RDM Legacy access and data reuse – transcript

Edinburgh University Data Library [Data Library]. (2013, November 8). *MANTRA – Jeff Haywood – RDM Legacy access and data reuse* [video file and transcript]. Retrieved from <http://youtu.be/BRwcPhavElI>

I think that one of the problems that you have as a researcher with respect to your data and time, is that you are very much in the here and now. And for many researchers, not for all, but for many researchers, each project is a new project, and you don’t really see it as a longitudinal process into the future, that at some point you yourself may wish to come back and re-use data.

Data re-use is often talked about as re-use by others, but isn’t necessarily the case, that that’s always, that will be your concern. Your concern may actually be going back to your old data and re-using it, and I have certainly accumulated, I accumulated over a period of twenty years longitudinal datasets without ever planning really to gather those datasets, as such. And at a certain point I wanted to go back and relook at old data that would have been gathered ten and more years ago, and I had firmly in my head that those data existed, and actually an assumption in my head that I could, I could re-read them, that I could recover them, and after many months of searching on old machines and in filing cabinet drawers, I sadly had to come to the conclusion that I had actually lost the data that predated the year 2000. And it wasn’t a Y2K problem; it’s just that I had never systematically ensured that I had gathered the copies, and that I had actually turned the data into a readable format for now. So some of the old files that I was able to find was on proprietary software that I could no longer get hold of, some of it was sitting on drives, zip drives of the kind that actually were unreadable now. And so, as a consequence just by accident and by forgetfulness, what was actually a 22-20 year-long longitudinal dataset, if reconstructed and worked up, actually could not be reconstructed.

And had I thought on it consistently as I went through, what will happen to this data in a few years’ time, how will I ensure that I am able to go back to it later, I would have actually taken actions at those points, but of course you can’t go back and re-run time. And each of those, each of those individual actions, where I failed to maintain the integrity of the data over time, were purely accidental. I just didn’t think about it sufficiently carefully at that point. So I think that data management planning has certainly got an immediate consequence, how do I look after these data now in the short term, but it does also make you think about how will these data continue to be available to me as well as to others, into the future, or indeed to other members of a research group when I have left it.