Speaker: Joy Davidson

Video Title: The Role of Data Stewards

**Transcript:**

**Slide 1: Where do Data Stewards fit into the landscape? (0:06 - 0:52)**

So first of all, I think this diagram hopefully can give you a sense that data stewards again sit slightly in the middle of the road in many ways. So you've got data scientists or people who are research specialists in a particular domain on one end, and you have people who are research data management and curation specialists on the other. Data Stewards tend to sit somewhere in between those and how far either side of that centre they sit really depends on the role that is being supported at the institution. So it's somewhere between the research community and the support community and it is a sliding scale.

**Slide 2: Pointers from the Dutch Landscape (0:52 - 3:28)**

This is a nice example from a report that came out of the Netherlands not too long ago, and it tries to cluster different data stewardship roles into three main headers. The first is one that looks at policy, and in that role, they would see the data steward as being somebody who understands the policy of the various different stakeholders, whether they're the funders or the publishers. But understands the drivers for FAIR at the policy side and research data management and interprets those and helps to translate it for the researchers. So their role is very much trying to ensure that researchers can work and be adhering to policies that govern the way they work.

Other roles within data stewardship tend to be more focused around research activity, and in some of these cases, they are very much engaged in the research itself. So from a very early stage, a research data steward would be involved in thinking about workflows, trying to think about standards that might be used and certain kinds of tools that could be used to carry out the work. So it's a much more hands on role. And that is very much focused at the research itself rather than the broader context.

Others are still looking at this in a more infrastructure provision sort of sense. So you've got data stewardship, and that tends to be the people who are developing the infrastructure and the tools that help people to do certain things. And this can be more at the kind of infrastructure and IT level. And I think there's a little bit of blurring lines between all of these roles. I think they don't work in isolation and so you can see, they all sort of intersect around this notion of FAIR. So there's these kind of slight nuances coming out. So data stewardship, could be kind of taking three different strands. For what we're doing we very much are kind of that first policy sort of focused, the kind of training we're doing is trying to help people to understand the bigger picture and to be able to then translate it for the researchers they work with.

So what we also see from this report is that there are kind of two different approaches for data stewardship at the institutional level. One is this notion of generic and the other is embedded. And I'll tell you what that means in a little bit more detail in this slide.

**Slide 3: Tasks and roles in Dutch research institutes (3:28 - 5:17)**

So generic data stewardship, is that sort of [generic] level and again I think it's very much the kind of session we're doing today, It’s aimed at those sort of generic data stewards. And this tends to be those people and as we've seen from the poll that we did earlier, there's a lot of people who are the only data stewards support in their organisations. So there is for the entire organisation across all of the faculties and sort of central level and they have to help all of the researchers, you know, at a sort of a generic level. They can give them enough of that information to help them understand the policies they have to align with, and the key messages and serve more of that sort of a signposting to further information. Again, they've got a really strong emphasis on providing information and training with regards to policy requirements and good practice guidelines. And they generally tend to be those people that help the researchers develop their data management plans.

The Embedded role are those people as I mentioned before, they are much more kind of working on specific projects. So they are part of the project team. From the outset. They are helping to define how the work gets carried out and making sure that it's kind of born FAIR data that people are generating. And I think you know that obviously implies a certain level of commitment. It's not something that every organisation can support. It does mean that you need a team of these people and they have to have expertise in different domain areas, to be able to get involved with research projects and to help make sure that this kind of activity is happening from the start. So those are the kind of main differences between generic and embedded support.

**Slide 4: Stewardship - overlapping roles and responsibilities (5:17 - 6:54)**

This diagram again, you're seeing I think, with most of this stuff, hopefully that there's lots of overlap. A lot of these things they're not binary, there's fluidity between a lot of these different roles. And the emphasis will kind of depend largely on where you're working. But I think that we're starting to see that there is a lot of overlap between points of contact. And the data steward is, in many cases, that intermediary role who tries to join up a lot of the other kind of stakeholders and make sure that they're all speaking to each other and they all understand what we're trying to achieve in terms of sharing the data that we're generating. I think it's important to remember too, that the data steward role can often be the first point of contact. So if a researcher is looking for advice and guidance and how to make their data FAIR, or for research data management, they'll often come to the data stewardship role to get that advice. So it's really important even if you don't have a lot of domain specific expertise in house, that you can at least sign post people to the right kinds of places to look for that support, whether it's within the institution or somewhere else.

And I think the other key thing that we're starting to see with this sort of diagram is, you can see the top left circle is all about developing software. And there is another strand of activity that's happening, where you've got data stewards who are becoming much more research software engineers, and that is another kind of direction that this role is taking.

**Slide 5: What is a Research Software Engineer? (6:54 - 8:08)**

So what does a research software engineer do? They kind of sit between software engineers and research. So they're people who tend to have discipline specific expertise, but also the technical ability. And this kind of stuff happens quite often when you've got researchers who are working and they need to develop special bits of software to support different kinds of analysis of the data, or to support different kinds of capture. So if you've got people developing specific code and things like that, for the research, that would tend to be this kind of research software engineering role. And the good thing about this sort of role is that they have domain specific expertise, the information technology background to be able to develop solutions, but also the awareness of research, data management, FAIR and curation. So it's quite a specific skill set in this and again, I think not every institution will be in a position to support this sort of a role, but it is something that is becoming more on the radar. So it's worth being aware of this as being something that is considered an aspect of data stewardship.

**Slide 6: No single blueprint for data stewardship provision (8:08 - 10:43)**

So in terms of providing support at your institution, there's no single way to do this. There's no right or wrong way to do it. I think a lot of it depends on how you're organised at your own institution and what sort of resources you have.

So two examples on this slide. One is the university at TU Delft, and they've invested quite heavily in trying to set up data stewards at the institution. So they've got, every faculty has a dedicated data steward. And I think there's eight of them. We'll look a little bit more detail at the next slide. But essentially, they are a team that is managed centrally, but they all work in the different faculties and they have a good level of discipline specific knowledge and are researchers in their own right. That is one model and it is quite resource intensive. I think you know, we'd all love to be able to do that. But I think the reality is that most research institutions probably can't support that level of support at this stage.

Other models that we're seeing that I think are much more promising for the bulk of the universities that need to provide this kind of support are things like the data curation network, and this is an example from the US but we're also seeing similar approaches in Sweden and in other countries. And this is a sort of a shared approach to data stewardship provisions. So while you might have expertise in the humanities in one research organisation, and another one has expertise in the medical sciences, you can start to see then how you can pool this together and try to leverage different kinds of support so that you're not having to have it all in one organisation, but rather you've got that peer network, and people can start to contribute and leverage information from each other. And I think that model does offer great potential, I think for especially, you know, we're starting to look at Ireland and this kind of approach to providing data stewardship, since there's a lot of people in early kind of roles and they've only been in post for about six months and there's only one of them in most organisations. I wouldn't really recommend trying to look at some of these models and seeing if it will potentially be something that would be suitable for the Irish context. Because it really will help you to build up a suite of resources that you can draw from without having to commit to that level of intense investment that we've seen at TU Delft.

**Slide 7: TU Delft example (10:43 - 11:25)**

So this is just a deeper look into what happens at TU Delft. As I mentioned, they've got eight data stewards, they also have a central coordinator. And it's a wide range of activities that they do around the circle. So you can see that they provide tools, they do the kind of policy level stuff, but also they do an awful lot about training and giving advice. So again, this model I think is worth having a look at there's a link to the blog, so you can see how they've gone about doing it. But it does require investment and buy in. So it's it would be nice if we could all do it, but we might be able to learn certain things that they're doing and borrow from this model.

**Slide 8: Currently, role largely depends on the needs of the employer (11:25 - 12:43)**

So I think one of the key things to get across is that there's really no professional role for data stewards at the moment. It is something that it's coming out as a profession. There was a recent study that looked at different job adverts, and this was in the UK but I think it stands pretty much across Europe, It found that there's a huge variety of the kinds of titles and the sorts of things that a data steward role would take on. Most often they were kind of tied to open access, or the repository service in the institution, or research data management. But it was common. I think that for many of these roles, there was an awful lot of overlap. So wasn't just a data steward. You were also kind of looking after open access publications or you might be the repository manager. So in a lot of cases, data stewards wear more than one hat. So I think it's worth kind of bearing in mind that currently, the role that you take on as a data steward will probably depend more on what your employer is looking for, what gaps they need to fill at the institution, than based on anything to do with that kind of professional set of skills, that is agreed as being data stewardship skills.

**Slide 9: Focus on the basics first and build from there (12:43 - 13:52)**

So I think key message is that it is a very kind of quickly evolving landscape. At the moment, you're probably, the role, is going to depend on what your employer wants, and what they need and what resources they have available. It may change over time, but for now, the best thing that you can do is really kind of focus on the basics and start to build up from there. So making sure that researchers know, to document their data to standards, importantly to use a repository to share their data, and to get a PID. And as we heard earlier to only assign one PID, every data set that they curate, and to assign an appropriate licence. And you'll hear a little bit more about the services tomorrow. So I won't go into too much detail there. But I think it is trying to think about how you can join up a little bit across your institution and leverage expertise and knowledge from different parts of the institution to help you, especially if you're the only person in the data stewardship role at your organisation. So start with the basics and try to think about how you can, how you work in your organisation and what other operational units might be involved in helping you to do this.