3rd LEARN Workshop | Helsinki | 28th June 2016

Make research data management policies work: report from the Breakout groups

The session was organised around three breakout groups each devoted to specific RDM related topics:

1) Research Data services' planning, implementation and governance;
2) Mechanisms for competencies and skills development; and
3) Making possible and encouraging the reuse of data: incentives needed.

The schedule for the session was set to have two consecutive rounds of the three simultaneous breakout groups. Participants were offered to choose two of the three break-out groups at registration, using a card set and a colour scheme for groups and rounds, thereby aiming to ease the process of choosing the group, and at the same time limiting the number of participants per round and achieve balanced discussion groups

Each breakout group was serviced by a chairperson and a rapporteur provided from one of the Project partners. All groups were given specific briefing notes about their topics and suggested questions for discussion (see Appendix C). Rapporteurs took notes during the session and compiled full reports after the event. Below is the summary of the rapporteurs' notes from the different breakout sessions.

Every group opened the session with the Chair asking the floor whether their institution had developed and released an RDM Policy, and if so, who had been involved in its development.

Representatives from at least ten institutions reported to have a RDM policy in place as well as RDM services to some degrees. These were mainly research institutions from Finland (University of Helsinki, Aalto University, University of Turku, Tampere University of Technology, Academy of Finland); United Kingdom (STFC Rutherford Appleton Laboratory, University of Oxford), as well as KU Leuven (Belgium), Leiden University (Netherlands), and the University of Tartu together with the Estonian Research Council (Estonia). Some representatives reported that their institutions were currently developing an RDM policy (e.g. London School of Economics and Political Science) or already have developed a policy which is not implemented yet.

Concerning Research Data services, all of the attendees were involved in Research Support Services to some extent, and some of them were currently designing and developing Data Management Services for their institutions. A common service provided for researchers is support for the development of Data Management Plans (DMP) that are requested for the projects financed under the H2020 program. It has to be noted that different approaches are possible, and it is not always the case that a service is created as a consequence of a RDM policy. On the contrary, it is common to find

institutions that are offering RDM services for their researchers although no policy has been developed yet. This approach prioritises the provision of solutions to current existing challenges.

Regarding the planning, implementation and governance of RDM services, the attendees agreed that university libraries should have a prominent role in this process . A multidisciplinary team including external experts should be established from the very beginning, on the design and implementation of RDM policies, the research administration and supporting the whole data management process. The importance of shared (technical) infrastructures was also addressed.

In this regard, a proposed solution was to establish "RDM groups/teams" with representatives from archives, libraries, ICT departments, researchers, research support and grant offices. Yet another approach is to establish services operating on two levels: Data Support (with the consultancy of IT teams) and Research Administration. On the whole, it can be observed that Support Services and libraries are increasingly converging at many organisations.

For the governance of these RDM support services, the importance of reaching researchers at the different departments, career stages and disciplines was highlighted. RDM services and staff should be available across the whole campus, and work on raising the researcher's awareness at the earliest stage possible, thus RDM training for doctoral students is considered crucial. Moreover, RDM services should address subject disciplinary differences and adapt their infrastructure and guidance to their specific needs.

Concerning the necessary resources to implement and maintain RDM services, several challenges were identified. In general, it was considered complex to estimate the budget needed to offer these services in the long-term. The main difficulties come from the costs of human resources, both in terms of training and service provision. Technical infrastructure and data storage facilities could be expensive but are more easily costed and justified. Shared services could be a solution for minimising costs at these different levels (technical and personnel). It was agreed that collaboration with other institutions for the creation of shared services is easier at a national level.

As it has been pointed out, one of the most significant challenges is the lack of necessary **skills for RDM services' implementation and offering**. The second breakout group was devoted to this issue, and analysed several options and strategies for competencies and skills development. The pros and cons of the different approaches were discussed. Some were considered to be more effective (as a summer course or other face-to-face training courses) but also too labour intensive and resources' consuming, while lighter solutions like webinars and online materials were considered to be more feasible but have a lower impact factor.

Some existing initiatives were mentioned, most of them addressed to researchers including PhD students, but just a few of them to librarians and other staff involved in RDM support services. At the moment this way of working frequently means a joint endeavour to learn together how to deal with RDM at an institutional level.

Several topics were considered relevant as elements of these training activities, including legal and technical issues, open data, linked data, semantics, among others. A general awareness about what data science is and its implications in different disciplines is needed. In regards to the training on

legal issues, a lack of legal clarity on research data was pointed as a generalised problem. Legal aspects are generally the area where support is more urgently needed. Participants suggested the creation of a network of legal experts that would able to provide support across institutions internationally.

The third breakout group was focused on the **reuse of data**, i.e what are the requirements and challenges to make it possible, and which incentives could contribute to its success. It was agreed that an environment has to be created wherein data reuse is possible both from a legal and a technical perspective.

The challenges are varied, but the most difficult to overcome are the cultural ones. At present, there is no widespread culture of data sharing and reuse. Researchers are not used to share and open up their data, and common trust about the quality of open data is not yet established. On the contrary, there is a common fear about the loss of control of their own data, and potential misuse by other researchers. There is a compelling need for raising awareness on the benefits of data reuse for researchers and for supporting them with the process. Trust must be established on the sharing and the reuse side.

Technical barriers are also tricky, as datasets are far more complex than publications, with a wide range of data formats and multiple standards. There is a high demand for support when handling data and RDM services are not always in position to provide this support in a timely manner. The technical focus should be on data organisation in order to facilitate multiple and different ways for accessing data. It requires a sustainable infrastructure which guarantees the quality of data over time. The trustworthy data sharing environment can enable and encourage data reuse.

For building trust on the reuse of data, the quality of metadata is essential. Traceability it not easy for data, and any transfer to a different database or storing server could challenge data citation and verification. Besides, there are problems of semantic interoperability related to different interpretations of terms at databases of different subject disciplines.

Again, legal advice is necessary because the researchers have to deal with a lot of questions about privacy, ownership and copyright of the data. Copyright clearance is compulsory for data reuse, thus a well-established and clear licensing framework is needed for the researcher to know the limits and obligations in the reuse of data. Attribution and data citation are also key to this issue.

In order to encourage researchers to make their data reusable, there is a need to highlight the potential benefits for the individual, in regards to acknowledgement, greater visibility and reputation. For this purpose, data citation policies and mechanisms are required, as well as new metrics adapted to the characteristics of data production and reuse.

To conclude the session, the importance of research integrity was discussed. Although it was agreed that it is not easy to reproduce research as there is no funding available for that purpose, at least underlying data should be available together with publications. Publishers are changing their policies in regards to data deposit and it has proved to have a greater impact than funders' requirements. Some questions yet to be addressed are the responsibilities of publishers on data curation and community management.