The RECODE Project & Recommendations

RECODE (recodeproject.eu) -an FP-7 project funded by the European Union- has leveraged existing networks, communities and projects to address challenges within the open access and data dissemination and preservation sector. The sector includes several different networks, initiatives, projects and communities that are fragmented by discipline, geography, and, stakeholder category, often working in isolation or with limited contact with one another. RECODE has provided a forum for European stakeholders to work together towards common solutions to shared challenges.

To this end, RECODE has used five disciplinary case studies in open access to research data (physics, health, bioengineering, environment and archaeology) to examine four grand challenges:

- stakeholders values and ecosystems,
- legal and ethical concerns,
- infrastructure and technology challenges, and
- institutional challenges.

On the basis of this work, RECODE identified two overarching issues in the mobilisation of open access to research data: a lack of a coherent open data ecosystem; and a lack of attention to the specificity of research practice, processes and data collections. These findings along with the horizontal analyses of the RECODE case studies in relation to the four grand challenges, have informed the following policy recommendations on open access to research data.

These policy recommendations are targeted at key stakeholders in the scholarly communication ecosystem, namely research funders, research institutions, data managers, and publishers. They will assist each of the stakeholders in furthering the goals of open access to research data by providing both over-arching and stakeholder-specific recommendations. These function, as suggestions to address and attend to central issues that RECODE identified through the research work.

The current report thus comprises:

- summary of project findings
- overarching recommendations
- targeted policy recommendations for funders, research institutions, data managers, and publishers
- practical guides for developing policies for funders, research institutions, data managers, and publishers
- resources to expedite the process of policy development and implementation among stakeholders

The current publication is a short version of the Report "Policy guidelines for open access and data dissemination and preservation" available at the RECODE project website, along with other reports produced in the framework of the project.

The RECODE project findings

While a consensus is observed amongst many policy makers on the benefits of open access for science, industry and civil society, there are still important barriers that need to be overcome. RECODE identified in particular two overarching issues in the mobilization of open access to research data: a lack of coherent open data ecosystem and a lack of attention to the specificity of research practices, processes and form of data collections.

The project performed literature review of policy documents, current research, reports and projects, and conducted interviews in five disciplinary case studies (Physics, Health, Bioengineering, Earth Sciences, Archaeology) in order to address four grand challenges in open access to research data. On the basis of this work it developed overarching and specific recommendations for funders, research institutions, data managers and publishers.

Stakeholder Values and Ecosystems

RECODE studied the diverse relevant stakeholders, specifically their functions and values. Stakeholders were identified through the following functions: (1) funding and initiating, (2) creating, (3) disseminating, (4) curating, (5) using. This community of stakeholders shares multiple and occasionally overlapping functions and an overarching consensus on the benefits of open access to research data. The latter relate to the increase in productivity and quality of scientific work, the economic and social benefits obtained, while there is a clear shared perception of open access to research data as a general public good. Despite this consensus, RECODE showed that the road towards open research data is not perceived in the same way by the various stakeholders. This results from conflicting value chains, parallel and disconnected processes, especially between the current disciplinary specific research practices and increasing funder and institutional demands for open access to the former. Concerns are raised about the costs of research data, while the participation of the research community emerges as a critical point in the success towards accessible, intelligible, assessable, and usable open research data.

Infrastructure and Technology Challenges

The main infrastructure and technology challenges identified by RECODE project were grouped into five broad categories: heterogeneity and interoperability; accessibility and discoverability; preservation and curation; quality and assessability; security. RECODE research concluded that technological challenges are not perceived as a concern in implementing open access to research data when compared to financial, cultural and legal ones. In addressing the above challenges the project assessed that it is necessary to adopt technical and infrastructural solutions that holistically address the above issues. Attention is drawn to: open and interoperable standards, harmonized discovery and services, persistent identifiers, promotion of a culture for data management, virtualization technologies, research data that are fit for use, technical solutions for security and legal issues around open research data. The different attitudes in various scientific fields also emerged as critical in relevant policy development.

Legal and Ethical Challenges

RECODE examined and analysed legal and ethical issues in open access to research data. Legal issues focused on intellectual property rights (including copyright, trade secrets and database rights) privacy and data protection, open access mandates. Ethical issues focused on unintended secondary uses, misappropriation and commercialization of research data, unequal distribution of scientific results and disproportionate impacts on scientific freedom as well as other economic, social and scientific costs. RECODE demonstrated that stakeholders are often subject to conflicted legal obligations, resulting in a drain of resources as well as efforts to establish creative ways of dealing with the challenges. Researchers and institutions have already adopted strategies and measures to address potential legal and ethical issues, such as access control mechanisms, licensing and 'soft law' measures, and many of these strategies are used to address both legal and ethical issues. RECODE recommends the extensive use of open licensing and implementing technical solutions for legal and ethical issues, systematically turning institutional attention to developing solutions for legal and ethical problems arising from open access to research data, including internal review processes. Understanding that not all data can be open, RECODE recommends focusing on addressing when it is lawful and appropriate to provide open access to personal data and establishing better reward systems for high-quality data.

Institutional Challenges

Financial support, evaluating and maintaining the quality, value and trustworthiness of research data, training of researchers and other relevant stakeholders as well as awareness-raising on the opportunities and limitations of open access to research data were identified as key challenges faced by institutions such as archives, libraries, universities, data centres, and research funders. Institutions need to address the issue of sustained funding for long-term research data curation as a distinct need and consider scalable collaborative efforts. Research data quality is essential for reuse and long-term preservation of the growing volume of research data. While technical quality is being addressed, more attention should be directed towards developing clear guidelines for scientific quality. In doing so, it is essential to provide researchers rewards by including research data in evaluations, to have clear responsibility lines among stakeholders, and further explore mechanisms that contribute to evaluation, such as data journals and peer review mechanisms. Institutions are also expected to play a key role in providing training to researchers and other relevant stakeholders, such as data managers. In developing appropriate training and educational courses institutions are faced with the diverse needs and knowledge levels between and within disciplines, established research cultures and the pace of technological developments. Closely related to the above is the need to raise awareness on the opportunities and limitations surrounding open access. Institutions can have an active role in this respect too through the adoption of different strategies which nonetheless necessitate collaboration with other stakeholders.

Overarching Recommendations

The RECODE overarching recommendations are intended to direct consensus-building and action towards ten broad areas that were identified by project research as significant in view of enabling open access to research data. The broad nature of these recommendations is also intended to be useful and accessible to both stakeholders with very developed open access policies that could be improved and stakeholders with less developed policies. As such, they are supplemented by more specific recommendations for each category of stakeholder below. Finally, these overarching policy recommendations are necessarily geared towards decision-making stakeholders, but in all cases, we encourage these decision-makers to consult, involve and take seriously the perspectives and needs of the research community before developing policies or programmes. The RECODE project findings suggest that the development of open access to research data needs to be informed by the research practices and processes in the different disciplines and characterised by a partnership approach among key stakeholders. This will help ensure the engagement from the wide range of research communities and the embedding of open access within research practice and process.

The RECODE ten overarching recommendations are the following:

1. Develop aligned and comprehensive policies for open access to research data

Funder, institutional and publisher policies setting open access to research data as the default practice are necessary in transitioning towards open science. Policies should be consistent with national priorities and aligned with the European framework for open access to research data (2012 Recommendation and Horizon 2020), while also complementing that for open government data. Provisions should be made for the necessary resources that will allow policy implementation.

2. Ensure appropriate funding for open access to research data

Policies and mandates for open access will bring the expected results if accompanied by appropriate funds. Particular attention should be

directed towards provisions for funding the development and long-term sustainability of necessary infrastructures; training of researchers, librarians and other technical staff; innovative actions.

3. Develop policies and initiatives that offer researchers rewards for open access to high quality data

   Funder and institutional policies that offer researchers rewards for providing open access to high quality data are central in the transition towards open science. Official measures and processes need to be put in place to include the open sharing of research data in funding and professional advancement decisions.

4. Identify key stakeholders and relevant networks and foster collaborative work for a sustainable ecosystem for open access to research data

   The open access ecosystem comprises a diverse group of stakeholders with multiple and often overlapping functions. To be sustainable, collaboration is essential as it affords the gradual development of a coherent view among stakeholders, an agreement on their roles and responsibilities, the allocation of resources and alignment of stakeholder policies, while avoiding the duplication of effort and loss of r esources, as well as capacity-building.

5. Plan for the long-term, sustainable curation and preservation of open access data

   Stakeholders should draw their attention specifically to the long-term availability of high-quality research data. A strategy for long-term, sustainable curation and preservation requires leveraging resources as well as developing appropriate services and infrastructure. In doing so, the use of collaborative models should be considered.

6. Develop comprehensive and collaborative technical and infrastructure solutions that afford open access to and long-term preservation of high-quality research data

   Existing infrastructures should be further collaboratively developed to address in a comprehensive way data harmonization, discovery and access, preservation, technological obsolescence, documentation and

metadata, quality and relevance indicators and security issues, among others. Approaches should address the diverse disciplinary requirements and data variety, as well as metadata and data standardization.

7. Develop technical and scientific quality standards for research data

Stakeholders should collaborate in developing shared quality standards that will ensure the proliferation of high-quality reusable research data. Consensus should be built on the technical quality standards of research data, as well as on their scientific quality in line with disciplinary practices and norms. Appropriate strategies should be developed for the evaluation of the scientific quality of data.

8. Require the use of harmonized open licensing frameworks

Open licenses, like creative commons, describe the terms under which research data should be accessed, shared, and re-used. Their popularity is an indication of their utility and efficacy, yet further options for licensing should be examined, along with identifying mechanisms to enforce these licenses and developing new, interoperable licenses.

9. Systematically address legal and ethical issues arising from open access to research data

Open access to research data raises important legal and ethical issues, which should be addressed systematically by stakeholders. This can be done through the institutionalization of processes, dedicated fora, training, the use of technological solutions (e.g. machine-readable licenses) and the systematic pursuit for new and more efficient solutions.

10. Support the transition to open research data through curriculum-development and training

The transition to an open science paradigm where research data plays a significant role requires training and education for researchers and for data managers who support open science. Courses for getting researchers and data managers up-to date with current relevant issues are necessary, as well as the development of curricula that contribute towards the development of data science and information management as distinct and legitimate career paths.