
Extending The Organisational Profile Document To Cover Research Data Management

- Gemma Ellis Welcome to the Jisc podcast. Over the last few months we've been conducting interviews for projects in our Research Data Spring programme for more about the work, their progress and how they've found the process so far. We also want to find out more about the people behind the ideas and get to know them and their aspirations. Here we find out more about one of the projects that's looking specifically at the problems with research data management.
- Joy Davidson I'm Joy Davidson and I'm Associate Director of the Digital Curation Centre. It's a partnership between the University of Edinburgh, the University of Bath and the University of Glasgow, we're a Jisc funded project and I am based at HATII at the University of Glasgow.
- Gemma Ellis What is your project called and what is it going to do?
- Joy Davidson It's called extending the organisational profile document or OPD to cover research data management. Essentially what we were trying to do – we're now in the second phase of funding through the Research Data Spring – in the first phase of the project and what we were really trying to do was to agree within the community a basic list of research data management infrastructure components, and this was in light of the Engineering and Physical Science Research Council's policy framework on research data which came into effect in May this year. Basically we were trying to help institutions to know what sort of infrastructure components they

need to have any chance of complying with this policy framework. That was one of the key goals.

The second was to make that sort of infrastructure more visible and easier to identify. I think the key thing to remember is that research data management is not a single service. It is a set of different and distributed services across the institution being delivered by different operational units and very difficult to kind of get a holistic view of what's going on. So the basic profile work was meant to make this easier for institutions to be able to identify and to promote to researchers, to make sure that it was visible and usable to them to help them in their grant applications and through their project work.

So we identified during that first phase 15 different components that would be needed to deliver a basic research data management service within an institution and I think it's important to stress here that it wasn't all technical infrastructure. We also were looking at training, support and policy as these are probably more important in many ways than the technical infrastructure.

The second part of the work was to try and make this more visible and what we decided to use there was something called the organisational profile document or OPD, and this was something that was developed by the Equipment.data group at the University of Southampton to help universities meet another of EPSRC's mandates, which was about exposing information on research equipment which was bought with public funds. It's basically just a simple RDF file and it enables this equipment data to be made more visible to both humans and is machine readable. Within this profile, this OPD, it's a very small set of mandatory fields that have to be completed, but when we were working with this group previously on other work we thought there was great potential to extend the profile document to cover research data management infrastructure. So part of the work that we had to do in the first phase was to extend the Linking-You vocabulary which was developed in an earlier Jisc funded project, so we came up with an additional set of Linking-You vocabulary terms which would be required to make the RDM infrastructure components visible and machine-readable. So that was the second part of the project.

We completed that up until – I believe it was June this year when we did put forward the Phase 2 pitch for the Research Data Spring and in the second phase what we're really trying to do is to take that basic profile and to really understand what it means in terms of delivering these components, these different services and bits of infrastructure. Instead of just being able to say there is something in place, we now want to know who delivers that service, which operational unit within the institution is responsible for delivering the service and how many staff are involved in delivering the services, what kind of skills do they have. We're really trying to get a much more in-depth view of the individual components and to try to build up a much more detailed picture of how these beta services are being offered within the institution. We only started in September so we're a little bit behind some of the other Research Data Spring projects. We'll be working with five universities. We've got a couple already identified and we'll be working to pin down the remaining three between late October and early November, so that will

be the University of Glasgow and the University of Lancaster have already agreed to take part and we'll be looking for a further three institutions to work with.

Gemma Ellis So why did you choose to tackle the particular problem of research data management?

Joy Davidson Well the DCC, we have been funded by Jisc to support research data management capacity with universities in the UK, and for the last few years we've been working in a more one-to-one capacity with a lot of institutions trying to help support the implementation of new services and developing policies and various different aspects. And the key thing that we've really kind of found over the last couple of years is that it is not a single service. There is no one way that any institution is approaching this problem. There's a variety of different approaches being used, lots of different bits of infrastructure. In some institutions it might be the research office taking the lead; in others it's the library, and in others it might be the IT group that's taking the lead. So there's really no single way to approach this. What we wanted to do, we've been trying to promote good practice within institutions, so that's one of the kind of key challenges we're hoping to address with developing the profile and helping people to be more visible about what they are implementing in their own institutions.

I think the second thing we realised over the last few years in the lead up to the EPSRC policy framework on research data coming into effect, a lot of institutions struggled a little bit to translate what these requirements meant in terms of infrastructure that they needed to have in place, so hopefully by having this list of 15 components that we've drafted within the community we can all agree that we think we know what we all mean when we say research data management infrastructure, and that will give us a good way to start to think about where there is currently good practice and good things happening within the institutions but also hopefully identify the gaps where people are still struggling and will need some additional support to try and move forward.

I think the third thing, again, is just that there is no really easy way for any institution to try and find out how they can learn from their peers, so without this sort of comparability it's very difficult to try and help people to progress learning from what's going on.

Gemma Ellis What are you hoping will change as a result of your project?

Joy Davidson One of the key things we're really hoping to try and take forward through the work – better awareness of the range of different infrastructure components being offered across the institution, so easier for the people who are signposting researcher to use various bits of infrastructure to know where they are and to be able to do that easily. Because I think the other thing we had to remember is that anybody who is taking this on quite often has to do this as part of their core duties in other various different activities across their job profile. There is no extra time really being built in to do this work so people need to make it as easy as possible. So hopefully this will make it much easier for people to be able to signpost what's going on in the institution. But we also want to make sure that the really good infrastructure development that's happened in the lead up to the EPSRC policy framework coming in to effect, we want to make

sure that institutions can maintain that because what we're seeing are an awful lot of pilot implementations going on where you might have guaranteed funding for two to three years, and what we don't want to see is at the end of that period there's no case to take this forward. So we want institutions to get a better picture of the individual infrastructure components that they have in place and what they're offering, and to be able to really know what sort of effort is being allocated to this, what kind of activities are being undertaken and to get from that a better idea of the costs for delivering these specific services and infrastructure components. This will, in turn, hopefully help them to develop better business cases and hopefully get better sustained funding for the services to be turned into more permanent services rather than pilots.

Gemma Ellis How have you found the process so far?

Joy Davidson The process has been quite different. We've been working with Jisc for well over a decade now and we were much more used to the previous way of doing things, but I do think this approach is very helpful in terms of avoiding duplication of effort. I think we're definitely getting more of an awareness of what different groups are doing. The sandpit events really help you to get a chance to talk to people and to see early on before you start any work where there might be potential for a collaboration, and I think that this has really kind of meant stronger consortiums going forward. There's been a lot of really great ideas being pitched and over the couple of days we have seen very small groups of people doing individual things coming together to form bigger groups that are doing much more aligned work and hopefully we'll have a much better impact as a result of that collaboration. So I definitely think the approach is moving in the right direction and we're definitely seeing some positive results from it.

Moderator And what are you most looking forward to in terms of things that are coming up and project developments?

Joy Davidson I think we're all aware of the fact that budgets are getting tighter within universities and competition for those funds really requires making a very good business case, and I really do hope that the work that we're doing here with the OPD and the profiling and understanding how the services are being delivered will help to make people better equipped to make that case for continued investment, because I think infrastructure is going to be key to having any kind of good research data management. This is going to be a long term challenge for institutions and we're just getting to grips with understanding what it's going to cost, and I think it will be an evolving picture and how that will be funded over the next 10-20 years will be something that probably changes over time, but hopefully we can start to get people to get to grips with the service delivery and understanding the implications in terms of costing and resourcing these services and better able to make a good case for further investment.

We're hoping that we can start to make the profiles more comparable so that institutions can learn from each other, can understand different approaches that are being taken to deliver different parts of an overall research data management service, so can I learn from my peers about how I might be delivering training and support, how I might be delivering a long term preservation archive component. So we really want to be able to make sure that there's this

external comparability as well and ultimately we want to be able to use those profiles to really build up a picture of what's happening nationally. We want to know where are people doing well, where do they need additional help, where particularly can Jisc help to make sure that the universities in the UK are able to meet these requirements, so where additional services might be needed we hope to be able to articulate that back up to Jisc.

Gemma Ellis Have you experienced any benefits in taking part in Research Data Spring?

Joy Davidson I think for me it's really helping to focus my understanding of the problem area, and I think it has been very beneficial for us to try and pin down exactly what we mean by research data management infrastructure. Having that basic list of components has really helped to kind of separate out some of these activities and has made me a little bit more effective in my own role within the DCC in trying to help people to start to address this, because it can be quite a daunting prospect, particularly for smaller institutions who don't have a lot resources to hand it can be difficult to know where to start. By breaking it down into the individual components we can at least help people to pick an area that might be feasible for them to make a start with, rather than this daunting kind of notion of research data management as a single thing we can help them to start to look at it in a more realistic and manageable way. So that has helped me in my role within the DCC.

Gemma Ellis What about the next steps for your project?

Joy Davidson So as I mentioned we started a little bit late, we started in September. So what we've been doing for the first five weeks or so now is really just trying to get the list of questions for each of the 15 components we identified in phase one. We're coming up with a fairly detailed list of questions around each of these components to really come to grips with how they're being delivered within the institution and understanding that kind of resource allocation, governance issues and costing. So we'll be finalising that list of questions over the next week or so.

What we'll be doing then, late October and into November, is working with our five institutions to really develop the more detailed profiles. In addition our partners at Southampton will be developing an OPD checker. They have an OPD checker so you can come to their websites and look for an institution - if they have an OPD it's exposed through that checker - but we realise that maybe there are other places that we might need to have a checker, perhaps on the Jisc website or maybe the DCC website, so they will be coming up with a wireframe sort of prototype of whether that might be feasible, and also helping to support institutions to develop their profiles to make them more visible both to humans and to machines, so that will be happening.

We'll also be preparing a pitch for Phase 3, and that will be taking place at the Research Data Spring sandpit in early December. In Phase 3 what we're really hoping to do is to move from the basic profile that we developed in Phase 1 to the more detailed understanding of what it means to deliver these service components in Phase 2. In Phase 3 what we're really hoping to try and develop are metrics because, as I mentioned, a lot of what we're really hoping to do is help people to make a business case and without developing metrics it's really hard to make that case.

So we want to try and come up with some starting points for helping people to think about what is a good service for data management planning or what is a good service for short term storage of research data, how do we start to build up a picture of how to measure these services that will be really essential for the longer term business planning activities. So that is something we'll be pitching for Phase 3, and hopefully again we'll be able to progress the basic profile as a CASRAI standard.

Gemma Ellis Are there any particular metrics or measures you'll be using to show your project's been a success?

Joy Davidson Yeah. Obviously we're only working with a small group of institutions in this phase, so if we can complete these more detailed profiles for each of those that will be obviously our first metric of success, but beyond that we really want to see other institutions being able to benefit from this basic profile and the list of questions to understand how they are delivering the services. We're currently looking to take this forward into some DCC work that's planned, which is helping institutions to self-assess their progress in terms of developing research data management, so that will be another measure of success, whether we can take some of this work forward, and hopefully we'll be able to start to see more of these basic profiles being exposed through the organisational profile document as well. So if we can increase the visibility of these that would be another great sign of success for us.

Gemma Ellis Some lighter hearted questions now. So if you could describe your project in three words what would they be?

Joy Davidson I think the key three things we really wanted to get across is, number one, awareness of the range of different infrastructure components within the institution, visibility of these components making sure that everybody, both the support staff and the researchers who are supposed to use them, know where they are and how to access them, and this notion of comparability – can we start to show different approaches and to help people to see how they can start to mix and match different approaches to deliver a service that fits their organisation most effectively.

Gemma Ellis If your project was an animal what would it be?

Joy Davidson At the moment it's a beautiful autumn day here in Glasgow. I'll probably go for a walk at lunch in the park and I'll see lots of squirrels working hard to build up a set of stores for the winter. They do all this kind of digging and burying their food for the winter and they have a unique ability to remember where all of these different food stores are and they can go back over the winter and find it when they need it. So in that respect I'm hoping that we'll be able to help make all of these different distributed bits of infrastructure, so the resources, more easily visible so people don't have to work quite so hard as the poor little squirrels do to find and use the research infrastructure that they have at their own institutions when they need to make use of it.

Gemma Ellis And finally what should your project's advertising slogan be?

Joy Davidson The DCC has a strapline, 'good research needs good data', and I would say it's an extension of that. We think that good research needs good research infrastructure but that also means that you have to know where that research infrastructure is, what you have and that others need to know what you have as well. So you need to make sure that researchers are able to find it and to access it when they need it. So good research needs good research infrastructure.

Gemma Ellis Thank you Joy.

[ENDS]