, provided it’s reasonable, rather than doing something fancy or something more ‘proper’, but that nobody knows about, because they will get stuck. So I tend to favour the informal support they can get, and ask them to take advantage of that – it’s better than what I can do. Plus, we don’t do hands-on stuff, so the informal support is great for hands-on, like “hey, I want to do use this database but it doesn’t work, can you show me the command”. That is what people actually need on a daily basis.

15. What would be the top 3 most frequently occurring subjects that researchers ask assistance for? So for example, asking for help with DMPs, but also if they want to deposit data into repositories, or shape their data into more shareable forms – would you say those are also frequently occurring reasons?

Actually, the last one, not so much at my faculty, because we don’t really, I haven’t really found decent standards to describe what we’re doing every year. [...] So metadata standards, we talk about that. I think the best one you can find is for interview data, and it describes what a question and an answer is, and that is kind of pointless. So actually, I would love to get requests on how to format the data, but that’s not happening because I wouldn’t know what to answer.

The motivation for the DMPs can be quite different, so they come because they are PhD-students that need to do that for their [*unclear*], or because they have to do it for HREC, or because of a project. So the DMP is kind of the very generic kind of request that everyone that talks about data eventually has to do, but the motivation for the DMPs will change a bit.

16. Do you think that researchers actually see the use of DMPs, or do you think they experience it more as cumbersome or a formality?

I have the impression that right now, at first it’s cumbersome, but when they’re done they actually got a bit out of it and they are a bit more satisfied. But for the moment, it really comes across as something unpleasant. It’s a bit of administrative stuff, that people are not very keen on doing, but if you do spend a bit of time, you can figure out some interesting thing on what will happen in your experiment. So, valuable.

17. What is your opinion on the current questions offered in the DMP?

We’re still missing the question about what data you have. I mean, how crazy is that. But the truth is, firstly, there are a bunch of different templates and they ask for very different things, and they organise their questions in different ways. I have some preferences; among all the templates that I’ve seen, I think the TU Delft one is the easiest one. It talks about data, most of all, except for what data you’ll store, but it is really organising in kind of the experimental workflow: you start at the beginning of the experiment, and try to reach a point where you archive things. For instance, I think the NWO template was organised around the FAIR principle, so 4 questions: how are you going to make your data findable, how are you going to make it accessible, etc. That is impossible to answer, because most of these concepts rely on the same stuff: you have to put a license in; that’s accessibility but also reusability. So the breakdown of the concepts is really important.

The TU Delft template will be updated; there’s been a lot of work to do on adding the question what data you have. Maybe, in the future, if we have better tools to describe what we’re doing, I would like to see something about data flows, where people can have this kind of raw data-stage and archival stage, and try to fill what goes in the middle, with this idea that you don’t master the raw data and the final data is not on your laptop anymore. So trying to think in terms of data flow and trying to express that in the DMP will help.

18. You just said that the TU Delft template is about to be updated; do you have any idea when, roughly?

No, I’m not following that. The person that is taking care of the communication with DMPOnline is works at the library. We try to update the template once in a while, but the support team of the DMPOnline is rather small, so they do what they can.

19. We have different data infrastructures, like the 4TU Centre for Research Data, but there is also Zenodo, Dryad, Figshare etc. How familiar are researchers with research data infrastructures, only in the context of uploading data to it? So, do they know which infrastructure to use for their subject, and do they know how to actually use it?

I haven’t had any requests regarding the favour to use 4TU or Zenodo, so I assume that the people who choose to use those platforms are familiar enough to actually succeed in uploading the data. Some people know about Zenodo, they are aware that 4TU Centre exists; picking one out of the two is kind of ehhh… they don’t know how to pick, but then, neither do I. I tend to go for 4TU for all the stuff that we do, and if there’s something that is more flashy or popular then I suggest Zenodo, because then you get a bit more exposure. But in general, I don’t make a strong difference between those two. So I usually say: pick the one that fits your domain; if you are already finding information for you on 4TU or Zenodo, then go there, because that is where all the researchers are putting their data. If you don’t have any preferences, 4TU is probably good enough, it will do the job so if you don’t care, that will work.

20. What would you say is the general attitude of researchers uploading their own data? Is it mostly to fit some policies, or do you think they also understand they can get some exposure from it?

It depends a bit on the domain. If they are in a domain where sharing data is already popular, then they are very much on top of that. Otherwise, [research discipline X] groups don’t share anything, just by default, they’re kind of annoying like that. [Researchers in discipline Y] don’t think they have anything to share, same for [researchers in discipline Z]. They tend to think “oh well, it’s all Excel spreadsheets so who cares”.

21. Now the same two questions for downloading data: do people know where they can find research data that might be useful for their own research, and do they know how to use those infrastructures?

My experience so far at our faculty is that the people that are using big datasets coming from the outside world, they know where to get it, but not research data. So they know where to get data from the UN, IMF or EU, for instance, so they are keenly aware of the open data aspect, but for pure research data or reusing other people’s data if it was not published through those large institutions, they tend to be less known.

22. What would you say their attitude towards downloading data is? Do you think they do not try because they do not know that there is that much research data is out there, or because they are incapable of finding the right data?

I think they don’t search in repositories. However, if a dataset is mentioned in a paper, they will follow the link. They’re not going to discover data on repositories yet, I don’t think that’s something they’re doing.

23. How could we stimulate that they will look for themselves in these repositories? There are a lot of workshops and stuff, but you must have people attend them, and people should know that the workshops are there.

[...] I think it is because the search engines for data actually suck. Discovering data, unless it was extremely precisely described using very formal language and metadata stuff, you’re not going to find anything because it is really hard to write queries that match a specific research question. The connections between the datasets, right now they’re established based on text comparison. I did a search on 4TU or Zenodo, I don’t remember, about something about the medieval potato crisis, something like that, and I ended up with poetry and all kinds of things that were completely unrelated. We don’t really know yet how to efficiently search and connect datasets to discover them. I think it’s a technical limitation, we just have too much stuff, and no one wants to go through every single one and say “Oh I’m going to look for research data”: it’s going to take you a month without that something.

At some point, we will really want to know, and I think we will lose a lot of data in the process because we won’t index everything. But if we have this search engine, and maybe papers that were not in clear format but were somewhat connected to the data, we can put some emphasis on the need to take care of that.

24. Onto the institutional arrangements for data sharing: of course institutional arrangements are policies or workshops offered by the library, and the support you as a data steward offer. There are also financial resources that researchers can appeal for, for example to cover the costs they make for making the data shareable. How well known would you say these options are amongst researchers?

Not that well known. That’s something we’re trying to push forward. The cost of data management is very hard to establish, and the options that are offered for support are kind of weirdly arranged. That actually matches how satisfied the researchers are with those resources, and I think they’re not really satisfied because they’re not really using them.

One of the practical reasons is for instance that the library offers data creation funds, so you can have a few thousand euros at the end of your project to wrap up data things. But the way that projects are organised is basically you have to be done by the end, or you don’t get the [*unclear*]. So there is no extension on the project themselves, and you need to wrap up everything. So you need to know at the beginning of the project that you will need this extension from the library to create your data, which you may not know. You cannot extend the budget of an existing project just like that, so even getting the fund I think is a bit complicated. I think I forwarded an e-mail about that to only one person in over a year and a half.

If you get in touch with people at the library, just ask how many people actually requested that, and how many people got them. I’d be curious to know!

25. On the funds that researchers can appeal for for making the data shareable: the researchers that actually find the funds and make use of them, do you think they find that it is enough money?

I have no idea because I’m not sure if anyone even managed to use them! I think the funding scheme is not ideal. Funding for data management is a recurring topic. The way the projects are handled and the lack of visibility on how much it is going to cost in general is kind of a pain. We do this by a kind of rough estimate.

26. What is your idea on these initiatives and incentives?

These are good ideas and initiatives, but kind of weirdly implemented. It’s going in the right direction, but on the wrong vehicle.

27. What do you think the general opinion of researchers is about the support they receive?

I think, well I hope they’re happy with it! I think they’re generally satisfied because I‘m getting more and more requests for training, so people are putting that on my to-do list. It’s not necessarily because I’m doing a great job, but because the pressure on data management coming from funders and journals is actually increasing all the time. They are eager to get it and they are happy to have someone to ask questions to, even if it’s not for formal training but just a couple of questions about “oh, I’ve been asked to put a name here or tick a box here, which one do I check?”, and having that is very nice.

I think compared to other universities: I’ve actually helped a cross-university project, and the other universities don’t necessarily have the kind of support we have and they kind of liked it, and they were very happy to have a contact in TU Delft to send e-mails to.

28. Could you elaborate on that project? I was looking at other Dutch universities and their RDM policies and found that the number of universities at the same level as TU Delft is rather low.

Yeah, it’s moving forward, everyone is getting to it. But it’s super fun because data management is becoming a kind of research topic on its own. You have conferences and journals, so people are getting together. The first conference I went to, people had no idea what they were supposed to do besides “oh yeah, we’re supposed to put data in the same place”, and like, you have 25 different universities and everyone is going in a completely different direction. And everyone says “oh, we all get together”, and “everyone go join me”. I see the whole data management part as well as a discussion on the possibilities: we are barely starting, we don’t really know what we’re talking about. People say “oh, I’ve read the stat for reproducibility and it says that everyone sucks”, but the next one says “well, you got your stat wrong, it actually shows that we’re good”. This kind of metadiscussion about how research is being done and organised is kind of the reverse of the discussion, and we’re going to take some for about ten years. The outcome will be hopefully better than where we are now, but it’s good to be reminded that this is the real starting point of a new discussion. All the stuff we’re putting in place right now, it will be replaced very soon with something better.

Transcripts policy makers / support staff: I4

***Interview date: 27-5-2020***

2. [...] When did you start with RDM policy making?

In our university, [...] we developed some efforts on open science for our university. I think it was already in 2014, so that’s 6 years ago, that we realised that it was important to do a bit more, not just in open access, where we already were active in, but also to expand that to the whole research process and we had several strategy requests which we then asked of the CVB, the board of the university, to accommodate so that we could start with the research data programme in the university, which resulted in having data stewards per faculty. We first got temporary budget for that, and now it’s something that the faculty pays for themselves and it’s coordinated by the library. We actually, I think it’s because we are a technical university, we actually first worked on doing things, getting our hands on the practice, and only the last two years we really worked on the framework programme so that we had a policy in place. So that actually followed having the data stewards and working on the data management plan et cetera. And also our research data services team in the library was I think also started around that period, so 2013-2014.

4. How big is the share of RDM policy making and open science policy making in your role?

It’s an important part, but of course I’m not only responsible for realising open science or having FAIR data. [...] Open science is I think perhaps 30 or 40% of my time [...].

[...] The Netherlands really were stronger in getting the open science topic on the table when we were chair of the EU in the first half year of 2016. This is a rotation thing, every country will have that half a year, we had that in the first half of 2016. One of the priorities was open science. [...] A congress on open science in the Netherlands in Amsterdam was organised. There was something called the Amsterdam Call for Action, if you look for it you will find it. In May 2016 the Science ministers came together and they announced the document on open science, in which they talked about open access, FAIR data, recognition and rewarding, and also advised all member states to have a roadmap on open science. In the summer of 2016 we as the Netherlands were chair of the EU. Sander Dekker [the state secretary then] also played an important role in the Netherlands. I think it was in 2013 that he issued a letter in which he asked all parties, so both the vendors, both the publishers, and the universities, all stakeholders, to make this step towards 100 percent open access. And he also said that he didn't want to add extra money in the system to achieve that. So that has been a crucial part to really a better uptake in open access in the Netherlands. And it was also, I think, the reason that, well, we were a bit in the forefront there, that's why we also have done the same now with open science. And the national coordinator of open science, Karel Luyben, so he used to be our rector, now also has a big role in Europe. It gives the Netherlands a good position, I think.

5. All right. Now on to the more formal part of the interview. I'd like to start out by talking a bit about RDM support that's offered at the TU Delft. Well, first off, of course, there are multiple types of support. Researchers can go to data stewards or the library and so on. And I did my own research, so I think I have a good impression of it, but could you elaborate a bit on it as well? So we get on the same level.

Yeah. So, you're specifically, of course, interested in RDM support.

6. Uh-uh.

I think what the university library tries to do is to give researchers integrated research support, and they do that with something we call the research cycle. You will see it if you visit the library website and you go to 'Support for researchers', you see these different blocks that we have. So you first start your research by having ideas and then you need to get funding or acceptation of your idea and then you are doing your experiments and then you often publish, disseminate the results of your experiments. And in all these blocks you need support of the university. And in some of these blocks it is the library helping, especially in the first block, so getting all this material, all these journals, all these databases, everything that the library acquires so that you can work on top of shoulders of others, and the fourth block, so disseminating. When the library realized that it had to be more active in the experimentation part, so helping researchers while they were doing research, the group Research Data Services was started. The library started this archive so that people who work in disciplines that do not have their own data repository can actually have a place where they can upload safely their data, because we noticed that still in a lot of departments, the data is not on floppy disks anymore, of course, but it is on drives or network drives or whatever, but not connected to a bigger picture. And if a PhD-student or a student leaves the university then a lot of the data is lost. And you have to realize that the whole reason of working in open science and open access has to do with the fact that for a big part, the university is paid for by public money. So we think that you should really be careful with what you do with your public money and that the public should also benefit from everything that goes out of the university. So that's why we have been able, I think, to expand our service to also research data support. There now is a group of 11, 12 people in the library. And now on top of that, we have the data stewards: they work with the departments, they work with project groups. They have the contacts. Of course, I do not have that personally. So it's all done by the people in the team. And we have a data archive. We have the data management plans to help them get going when they do a grant proposal. We help them with metadata, to describe their data, if they need help with that. We help with ingesting their data in the repository if it's needed. That's on the level of the researcher and the research groups. And then, of course, we helped the faculties and the board with having this framework policy so that it also is connected. So that's what we do.

7. And would you say to the broad availability of support, so on the one side by the library and on the other side by data stewards, is well known with researchers?

Of course, I cannot just say yes or no. We have been doing a survey a few years ago and also last year, where we saw that with the addition of the data stewards at the faculties, the uptake of the knowledge of what to do with your data has, I think, increased. I mean, for me, it's not so important that the library is visible, but that the support that the people really need, that they realize it's there and that they can find it. I think it has improved because of all the things we have been doing. But I also think that there is stil... Because there is mobility. So you constantly have new PhDs, people come in. I think we can improve on making our services even more visible. So, yes, I think a lot of people can find us. The best thing we have done is to have this person in the faculty: they know of a lot of other services that they can tap into and help the researchers in that way. So I think that has been very effective, but I also think we can improve from that.

8. So you would say the most important part is that people can find a data steward, who then could eventually forward them to other resources?

Yeah.

9. And have you heard anything about the demand for the support services? So do you think that the people that actually know about the support services also use them?

Um, yeah, I think so. I can't give you the right quantitative numbers. Again, I know that we have been doing surveys and I've seen them and I thought, yes, it's going in the right way. But I think what's important is that you are there at the moment the researcher needs support. That's the purpose. So if you come to a lunch talk at the department and there is someone to say something about the support services, not just, of course, research data services, but also publishing services and a lot of other things we do. But you are not at the moment involved in... You're doing something else, you just started or perhaps you're about to leave or whatever. You're not at the right time and people forget it. So what is essential? That you combine it with the right moment. I always say: be there when there is a new employee, or perhaps when somebody just got a grant approved from NWO or from Europe. Explain then what the steps are. So try to be there at the place and moment somebody needs you. That's easier said than done.

What also helped us, apart from, of course, having a very supportive board who also financed some of the strategic activities that we're doing, what also helps are the requirements of the funders. In that sense, the demand is there because you can't do without a proper data management paragraph if you want to have a grant. You can't do without publishing your article open access, otherwise you will not get the full sum of the funder. So I think that's also creating the demand.

10. So would you then say that most of the people looking for support are doing so because they are obligated to, either because of their funder's requirements or because of the DMPs necessary?

Yeah, and I think that, in a way, that's a good thing. The whole movement of open access originated, I think, with the libraries, and it was almost a belief: it's something that's good. So we are paid for by public money, so we have to give back our results to the public. And it's something like, it's almost preaching. I think you can't get all people to believe in this, for whatever reason. But I think the fact that there is also some obligation there because of just guidelines for funders, I think that that's good because it helps. And it not only then convinces the leaders, the ambassadors, but also those that are a bit critical. So I think that that does help.

11. I can imagine, because what I heard in another interview is that it's mostly about changing the mindset of researchers to actually get involved and start doing it. And I also think that by using those policies that people will start to think about it and then maybe change their mindset throughout the process. Is that what the policy is also aiming for? That they not only see it as an obligation but also see the use in it?

Yes, of course. What I said, we are a pragmatic university or a technical university. So a lot of the things we have been doing without having a proper data policy in place. And what helps is if people are a bit critical and they say, "OK, and where can I find it then? Is there a policy?" And you have to say no, no, then that doesn't sound professional. So you need all layers involved. So you can't, as a support unit, as a library, tell people what to do. Right? The support unit is supporting the department. Supporting the faculty. Supporting the university because they think it's important that we have a proper research data management protocol because we want our research to be transparent, to be reproducible. Integrity. These are the values we have. And what you do as a supporting unit is supporting the researchers in helping them. But the people who should say to the researchers that it's important should be the departmental chairs, should be the deans, should be the board, and not us. What we need, you know, in our face-to-face talk, if I talk with a researcher, what I need is that I can say, "no, look, but this is supported by the board or by your dean", and then the policy helps. That's the way it's connected, I think.

12. And how well-known would you say that the policies are with researchers, and then especially for the faculties that do not have their own specific policy in place yet?

No, so that's what you heard, I think, of course, also in the talks you've had so far. I'm not sure whether you also talked to [...], because they have far more in-depth knowledge how things now actually work in the research unit. But they're not there that long. So only two or three years ago, we got the generic framework approved of by the board. And we are almost, I think the board will discuss the approved faculty data poliy documents this year, so we're not yet there. So, yeah.

13. So it's really a bit too early to say whether people know about it?

Yeah, I think so. Yeah.

14. Right, I see.

I think Applied Sciences is a good example, because there is the quantum group there, there are some people working who actually took the policy and made it their own, and that's the way of course, that we would like it. So not that you have a document that somebody else says "yeah, you're required to do this" or "you should comply with that", but that it's something that you, from within the department actually, think is important.

15. Yeah, so it is more like intrinsic motivation rather than an obligation.

Yeah, yeah. But I don't think the policies are that known yet. I think.

16. All right. Now a bit more of an open question, is there anything you would like to see changed in the data management policies right now?

Not I think in the policy, I mean, it's just there, but it's a growing and living document. What you see is, there are a lot of things happening nationally, internationally. So I wonder what the connection is between... It helps if we would have a standard data management lens or protocol for certain disciplines. And that's being discussed, it's not yet in place. I think that would help because researchers, of course, work together. Not just within the university and not just nationally, but of course, internationally, so it helps the groups if there are more standards also for these data management plans. But we are discussing that also with the funders. I think that's something we could improve. And I guess also policy alignment would help, so that you're not working in a university in Delft and you're working with somebody in Melbourne and compliance is completely different. That would not be helpful. So I think with the movement in Europe with what's happening also at the level of EOSC, European Open Science Cloud, and a lot of the working groups that are working on standardization, what does FAIR mean, FAIR principles, I think that all will help, but it's for a researcher pretty chaotic. I also do not think that the researcher needs to know about all this initiative as long as the researcher knows that he can reach out to the data stewards, and to the library and to the ICT department, and that we know how to support them. So I think more interaction with the discipline-specific groups to align policies is important.

17. Right. And would you say that aligning policies with other universities, and let's keep it in The Netherlands for simplicity's sake, would you say that that is a role that the library or just the general board of the TU Delft should take on? Or should it be faculty-specific? Because, for example, I can imagine that the faculty of Electrical Engineering, Mathematics and Computer Science on the one hand and Technology, Policy and Management on the other are rather different in which universities they work with.

I think it's what I said, you need all layers, it's the same here. I think you would both need the disciplines, so the people working on artificial intelligence or people working in the medical field, working together and you already have that bit, you have some discipline-specific initiatives, you have the life sciences and DTL [*Dutch Techcentre for Life Sciences*], you have DANS with the social sciences and humanities, you have our own 4TU network for the universities, the technical universities. And then you should have the people with earth and climate, and then you should actually try to have them think about their specific challenges. But you also need the administrators, so the deans and the departmental chairs and the boards. So you should actually approach it from all layers, I think. And that's also more or less something that's been taken up now. There's talk about DCCs, the Data Competence Centres [*see* [*https://www.surf.nl/data-competence-centers-verbinden-mensen-en-techniek*](https://www.surf.nl/data-competence-centers-verbinden-mensen-en-techniek)]. It's the Data Competence Centre where we are trying to see if we want to have something on a national level and then on local areas. So it is a lot of things being discussed at the moment. And I think the most risky part is that it's only a discussion taking place at the administrator level and not with the researchers. So that's something we should improve on, I think.

18. Yeah, we can say that, I think TU Delft is rather far in developing their policies on RDM and related stuff, especially compared to other Dutch universities. I did some research and only three other universities made similar progress with their policies as TU Delft. Is that being picked up now by other universities or does that hinder the standardization of policies?

No, no. It's being picked up for sure. Yeah, it's not always done and led by the libraries, it's sometimes more at the sort of strategic unit of the university or like in Rotterdam, where you have the academic services group where this is being part of, or sometimes even ICT that takes the lead. But it's a topic that's on all desks of universities now because this open science, this national program of open science with these pillars, and one of the pillars is of course FAIR data, they are really working on coordinating how we deal with data in the Netherlands and also giving advice to the universities and research institutes how to do it in their local environment. So you can't stay behind. You have to go into it. Yeah.

19. All right. Well, I'd like to move on to the next subject of research data infrastructures. We've talked a bit about the 4TU Data Centre. And of course, there's also Zenodo, Dryad, Figshare and other repositories. How familiar do you think researchers are with the appropriate infrastructures? So do you think that they can find the fitting infrastructure for their domain?

That depends a bit. In some areas, like in the medical field or astronomers, they've also always been forced to share data because data was huge. So there I think it's far more common to use shared infrastructures and sometimes perhaps people don't know exactly what they're using, they're just copying behavior, and are not realizing that it's perhaps hosted by Figshare or that it's hosted by CERN. Because you know, that's not really relevant. I don't think that our 4TU research data archive is very well known. It's not our intention that everybody should know it. It's a pretty good alternative if you don't have a data repository in your own discipline. And I think it's growing, the knowledge, so it's no one answer, I think.

20. And in general, would you say that they know how to use it? Because I spoke with a data steward earlier, who didn't say a lot of researchers were having trouble with depositing or downloading data. So in terms of the infrastructure itself, do you think the usability is understandable for most researchers?

That depends really on the type of infrastructure, I think. It's like with all systems and software. But people are now so media literate. A lot of researchers are very literate, I think. They are very capable. We are now just going to move to Figshare ourselves. So the ingest, the front end of our archive will launch in September and then you have this product requirements document, functionalities, and it's all about user friendliness of course. So user experience, service design is also important because if you have somebody who at least is able to find your service, but then actually is stuck because it's completely unfriendly to use, then you've lost this person. So every vendor finds it very important. So I think the basics is... But I think there are often some advanced usage possibilities that people probably will not realize. That's something perhaps that all the services could improve on. But you would really need to talk with a few more data stewards I think, and of course researchers themselves to really get a proper answer to this question.

21. Did you just say the 4TU Data Repository was collaborating with Figshare? Could you elaborate a bit on that?

We just signed a contract, so we are just partnering up with Figshare now and launching that in September. So the person who is really implementing that now for the universities is [...]. So if you are really interested in seeing that, what sort of requirements we had and how that is implemented, you really should ask [...].

22. Thanks! We just talked about the uploading aspect of those infrastructures, so people depositing their data into these infrastructures. Now, if we ask ourselves the same but for the downloading aspect; do you think that researchers are aware of the possibility that they can use all these resources?

No, I think, but again, I'm not basing this now on really quantitative numbers. As far as I know, this is exactly why the national plan has 'reuse of data' as the third ambition. Because the whole idea of sharing is that other people can find it and can use it, and that we reduce double research being done and improve reproducibility. [...] Of course, now in this whole covid-19 epidemic, you see how important it is that data is shared and that people can reuse data, but there are not that many good examples, outside of the medical area. So I think this is really something we should focus on in the coming years. It's not just the data, but also the articles, how we can actually make sure that we do what we have as an ambition so that there is perhaps less double research being done and research is reproducible, but also being used then. So I think that's really something we can improve on.

23. And would you say that would be more the role of the data stewards to inform people about it, or more a task of the library?

Both, both. I mean, I see the data stewards and the library very connected to each other. I think it is peer behaviour what you need. If you see that people in your research groups or research area, nationally or internationally, are doing that, then you probably will follow that behavior. So you need some good champions, people who are known in the field, for making that first move and it can help if data stewards could identify such people. That will actually make it flow and go, I think that's the way we could do it.

24. Could you name any other reasons for researchers to either use or not use certain infrastructures, besides what we already talked about? So, more or less peer experience, and just the habits of people in the research domain; are there any other reasons for people to use or not use infrastructures?

Well,the most common objection we had when we talked about open data - we now prefer to refer to it as FAIR data - is that people see it as their own baby and they are afraid that either they put it out on an infrastructure and it's away, it's not connected to them as the creator anymore so they do not get the credit, or that it's perhaps used in a way that they don't think it should be used. That's their concern. And you can, of course, help them by explaining how these infrastructures work and that in all licenses that we have, of course, the creator should always be mentioned. And you could argue perhaps people don't do that, but then in a way you talk about research fraud, because that's something that should not be the default behavior, I think. And you can work with proper metadata to make sure that people realize the context of your data. You can also have sometimes not an open infrastructure, but always asking approval of the creator for use. There are different ways to deal with that. So the most important thing is not to laugh them away or to not take them serious, but really talk through and understand where the concern exactly is. So, yeah, I think those are the most important ones that we came across.

25. I see. How do you think we could use alternative metrics, so rather than just the impact factor or citation factor of articles, could help in this aspect?

The incentive for the researchers to share, not the concern but the incentive is, of course, that it improves their visibility and that their research is becoming more known. The traditional metrics are perhaps still appreciated by the researchers, but we as administrators are a bit more critical because it not always gives the right credit to people really working on the research. And it focuses a lot on the bigger journals. It's more a journal metric than an article or research metric so that's why we are a bit skeptical. Altmetric, of course, gives other means of being visible, being cited; there's a focus on social media so that can help in making research also visible in another way. And I think we've experimented with the tool, it's also just the tool of metrics, you can buy it from Digital Science. It's another group that we have in the library that's headed by [...] who can tell you a lot about that, because we did not continue! We thought the tool was a bit too expensive for what it actually delivered. That's also interesting to know, I think. But I think that's still on the level of the articles. So data citation is another one. [...] What is crucial with every research object, whatever it is, is that you connect it with a persistent identifier so that you never lose it. Of course, there are also citation regulations now for data, but I think the Altmetrics part is not yet so available as for articles, I think.

26. Yeah, that makes sense. We've talked about the support that researchers can get in terms of help with making their data shareable, or help with the infrastructure or whatever. But there's also the financial aspect of making your data shareable. So I'd like to zoom in on the data funds that the library offers. How well known would you say these options are amongst researchers, that they can make use of this one.

Actually, I'm not really sure. I know we introduced it just two, three years ago, I think. We also have an open access fund; that's pretty known because it's always exhausted before the year ends. So that's a very popular fund. And the fund for data is, I think, you know, it's known by the data stewards and they should be able to point researchers to it, in the case that they struggle with finance and don't get anything from a funder or whatever, so they can help making data FAIR and open. But I don't think it's really known with the researchers. But we also should be a bit careful because it's not a huge amount of money. So it's actually the way it works, you know? If the data stewards are aware, they can actually pinpoint researchers who they think are compliant with the fund guidelines, to it. So I think it's also not a big problem that not every individual researcher is aware.

27. Yeah. And you just said that the money is a bit low, would you say that it is too low?

I don't know. We always call this sort of 'transitional money', that's the same also with the open access fund; in a way, it should slowly fade out because we want everybody to publish open access and we require publishers to offer that. So in a way, this is transitional money until we all have transferred to one hundred percent open access. So it's the same for the data fund, it should be there just for a limited period of time to give some incentives to be able to make your data FAIR or open. But it's not something that's sustainable to do like that. But I think the whole transition to open science, there is a document on finances being written by the national program open science (*see* [*https://www.openscience.nl/files/openscience/2019-12/Report%20Transition%20Costs%20for%20OS%20in%20the%20Netherlands.pdf*](https://www.openscience.nl/files/openscience/2019-12/Report%20Transition%20Costs%20for%20OS%20in%20the%20Netherlands.pdf)), I think that's where you see there are a few scenarios sketched and they made a comparison with what Germany is doing with their national data infrastructures and how are they funding it and what size are they funding it, quantity, they make a comparison with the Netherlands, what that could mean for us. So I'm not saying that that's the money we will get from the ministry or... But it gives an indication of what we're talking about. And then this data fund is really a very small part of realizing the change, I think.

28. And just a bit more for my own image. What is your role in the way that these funds come together?

How can I explain it to you... The way the library is structured, the other university services are also constructed that way, there is a director and you then have the departments. Like I said, the library has the research services department, education services department and resources department. Together there are four managers. And then you have per department, you have a team, with a team lead. So for research data services you have a team lead, for research support, for educational open spaces, you have a team lead, called 'head'. They do that together. And my role is often to put their discussion on the strategic level with the board or with the deans or with the other directors. And so, for instance, for open science, I've been really involved. And now there is a program, there was a strategy plan approved at the end of December last year, there is a program manager, there are project teams, we’re all online. The more we implement things, the less I get involved because then the money is arranged, the people are there, and I trust that they’re doing what they should be doing. What my role has been, is to make a decision to start a services team, to hire somebody as a head, to allow the head to hire people, to have some strategy funding from the board so that we could move on. But the moment they’re there, I’m a bit more disconnected. I know them, because I sometimes go to some meetings they go to and I think they might know me, but it’s also not very important that they know me, because then you should know the researchers and the dean, you know?

29. So if I understand correctly, you had a bit of a bigger role once it was all starting to develop, but now that it’s all in place you can leave them alone?

Yeah, I take a step back actually, yeah. And I move on to the next topic, so I’m thinking of… We have research data services, what else do we need to support the people? And open science, we now talk about open hardware, there is an open hardware community that we can perhaps support. We are discussing citizen science, whether that could be something that we also take up in the university. So renew, and then what I do is that I connect people that are actually working on it in the research group, with [...] who is in charge, I report to him for this topic to see if he is enthusiastic. So I’m moving forward, so that’s the role, but not of course only on this topic but also on other topics.