



Policy
recommendations
for open access
to research data

www.recodeproject.eu

Copyright© 2014 RECODE project consortium

Editing English version & design: **National Documentation Centre**
Publication is available at: **recodeproject.eu**



This publication is licensed under
a Creative Commons Attribution 4.0 license.
To view this license deed, visit:
<http://creativecommons.org/licenses/by/4.0>



Policy
recommendations
for open access
to research data

Table of Contents

- The RECODE Project and Recommendations
- The RECODE project findings
- Overarching Recommendations
- Stakeholder-specific recommendations
 - | Funders
 - | Research Institutions
 - | Data Managers
 - | Publishers
- A Practical Guide for developing policies for Research Funders
- A Practical Guide for developing policies for Research Institutions
- A Practical Guide for developing policies for Publishers
- Resources

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 resources, 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.

Stakeholder - specific recommendations

Funders

Funding bodies are key stakeholders in the open access ecosystem: they develop and mandate policies that affect how data is managed, accessed, disseminated and preserved and how funds are allocated in the various phases foreseen in the process of making research data open. Research funders include the European Union (EU) and national governments, individual public funders that distribute competitive funding, non-profit institutions and private funders. This variability in the types of research funders, depending -among others- on their public or private nature, the size and effect of funding they mobilize for research and the country circumstances impacts on the measures and strategies they adopt.

The drive for open access to research data, especially those produced as a result of public funding, is justified by reference to the public interest, yet funder policies for open access to research data remain limited, especially when compared to those for peer-reviewed publications. At the EU level, the most prominent funder is the European Commission (EC), representing an important source of competitive funding for some member states. Thus, the EC can have a catalyst role in the formulation of open access policies for publications and research data among member states. Setting the example as a major European public funder, the EC has elaborated a comprehensive framework to support open access to scientific information, including research data. In 2012 it passed the "Recommendation on access to, preservation of and dissemination of scientific information" and formulated a pilot action on open access to research data in the context of Horizon 2020, the main EC funding program for research for the period 2014-2020. The Recommendation calls on member states to develop comprehensive and aligned policies and strategies that will ensure open access to publications and research data from publicly funded research. The Open Data Pilot is implemented in seven areas for 2014 and 2015 and requires open access to research data generated by the projects.

At member state level, UK research funders, the Research Councils UK and the Wellcome Trust, are global pace-setters in policy development for research data and in comprehensively developing relevant supporting services. In the rest of Europe, a great number of funding bodies have yet to develop policies on open access to research data or have no immediate intention of doing so, while most governmental policies and strategies concentrate in the field of governmental rather than research data. Beyond the EU, the White House issued a Directive in 2013, whereby all federal funding agencies with a \$100 million/year funding for extramural research or development should require open access in their policies, both for research publications and research data.

The most significant and effective funder policies set open access to research data as the default requirement for the funded research with provision for possible exceptions for legal and ethical reasons. They require deposit of research data supporting publications and other important research data in certified repositories. They require researchers to describe these and other provisions (e.g. evaluation of their data; long term preservation provisions) in mandatory Data Management Plans (DMPs), which are submitted with the grant proposals and evaluated. The costs for data management are usually eligible for projects. To secure the reusability of research data and the ability to identify and measure policy compliance, funders have introduced technical specifications in their policies (e.g. digital object identifiers (DOI), specific metadata standards etc.) as well as provisions on appropriate licensing. Most importantly, efficient policies include clear descriptions of responsibilities/ expectations for the main stakeholders involved: funders, researchers (either under their capacity of grant applicants or grant holders), research institutions, data centers and repositories, and publishers. With regard to monitoring some funders include provisions on the monitoring of their policies.

Current practices demonstrate that there is no one-size-fits-all solution: different countries have different approaches towards developing such strategies and policies, dependent upon local conditions. In developing related policies, research funders are encouraged to study the policies and practices of other countries and have a solid knowledge of important issues in their own country such as (but not limited to) the available infrastructures and support services, the diversity of scientific and scholarly practices.

Recommendations

1. **Develop explicit policies for open access to research data with clear roles and responsibilities**

Funder policies should set open access as the default for research data. Explicit policies with clear description of roles and responsibilities for each stakeholder (i.e. funders, grantees, repositories/data centres that curate the research data) are key in fostering change through their impact on research cultures.

2. **Adopt a comprehensive approach in funding the implementation of open access to and preservation of research data**

Appropriate financing and comprehensive planning is necessary for the following: collaborative and scalable infrastructures and services for access to and long-term preservation of research data; innovative actions that boost data-reuse in the research and innovation sector; development of skills among researchers and information specialists, both formal (curriculum development) and informal (training activities). In achieving this comprehensive approach, they are encouraged to mobilize complementary funding instruments.

3. **Reinforce the significance of the Data Management Plan (DMP) to embed and promote data management as a distinct activity within the research process**

Funders are encouraged to acknowledge the DMP as a distinct activity within the project and appoint data management experts to review and monitor their implementation. DMPs should be accompanied by the allocation of appropriate resources for the delivery of such plans and for monitoring researchers' compliance.

4. **Raise awareness and promote open research data in view of leading an open science paradigm**

Funders should engage in activities such as the promotion of good practices by specific researchers and research groups and/or establishing prizes for good practices in sharing research data, in view of leading

cultural change towards the open science paradigm.

5. Foster collaboration with relevant stakeholders and networks

Funders should take the lead in bringing together researchers, research institutions policy makers, data managers, publishers, in view of developing aligned policies and sustainable strategies and infrastructures for open access to research data.

Research Institutions

Research institutions refer to universities and higher education organizations engaged in primary or secondary research and to publicly and privately funded research institutes/centres. Research institutions hold a focal role in transitioning to open access practices, as the primary loci where researchers carry out and publish their work. In recent years, research institutions around the world have been promoting the uptake of open access practices, as shown in the steadily increasing number of relevant policies. Nonetheless, the main focus thus far has been on open access to publications rather than research data and a comparatively small number of institutions has developed policies for research data management. Motivation to develop policies derives from institutions' need to safeguard their intellectual, financial, human and material investment, as well as the increasing pressure from research funders who require that research data produced with their funding is properly managed and openly accessible. In some cases, the motivation for developing a sound institutional data strategy derives from researchers, who acknowledge the significance of research data and the need for better management.

The most consistent progress in research data management is observed in the UK, the USA and Australia. Rapid developments both in the UK and the USA are mostly the result of funder mandates: Research Councils UK and the National Science Foundation and the National Institutes of Health in the USA. In Australia, while the policy of the main funding agency is not mandatory for research data, universities have made significant progress in addressing research data management under the influence of the Australian Code for Responsible Conduct of Research, requiring an institutional data management policy.

An effective data policy sets the pace and the requirements by which the research community within the institution is to abide. Such policy should allocate in a clear way responsibilities and tasks to the different actors within the institution, with researchers carrying the obligation to manage their research data to specific standards and the institution assuming the obligation to provide the services (infrastructure, training etc.) that will in turn allow researchers to comply with the policy requirements. While the allocation of responsibilities for each stakeholder is important, policies should be flexible enough to accommodate for the changes in researchers' needs and keep pace with technological developments. Institutional policies share a number of other common elements: they recognize the significance and value of research data and high standards for their management; they set open access to research data as the default, where this is appropriate and legally possible; they require researchers to develop a DMP; they render researchers responsible for the data management within their project; they acknowledge the need to respect funder requirements. Furthermore, they set requirements regarding where to deposit research data and outline broadly the data retention policy/strategy of the institution.

Developing and implementing a data management policy and developing relevant services is a team effort requiring the collaboration of multiple actors. The main ones are the research office, the IT departments, the academic units, the libraries and the researchers. When it comes to developing services, the university library and the IT department are those mostly involved in operationalizing policies: i.e. the development of the technical infrastructure and its services, the training for the researchers and advocacy services. It is common that IT departments undertake the software and infrastructure development, while the library supports archiving, training and advocacy activities. In developing data management services institutions need to consider which services should be developed in-house and which may be outsourced, on the basis of an assessment of their needs and resources. With respect to infrastructures, while in general they are more developed as compared to the associated policy frameworks, dedicated research data repositories are not widespread among research institutions.

Institutional policies for data management and open access to research data should be accompanied by relevant funds. In particular, funding is necessary both for data management during the life cycle as well as for the curation and preservation of data in the long term as in some cases research

institutions are seen as the 'obvious' place to host data, while in others they might constitute the only viable option given the patchy coverage of subject-specific data repositories or other data services. Yet, as external funding is usually limited to the lifetime of research projects, research institutions must increasingly turn towards finding resources for the long-term management and preservation of their output in research data.

In terms of training, formal training is necessary for researchers, as well as for librarians and information professionals in order to transition to open access to research data and a culture of open science more generally. While researchers in some fields may require training because they lack the knowledge and the skills on how to make their research data available and accessible, or how to reuse data and incorporate data in their research process, librarians and information experts require training for providing research data services that are necessary in an increasingly data-intensive research environment. Thus, workshops, as well as formal training programmes and curricula that enable data management skills, data-intensive research, and the gradual development of data-scientists are important activities for research institutions to engage in.

Finally, further progress is needed in terms of rewarding researchers for good data management and providing open access to research data. Currently there is little, if any, formal recognition for data outputs in academic promotion or other assessment processes, which inhibits progress towards open access to research data.

Recommendations

1. **Develop an explicit institutional research data strategy with open access as the default position**

Consultation and collaboration with the research community is critical in understanding their needs and in developing the necessary infrastructure and services. The establishment of committees within institutions that will work in close collaboration with funders and the research communities will alleviate significant pressure from researchers and accommodate disciplinary practices.

2. **Actively pursue collaborations between and within institutions in fostering a sustainable ecosystem and infrastructure for open access to and long-term preservation of research data**

Developing relevant services requires the collaboration among different institutional departments within an institution. It further requires research institutions to evaluate their current capacities and collaborate with other institutions and centers of expertise in providing services and enabling a sustainable and scalable scholarly communications ecosystem.

3. **Include open access to high quality research data as a formal criterion for career progression**

Formal acknowledgement of research data as a legitimate output is expected to bring gradual change in practices. Such formal recognition should be accompanied by the development and use of metrics that allow the collection and tracking of data use and impact.

4. **Develop educational and training programmes for researchers and staff to improve data management skills and to enhance data-intensive research**

In designing such programmes research institutions should pay attention to disciplinary specificity and practices, while avoiding one-size-fits-all solutions. In doing so, research institutions can explore the possibility of developing joint courses with data managers, especially data

centers, and across different specialties.

5. Raise awareness about the benefits of open access to research data and provide rewards

Focusing on awareness-raising and advocacy activities, as well as rewarding researchers are necessary tools to this end. Awareness and advocacy activities can have different formats, such as seminars, webinars, brochures, leaflets etc., and should be explored in combination with the development of training programmes for researchers.

6. Support the research community through the provision of legal and ethical advisory services

Research institutions may systematically support their researchers in addressing legal and ethical challenges raised by open access to research data by deploying specific instruments (e.g. committees, formal training) to develop new and common solutions to issues such as licensing, privacy and confidentiality, among others.

Data Managers

The term data manager refers to those stakeholders within the open access ecosystem that are charged with the management of the scientific and/or cultural digital output in data: data centres, which are mainly government financed operations for making datasets available, libraries, archives and memory institutions that maintain collections of content. Some of them have developed strength in relevant technological infrastructures for the storage, curation and long-term preservation of digital data, while others are still lagging behind.

Data centres come in different forms and sizes and often emerge from a disciplinary community. Their most basic function relates to the storing of research datasets for a defined community and making them accessible for other researchers to discover and use. This entails two roles: firstly, ensuring that data is discoverable as well as ensuring the tools allowing other researchers to find and access them and, secondly, providing support services for researchers who need to get their data and metadata into shape prior to deposit.

Libraries traditionally provide access to resources and publications through subscription. With reference to research data, and under the pressure of research funders' mandates, they are gradually becoming involved in data curation, while being the primary training and information locus on this topic for the researchers, offering awareness-raising and advocacy services. Despite their eagerness to acquire an important role in the transition to an open access research culture, libraries currently fall short both in terms of their current practices, as well as in terms of meeting the demands of researchers and users in relation to the provision of data management and support services.

Irrespective of their character, data managers should address open access to research data as an important development towards open science and develop services to support the needs of their patrons. These services can be defined on the basis of their mission and context, and by establishing extensive collaborations with the research community and other important stakeholders in the scholarly communication ecosystem (research institutions, publishers, funders), as well as relying on current best practices and resources.

The costs of data management and curation services are a further issue that data managers should address. Data management costs are incurred by the acquisition, ingestion and access to data, personnel wages, training costs for researchers and (data) librarians, the technical infrastructure and outreach programmes. Reliance exclusively on project funding is nonetheless problematic, as it does not guarantee long-term funds and, thus, operations. Consequently, developing sustainable funding models on the principles of diversifying sources of income and establishing collaborations should be addressed with particular care.

A further important contribution of data managers is towards the proliferation of high quality research data, i.e. securing the technical quality of research data. Data needs to be presented in standardized formats and accompanied by appropriate metadata; if these conditions are not met data are hard to work with and require additional time and financial resources to make them accessible and usable. Several repositories and data centres have developed quality assurance measures and offer a range of services to evaluate the technical quality of data sets. These include providing process documentation, completeness/consistency checks, training on data management and sharing, file format validation, metadata checks, storage integrity verification and tools for annotating the quality information. In addition, numerous libraries and data centers have been experimenting with new mechanisms to enhance data quality through platforms for discussing data sets or offering tools for alternative metrics (altmetrics).

Data managers also have a role in the selection of data for long-term preservation and retention. The gap between short-term access and long-term preservation to research data needs to be addressed, and emphasis needs to be placed on long-term preservation. The value of data is assessed both in terms of its technical as well as of its scientific quality.

Aside from the quality of the research data, the quality of services offered by data centres and repositories is becoming a cutting edge issue. Furthermore, research funders and publishers are putting additional pressure by inquiring deposit in certified and accredited repositories, in an effort to secure the reusability and long-term preservation of research data. In such context, obtaining accreditation or certification to appropriate standards is a way for ensuring both the quality of data repositories and of the quality assurance process.

Finally, data managers have a significant role in providing training to researchers for meeting technical quality standards with their data sets, as well as in developing disciplinary standards. Data centers with expertise in data curation have an important role in enhancing the skills of research library staff in data managements, data quality and developing data services.

Recommendations

1. Assess their position within the open access ecosystem in view of developing collaborative infrastructures and services

Research libraries and data centers are encouraged to evaluate their overall capacity and positioning within the open access ecosystem and assess the types of services to be provided, in collaboration with other stakeholders (research institutions, publishers and funders) and the research community.

2. Develop sustainable business models to ensure long-term service provision

Planning for sources of income should be addressed efficiently and, as much as possible at the outset of service development, while the strategy should be reviewed at regular intervals. Acquiring income may require the diversification of income resources and the layering of the services offered, whereby some services incur charges for the users.

3. Establish mechanisms for data quality that ensure re-use and long-term preservation through collaborative work

To ensure data quality for re-use and long-term preservation data managers are equipped with a range of quality assurance and control strategies. While they are expected to take the leading role in close collaboration with research communities (scholarly societies, research institutions and researchers) in establishing citation standards, their collaboration with funders, publishers and journal editors is central in ensuring the enforcement of relevant policies.

4. Acquire certification/accreditation to guarantee high quality services in the long term

Establishing quality assurance mechanisms is important not only for the trustworthiness of research data but for the data centres hosting them. Data centres are thus encouraged to seek appropriate certification and accreditation guaranteeing the quality of their services, such as the Data Seal of Approval and/or other appropriate ISO certification.

5. Support data management through the development of training programmes for researchers and librarians/ technical staff

Libraries should be minimally able to deliver training courses on DMP of general or discipline-specific nature to serve the particular needs of their research communities and librarians as well as more specialized topics like intellectual property rights, licensing, re-use of research data and ethical issues.

Publishers

As a result of the journal-based dissemination structure of research, publishers are key stakeholders in the open access ecosystem. They are, thus, in a unique position, in cooperation with the rest of stakeholders to contribute towards a culture of openly sharing research data of high quality, linked to the publications they support, and fit for re-use. The publisher ecosystem is diverse, comprising institutions which are very small non-profit scholarly led, university-based operations, and small entrepreneurial ventures, as well as giant multinational enterprises that are central to the market.

Whereas publishers have placed a strong focus on open access to publications and open access as a business model, their engagement with research data and open research data in particular is relatively recent. Publishers are interested in research data and open research data because they add value to their main products (publications) by enhancing the trustworthiness of the published research through the ability to verify it, which lies at the heart of ethical conduct of research. Publishers are also increasingly developing policies as a response to the pressure from funders' policies in relation to open access.

This recent attention to research data is leading publishers to exploit the possibilities of research data in new data-based products and services, such as publishing data journals, extending peer-review to research data, and offering services to enhance data quality. The emergence of data journals should be linked to the effort of publishing data separately that allow essential parts of the scientific record to be made available in an intelligible form to the scientific community. Data journals are community peer-reviewed open access platforms for publishing, sharing and disseminating data that cover a wide range of disciplines. The papers published contain information on the acquisition, methods, and processing of specific data sets. The published papers are cross-linked with approved repositories, citing data sets that have been deposited in such repositories or data centres. The publication of data papers can be considered as a good practice example of data management as it includes an element of peer review to the dataset, maximizes the opportunities for data re-use and provides academic accreditation to researchers. As data papers are becoming distinct publishing products, a number of data journals are also supporting alternative metrics (altmet-

rics), thereby enhancing further data publication. Recent emphasis on open access to research data and data publications brings to the fore the scientific quality of research data and the significance of research data peer-review. A further related topic is the citation of research data. Apart from data peer-review, publishers may contribute to the standardization of research data by gradually introducing policies that are compliant with current best practices.

Publishers are also turning their attention to include content discovery and linking services, as well as services that focus on exploiting content with text and data mining (TDM) tools. Increasing attention on TDM is a direct result of researchers' need to explore large databases of content, data and publications. Despite the estimated economic opportunities TDM can bring, the perceived threat by publishers towards allowing fully unobstructed TDM to be performed in their content has resulted in restrictive measures that limit researchers' abilities for cutting-edge computer-aided research.

In developing policies for open access to research data, peer-review of research data, and products/services such as data journals, it is understandable that publishers are required to collaborate closely with other important stakeholders. Close collaboration with data centers and repositories (data managers) is necessary, since the latter are the primary content holders in the case of research data, and thus the destination to which the publications provide links to for access to research data. Data managers are the guarantors for the technical quality, security, curation and preservation of research data. As publications increasingly involve mixing and linking papers and data, collaboration is required in establishing principles for standards that will guarantee the long-term access to high quality data. Finally, close collaboration is required between the publishers and the scientific community, such as scholarly societies and journal editors, in developing those editorial principles that promote citation of research data through the development of disciplinary-specific standards alongside internationally accepted principles as well as data review processes. Collaboration between publishers and funders is also essential, in view of the development of products and services that align to funder requirements.

Recommendations

1. Gradually develop mandatory policies for open access to research data supporting publications

Editorial policies should address issues like documentation, metadata and format of published data, licensing, and citation. Editorial policies should be enhanced further through data availability statements provided both during the article submission process and the peer-review process. Policies should provision measures in cases of non-compliance brought to light after publication (such as retracting the published article).

2. Collaborate with certified repositories and data centers to streamline data submission

Publishers are encouraged to collaborate with repositories and data centers that meet accepted criteria regarding their trustworthiness. For disciplines without community endorsed data centers/repositories, publishers can assist researchers by providing guidance and assistance on appropriate institutional repositories or commercial data services may be designated for deposit and access.

3. Support data as a first-class scholarly output through the establishment of peer-review processes

Establishing peer-review processes for research data is a measure that contributes to the further enhancement of products of high quality. Peer-review processes should specify the criteria used relating to the technical aspects and quality of research data (completeness and consistency of dataset, appropriate standards, software used), while their scientific quality is assessed by the research community through pre- and post-publication peer-review.

4. Develop policies requiring citations for research data

Publishers should require that data accompanying their publications are citeable, and provide clear guidelines for data citation. Data cita-

tion should include DOIs, as well as licensing information (e.g. Creative Commons licenses), preferably machine actionable, that informs users about what they are able to do with research data.

5. Establish licensing policies that encourage the use of TDM

Editorial policies should clearly state the licenses (default and recommended) under which the data are published and re-used. Taking into consideration the significant economic benefits that can be derived from the use of TDM tools publishers are encouraged to adapt their policies to allow for an increases use of such techniques in research.

A Practical Guide for developing policies for Research Funders

Preparing and implementing a policy

The following key points should be addressed by funders in developing and implementing a policy for open access to research data:

- **Knowledge of international policies** to assess position and standing in terms of policies, infrastructures, practices and degree of participation in international fora
- **Participation in dialogue and collaboration** among stakeholders at the national level, and minimally, with research institution administration, researchers and particular disciplinary communities (e.g. scholarly societies), data managers, publishers.
- **Assessment of existing and required infrastructure** to support policy implementation.
- **Assessment of related policy-implementation costs** for research data management during projects, long-term curation and preservation, infrastructure development, funding for innovative, disciplinary, education, training and awareness activities and earmarking of funds.
- **Policy content development** with open access as the default (cf. below on policy content).
- **Data Management Plans (DPMs)** as essential components of grant proposals, where data is generated.

- **Provision for relevant open access clause in grant agreements,** accompanied by the description of sanction mechanisms in cases of non-compliance, as well as clarification of eligible costs.
- **Guidance to researchers** through the development of appropriate tools such as templates for data management and resources on data management and DMPs, and criteria for eligible repositories/data centers for data deposit.
- **Rewards to researchers** through measures that can assist in changing research cultures such as the award of prizes for high-quality data or through events that focus on highlighting and communicating success stories on data sharing and re-use.
- **Policy monitoring mechanisms** to assess and measure compliance and efficiency and revise policy, where necessary.

Policy content

A policy should address the following issues:

- **Open access as default.** The policy should set open access for research data as the default and mandatory requirement and provide appropriate support and funding (e.g. expenses for storage). The possibility for closed data should be accommodated when ethical, copyright, confidentiality, security and similar issues are demonstrably of key concern.
- **Responsibilities.** The policy should assign responsibilities and set out the expectations for the main stakeholders involved, namely: funders, researchers (either under their capacity of grant applicants or grant holders), research institutions, data centers and repositories.
- **Target content.** The policy should be explicit on which data should be open. Open access should be required for research data used to validate scientific claims in publications, while open access to other data produced in the project may be required to be open as well, including associated metadata. While open access to the research data itself may not always be possible, deposit in repositories/data centers with open metadata should be required.
- **Data Management Plan.** The policy should require grant applicants who will generate data to provide a DMP as the main tool through which to address comprehensively data management, including access to data. Templates for DMPs should be provided along with resources.
- **Time of deposit.** The policy should require data supporting publications to be made open ideally at the latest at the same time with the publications and link to it, while other data by the end of the project.
- **Locus of deposit.** The policy should require deposit in certified and trusted repositories and/or data centers that are of relevance to the scientific communities. Funders recommend or require deposit with specific data centers or repositories.
- **Technical specifications to allow reuse.** To enable research data reuse and citation funders should require information on metadata, DOI, interoperability of systems, machine readability and mineability

and software in the policy.

- **Licensing research data.** The policy should require that research data is accompanied by licensing describing the terms of use, such as Creative Commons licenses. Preferably licensing information should be machine-actionable.
- **Provisions for long-term availability.** Policies should include provisions for the long-term availability of data, since re-use and availability are primary reasons for open access to research data.
- **Compliance with policy.** The policy should make statements regarding compliance to it by the researchers and clarify measures for non-compliance (e.g. funder may refrain from delivering the full amount of funding in cases of non-compliance)

Practical checklist for funders

- Have you mapped relevant international policies for open access to research data?
- Have you involved stakeholders and the research community in developing the policy?
- Have you assessed the available infrastructures that are necessary for the implementation of your policy?
- Have you estimated the costs for data management and preservation?
- Does your policy include statements on:
 - Open access as the default and mandatory position and possibility for closed access is offered when necessary
 - Distribution of responsibilities to involved parties
 - Target data for open access
 - Time of deposit
 - Locus of deposit
 - Technical specifications
 - Licensing

- Requirement of Data Management Plan
- Compliance and monitoring statement

- Do you require grant applicants to offer information regarding data management at the application stage?
- Do you include open access to research data as a clause in your grant agreements?
- Do you offer guidance to researchers in your website and otherwise to enable them to comply with your policy?
- Have you made provisions to provide incentives to researchers for making their research data open?
- Have you established a monitoring and compliance mechanism?
- Have you decided how and when to evaluate the efficacy of your policy?

A Practical Guide for developing policies for Research Institutions

Preparing and implementing a policy

The following key points should be addressed by research institutions in developing and implementing a policy for data management and open access to research data:

- **Knowledge of international institutional policies** to assess institution's position, participate international fora
- **Participation in dialogue and collaboration** among stakeholders within the institution and outside of it (e.g. funders, scholarly societies, data managers) for policy development
- **Assessment of state of existing and necessary infrastructure** to support policy implementation through economies of scale and collaborative initiatives
- **Cost assessment for policy implementation** for research data management (especially for long-term provisions), infrastructure and service development, training and education and awareness activities, and earmarking of funds
- **Policy content development** with clear description of roles and responsibilities of stakeholders involved
- **Data Management in research practice.** Where data is generated, data management should form an essential element of research practice by providing appropriate resources, reviewing and monitoring of related practices

- **Guidance to researchers.** Development of appropriate tools such as templates for data management and resources on data management and DMP, and relevant training to researchers
- **Rewards for researchers** through the formal acknowledgment of research data as a criterion for career progression
- **Policy monitoring mechanisms** to assess and measure compliance and efficiency and revise policy, where necessary.

Policy content

A policy should address the following issues:

- **Open access as default.** The policy should set open access for research data as the default requirement and provide appropriate support and funding (e.g. expenses for storage). Such policy should be mandatory and not voluntary. The possibility for closed data should be accommodated when ethical, copyright, confidentiality, security and similar issues are demonstrably of key concern.
- **Responsibilities.** The policy should define in a clear way the responsibilities of the institution and its researchers. Researchers carry the obligation to manage their research data according to specific standards and the institution assuming the responsibility of providing the necessary services (infrastructure, training etc.).
- **Locus of deposit.** The policy should specify that data are to be deposited in the institutional repository. In the case of absence of an institutional repository the related policy should provide guidance on deposit in trusted repositories (list of trusted repositories or criteria that researchers can use for selecting the appropriate repository).
- **Time of deposit.** The policy should require data supporting publications to be made open ideally at the latest at the same time with the publications and link to it, while other data by the end of the project.
- **Licensing.** The policy should require that research data is accompanied by licensing describing the terms of use, such as Creative Commons licenses. Preferably licensing information should be machine-actionable.

Practical checklist for research institutions

- Does your policy include statements on:
 - Open access as the default and mandatory position and the possibility for closed access when necessary
 - Distribution of responsibilities to involved parties
 - Target data for open access
 - Time of deposit
 - Locus of deposit
 - Technical specifications
 - Licensing
 - Requirement of Data Management Plan
 - Compliance and monitoring statement
- Have you involved stakeholders both within and outside the institution in developing the policy?
- Have you assessed your infrastructure and services and have you considered potential collaborations with data centres?
- Do you offer guidance and support to researchers at your institution to enable researchers to comply with your policy?
- Have you made provisions to provide incentives to researchers for making their research data open? (e.g. open data as a formal criterion for career progression?)
- Have you established a monitoring and compliance mechanism?
- Have you decided how and when to evaluate the efficacy of your policy?

A Practical Guide for developing policies for Publishers

Developing a Policy

The following key issues should be addressed by Publishers in the process of developing a policy for open access to research data:

- **Collaboration and consultation** with publisher collective instruments and researcher communities for policy and standard development
- **Definition of policy with open access as the default position**, accommodating closed access for legal or ethical reasons, and including possibility of retraction of publications for non-compliance.
- **Editorial policies for research data** and in particular consideration of adoption of peer-review for research data and standardization of data citation (including DOIs and licensing requirements) in collaboration with research community
- **Guidelines and support for authors** to comply with open access research data policy and data editorial policies.
- **Choice of accredited repositories or data centers** for deposit of research data in collaboration with the research community
- **Indicators for measuring data impact** such as data downloads, use, re-use, citation etc., enhancing the recognition of research data as a first-class scholarly output.
- **Promoting data sharing** through prizes or competitions rewarding high-quality data sharing.

Practical checklist for publishers

- Do you consult with publisher collective instruments and researcher communities in addressing/developing policies for open access to research data?
- Does your editorial policy include:
 - Open access as the default
 - Provisions for cases of closed access
 - Statement on licenses (default, alternative)
 - Sanctions for non-compliance
 - A data availability statement
- Have you developed editorial policies for research data that cover peer-review and standardization of citation requirements?
- Do you provide a list of recommended repositories for data submission?
- Do you require open licenses (such as Creative Commons) for research data accompanying publications?
- Do you offer clear guidance and support to authors to comply with the aforementioned policies?
- Have you developed indicators for measuring data impact from your publications?
- Do you encourage data sharing through specific actions, such as prizes?

Resources

Funder Policies

RCUK Common Principles on Data Policy

<http://www.rcuk.ac.uk/research/datapolicy/>

The ESRC Research Data Policy

<http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx>

The EPSRC policy framework on research data

<http://www.epsrc.ac.uk/about/standards/researchdata/expectations/>

The White House Open Access policy directive

http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

The NSF Data Sharing Policy

(including specific department guidelines)

<http://www.nsf.gov/bfa/dias/policy/dmp.jsp>

European Commission policies for Open Access

Recommendation on open access to, dissemination of and preservation of scientific information (2012)

http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

Horizon Model Grant Agreement article 29

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

The Open Access guidelines for Horizon 2020

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Funding initiatives for data-intensive research

The NIH Big Data2K initiative for intensive data research

<http://bd2k.nih.gov/index.html#sthash.AQOOxJfr.dpbs>

The NEH and funder alliance Digging into the Data Challenge in the Humanities

<http://diggingintodata.org/>

Research Institution Policies

The University of Oxford Policy on the Management of Research Data and Records

http://researchdata.ox.ac.uk/files/2014/01/Policy_on_the_Management_of_Research_Data_and_Records.pdf

Guide for developing institutional research data policy

<http://www.dcc.ac.uk/resources/policy-and-legal/five-steps-developing-research-data-policy/five-steps-developing-research>

List of institutional policies in the UK

<http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies>

Collaborations

RDNL (national institutional collaboration for research data, the Netherlands)

<http://www.researchdata.nl/>

UK Data Archive (funder data-center partnership)

<http://www.data-archive.ac.uk/>

Geoscience Data Journal

(publisher and repositories collaboration)

<http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%292049-6060>

3TU.Datacentrum

(multi-institutional collaboration for research data services)

<http://datacentrum.3tu.nl/en/home/>

Data Centers/Data services

Data Archiving and Networking Services in the Netherlands

<http://www.dans.knaw.nl/en>

The National Oceanographic Data Centre

<http://www.nodc.noaa.gov/about/overview.html>

Expertise and Resources on Research data management and curation

The Digital Curation Centre

Home

<http://www.dcc.ac.uk/>

Training and Expertise

<http://www.dcc.ac.uk/training>

DMPs

<http://www.dcc.ac.uk/resources/data-management-plans>

Developing services

<http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services>

Accreditation

Data Seal of Approval

<http://www.datasealofapproval.org>

Publisher Policies

The PLOS mandatory policy for open access to research data

<http://www.plosone.org/static/policies.action#sharing>

The Journal of Open Archaeology

<http://openarchaeologydata.metajnl.com/>

Data Citation Principles

The FORCE11 Data Citation Principles

<https://www.force11.org/datacitation>

List of Project Partners

UK TRI

Trilateral Research and Consulting | trilateralresearch.com

NL KNAW

Koninklijke Nederlandse Akademie van Wetenschappen-KNAW
| www.knaw.nl

UK USFD

The University of Sheffield | www.sheffield.ac.uk

NL LIBER

Stichting LIBER | libereurope.eu

EL EKT

National Documentation Centre/NHRF | www.ekt.gr

IT CNR-IIA

Consiglio Nazionale delle Ricerche | www.cnr.it/sitocnr

SW BTH

Blekinge Tekniska Hogskola | www.bth.se

NL AUP

Amsterdam University Press | en.aup.nl

Policy
recommendations
for open access
to research data

www.recodeproject.eu