

# 1st LEARN Workshop, Embedding Research Data as part of the research cycle

London, 29th January 2016

## Breakout Session Group 3

Chair: **Ignasi Labastida** (University of Barcelona)

Rapporteurs: **Barbara Sanchez Solis & Raman Ganuly** (University of Vienna)

### Breakout group initial participants list:

Barbara Sanchez Solis	University of Vienna
Frances Madden	Royal Holloway , University of London
Geoffrey Boulton	University of Edinburgh
Hiten Vaghmaria	University of Westminster
Ignasi Labastida	University of Barcelona
Jonathan Eaton	Royal London Hospital
Lucie Burgess	University of Oxford
Lydia Johnson	University of Nottingham
Marta Teperek	Cambridge Big Data
Michael Ball	Biotechnology & Bio. Sciences Research Council
Mura Ghosh	Senate House Library London
Nele Vermaete	Katholieke Universiteit Leuven
Rad Shahbazi	University of London
Raman Gagnuly	University of Vienna
René Schneider	Haute Ecole de Gestion, Genève
Rodney Amis	University College London
Sarah Molloy	Queen Mary University of London
Stephen Grace	London South Bank University
Vicky Wallace	University of Birmingham

### Policy development at national/regional/research institution level

In the UK, many institutions have issued a RDM policy in the last years. One participant from the University of Leuven/Belgium informed the group that an RDM policy had not yet been implemented there. Two participants from the University of Vienna reported that in Austria, no university had issued a RDM policy so far. In Catalonia, university libraries are working together to develop a white paper aimed at asking for institutional policies.

### Engagement with researchers

Young researchers need to change their approach. They are often disappointed that they cannot publish in journals like Science, Cell etc. One way they can be encouraged to share their data is through recognition. They should be motivated to share their research data in a repository, with the dataset having its own persistent identifier. In this way, they can get recognition for data that does not appear easily in journals, and for software, too. For data and papers, different descriptors should be used. Researchers should get citable data onto systems. Ideally, RDM should be an integral part of research. What has to be kept in mind is, what is the researcher's requirement for their data? Funders say one thing, but what about the researchers themselves? After all, they have got a lot of rules and administration to consider.

### **Incentives to help encourage researchers to share their research data**

Report from the Library University of Oxford: They have funding from JISC (data papers, publisher submission systems). Citation is one of the things that gives researchers a benefit. The problem: Support is not done systematically.

Report from Cambridge: Data services are currently being developed. When applying for a grant, data need to be open. For these benefits, researchers have to share.

The question was raised whether there needed to be a massive shift in organisation. Maybe, instead of saying incentives, universities should make judgements on what (which data) matters in the future. Yet we have to ask ourselves to what extent does a university feel free to make its own judgements? From the IT perspective, the best benefit for the researcher would be to provide an adequate infrastructure. Storage on hard disc is very common and in this way researchers would be encouraged to use more secure systems from the beginning of their research processes.

### **How do researchers feel about the Open agenda and Open research data?**

Citation seems to be the most important thing that matters. The existing way of doing things (publication in high impact journals) is prioritised. However, this might have to be scrutinized. It was expressed by some of the participants that metrics have to be treated with caution. In the UK, the Research Excellence Framework (REF) has a huge impact.

### **What are the training and CPD needs of researchers for RDM?**

There are RDM activities in certain academic disciplines. Within the institutions, services have to be adapted to meet researchers' needs. The value of what they do has to be recognized.

### **Costs of providing RDM infrastructure and support**

There was a discussion whether storage infrastructure should be local or national. Is it worthwhile that all institutions install their own systems? An argument in favour of institutional repositories is visibility (also of the institution) and the question of ownership of research data and exploitation rights. The University of Oxford just started running a data repository. A good option often seems to be discipline-specific repositories. Maybe institutions should concentrate on areas of deficit?

One participant from IT Services pointed out that many researchers have their own subject disciplinary data repositories to store raw data, but they do not always have a platform for processing the data. This is something an institution could provide for.

It was mentioned that disciplinary repositories can be precarious, also in terms of persistence. Cross-disciplinary archives like *Figshare* are not necessarily the safest place to share data. Universities, on the other hand, are stable and tend to persist in time. Yet they cannot fill all the disciplinary gaps. Moreover, there are considerable costs involved (not only hardware, but also people and processes behind and costs of curation).

Although costs are part of the Data Management Plan, how can people know about the costs? It is often difficult to forecast costs for projects. In the discussion, it was deemed important that costs need to be justified. The University of Oxford budgets costs for 5 years. Not everything can be put into overheads. The question is how direct costs can be recovered. There was the general observation that researchers do not ask funders for monies to meet RDM costs.

It was also mentioned that this cost should be included as a general cost of research (Boulton)

### **Barriers to moving towards a research environment where research data is available for sharing and re-use**

The global problem is how to develop from a small community to a big community which shares data. Young people care about their visibility, but they become worried about what happens to their data. The sense of community responsibility is also important. All disciplines have their history, how they did things. Maybe a general recommendation cannot be made.