Managing The Data Deluge

Before you start working with data, you have to create a plan for managing that data. Paul Jeffreys discusses what it takes to effectively manage today's deluge of data.

It has been said that more data has been created in the last year digitally than in the history of mankind before. And this, of course, affects everybody across society. Within the university world, it affects the way we do research, and it affects our teaching and learning.

I think you have to create a data management plan before you take any data. You have to decide how you're going to collect your data and curate it. And this has to be in collaboration with a person who's funding it: the research services, the group in your organization that looks after the research that you do and has an overview of the research that you do; the researcher himself, or herself, of course; but also those who, in due course, assess the quality of the research. And this data management plan becomes the thing that drives the life cycle of the data collection.

So, initially, you provide the researcher with a way of accessing his or her data effectively so that they see a current right at the very start, something that will be a benefit to them. Then you have to capture the metadata from the beginning. You can't retrospectively capture that additional data that tells you about the data so that you can find it, use it, or reuse it. You need some form of global register that you put this metadata into with agreed-upon standards, so that the data you collect doesn't have to be held in the same place. It won't be held in the same place.

And so, it's actually a very wide collaboration. It involves the researchers; the IT people; the librarians--the library people have the best expertise and understanding of metadata--the research services, the groups that manage the overall research that you do in your organization; and the funders; the people who measure your output. So it's a very broad activity. This is the biggest challenge facing higher education today.

This is applicable to all sorts of organizations. Any group that is beginning to face the data deluge, as people call it, the vast amounts of data that they will be collecting--intelligence information or senses of the services they provide--are going to have to face the issue how to deal with it. You can't just let it flow in without having thought about it.