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

London, 29th January 2016

Breakout Session Group 2

Chair: **Wouter Schallier** (U.N. Economic Commission for Latin America and the Caribbean, Chief Librarian)

Rapporteurs: **Paolo Budroni** (Vienna University Library and Archive Services, Managing Director e-Infrastructures)

Breakout group initial participants list:

Aneesha Anand Mohan	Brunel University
Beverly Jones	University of Lincoln
Gemma Vart	University of Roehampton
Hannelore Vanhaverbeke	Katholieke Universiteit Leuven
Ilaria Marsili	University College London
James Wilson	University College London
Jennifer Quah	University of the West of England
Jo Lake	Economic and Social Research Council
Manoj Nanji	Queen Mary University of London
Martin Moyle	University College London
Matthew Siekierkowski	University of Huddersfield
Muriel Reigersberg	University of London
Naomi Mackay	University College London
Paolo Budroni	University of Vienna
Peter Murray-Rust	University of Cambridge
Rachel Bruce	JISC
Roger Newham	Buckinghamshire New University
Stuart Lewis	University of Edinburgh
Wouter Schallier	UN Economic Commission for Latin America and the Caribbean

Topic: Policies and terminology
Session II opened with the question, "Who among the participants has an RDM-policy?"
Representatives from five universities, including Oxford, Brunel, Edinburgh, Goldsmiths, and Leuven, took turns illustrating the current status of their respective policies/guidelines. Those guidelines often make individual researchers responsible for making data available (NOT: sharable!) for a certain amount of time (3-5 years) but do not go into further details of infrastructure, format, metadata etc.
The first key topic, "What kind of data should be preserved?" was followed by the question, "Who makes that decision? "
The next question asked was "Who should advise principal investigators? "

The participants answered variously "libraries", "IT-services"(in the case of Oxford), and "research services". Some participants explained that they avoided the term "policy" and instead preferred to use the term "guidelines" (policy vs. guidelines debate was a theme), because of its looser and less formal definition.

Some also avoided using the term "open"(as in "open data" and "open science"), because in some cases researchers oppose the term "open". It is also important to clarify towards researchers that "open data" does not mean that *all* data should be open.

The consensus was that there is a need for clearer definitions and that more attention must be paid to terminology.

Named terms: open, data, research materials, et al.

It was generally agreed that semantics and terms were a source of confusion.

Topic: Assignment of roles and competencies

There is a need to define the roles and competencies of the actors/stakeholders identified during the session:

> a) Individual researchers
> b) Institutions
> c) Funders
> d) Publishers
> e) Companies
> f) Libraries and IT – services

Topic: Sharing of data, legal issues, and institutional frameworks

During a question and answer period, much attention was devoted to data sharing and its inherent legal issues.

There were three main drivers identified for sharing data: funders' requirements, researchers' needs and institutional policy (for example to enhance the research profile of the institution).

Participants discussed the use of licenses and general legal issues. Particular attention was paid to the notion of agreements which could support the sharing of data, as well as templates which could help establish a legal framework for the sharing of data.

The University of Leuven has already introduced a "data access committee" to address this topic. Since this committee was created on the request of a commercial publisher, this information led to the question "Who controls the data publication processes?" The agreement was that researchers and institutions should take their responsibility, instead of leaving it to third parties.

The subsequent discussion concerned Human Resources Development and yielded the question, "What kind of profiles are required in RDM?"

It was reported that the University of Leuven offers a "course on scientific integrity."

Participants generally agreed that a need for "services for RDM" as well as so-called "central points of contact, " exists. Generally there seems to be a lack of people within the organization who are formally in charge of RDM.

Topic: Funders

This topic was of little relevance.

Further topics

The issues of a "costs model" and awareness about the "value of data," as well as a brief discussion about "models of basic services" concluded the session.

On the topic of cost, the agreement was that institutions should provide – "free of charge"- a set of basic services to its researchers for preserving/sharing data, along with a set of additional services that should

be charged separately. These basic services should include: a help desk and a basic infrastructure for storing data.

Most researchers seem to underestimate the value of their data. Institutions use different strategies to raise awareness: some keep research data separate from Open Access, others do exactly the opposite and sell (open) research data management as part of an Open Access policy. At least Open Access seems to be widely accepted by researchers nowadays.

Conclusions/Findings

1. With regard to the research life cycle and the definition of stakeholders, definitions and identification are needed.
2. Concerning the question, "What kind of data should be preserved?", the key to answering this question are: a) policies, b) guidelines, and c) guiding principles.
3. Guidelines/policies should be published and made available to the public.
4. The suggested length of such documents is one page and should be kept brief and concise, with only some general principles. Longer documents easily get outdated.
5. Terminology: Improvements in terminology and definitions are needed, especially for key terms like "data" and "research data".
6. It is important to clarify towards researchers that "open data" does not mean that ALL their data should be open.
7. Concerning the question, "Who are the drivers of such policies?" the participants expressed a need for identification and clarity. Participants suggested the identification of three categories of drivers: a) external drivers (e.g. funders' requirements), b) internal drivers (e.g. the researchers themselves), and c) institutional drivers
8. On the topics of awareness and building awareness, there is a need for increased awareness concerning policies.
9. RDM cannot be completely left to third parties. Researchers and their institutions should shoulder their responsibilities.
10. Human Resources Development: more detailed discussion of the human development sector is needed. Such a discussion could lead to the definition of profiles, training activities, courses on scientific integrity and the creation of central points of contact. Generally there seems to be a lack of people within the organization who are formally in charge of RDM.
11. Costs: A definition for the digital workflows in the research life cycle and an assignment of roles and competencies for the stakeholders involved could contribute to the definition of costs generated by RDM.
12. On the topic of cost, the agreement was that institutions should provide – "free of charge" – a set of basic services to its researchers for preserving/sharing data, along with a set of additional services that should be charged separately.
13. Most researchers seem to underestimate the value of their data. Institutions use different strategies to raise awareness.

Topics addressed:

Guiding principles, cost models, model of basic services, terminology work, definitions, definitions of stakeholders, actors, legal services (licenses, agreements, advice, consultation, the issue of "ownership"), dialogue with funders, training, dissemination, re-organization of processes (central point of contact).