2nd LEARN Workshop | Vienna | 6th April 2016

Research Data Management towards Open Science – Round Tables

Notes of Round Table 2

Role of policies concerning research data

Moderator: *Ignasi Labastida*

Rapporteur: *Gema Bueno de la Fuente*

Main questions:

a) What data should be assessed for deposit and retention?
b) What is "research data"?
c) Should open educational resources be considered in a research data policy?
d) Are metadata research data?
e) Should research data be open by default?
f) Are there exceptions?

b) What is research data // a) What data should be assessed for deposit and retention

Participants agreed that in order to fit all the possibilities in every discipline a wide and open definition for research data is required. It may be counterproductive to state a restricted and overly specific definition of research data at the institutional data policy.

Everything that is used or produced, or collected by researchers along a research process could be considered research data, and thus potentially assessed for deposit and retention. It may include digital and non-digital assets, in any medium (text, image, audio, video, software).

The researchers should decide what their research data are. What defines its status of research data are the use and the purpose given to it.

Some participants commented on the issue of the data and information lifecycle. If there is a final product as a result of the research process (paper, video, graphic materials), it should be considered as a publication and therefore could be deposited on a publications' repository, and not in a research data repository. It could be argued, however, that the research data, and the final results as publications, are closely related and their links should be preserved.

The dichotomy between institutional and subject repositories was also briefly discussed. This issue that led to extensive and intensive debates at the beginnings of the publications

repositories is now being brought back to the table. The difficulty of implementing research data repositories at universities that are multidisciplinary by nature was highlighted. If a one-size-fits-all solution is in place, some functionalities of research data may be lost.

One participant stated that institutional policies should not be mandatory, but they should offer researchers a place to deposit; in case researchers do not have any other place to deposit their data, as a subject repository.

It was also argued that if there is an institutional open access policy for publications, it should apply to research data.

Finally, participants agreed on the need to consider the whole research lifecycle and not only the deposit of research data at repositories. Services supporting and accompanying this lifecycle should be put in place.

c) Should open educational resources (OER) be considered in a research data policy?

There is no clear position in regards to OER to be considered research data, it will depend on the nature of these OER, and its potential use for research. There would be different approaches depending on the institution, and the discipline. Attendees agreed that it could be included in a data policy as a recommendation, subject to professorial decision. The need for clear licences and authorship metadata was discussed. The issue of intellectual property ownership for lectures and syllabus at an academic institution was also raised.

Participants from research institutions without specific teaching tasks were the most reluctant to consider OER in such a policy.

d) Are metadata research data?

There was a clear agreement that metadata was a part of the research data, and that it should be included by a research data policy.

The permissions and limits about access and reuse of metadata should be clearly stated within this policy, and a licence attached to metadata records.

e) Should research data be open by default? f) Are there exceptions?

There was a general agreement on the need of research data to be open by default, as long as there is an opt-out clause, although participants shared their concerns about who and how could decide if a dataset could apply for an opt-out. The research data policy should

include a list of potential reasons to be considered for opt-out (such as privacy, confidentiality, or even commercial purposes). These considerations should be taken into account from the beginning of the research project, as part of the research data management plan. The intention of applying for an opt-out should be stated in advance, along with a sound and reasonable justification.

Another key ideas and comments from round table participants were:

- Is there a need to assess the quality of research data to be deposited on an institutional repository? Is it reasonable and/or possible to establish some quality filters? Should there be a reference to data quality at the data policy level?
- The research data repository should not be a data dump, so a minimum filter and requirements should be implemented in order to have curated data. It is important to make it clear to researchers that the repository is not just an storage service, but a curatorial one
- Institutions should be responsible for research data created by their community along time, so the data policy must specify the procedures and measures to be taken in the case of researchers leaving the institution, as it is often the case for PhD students once they finish their pre-doctoral fellowship.
- Researchers should be trained from the beginning of their careers or when joining the institution, on research data policies and plans. They should be aware of what they are expected to do with their research data at the institution and which services and tools are available to them.
- There is a clear need for advocacy that shows benefits of research data management to researchers individually. It is important to make it easy for them, providing infrastructure and services.