Research Data Management Case Studies: Video Transcripts

# Professor Anna Phillips, Panini Project

## Video 1

Hi I am Professor Anna Phillips, I’m a Professor of behavioural medicine in the School of Sport, Exercise and Rehabilitation Sciences. I’m the Principal Investigator of the Panini Project, which stands for Physical Activity and Nutrition Influences in Ageing. The Panini project is multidisciplinary so its going to actually generate quite a lot of different types of data. First of all there will be questionnaire data such as perceived stress, social support, there will be diet diaries, there will be actigraphy which is activity monitoring, there are going to be blood samples taken so there will actually be gene sequencing and epigenetics information, immunassays, a whole range of physical, biological and psychological measures. The data will be stored for each individual project, and there will be 11 projects within Panini, at the actual site where each student is going to be trained initially, but what we are going to do is hopefully have a gold standard of different measures that each project tries to use, and then brings all that together in a shared dataset and that will be stored initially in the University of Amsterdam, and then brought over to the University of Birmingham for long term storage and analysis. Some of the real challenges with the data sharing and data management within Panini are that we are working across 8 different organisations, and some of the project at each of the organisations will be collecting new data, some of them will be analysing data or doing follow-ups on existing projects where they already have consent to collect the measures that they planned to collect but now because we want to share the data across the Panini network they are going to have to get new consent and permission to actually share their existing data more broadly than their original projects. So trying to fit those two things together the new data and existing data access is a bit of challenge. Also, just agreeing how we are going to identify the different types of data and name it in order to blend it all together in a way that’s identifiable, in terms of who generated this data, when, what is the population, what are the different measures, this all needs to be labelled in a way that we can actually understand what we are looking at for our future analysis and also for when we make the data more open access.

## Video 2: Horizon 2020 open data

It’s funded as part of the Horizon 2020 project that the European Commission fund, and Birmingham is the lead organisation running this particular project. When I was applying for the Panini project initially, one of the boxes on the form asked whether we would be willing to be part of a data sharing data management pilot. And I thought that this was probably something that you would see a lot more in research as time moves along, from a whole range of different funders, who would want to know exactly how we are going to manage our data and whether it is going to be open access and shared at the end. So when I saw that option I thought we might as well take this choice now and learn how to do this rather than wait until it’s obligatory and then have a very steep learning curve. So the first DMP this had to be done within the first 6 months of the project and the project started in January this year. So we started writing that, drafting that up and sending it round all the partners straight away for their comments and edits. We are then expected to at least review this annually, and definitely have 2 new versions during the life of the project (which is 4 years) so we will be doing that as part of our supervisory board management team activities during the whole project. Within Horizon 2020 they have a European Commission portal where you can submit all the deliverables on your project and deliverables can be everything from the science and the reports that you have done during the project, right through to more administrative type things, such as finances or the data management plan. When they started the data management pilot the European Commission came up with their own online website and advice for how you might structure your data management plan and then somewhere to store it online as well, so we have completed that in their own online database but also we have to submit it as a deliverable directly to the European Commission to show them that we have thought through these issues.

## Video 3: Moving towards open data

I think if you haven’t ever thought about Research Data Management one thing to do would be to go on at least the online training course that we have here at the University of Birmingham (if that’s where you are) or at similar organisations. That was very helpful to me just for things like the terminology and that you need to think about file naming protocols, but also that there are a lot of other resources out there on the web in terms of the FAIR principles of freedom of access and how accessible data needs to be and where you would store it online and how you would open that up to researchers worldwide. So I’d certainly use those online resources to try and learn exactly what is needed. I think open data will affect my research in possibly negative and positive ways - the negative might be that I’ve opened up all my data and haven’t quite finished completely mining it and getting the most from it myself before somebody else has access to it to do that. But on the other hand you are never working on just one project at once so you probably will never completely finish with your own data – just because its open access doesn’t mean that I can’t carry on working with it but it also means that other people with other ideas might be able to work with me or separately on it and get the best from it. I think the benefits from open data for me are access to data that I don’t currently have myself where I can answer questions that are related to my research but I don’t have any data collected on that. And for academics generally I think it will hopefully increase the chances of collaboration globally so that we can work together on bigger projects rather than doing small projects and then piece by piece publishing the results of those. I think open access means that we can really pool data together to try and get to the bottom of our research questions.

<http://intranet.birmingham.ac.uk/rdm>

<http://www.birmingham.ac.uk/generic/panini/index.aspx>