



Briefing sheet for RECODE¹ Workshop WP4

Institutional support for open access to research data

1 July 2014, Riga, Latvia.

Background

The Policy Recommendations for Open Access to Research Data in Europe (RECODE) project addresses the drivers and barriers in developing open access to research data in Europe to draft recommendations for policy. Institutions, such as universities, libraries, data centers, publishers, professional associations and funding bodies play an important role in making open research data possible. They support researchers, provide infrastructure and funding, and set guidelines. In performing supporting open access to research data they face several organizational challenges, such as finding the financial resources to make open access possible or maintaining the quality and integrity of data. In this workshop we will examine some of these organizational challenges and various ways of addressing them.

The workshop will be a multidisciplinary space where a range of different types of stakeholders will provide their perspective on some of the challenges that we have identified in our research, suggest additional or alternate challenges and identify and evaluate proposed good practice solutions to address these challenges. Workshop participants will add their unique stakeholder and disciplinary perspective to the project's initial findings in this area. Through their participation in the workshop discussions, participants will also provide input into the policy recommendations produced in the RECODE project.

The aims of the workshop are to:

- Present our key findings from a review of select policy documents, academic literature and related documents about open access to research data and findings from the five case studies of research practice;
- solicit participants' feedback on the effectiveness of current institutional policies and strategies in enabling open access to research data;
- come to a better understanding of the challenges that institutions face and explore and develop potential solutions in different disciplinary contexts.

¹ The RECODE project is funded by the European Commission under the 7th Framework Programme, Grant no. 321463

In preparation of the workshop

During the workshop there will be plenty of time for discussion and feedback. To encourage the discussion we have provided a very brief overview of the challenges we have identified and some of the current developments that address these challenges. In addition we have formulated three idealized scenarios of the further development of open access to research data. The scenarios are intended to stimulate discussion during the breakout session about the roles and responsibilities of institutions in giving shape to open access to research data.

We ask you to read the overview and the scenarios and consider the possible benefits and drawbacks of the different approaches to addressing the challenges. In addition, we encourage you to think of good practices and alternative solutions that you are aware of and that you would like to share with us and the other workshop participants.

Challenges facing institutions

1. Evaluating and maintaining the quality, value and trustworthiness of research data:

- Different ideas about what quality is and how it should be evaluated
- Lack of time and resources for data peer review
- Developing strategies to evaluate and reward high-quality data management
- Selection and retention of data

Examples of current developments:

- Developing standards and formats regarding the quality of data in terms of interoperability and re-usability.
- Strategies for evaluating the quality of data, for instance through automated or human quality assessment at the ingestion stage.
- Strategies to ensure the scientific value of data sets, for instance through peer review at the time of publication or developing computer-supported methods that provide peer review at the early stages of development.
- Certification of repositories and datasets and repositories
- Development of digital object identifier (DOI) system

2. Securing funding for open research data

- Resources are scarce and although open access might reduce costs in the long term, significant investments are needed to make open research data possible.
- Strategies for the funding of data management, the data infrastructure and the long-term preservation of research data are still in the early stages of development.
- Data management is mostly funded through project grants and strategies for long-term preservation of data are lacking, both in terms of securing funding and selecting data to be kept accessible.

- Open access can generate new costs, such as maintaining disciplinary expertise and technology infrastructure.

Examples of current developments

- Collecting and sharing cost/benefit analyses
- Investments in data infrastructure and data management
- Funding agencies requirements for data management plans in project proposals
- Collaborations between institutions to achieve economies of scale

3. Training researchers and other relevant stakeholders in the possibilities of open research data

- Disciplines have very different needs and requirements
- The technical expertise is not always readily available in certain disciplines.
- Although libraries are assigned an important role in data management support, the necessary knowledge and skills for effective support are generally under-developed within these institutions.
- It can be difficult to translate the knowledge about data management to daily practices.
- There are relatively few 'data scientists' with the knowledge and skills to support researchers and research groups. There is a need for more training programs or career development opportunities for these types of scientists.
- Training and skill development for data management and open access require input from different stakeholders, including libraries, research communities and IT experts. Yet, it is currently not clear which institution is responsible for what.

Example of current developments:

- Development of data curation courses and programs for researchers, librarians and other staff
- Training provided by consortia and within projects
- Mandates for data management plans
- Data centres and libraries developing training courses in collaboration with research communities

4. Creating awareness

- Awareness about the possibilities and limitations of open access to research data is still relatively low, which hampers further development of open research data.
- Advocacy of open research data can generate considerable resistance within some research communities or from some institutions
- Engaging researchers is a challenge because of the amount of work involved in making data open, the lack of recognition, concerns about privacy, confidentiality, and misinterpretation of data, concerns about being scooped etc.

Examples of current developments

- Workshops, seminars, brochures, and flyers
- Mandates for data management plans
- Pilot programs
- Developing and stimulating data citation practices

Three scenarios

Most institutions are still in the early stages of development when it comes to open research data and they have various options in addressing the four challenges and giving shape to their roles and responsibilities. In making their choices they influence what open research ecosystems will become. We have therefore developed the following three scenarios to explore how institutions can participate in giving shape to these ecosystems. These scenarios are idealized and in practice will probably overlap. However, using these scenarios we want to examine, together with the workshop participant, the various choices and options that institutions have and what they will entail.

Scenario 1: Fully open linked data

Open research data is fully open and can be easily linked. Data is collected in certified disciplinary data repositories, institutional repositories and national data centers according to general standards and a limited number of formats.

Scenario 2: Long-term curation and preservation

A selection of research data is made openly available in certified data centers and repositories that guarantee long-term storage and availability of high-quality data sets.

Scenario 3: Disciplinary integrity

Research data are openly accessible in distributed temporary subject or project repositories, disciplinary repositories or data centres, and institutional repositories according to the standards agreed upon within research communities.

Table 1 on the next page outlines what the three scenarios mean with regard to addressing the four challenges.

Table 1: Addressing the four challenges in the three scenarios

Challenge	Scenario 1	Scenario 2	Scenario 3
1	<ul style="list-style-type: none"> - Key role for funders, data centers and repositories in setting standards - Emphasis is on interoperability and re-usability - Development of automated quality assurance tools - Centralized certification - Mandates for data management plans 	<ul style="list-style-type: none"> - Key role for libraries (institutional repositories) and data centers in setting standards - Emphasis is on data preservation and accessibility - Data librarians and data scientists perform quality control - Data curators stimulate and offer support for data management plans 	<ul style="list-style-type: none"> - Key role for scholarly societies, research institutes and publishers in setting standards - Emphasis is on scientific value - Various discipline specific strategies for quality assurance - Incentives for data management plans
2	<ul style="list-style-type: none"> - Funding for centralized data infrastructures and certified data centres and repositories - Research institutes, universities and funders contribute to data centres and repositories 	<ul style="list-style-type: none"> - Data centres and libraries receive long-term funding from funders, research producing institutions and other institutions 	<ul style="list-style-type: none"> - Funding comes from project grants, research societies and cost recovery strategies
3	<ul style="list-style-type: none"> - General training programs provided by data centers and repositories 	<ul style="list-style-type: none"> - Data centres and libraries are responsible for training and skill development of researchers 	<ul style="list-style-type: none"> - Research producing institutes are responsible for training and skill development
4	<ul style="list-style-type: none"> - Funders in collaboration with IT experts and data centres and repositories advocate the possibilities of open access - Funders create incentives for open access through mandates and reward strategies 	<ul style="list-style-type: none"> - Data centres and libraries inform research communities about data curation and open access - These institutions support research communities in data citation practices 	<ul style="list-style-type: none"> - Scholarly societies, research producing institutions and publishers stimulate open access - These institutions also develop and stimulate data citation practices - Research producing institutes develop reward systems that recognize open research data as valuable